

ISSN 1998-6688 (Print)
ISSN 2959-8109 (Online)

**ҚАЗАҚСТАН-БРИТАН ТЕХНИКАЛЫҚ
УНИВЕРСИТЕТІНІҢ
ХАБАРШЫСЫ**

**HERALD
OF THE KAZAKH-BRITISH
TECHNICAL UNIVERSITY**

**ВЕСТНИК
КАЗАХСТАНСКО-БРИТАНСКОГО
ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА**

Volume 23, Issue 2
April–June, 2026

**ҚАЗАҚСТАН-БРИТАН
ТЕХНИКАЛЫҚ УНИВЕРСИТЕТІНІҢ
ХАБАРШЫСЫ**

**HERALD
OF THE KAZAKH-BRITISH
TECHNICAL UNIVERSITY**

**ВЕСТНИК
КАЗАХСТАНСКО-БРИТАНСКОГО
ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА**

Алматы
№ 2 (77) 2026 г.

Издание зарегистрировано Министерством культуры и информации Республики Казахстан. Свидетельство о постановке на учет СМИ № 9757 – Ж от 03.12.2008 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

В журнале публикуются научные статьи на английском, казахском и русском языках по направлениям: компьютерные науки, математические науки, физические науки, нефтегазовая инженерия и геология, экономика и бизнес.

Подписной индекс – 74206

Издается с 2004 года. Выходит 4 раза в год.

УЧРЕДИТЕЛЬ
Казахстанско-Британский технический университет

ISSN 1998-6688 (Print)
ISSN 2959-8109 (Online)

© Казахстанско-Британский технический университет, 2026

Editor-in-chief – **B. Kulpeshov**, Doctor of Physical and Mathematical Sciences, academician of the National Academy of Sciences of the Republic of Kazakhstan, Professor, School of Applied Mathematics of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan

EDITORIAL BOARD

Section “Computer science”

- Akzhalova A.Zh.** - Candidate of Physical and Mathematical sciences, PhD, Professor, School of Information Technology and Engineering of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Pak A.A.** - Candidate of Technical Sciences, Associate professor, School of Information Technology and Engineering of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Samigulina Z.I.** - PhD, Professor, School of Information Technology and Engineering of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Sadok Ben Yahia** - Professor, University of Southern Denmark, Denmark;
- Eduard Siemens** - Professor, Anhalt University of Applied Sciences, Germany;
- Uvalieva I.M.** - PhD, Professor at the Faculty of Computer Science, EKTU named after D. Serikbayev, Ust-Kamenogorsk, Kazakhstan;
- Serek A.G.** - PhD, Associate Professor, «Astana IT University» LLP, Astana, Kazakhstan.

Section “Mathematical Sciences”

- Kalmurzayev B.S.** - PhD, Associate professor, School of Applied Mathematics of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Sudoplatov S.V.** - Doctor of Physical and Mathematical Sciences, Chief Researcher, Institute of Mathematics SB RAS; Head of Department, Novosibirsk State Technical University, Novosibirsk, Russia;
- Bazhenov N.A.** - Candidate of Physical and Mathematical Sciences, Senior Researcher, Institute of Mathematics SB RAS, Novosibirsk, Russia;
- Nurkhaidarov E.** - PhD, Associate Professor, Department of Mathematics, The Pennsylvania State University, University Park, USA.

Section “Physical Sciences”

- Abdirakhmanov A.R.** - PhD, Assistant Professor JSC “KBTU”, Almaty, Kazakhstan; University of Mons, Mons, Belgium;
- Beysenkhanov N.B.** - Doctor of Physical and Mathematical Sciences, Dean, Professor, School of Materials Science and Green Technologies JSC “KBTU”, Almaty, Kazakhstan;
- Laifa Boufendi** - PhD, Professor, Orlean University, France

Section “Oil and Gas Engineering and Geology”

- Ismailov A.A.** - Candidate of Technical Sciences, Dean, Professor, School of Energy and Oil and Gas Industry of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Korobkin V.V.** - Candidate of Geological and Mineralogical Sciences, Professor, School of Geology of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Kang Wanli** - Professor, School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, PR China;
- Sarsenbekuly B.** - PhD, Associate professor, School of Energy and Petroleum Engineering of JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Moldabaeva G.Zh.** - Doctor of Technical Sciences, Professor, Department of Petroleum Engineering of K. Turysov Institute of Geology and Petroleum Engineering, KazNTU, Almaty, Kazakhstan.

Section “Economics and Business”

- Narbaev T.S.** - PhD, Professor, Business School JSC “KBTU”, Almaty, Kazakhstan;
- Kozhakhmetova A.K.** - PhD, Assistant Professor, Business School JSC “KBTU”, Almaty, Kazakhstan;
- Aldashev A.** - Professor, International School of Economics JSC “KBTU”, Almaty, Kazakhstan;
- Akhmedov E.T.** - PhD, Associate Professor, Business School, JSC “Kazakh-British Technical University”, Almaty, Kazakhstan;
- Seitkazieva A.M.** - Doctor of Economics, Professor/Business School JSC “KBTU”, Almaty, Kazakhstan;
- Gabriel Castelblanco** - PhD, Assistant-Professor, University of Florida, Florida, USA;
- Dubina I.** - Doctor of Economic Sciences, Professor, Novosibirsk National Research State University (NSU);
- Osei-Kyei Robert** - PhD, Associate Professor, University of Western Sydney, Penrith, Australia;
- Montag Josef** - PhD, Senior Lecturer, Charles University, Prague, Czech Republic..

Бас редактор – **Кулпешов Б.Ш.**, ф.-м.ғ.д., ҚР ҰҒА академигі, профессор,
«ҚБТУ» АҚ Қолданбалы математика мектебі, Алматы қ., Қазақстан

РЕДАКЦИЯ АЛҚАСЫ

«Компьютерлік ғылымдар» бөлімі

- Акжалова А.Ж.** - ф.-м.ғ.к, PhD, профессор, «ҚБТУ» АҚ Ақпараттық технологиялар және инженерия мектебі, Алматы қ., Қазақстан;
Пак А.А. - т.ғ.к., қауымдастырылған профессор, «ҚБТУ» АҚ Ақпараттық технологиялар және инженерия мектебі, Алматы қ., Қазақстан;
Самигулина З.И. - PhD, профессор, «ҚБТУ» АҚ Ақпараттық технологиялар және инженерия мектебі, Алматы қ., Қазақстан;
Садок Бен Яхия - профессор, Оңтүстік Дания университеті, Дания;
Эдуард Сименс - профессор, Анхальт қолданбалы ғылымдар университеті, Германия;
Увалиева И.М. - PhD, компьютерлік ғылымдар факультетінің профессоры, Д. Серікбаев атындағы ШҚТУ, Өскемен қ., Қазақстан;
Серік А.Ғ. - PhD, қауымдастырылған профессор, «Astana IT University» ЖШС, Астана қаласы, Қазақстан.

«Математикалық ғылымдар» бөлімі

- Калмурзаев Б.С.** - PhD, қауымдастырылған профессор, «ҚБТУ» АҚ Қолданбалы математика мектебі, Алматы қ., Қазақстан;
Судоплатов С.В. - ф.-м.ғ.д., профессор, РҒА СБ математика институтының бас ғылыми қызметкері; Новосібір мемлекеттік техникалық университеті, кафедра меңгерушісі, Новосібір қ., Ресей;
Баженов Н.А. - ф.-м.ғ.к., РҒА СБ математика институтының аға ғылыми қызметкері, Новосібір қ., Ресей;
Нурхайдаров Е. - PhD, қауымдастырылған профессор, Математика департаменті, Пенсильвания штатының университеті, Юниверсити-Парк, АҚШ.

«Физикалық ғылымдар» бөлімі

- Әбдірахманов А.Р.** - PhD, ассистент-профессор, «ҚБТУ» АҚ, Алматы қ., Қазақстан; Монс университеті, Монс қ., Бельгия;
Бейсенханов Н.Б. - ф.-м.ғ.д., декан, профессор, Материалтану және жасыл технологиялар мектебі «ҚБТУ» АҚ, Алматы қ., Қазақстан;
Laifa Boufendi - PhD, профессор, Орлеан университеті, Франция.

«Мұнай газ инженериясы және геология» бөлімі

- Исмаилов А.А.** - т.ғ.к., декан, профессор, «ҚБТУ» АҚ Энергетика және мұнай-газ индустриясы мектебі, Алматы қ., Қазақстан;
Коробкин В.В. - геол.-минерал.ғ.к., профессор, «ҚБТУ» АҚ Геология мектебі, Алматы қ., Қазақстан;
Канг Ванли - профессор, Мұнай инженериясы мектебі, Қытай мұнай университеті (Шығыс Қытай), Циндао қ., ҚХР;
Сарсенбекұлы Б. - PhD, қауымдастырылған профессор, «ҚБТУ» АҚ Энергетика және мұнай-газ индустриясы мектебі, Алматы қ., Қазақстан;
Молдабаева Г.Ж. - тех.ғ.д., профессор, мұнай инженериясы кафедрасы, Қ. Тұрысов атындағы геология және мұнай-газ институты, ҚазҰТУ, Алматы қ., Қазақстан.

«Экономика және бизнес» бөлімі

- Нарбаев Т.С.** - PhD, профессор, Бизнес мектебі «ҚБТУ» АҚ, Алматы қ., Қазақстан;
Кожаметова А.К. - PhD, ассистент-профессор, Бизнес мектебі, «ҚБТУ» АҚ, Алматы қ., Қазақстан;
Алдашев А. - PhD, профессор, Халықаралық экономика мектебі, «ҚБТУ» АҚ, Алматы қ., Қазақстан;
Ахмедов Е.Т. - PhD, қауымдастырылған профессор, Бизнес мектебі, «Қазақстан-Британ техникалық университеті» АҚ, Алматы қ., Қазақстан;
Сейтказиева А.М. - э.ғ.к., профессор, Бизнес мектебі «ҚБТУ» АҚ, Алматы қ., Қазақстан;
Gabriel Castelblanco - PhD, ассистент-профессор, Флорида университеті, Флорида, АҚШ;
Дубина И. - э.ғ.д., профессор, Новосибирск ұлттық зерттеу мемлекеттік университеті (НМУ);
Osei-Kyei Robert - PhD, қауымдастырылған профессор, Батыс Сидней университеті, Пенрит қ., Австралия;
Montag Josef - PhD, сениор-лектор, Карлов университеті, Прага қ., Чехия Республикасы.

Главный редактор – **Кулпешов Б.Ш.**, д.ф.-м.н., академик НАН РК, профессор,
Школа прикладной математики АО «КБТУ», г. Алматы, Казахстан

ЧЛЕНЫ РЕДКОЛЛЕГИИ

Раздел «Компьютерные науки»

- Акжалова А.Ж.** - к.ф.-м.н., PhD, профессор, Школа информационных технологий и инженерии АО «КБТУ», г. Алматы, Казахстан;
- Пак А.А.** - к.т.н., ассоциированный профессор, Школа информационных технологий и инженерии АО «КБТУ», г. Алматы, Казахстан;
- Самигулина З.И.** - PhD, профессор, Школа информационных технологий и инженерии АО «КБТУ», г. Алматы, Казахстан;
- Садок Бен Яхия** - профессор, Университет Южной Дании, Дания;
- Эдуард Сименс** - профессор, Анхальтский университет прикладных наук, Германия;
- Увалиева И.М.** - PhD, профессор факультета компьютерных наук, ВКТУ им. Д.Серикбаева, г. Усть-Каменогорск, Казахстан;
- Серек А.Г.** - PhD, ассоциированный профессор, ТОО «Astana IT University», г. Астана, Казахстан.

Раздел «Математические науки»

- Калмурзаев Б.С.** - PhD, ассоциированный профессор, Школа прикладной математики АО «КБТУ», г. Алматы, Казахстан;
- Судоплатов С.В.** - д.ф.-м.н., профессор, главный научный сотрудник, Институт математики СО РАН; зав. кафедрой, Новосибирский государственный технический университет, г. Новосибирск, Россия;
- Баженов Н.А.** - к.ф.-м.н., старший научный сотрудник, Институт математики СО РАН, г. Новосибирск, Россия;
- Нурхайдаров Е.** - PhD, ассоциированный профессор, Департамент математики, Университет штата Пенсильвания, Университи-Парк, США.

Раздел «Физические науки»

- Әбдірахманов А.Р.** - PhD, ассистент-профессор АО «КБТУ», г. Алматы, Казахстан; Университет Монса, г. Монс, Бельгия;
- Бейсенханов Н.Б.** - д.ф.-м.н., декан, профессор, Школа материаловедения и зеленых технологий АО «КБТУ», г. Алматы, Казахстан;
- Laifa Boufendi** - PhD, профессор, Университет Орлеана, Франция.

Раздел «Нефтегазовая инженерия и геология»

- Исмаилов А.А.** - к.т.н., декан, профессор, Школа энергетики и нефтегазовой индустрии АО «КБТУ», г. Алматы, Казахстан;
- Коробкин В.В.** - к.геол.-минерал.н., профессор, Школа геологии АО «КБТУ», г. Алматы, Казахстан;
- Канг Ванли** - профессор, Школа нефтяной инженерии, Китайский нефтяной университет (Восточный Китай), г. Циндао, КНР;
- Сарсенбекұлы Б.** - PhD, ассоциированный профессор, Школа энергетики и нефтегазовой индустрии АО «КБТУ», г. Алматы, Казахстан;
- Молдабаева Г.Ж.** - д.тех.н., профессор, кафедра нефтяной инженерии, Институт геологии и нефтегазового дела им. К. Турысова, КазНТУ, г. Алматы, Казахстан.

Раздел «Экономика и бизнес»

- Нарбаев Т.С.** - PhD, профессор, Бизнес-школа АО «КБТУ», г. Алматы, Казахстан;
- Кожаметова А.К.** - PhD, ассистент-профессор, Бизнес-школа АО «КБТУ», г. Алматы, Казахстан;
- Алдашев А.** - PhD, профессор, Международная школа экономики АО «КБТУ», г. Алматы, Казахстан;
- Ахмедов Е.Т.** - PhD, ассоциированный профессор, Бизнес-школа АО «Казахстанско-Британский технический университет», г. Алматы, Казахстан;
- Сейтказиева А.М.** - д.э.н., профессор, Бизнес-школа АО «КБТУ», г. Алматы, Казахстан;
- Gabriel Castelblanco** - PhD, ассистент профессора, Университет Флориды, Флорида, США.
- Дубина И.** - д.э.н., профессор, Новосибирский национальный исследовательский государственный университет (НГУ);
- Osei-Kyei Robert** - PhD, ассоциированный профессор, Университет Западного Сиднея, г. Пенрит, Австралия;
- Montag Josef** - PhD, сениор-лектор, Карлов университет, Прага, Чешская Республика.

МАЗМҰНЫ – СОДЕРЖАНИЕ

**1 МАТЕМАТИКАЛЫҚ ҒЫЛЫМДАР
МАТЕМАТИЧЕСКИЕ НАУКИ**

Сейлова Р.Д., Талипова М.Ж., Бекбауова А.У., Еримбетова Г.Г. Управление квазилинейными интегро-дифференциальными системами с импульсным воздействием	12
Shaikhova G.N. Analytical solutions of the (2+1)-dimensional generalized Benjamin-Ono equation	26
Abdikalikova G.A., Sartabanov Zh.A., Aitenova G.M., Zhumagazyev A.Kh. Investigation of a multiperiodic problem for a finite-hereditary system of integro-differential equations.....	35
Kalmurzaev B.S., Nurlanbek D.D., Bazhenov N.A. On universal numberings of two-element families in the Ershov hierarchy	46
Иманбаев Н.С., Сайрам Н.Н. Локалды емес «серпімді» шеттік шарттармен берілген еселі дифференциалдау Штурм-Лиувилл операторының меншікті және қосарланған функциялар жүйесінің базистілігі туралы	53
Молыбайқызы А., Кабдрахова С.С., Искакова Н.Б., Минглибаева Б.Б. Бөлікті тұрақты аргументті импульсті гиперболалық тендеу үшін шеттік есептің шешімі туралы	61
Yersaliyeva A. Lie polynomials in free special Tortken algebras	75
Исахов А.А., Нугметова А., Абылкасымова А.Б., Аханова Н.Е. Прогнозирование загрязнения воздушной среды от выбросов в городских каньонах с использованием MI-CFD моделирования.....	83
Sinita A.V., Tskhay Yu.A., Shkorko A.K., Karduck A.P. Numerical solution of the inverse coefficient acoustic problem using Laplace transform and differential evolution method.....	108
Григоренко О.В., Кабаева А.М., Логачев А.В., Логачева О.М., Шевчук Е.В. Закон больших чисел для случайного блуждания в случайном пейзаже с некоррелированными слагаемыми	124

**2 КОМПЬЮТЕРЛІК ҒЫЛЫМДАР
КОМПЬЮТЕРНЫЕ НАУКИ**

Tokhmetov A., Serikbayeva S., Tanchenko L., Kenesbay M. AI-based energy forecasting and improved demand management in smart homes.....	133
Abilmazhinova T.M., Kuatbayeva A.A. Hybrid AI Model for HL7 Data Processing and Semantic Interoperability	150
Onayeva A.E. A three-tier information lifecycle management mechanism for hot-tier memory control in containerized Kafka–Flink streaming pipelines	159
Самбетбаева М.А., Серикбаева С.К., Сұлтангазиева А.Н., Мукажанов Н.К., Абдығалым Б.Х. Генетикалық аурулар туралы ақпаратты алу үшін аннотацияланған медициналық мәтіндер корпусын құру	171

Samigulina Z.I., Dyussenkulova B.Z., Butakova D.A. Reinforcement learning–driven control strategies for smart manufacturing systems	187
Рахимова Д., Жігер А., Малых В., Мансурова М., Карюкин В.И. Қазақ тіліндегі медициналық мәтіндерді постредакторлеудің әдістері мен сапалық талдауы.....	205
Kapparova A., Zholamanov B., Bolatbek A., Kuttybay N., Dosymbetova G., Zhumagaliyev Ye. Implementation of multi-agent framework of impala for a single zone temperature control of a simulated thermal zone	218
Akhmetov R.R., Kuatbayeva A.A. Generative ai for fintech.....	233
Azizov T., Abdiakhmetova Z. An smote-based method for enhancing credit card fraud detection	242
Bektemyssova G., Sabdenov A., Satybaldiyeva R., Bykov A., Nor'ashikin Binti Ali Optimizing syntactic-semantic relation extraction for the kazakh language with transformer architectures and synthetic corpora	250
Сербин В.В., Абдуллаева А.С., Абылханова А.А., Батырхан Ж.Ғ., Өтемис Ғ.А. Арнайы білім беру қажеттіліктері бар балаларға арналған қашықтықтан оқытуға қауіпсіз онлайн-платформаны жобалау.....	262
Кемелбеков Н.К., Рахимова Д.Р., Adali E. Service Mesh: архитектурные принципы, производительность и практика применения в микросервисных системах	276
Kulyal M., Dr. Umang, Dr. Saxena P., Dr. Pant J. Machine Learning-Driven Paddy Yield Prediction: Comparative Evaluation of Baseline vs. Ensemble Models in Udham Singh Nagar, Uttarakhand	290
Beishekeyev A., Umarov T. Formal verification of decision tree fairness and robustness via smt solver.....	312
Самбетбаева М., Еримбетова А., Абдығалым Б., Дайырбаева Э., Тұрғанбаев А. Онтологиялық және корпус деректерін біріктірген тілдік модель.....	326
Samigulina G.A., Samigulina Z.I., Bekeshev D.D., Kucherbaeva S.Z., Yarmukhamedova A.B. Artificial intelligence: a new bio-inspired approach in engineering education	340
Молдагулова А.Н., Касенхан А.М., Утебаев К.А., Изтурган Н.Б. Исследование моделей машинного обучения для прогнозирования навыков критического мышления учащихся.....	353

3 ФИЗИКАЛЫҚ ҒЫЛЫМДАР ФИЗИЧЕСКИЕ НАУКИ

Utegenov A.U., Ussenkhan S.S., Ussenov Ye.A., Orazbayev S.A. Automation of nanoparticle synthesis processes in a plasma environment using LabVIEW.....	365
Kanatov Zh.S., Kalkozova Zh.K., Mukash Zh.O., Mirzaeian M., Abdullin Kh.A. Surface transformation and enhanced electrochemical performance of NI3S2/NI foam electrodes for hybrid supercapacitors.....	381
Rakhymbay L., Myrzakhmetov B., Dauletbay A., Konarov A. Optimizing ZnSO ₄ -Li ₂ SO ₄ electrolytes for stable aqueous zinc-ion batteries	392
Aryngazin A., Kurbanova B., Sizyuk Y., Temirbayeva A., Alsar Zh., Insepov Z. Neutronics and thermal-hydraulics of pressurized water reactor with thorium-based fuel	401

4 МҰНАЙ ГАЗ ИНЖЕНЕРИЯСЫ ЖӘНЕ ГЕОЛОГИЯ НЕФТЕГАЗОВАЯ ИНЖЕНЕРИЯ И ГЕОЛОГИЯ

Abiken B.I., Stokes O., Gambal M., Tulemissova Zh.S.

Blast movement monitoring in grade control
of zhairem minesite considering ore losses and dilution 418

Saduov A.

Unsupervised Delineation of Prospectivity Zones
for Stratiform Cu-Co in the Southern Copperbelt Margin (Zambia) 435

**Сулейменова Р.Т., Мукамбеткалиева А.Н., Жанай М.А., Құлмағанбет Қ.,
Жилисбаев Г.Б., Амиров С.С.**

Z кен орны мысалында мұнай қабаттарындағы қысымның таралуын
және деформациялық сипаттамалардың әсерін зерттеу 451

Nygmanova A.S., Korobkin V.V., Buslov M.M., Chaklikov A.Ye.

Tulemissova Zh.S., Koishibayev S.M.

Geological and mineralogical controls on iron skarn formation in the Northwestern
Balkhash region (Central Kazakhstan) 464

**Umarbekova Z.T., Gadeev R., Togizov K., Mashrapova M.A.,
Miniskul Sh.D., Amanbaev R.A.**

Phase forms of gold and their influence on recovery
in the oxidation zone of the arkharly gold-silver deposit (Kazakhstan) 481

Nygmanova A.S., Korobkin V.V., Buslov M.M., Chaklikov A.Ye.

Geological characteristics and environmental aspects
of the development of the Zhuantobe skarn iron ore deposit 494

5 ЭКОНОМИКА ЖӘНЕ БИЗНЕС ЭКОНОМИКА И БИЗНЕС

Aralbayeva Zh.M., Bakdaulet A.

Analyzing Critical Success Factors and Risks in Project
Management Practices in Developing Countries 504

Ryskulbekova M.M.

Project-level determinants of total investment in greenfield infrastructure projects 524

Israzhdinova T.S.

Drivers and Barriers to ESG Implementation in Organisational and Project-Related
Management Practices in Emerging Countries 534

Nurman A.D., Tarkhanov P.K.

Turning diversity into performance: how cohesion drives Kazakhstani project teams 551

Ahmed Sait S., Vijesh P.V.

Next-Gen Sustainability: Unveiling the Financial Intentions, Challenges,
and Transformative Power of Generation Z Investors 560

CONTENTS

1 MATHEMATICAL SCIENCES

Seilova R.D., Talipova M.Zh., Bekbauova A.U., Yerimbetova G.G. Control of quasilinear integro-differential systems with impulse effects	12
Shaikhova G.N. Analytical solutions of the (2+1)-dimensional generalized Benjamin-Ono equation	26
Abdikalikova G.A., Sartabanov Zh.A., Aitenova G.M., Zhumagazyev A.Kh. Investigation of a multiperiodic problem for a finite-hereditary system of integro-differential equations.....	35
Kalmurzayev B.S., Nurlanbek D.D., Bazhenov N.A. On universal numberings of two-element families in the Ershov hierarchy	46
Imanbaev N.S., Sairam N.N. Basis property of systems of eigen and associated functions of a shturm-liouville multiple differentiation operator with nonlocal «perturbed» boundary conditions.....	53
Molybaikyzy A., Kabdrakhova S.S., Iskakova N.B., Minglibayeva B.B. On the solution of a boundary value problem for an impulsive hyperbolic equation with a piecewise-constant argument	61
Yersaliyeva A. Lie polynomials in free special Tortken algebras	75
Issakhov A.A., Nygmetova A., Abylkassymova A.B., Akhanova N.Ye. Forecasting air pollution from emissions in urban canyons using ML-CFD modeling	83
Sinita A.V., Tskhay Yu.A., Shkorko A.K., Karduck A.P. Numerical solution of the inverse coefficient acoustic problem using Laplace transform and differential evolution method	108
Grigorenko O.V., Kabaeva A.M., Logachov A.V., Logachova O.M., Shevchuk E.V. The law of large numbers for random walks in random scenery with uncorrelated terms	124

2 COMPUTER SCIENCE

Tokhmetov A., Serikbayeva S., Tanchenko L., Kenesbay M. AI-based energy forecasting and improved demand management in smart homes.....	133
Abilmazhinova T.M., Kuatbayeva A.A. Hybrid AI Model for HL7 Data Processing and Semantic Interoperability	150
Onayeva A.E. A three-tier information lifecycle management mechanism for hot-tier memory control in containerized Kafka–Flink streaming pipelines	159
Sambetbayeva M., Serikbaeyva S., Sultangazyeva A., Mukazhanov N., Abdy-galym B. Construction of an Annotated Medical Text Corpus for Information Extraction on Genetic Diseases.....	171
Samigulina Z.I., Dyussenkulova B.Z., Butakova D.A. Reinforcement learning–driven control strategies for smart manufacturing systems	187
Rakhimova D., Zhiger A., Malykh V., Mansurova M., Karyukin V. Methods of post-editing medical texts in the kazakh language and qualitative analysis.....	205

Kapparova A., Zholamanov B., Bolatbek A., Kuttybay N., Dosymbetova G., Zhumagaliyev Ye.	
Implementation of multi-agent framework of impala for a single zone temperature control of a simulated thermal zone	218
Akhmetov R.R., Kuatbayeva A.A.	
Generative ai for fintech.....	233
Azizov T., Abdiakhmetova Z.	
An smote-based method for enhancing credit card fraud detection	242
Bektemyssova G., Sabdenov A., Satybaldiyeva R., Bykov A., Nor'ashikin Binti Ali	
Optimizing syntactic-semantic relation extraction for the kazakh language with transformer architectures and synthetic corpora	250
Serbin V.V., Abdullayeva A.S., Abylkhanova A.A., Batyrkhan Zh.G., Utemis G.A.	
Development of a Secure Online Distance Learning Platform for Children with Special Educational Needs (SEN).....	262
Kemelbekov N.K., Rakhimova D.R., Adali E.	
Service Mesh: Architectural Principles, Performance and Practical Use in Microservice Systems	288
Kulyal M., Dr. Umang, Dr. Saxena P., Dr. Pant J.	
Machine Learning-Driven Paddy Yield Prediction: Comparative Evaluation of Baseline vs. Ensemble Models in Udham Singh Nagar, Uttarakhand	290
Beishekeyev A., Umarov T.	
Formal verification of decision tree fairness and robustness via smt solver.....	312
Sambetbayeva M., Yerimbetova A., Abdygalym B., Daiyrbayeva E., Turganbayev A.	
A Language Model Integrating Ontological and Corpus-Based Data	326
Samigulina G.A., Samigulina Z.I., Bekeshev D.D., Kucherbaeva S.Z., Yarmukhamedova A.B.	
Artificial intelligence: a new bio-inspired approach in engineering education	340
Moldagulova A.N., Kassenkhan A.M., Utebaev K.A., Izturgan N.B.	
Research of machine learning models for predicting students' critical thinking skills	353

3 PHYSICAL SCIENCES

Utegenov A.U., Ussenkhan S.S., Ussenov Ye.A., Orazbayev S.A.	
Automation of nanoparticle synthesis processes in a plasma environment using LabVIEW.....	365
Kanatov Zh.S., Kalkozova Zh.K., Mukash Zh.O., Mirzaeian M., Abdullin Kh.A.	
Surface transformation and enhanced electrochemical performance of NI ₃ S ₂ /NI foam electrodes for hybrid supercapacitors	381
Rakhymbay L., Myrzakhmetov B., Dauletbay A., Konarov A.	
Optimizing ZnSO ₄ -Li ₂ SO ₄ electrolytes for stable aqueous zinc-ion batteries	392
Aryngazin A., Kurbanova B., Sizyuk Y., Temirbayeva A., Alsar Zh., Insepov Z.	
Neutronics and Thermal-Hydraulics Of Pressurized Water Reactor With Thorium-Based Fuel	401

4 OIL AND GAS ENGINEERING, GEOLOGY

Abiken B.I., Stokes O., Gambal M., Tulemissova Zh.S.	
Blast movement monitoring in grade control of zhairesm minesite considering ore losses and dilution	418
Saduov A.	
Unsupervised Delineation of Prospectivity Zones for Stratiform Cu-Co in the Southern Copperbelt Margin (Zambia).....	435

Suleimenova R.T., Mukambetkaliyeva A.N., Zhanai M.A., Kulmaganbet K., Zhilisbayev G.B., Amirov S.S.

Study of pressure distribution in oil reservoirs and the influence of deformation characteristics using the example of field Z.....451

Nygmanova A.S., Korobkin V.V., Buslov M.M., Chaklikov A.Ye.

Tulemissova Zh.S., Koishibayev S.M.

Geological and mineralogical controls on iron skarn formation in the Northwestern Balkhash region (Central Kazakhstan) 464

Umarbekova Z.T., Gadeev R., Togizov K., Mashrapova M.A.,

Miniskul Sh.D., Amanbaev R.A.

Phase forms of gold and their influence on recovery in the oxidation zone of the arkharly gold-silver deposit (Kazakhstan) 481

Nygmanova A.S., Korobkin V.V., Buslov M.M., Chaklikov A.Ye.

Geological characteristics and environmental aspects of the development of the Zhuantobe skarn iron ore deposit.....494

5 ECONOMY AND BUSINESS

Aralbayeva Zh.M., Bakdaulet A.

Analyzing Critical Success Factors and Risks in Project Management Practices in Developing Countries 504

Ryskulbekova M.M.

Project-level determinants of total investment in greenfield infrastructure projects 524

Israzhdinova T.S.

Drivers and Barriers to ESG Implementation in Organisational and Project-Related Management Practices in Emerging Countries 534

Nurman A.D., Tarkhanov P.K.

Turning diversity into performance: how cohesion drives Kazakhstani project teams 551

Ahmed Sait S., Vijesh P.V.

Next-Gen Sustainability: Unveiling the Financial Intentions, Challenges, and Transformative Power of Generation Z Investors 560

**МАТЕМАТИКАЛЫҚ ҒЫЛЫМДАР
MATHEMATICAL SCIENCES
МАТЕМАТИЧЕСКИЕ НАУКИ**

УДК 517.938:519.7
МРНТИ 27.29.21, 28.15.15

<https://doi.org/10.55452/1998-6688-2026-23-2-12-25>

¹Сейлова Р.Д.,

к.ф.-м.н., доцент, ORCID ID: 0000-0002-4666-8001,
e-mail: roza_seilova@mail.ru

¹Талипова М.Ж.,

к.ф.-м.н., доцент, ORCID ID: 0000-0001-9728-8378,
e-mail: mira_talipova@mail.ru

¹Бекбауова А.У.,

к.ф.-м.н., доцент, ORCID ID: 0000-0002-5847-9881,
e-mail: mirra478@mail.ru

^{1*}Еримбетова Г.Г.,

докторант, ORCID ID: 0000-0002-7352-5997
*e-mail: gulmira.yerimbetova1982@gmail.com;

¹Актюбинский региональный университет им. К. Жубанова
г. Актобе, Казахстан

**УПРАВЛЕНИЕ КВАЗИЛИНЕЙНЫМИ ИНТЕГРО-ДИФФЕРЕНЦИАЛЬНЫМИ
СИСТЕМАМИ С ИМПУЛЬСНЫМ ВОЗДЕЙСТВИЕМ**

Аннотация

Современная теория управления активно расширяет границы исследований, выходя за рамки классических дифференциальных уравнений без особых воздействий. В данной статье рассматривается задача управляемости для квазилинейных интегро-дифференциальных систем с импульсным воздействием. Исследуется влияние слабых нелинейных возмущений, включенных в уравнения, определяющие моменты импульсного воздействия, что приводит к необходимости анализа систем с переменными моментами импульсов. Применяется метод сведения к уравнениям с фиксированными моментами импульсного воздействия, что позволяет использовать традиционные подходы к управлению. Разработаны методы построения разрешающего управления, обеспечивающего перевод системы из начального состояния в заданное конечное. Рассмотрены условия существования и единственности решений, а также оптимизация управления в среднем, направленная на минимизацию заданного функционала затрат. Полученные результаты могут быть применены в задачах автоматического регулирования, моделирования динамических систем с дискретными и непрерывными воздействиями, а также в управлении технологическими и экономическими процессами. Работа будет интересна для специалистов в области дифференциальных уравнений, теории управления и прикладной математики.

Ключевые слова: управляемость, квазилинейные интегро-дифференциальные уравнения, импульсное воздействие, методы оптимального управления, краевая задача управления, слабые нелинейные возмущения.

Введение

Современная теория управления активно расширяет границы исследований, выходя за рамки классических дифференциальных уравнений без особых воздействий. Динамические системы, развитие которых сопровождается внезапными скачкообразными (импульсными) изменениями, становятся все более востребованными в различных сферах науки и технологий. Такие процессы часто встречаются в механике, электроэнергетике, биологии и экономике, где учет дискретных изменений является критически важным. Интегро-дифференциальные системы обыкновенных дифференциальных уравнений и уравнений в частных производных с импульсными воздействиями и краевые задачи для таких систем, применения в прикладных задачах исследованы многими учеными [1–16].

В этой статье рассматривается задача управления для квазилинейных интегро-дифференциальных уравнений с импульсным воздействием, при котором моменты импульсов могут быть либо заранее заданными, либо определяться текущим состоянием системы. Предлагаемый подход основан на сочетании двух типов управления: классического, описываемого интегрируемой функцией, и импульсного, действующего в дискретные моменты времени.

Материалы и методы

В данной работе исследуется задача управления краевой задачей для квазилинейной системы интегро-дифференциальных уравнений с импульсным воздействием. Особенность рассматриваемого случая заключается в наличии слабых нелинейных возмущений, которые включены не только в правую часть уравнений, но и в уравнения, определяющие моменты импульсного воздействия. Это приводит к необходимости анализа систем с переменными моментами импульсов, что существенно усложняет задачу.

Для решения поставленной задачи используется метод сведения к уравнениям с фиксированными моментами импульсного воздействия, разработанный в работах ученых [17–20]. Данный метод позволяет преобразовать исходную систему к эквивалентному представлению, где моменты импульсных воздействий фиксированы, что существенно упрощает анализ управляемости и построение управляющих воздействий.

Исследование базируется на классических методах математической теории управления, включая:

- ♦ анализ существования и единственности решений для квазилинейных интегро-дифференциальных уравнений с импульсным воздействием;
- ♦ построение краевых условий, необходимых для нахождения разрешающего управления;
- ♦ оптимизацию управления в среднем, основанную на минимизации заданного функционала затрат.

Кроме того, учитывается влияние импульсных воздействий, происходящих в моменты времени, зависящие от состояния системы. В данной ситуации неприменим метод последовательных приближений, так как каждое приближение может содержать различные точки разрыва. В связи с этим используются альтернативные методы приведения системы к представлению с фиксированными моментами импульсного воздействия.

Сопоставление полученных результатов с работами в области краевых задач для импульсных интегро-дифференциальных уравнений, в частности [21, 22], позволяет выделить каче-

ственное развитие математического аппарата. Опираясь на фундамент, заложенный указанными авторами, данное исследование предлагает ряд уточняющих модификаций:

- ♦ если в работах [21, 22] основное внимание уделено системам с фиксированными моментами импульсов $t = \theta_i$, то в настоящей статье рассматривается случай, когда под влиянием слабых нелинейных возмущений $\mu\tau_i(x, \mu)$ моменты скачков $t = \zeta_i$ становятся переменными. Это потребовало введения дополнительных условий (леммы 1 и 2) для обеспечения единственности пересечения траекторией поверхности разрыва;

- ♦ в дополнение к численным подходам и методам параметризации, представленным в [21, 22], в данной работе акцент сделан на аналитическом методе. Использование специальных функционалов позволило свести исходную задачу к эквивалентной системе с фиксированными моментами, что упрощает дальнейший качественный анализ;

- ♦ на основе подходов к обеспечению существования решений, развитых в предшествующих работах, нами предложена структура двухкомпонентного управления (сочетание непрерывного $u(t)$ и дискретного воздействий v_i). Такой подход позволяет не только решать задачу разрешимости, но и оптимизировать функционал затрат;

- ♦ модель была дополнена учетом зависимости скачка $\Delta x(\zeta_i)$ от предыстории системы. Это развитие идей, заложенных в [21, 22], позволяет более детально описывать динамику сложных технологических процессов, где мгновенные изменения состояния обусловлены накопленным эффектом.

Результаты и обсуждения

Рассмотрим систему интегро-дифференциальных уравнений с импульсным воздействием, имеющую вид

$$\begin{aligned} \frac{dx}{dt} &= A(t)x + \int_{\alpha}^t K(t,s)x(s)ds + C(t)u(t) + f(t) + \mu g(t, x, u, \mu), \\ t \neq \theta_i + \mu\tau_i(x, \mu), \Delta x(\zeta_i) &= B_i x(\zeta_i) + \sum_{\alpha < \zeta_j < \zeta_i} D_{ij} x(\zeta_j) + Q_i v_i + I_i + \mu W_i(x(\zeta_i), v_i, \mu), i = \overline{1, p}, \end{aligned} \quad (1)$$

и краевое условие

$$x(\alpha) = a, x(\beta) = b, \quad (2)$$

где $\mu > 0$ – малый параметр,

$x \in R^n$, $\Delta x(\theta_i) \equiv x(\theta_i +) - x(\theta_i)$, $t \in [\alpha, \beta]$, $A(t), K(t, s), i = \overline{1, p}$ – $(n \times n)$ матрицы, столбцы матрицы $A(t)$ – элементы пространства $L_2^n[\alpha, \beta]$, $\{f, I\} \in \Pi^n[\alpha, \beta]$, $D_{ij}, B_i, i, j = \overline{1, p}$ – постоянные $(n \times n)$ матрицы, матрица $K(t, s)$, интегрируемая с квадратом на множестве $[\alpha, \beta] \times [\alpha, \beta]$, матрицы $C(t)$ и $Q_i, i = \overline{1, p}$, размера $n \times m$, m – фиксированное натуральное число, столбцы матрицы $C(t)$ являются функциями из $L_2^n[\alpha, \beta]$, $Q_i, i = \overline{1, p}$, где $L_2^n[\alpha, \beta]$ – пространство функций, интегрируемых с квадратом на отрезке $[\alpha, \beta]$; $D^r[1, p]$ – множество конечных последовательностей $\{\xi_i\}, i = \overline{1, p}$, $(,)$ – скалярное произведение.

Предполагается, что функции $g, W_i, i = \overline{1, p}$ обладают свойством непрерывности по всем аргументам и являются гладкими, имея непрерывные частные производные по каждой координате векторов x, u, v .

Движение, определенное уравнением (1), при фиксированных $\mu, \{u, v\} \in \Pi^m[\alpha, \beta]$ происходит следующим образом. Изображающая точка $P_t(t, x(t))$ решения $x(t) = x(t, \alpha, a)$ уравнения

$$\begin{aligned} \frac{dx}{dt} &= A(t)x(t) + \int_{\alpha}^t K(t,s)x(s)ds + C(t)u(t) + f(t) + \mu g(t, x, u, \mu), \\ t \neq \theta_i + \mu\tau_i(x, \mu) \end{aligned} \quad (3)$$

начинает свое движение в точке (α, a) и продолжает вдоль интегральной кривой этого решения до момента $t = \zeta_1$, когда P_t встречает первую поверхность разрыва $t = \theta_1 + \mu\tau_1(x, \mu)$. Таким образом, значение $t = \zeta_1$ определяется как удовлетворяющее равенству $\zeta_1 = \theta_1 + \mu\tau_1(x(\zeta_1), \mu)$. В этот момент P_t испытывает скачок на величину

$$\Delta x|_{t=\zeta_1} = B_1 x(\zeta_1) + \sum D_{11} x(\zeta_1) + Q_1 v_1 + I_1 + \mu W_1(x(\zeta_1), v_1, \mu)$$

и продолжает двигаться вдоль интегральной кривой решения $x(t, \zeta_1, x(\zeta_1))$ уравнения (3), пока P_t не встретит следующую поверхность разрыва и т.д.

Заметим, что каждое решение уравнения (1) является функцией из РАС $[\alpha, \beta]$, если оно определено на $[\alpha, \beta]$, где $РАС[\alpha, \beta]$ – множество функций $x(t)$ кусочно абсолютно непрерывных, непрерывных слева во всех точках $[\alpha, \beta]$ и испытывающих разрывы первого рода в точках $\{\theta_i\}, i = \overline{1, p}$;

Определение 1. Будем говорить, что разрешима задача управления краевой задачей γ_μ , если для каждого ограниченного множества $G \subset R^n$ существует $\mu_0, \mu_0 \in R^1, \mu_0 > 0$, такое, что для каждых $\{a, b\} \subset G$ и $\mu < \mu_0$ существует управление $\{u, v\} \in \Pi^m[\alpha, \beta]$, при котором уравнение (1) допускает решение, удовлетворяющее краевому условию (2).

Пусть $s \in R, s > 0$ и пусть Π_s есть множество элементов (x, u, v) таких, что $\|x\| + \|u\| + \|v\| \leq s$, где $\|\cdot\|$ есть евклидова норма в $R^k, k \in N$.

Для фиксированного положительного числа $\mu_1 \in R$, определим множество

$$G_s = \{(x, u, v, t, i, \mu)(x, u, v) \in \Pi_s, \alpha \leq t \leq \beta, i = 1, 2, 3, \dots, p, \mu \leq \mu_1\}.$$

Фиксируем $H \in R, H > 0$ и пусть

$$m_1 = \max \{ \sup_t \|A(t)\|, \sup_t \|C(t)\|, \sup_{t,s} \|K(t, s)\|, \max_i \|B_i\|, \max_{i,j} \|D_{ij}\| \},$$

$$m_2 = \max \{ \sup_t \|f(t)\|, \max_i \|I_i\| \},$$

$$m_3 = \max \{ \max_{G_H} \|g\|, \max_{G_H} \|W\|, \max_{G_H} \|\tau\| \}.$$

Лемма 1. Если

$$\mu_1 m_3 < \min(\theta_1 - \alpha, \beta - \theta_p) \quad (4)$$

и для всех x, μ, i и G_H справедливо, что

$$\theta_{i+1} + \mu\tau_{i+1}(x, \mu) < \theta_i + \mu\tau_i(x, \mu), \quad (5)$$

тогда каждое решение уравнения (1), определенное на $[\alpha, \beta]$ и принимающее значения в Π_H пересекает каждую из поверхностей $t = \theta_i + \mu\tau_i(x, \mu), i = 1, \dots, p$, по меньшей мере один раз.

Лемма 2. Пусть система (1) определена в области G_H и $\mu < \mu_2$, где

$$\mu_2 \leq \mu_1 \text{ и } \mu_2 L\gamma(\mu_2) < 1 \quad (6)$$

Тогда, если

$$\tau_i(x, \mu) \geq \tau_i((E + B_i + D_{ii})x + \sum_{j=1}^{i-1} D_{ij} x^{(j)} + Q_i v_i + I_i + \mu W_i(x, v, \mu), \mu) \quad (7)$$

для всех $\|x\| < H, \|x^{(j)}\| < H, j = 1, 2, \dots, i-1$, и $\|v\| < H$, тогда каждое решение $x(t)$ уравнения (1) встречает любую из поверхностей разрыва $t = \theta_i + \mu\tau_i(x, \mu), i = \overline{1, p}$, не больше чем один раз.

Для анализа задачи управления γ_μ применяется метод сравнения [1, 2], суть которого состоит в построении системы дифференциальных уравнений с импульсным воздействием в заранее заданные моменты времени, эквивалентной в определенном смысле исходной системе, где импульсы возникают на поверхностях.

Рассмотрим следующую систему

$$\begin{aligned} \frac{dy}{dt} &= A(t)y + \int_{\alpha}^t K(t,s)y(s)ds + C(t)u + f(t) \sum_{x < \theta_i < t} F_i(t, y, u, \mu), t \neq \theta_i, \\ \Delta y(\theta_i) &= B_i y(\theta_i) + \sum_{\alpha < \theta_j < \theta_i} D_{ij} y(\theta_j) + Q_i v_i + I_i + S_i(y, u, v_i, \mu), \end{aligned} \quad (8)$$

в которой все элементы такие же, как и в системе (1). За исключением F_i и $S_i, i = \overline{1, p}$. Функционалы F_i и S_i будут определены ниже.

Пусть $x(t)$ – решение (1) и $t = \xi_i, i = \overline{1, p}$, точки разрыва этого решения.

В дальнейшем, не нарушая общности, будем считать, что $\tau_i(x, \mu) \geq 0$ для всех $(x, i, \mu) \in G_H$. Таким образом, $\xi_i \geq \theta_i, i = \overline{1, p}$.

Определение 2. Системы (1) и (8) называются В-эквивалентными в G_H , если для некоторого числа $h \in R, 0 < h < H$ и достаточно малого μ справедливо, что:

а) для каждого решения $x(t)$ уравнения (1) $\|x(t)\| < h, t \in [\alpha, \beta]$ существует решение $y(t)$ решения (8), $\|y(t)\| < H$, такое, что $x(t) = y(t)$ для всех $t \in [\alpha, \beta]$, исключая, возможно, точки $t \in [\theta_i, \xi_i], i = \overline{1, p}$;

б) для каждого решения $y(t)$ уравнения (8), $\|y(t)\| < h, t \in [\alpha, \beta]$, существует решение $x(t)$ системы (1), $\|x(t)\| < H$ такое, что $x(t) = y(t)$, для всех $t \in [\alpha, \beta]$, за исключением, возможно, $t \in [\theta_i, \xi_i], i = \overline{1, p}$.

Ниже покажем, что можно подобрать функционалы F_i и S_i так, что (1) и (8) будут В-эквивалентны.

Фиксируем $h, 0 < h < H$ и выберем $\mu_3, \mu_3 > 0$, так, что

$$\mu_3 \leq \mu_2, \mu_3 m_3 \gamma(\mu_3) < H - h.$$

Пусть $\phi(t)$ есть решение (1), удовлетворяющее условию $\phi(\alpha) = a$ и $\|\phi(t)\| < h$, если $t \in [\alpha, \beta]$. Обозначим $t = \xi_i, i = \overline{1, p}$, точки разрыва решения $\phi(t)$.

Фиксируем $K, 1 < k \leq p$, и предположим, что $F_i, S_i, i = 1, 2, \dots, k-1$, уже определены так же, как и решение $\eta(t)$ уравнения (8), которое определено для $t \in [\alpha, \theta_k], \|\eta(t)\| < H, t \in [\alpha, \theta_k]$ и $\eta(t) = \phi(t), t \in [\alpha, \theta_k] \setminus \bigcup_{i=1}^{k-1} (\theta_i, \xi_i)$ и, кроме того, для всех $t \in [\alpha, \theta_k]$,

$$\begin{aligned} \int_{\alpha}^t [A(s)\eta(s) + \int_{\alpha}^s K(s, \sigma)\eta(\sigma)d\sigma + C(s)u + f(s) + \\ \sum_{\alpha < \theta_i < s} F_i(s, \eta, u, \mu)]ds = \int_{\alpha}^t [A(s)\phi(s) + \int_{\alpha}^s K(s, \sigma)\phi(\sigma)d\sigma + C(s)u + f(s) + \\ \mu g(s, \phi(s), u(s), \mu)]ds. \end{aligned} \quad (9)$$

Поскольку $\xi_k \geq \theta_k$, то $\eta(\theta_k) = \xi(\theta_k)$.

Пусть

$$\eta(\theta_k+) = (E + B_k)\eta(\theta_k) + \sum_{\alpha < \theta_j \leq \theta_k} D_{kj} \eta(\theta_j) + Q_k v_k + I_k + z, \quad (10)$$

где z будет определено ниже.

Продолжим решение $\eta(t)$ для всех $t \in [\theta_k, \xi_k]$, как решение следующей начальной задачи Коши.

$$\frac{d\psi}{dt} = F(t), \quad \psi(\xi_k) = \phi(\xi_k +), \quad (11)$$

где

$$F(t) = A(t)\phi(t) + \int_a^t K(t, \sigma)\phi(\sigma)d\sigma + C(t)u + f(t) + \mu g(t, \phi(t), u(t), \mu).$$

Имеем, что $\eta(t) = \phi(\xi_k +) + \int_{\xi_k}^t F(s)ds$, $t \in [\theta_k, \xi_k]$.

Таким образом,

$$\|\eta(t) - \phi(\xi_k +)\| \leq \max_{\theta_k \leq t \leq \xi_k} \left\| \int_{\theta_k}^t F(s)ds \right\| \leq \mu_3 m_3 \gamma(\mu_3) < H - h,$$

и, следовательно, $\|\eta(t)\| < H$, $t \in [\theta_k, \xi_k]$.

Больше того, из (10) и

$$\eta(\theta_k +) = \int_{\xi_k}^{\theta_k} F(s)ds + \phi(\xi_k +)$$

следует, что

$$z = \phi(\xi_k +) + \int_{\xi_k}^{\theta_k} F(s)ds - (E + B_k)\eta(\theta_k) - \sum_{\alpha < \theta_j \leq \theta_k} D_{kj} \eta(\theta_j) - Q_k v_k - I_k.$$

Обозначим значение z как $S_k(\eta, u, v_k, \mu)$. Нетрудно проверить, что

$$S_k(\eta, u, v_k, \mu) = B_k[\phi(\xi_k) - \eta(\theta_k)] + \sum_{\alpha < \xi_j \leq \xi_k} D_{kj} [\phi(\xi_j) - \eta(\theta_k)] + \mu W_k(\phi(\xi_k), v_k, \mu). \quad (12)$$

Пусть также

$$F_k(t, \eta, u, \mu) = \begin{cases} A(t)[\phi(t) - \eta(t)] + \int_a^t K(t, s)[\phi(s) - \eta(s)]ds + \mu g(t, \phi(t), u(t), \mu), & t \in [\theta_k, \xi_k], \\ 0, & \text{если } t \notin [\theta_k, \xi_k]. \end{cases} \quad (13)$$

Если подставить определенные по формулам (12) и (13) функции S_k и F_k в (8), то согласно построению уравнения (1) и (8) В-эквивалентны на промежутке $[\alpha, \theta_{k+1}]$.

Продолжая таким же образом для всех $k+1, k+2, \dots, p$ можно показать, что (1) и (8) В-эквивалентны в G_H .

Система (8) отличается от системы общего вида (1) тем, что моменты импульсного воздействия фиксированы. Так, (8), вообще говоря, относятся к уравнениям вида (1), то для нее тоже можно рассматривать задачу управления краевой задачей (2). Назовем ее краевой задачей λ_μ .

Обозначим

$$\Psi(t) = \int_a^t E(s)E^T(s)ds + \sum_{\alpha < \theta_i < t} P_i P_i^T, \quad (14)$$

где $E(t) = H(\beta, t)C(t)$, $P_i = H(\beta, \theta_i +)Q_i$.

Лемма 3. Пусть выполняются ранее сформулированные условия, а матрица $\Psi(\beta)$ является невырожденной. В таком случае задача управления λ_μ имеет решение, которое представляется в виде предельной функции равномерно сходящейся последовательности, построенной с использованием метода последовательных приближений.

$$\frac{d\psi}{dt} = F(t), \quad \psi(\xi_k) = \phi(\xi_k +), \quad (11)$$

где

$$F(t) = A(t)\phi(t) + \int_a^t K(t, \sigma)\phi(\sigma)d\sigma + C(t)u + f(t) + \mu g(t, \phi(t), u(t), \mu).$$

Имеем, что $\eta(t) = \phi(\xi_k +) + \int_{\xi_k}^t F(s)ds$, $t \in [\theta_k, \xi_k]$.

Таким образом,

$$\|\eta(t) - \phi(\xi_k +)\| \leq \max_{\theta_k \leq t \leq \xi_k} \left\| \int_{\theta_k}^t F(s)ds \right\| \leq \mu_3 m_3 \gamma(\mu_3) < H - h,$$

и, следовательно, $\|\eta(t)\| < H$, $t \in [\theta_k, \xi_k]$.

Больше того, из (10) и

$$\eta(\theta_k +) = \int_{\xi_k}^{\theta_k} F(s)ds + \phi(\xi_k +)$$

следует, что

$$z = \phi(\xi_k +) + \int_{\xi_k}^{\theta_k} F(s)ds - (E + B_k)\eta(\theta_k) - \sum_{\alpha < \theta_j \leq \theta_k} D_{kj} \eta(\theta_j) - Q_k v_k - I_k.$$

Обозначим значение z как $S_k(\eta, u, v_k, \mu)$. Нетрудно проверить, что

$$S_k(\eta, u, v_k, \mu) = B_k[\phi(\xi_k) - \eta(\theta_k)] + \sum_{\alpha < \xi_j \leq \xi_k} D_{kj} [\phi(\xi_j) - \eta(\theta_k)] + \mu W_k(\phi(\xi_k), v_k, \mu). \quad (12)$$

Пусть также

$$F_k(t, \eta, u, \mu) = \begin{cases} A(t)[\phi(t) - \eta(t)] + \int_a^t K(t, s)[\phi(s) - \eta(s)]ds + \mu g(t, \phi(t), u(t), \mu), & t \in [\theta_k, \xi_k], \\ 0, & \text{если } t \notin [\theta_k, \xi_k]. \end{cases} \quad (13)$$

Если подставить определенные по формулам (12) и (13) функции S_k и F_k в (8), то согласно построению уравнения (1) и (8) В-эквивалентны на промежутке $[\alpha, \theta_{k+1}]$.

Продолжая таким же образом для всех $k+1, k+2, \dots, p$ можно показать, что (1) и (8) В-эквивалентны в G_H .

Система (8) отличается от системы общего вида (1) тем, что моменты импульсного воздействия фиксированы. Так, (8), вообще говоря, относятся к уравнениям вида (1), то для нее тоже можно рассматривать задачу управления краевой задачей (2). Назовем ее краевой задачей λ_μ .

Обозначим

$$\Psi(t) = \int_a^t E(s)E^T(s)ds + \sum_{\alpha < \theta_i < t} P_i P_i^T, \quad (14)$$

где $E(t) = H(\beta, t)C(t)$, $P_i = H(\beta, \theta_i +)Q_i$.

Лемма 3. Пусть выполняются ранее сформулированные условия, а матрица $\Psi(\beta)$ является невырожденной. В таком случае задача управления λ_μ имеет решение, которое представляется в виде предельной функции равномерно сходящейся последовательности, построенной с использованием метода последовательных приближений.

Доказательство. Докажем существование решения для задачи λ_μ . В качестве разрешающего управления выступает пара $\{u, v\}$, имеющая следующий вид:

$$u(t) = E^T(t)c + \hat{u}(t), \quad t \in [\alpha, \beta], \quad v_i = P_i^T c + \hat{v}_i, \quad i = 1, 2, \dots, p, \quad (15)$$

где $c \in R^n$ постоянный вектор, и $\{\hat{u}, \hat{v}\}$ из $\Pi^m[\alpha, \beta]$ и ортогонально всем столбцам матрицы $[E, P_i^T]$.

Очевидно, что задача λ_μ сводится к нахождению решения следующего уравнения, эквивалентного исходной задаче.

$$y(t) = H(t, \alpha)a + \int_{\alpha}^t H(t, s) [C(s)u(s) + f(s) + \sum_{\alpha < \theta_i < s} F_i(s, y, u, \mu)] ds + \\ + \sum_{\alpha < \theta_i < t} H(t, \theta_i +) [Q_i v_i + I_i + S_i(y, u, v_i, \mu)], \quad x(\beta) = b \quad (16)$$

Подставив (15) в (16), найдем, что

$$c = \Psi^{-1}(\beta) \{ b - H(\beta, \alpha)a - \int_{\alpha}^{\beta} H(\beta, s) [f(s) + \sum_{\alpha < \theta_i < s} F_i(s, y, u, \mu)] ds - \\ - \sum_{i=1}^p H(\beta, \theta_i +) [I_i + S_i(y, u, v_i, \mu)] \}. \quad (17)$$

Обозначим

$$K = b - H(\beta, \alpha)a - \int_{\alpha}^{\beta} H(\beta, s) f(s) ds - \sum_{i=1}^p H(\beta, \theta_i +) I_i \\ u_0(t) = E^T(t) \Psi^{-1}(\beta) K + \hat{u}(t), \quad v_i^0 = P_i^T \Psi^{-1}(\beta) K + \hat{v}_i, \\ y_0(t) = H(t, \alpha)a + \int_{\alpha}^t H(t, s) [C(s)u_0(s) + f(s)] ds + \sum_{\alpha < \theta_i < t} H(t, \theta_i +) [Q_i v_i^0 + I_i], \\ \phi = (y(t), u(t), v_i), \quad \phi_0 = (y_0(t), u_0(t), v_i^0), \\ \varkappa(t, \phi, \mu) = \mu^{-1} \int_{\alpha}^t H(t, s) \sum_{\alpha < \theta_i < s} F_i(s, y, u, \mu) ds, \\ \psi(t, \phi, \mu) = \mu^{-1} \sum_{\alpha < \theta_i < t} H(t, \theta_i +) S_i(y, u, v_i, \mu).$$

Из (16) и (17) следует, что

$$\phi = \phi_0 + \mu E(\phi, \mu), \quad (18)$$

где $E = (E_1, E_2, E^i)$,

$$E_1(t, \phi) = \varkappa(t, \phi, \mu) + \psi(t, \phi, \mu) - \Psi(t) \Psi^{-1}(\beta) [\varkappa(\beta, \phi, \mu) + \psi(\beta, \phi, \mu)],$$

$$E_2(t, \phi) = E(t)^T \Psi^{-1}(\beta) [\varkappa(\beta, \phi, \mu) + \psi(\beta, \phi, \mu)],$$

$$E^i(t, \phi) = E_i^T \Psi^{-1}(\beta) [\varkappa(\beta, \phi, \mu) + \psi(\beta, \phi, \mu)]$$

Рассмотрим пространство B функций ϕ , имеющих вид $\phi = (y(t), u(t), v_i)$ с нормой

$$\|\phi\| = \max_t |y(t)| + \max_t |u(t)| + \max_t |v_i|, \quad \}$$

где $y(t) \in \text{PAC}[\alpha, \beta]$ и $\{u, v\} \in \Pi^m[\alpha, \beta]$.

Предположим теперь, что величина h , определенная ранее, также удовлетворяет условию $\|\phi_0\| < h$ и существует число $\mu_4 \in R$, для которого выполняется неравенство $\mu_4 \leq \mu_3$, такое, что $\mu_4 \max_{\substack{\|\phi\| \leq h \\ 0 < \mu \leq \mu_1}} \|\varepsilon(\phi, \mu)\| < h - \|\phi_0\|$.

Пусть $\chi = \{\phi: \phi \in B, \|\phi\| < h\}$.

Легко убедиться, что при $\mu \leq \mu_4$ оператор $\phi_0 + \mu E(\phi, \mu)$ отображает пространство χ в себя. Далее докажем, что при достаточно малом значении μ оператор μE является сжимающим на χ . Для этого сначала покажем, что функции $\pi_j(t, \phi, \mu)$ и $\psi(t, \phi, \mu)$ удовлетворяют условию Липшица по переменной ϕ . Введем обозначения $\phi_j = (\eta_j, u_j, v_i^{(j)}) \in \chi$, $j = 1, 2$. Фиксируем k , $1 < k \leq p$ и пусть $\pi_j, j = 1, 2$, функции, определенные на $[\alpha, \theta_k]$ такие, что $\pi_j(t) = \eta_j(t)$, если $t \in [\alpha, \theta_k] \setminus \bigcup_{i=1}^{k-1} (\theta_i, \zeta_k^{(j)})$, где $\zeta_k^{(j)}$ – точка разрыва функции π_j и

$$\begin{aligned} & \int_{\alpha}^t [A(s)\pi_j(s) + \int_{\alpha}^s K(s, \sigma)\pi_j(\sigma)d\sigma + \mu g(s, \pi_j(s), u_j(s), \mu)]ds = \\ & = \int_{\alpha}^t [A(s)\eta_j(s) + \int_{\alpha}^s K(s, \sigma)\eta_j(\sigma)d\sigma + \sum_{\alpha < \theta_i < s} F_i(s, \eta_j, u_j, \mu)]ds \end{aligned} \quad (19)$$

для всех $t \in [\alpha, \theta_k]$.

Предположим, что

$$\|\pi_1(t) - \pi_2(t)\| \leq l_1(\mu)\|\phi_1 - \phi_2\| \text{ для всех } t \in [\alpha, \theta_k] \setminus \bigcup_{i=1}^{k-1} (\zeta_i^{(1)}, \zeta_i^{(2)}) \quad (20)$$

и

$$\zeta_i^{(2)} - \zeta_i^{(1)} \leq \mu l_2(\mu)\|\phi_1 - \phi_2\| \quad (21)$$

где $l_1(\mu)$ и $l_2(\mu)$ – ограниченные функции и без ограничения общности $\zeta_i^{(2)} \geq \zeta_i^{(1)}$ $i = \overline{1, p}$.

Теперь продолжим функцию $\pi_j(t)$, рассматривая ее как решение следующей системы

$$\frac{d\pi}{dt} = A(t)\pi(t) + \int_{\alpha}^t K(t, s)\pi(s)ds + C(t)u_j(t) + f(t) + \mu g(t, \pi, u_j, \mu),$$

с начальным условием $\pi(\theta_k) = \eta_j(\theta_k)$, пока это решение не пересечет поверхность $t = \theta_k + \mu\tau_k(x, \mu)$ в момент $t = \zeta_k^{(j)}$.

Очевидно, что для $t \in [\theta_k, \zeta_k^{(j)}]$,

$$\begin{aligned} \pi_j(t) = \eta_j(\theta_k) + \int_{\theta_k}^t [A(s)\pi_j(s) + \int_{\alpha}^s K(t, \sigma)\pi_j(\sigma)d\sigma + C(t)u_j(t) + \\ + f(t) + \mu g(t, \pi_j(s), u_j(s), \mu)]ds. \end{aligned} \quad (22)$$

Предположим, что $|\pi_j(t)| < H$, $t \in [\alpha, \theta_k]$. Покажем, что это неравенство верно также для $t \in [\theta_k, \zeta_k^{(j)}]$. Допустим, что утверждение неверно, и пусть $t = t^*$ – это первая точка, в которой неравенство перестает выполняться, то есть выполняется условие $|\pi_j(t^*)| \geq H$.

Но тогда, используя (22) для $t \in [\theta_k, t^*]$, имеем, что $|\pi_j(t)| \leq h + \mu_3 m_3 \gamma(\mu_3) < H$, и в частности $|\pi_j(t^*)| < H$. Следовательно, полученное противоречие подтверждает справедливость данного утверждения.

Теперь, используя соотношения (20)–(22), приходим к следующему результату.

$$\|\pi_1(t) - \pi_2(t)\| \leq l_3(\mu)\|\phi_1 - \phi_2\| \text{ для } t \in [\theta_k, \zeta_k^{(1)}], \quad (23)$$

где

$$l_3(\mu) = (1 + l_1(\mu)(\beta - \alpha) + pl_2(\mu)\gamma(\mu) + m_1 + \mu L)(1 - \mu t_3(m_1(1 + \mu t_3) + \mu L))^{-1}.$$

Кроме того, используя

$$\begin{aligned}\zeta_k^{(2)} - \zeta_k^{(1)} &= \mu \tau_k(\pi_2(\zeta_k^{(2)}), \mu) - \mu \tau_k(\pi_1(\zeta_k^{(1)}), \mu) \leq \\ &\leq \mu L |\pi_2(\zeta_k^{(2)}) - \pi_2(\zeta_k^{(1)}) - \pi_2(\zeta_k^{(1)}) + \pi_1(\zeta_k^{(1)})|,\end{aligned}$$

найдем, что

$$\zeta_k^{(2)} - \zeta_k^{(1)} \leq \mu l_4(\mu) \|\phi_1 - \phi_2\| \quad (24)$$

где $l_4(\mu) = \mu L l_3(\mu)(1 - \mu L \gamma(\mu))^{-1}$.

Теперь можно рассмотреть S_k .

$$\begin{aligned}|S_k(\eta_1, u_1, v_k^{(1)}, \mu) - S_k(\eta_2, u_2, v_k^{(2)}, \mu)| &= B_k(\pi_1(\zeta_k^{(1)}) - \eta_1(\theta_k) - \pi_2(\zeta_k^{(1)}) + \eta_2(\theta_k) + \\ &+ B_k(\pi_2(\zeta_k^{(1)}) - \pi_2(\zeta_k^{(2)}) + \sum_{\alpha < \xi_j \leq \xi_k} [D_{kj}(\pi_1(\zeta_j^{(1)}) - \eta_1(\theta_j) - \pi_2(\zeta_j^{(1)}) + \\ &+ \eta_2(\theta_j) + D_{kj}(\pi_2(\zeta_j^{(1)}) - \pi_2(\zeta_j^{(2)}))] + \mu(W_k(\pi_1(\zeta_k^{(1)}), v_k^{(1)}, \mu) - W_k(\pi_2(\zeta_k^{(2)}), v_k^{(2)}, \mu))\end{aligned} \quad (25)$$

Применив (22), получим

$$\begin{aligned}&\pi_1(\zeta_j^{(1)}) - \eta_1(\theta_j) - \pi_2(\zeta_j^{(2)}) + \eta_2(\theta_j) = \\ &= \int_{\theta_k}^{\zeta_k} A(s) [\pi_1(s) - \pi_2(s)] ds + \int_{\alpha}^s K(s, \sigma) [\pi_1(\sigma) - \pi_2(\sigma)] d\sigma + C(s) [u_1(s) - u_2(s)] \\ &\quad + \mu [g(s, \pi_1(s), u_1(s), \mu) - g(s, \pi_2(s), u_2(s), \mu)] ds + \\ &+ \int_{\zeta_k^{(1)}}^{\zeta_k^{(2)}} \{A(s) \pi_2(t) + \int_{\alpha}^s K(t, s) \pi_2(s) ds + C(t) u_2(t) + f(t) + \mu g(t, \pi_2(t), u_2(t), \mu)\} dt.\end{aligned}$$

Поэтому, используя (23) и (24), мы найдем

$$|\pi_1(\zeta_j^{(1)}) - \eta_1(\theta_j) - \pi_2(\zeta_j^{(2)}) + \eta_2(\theta_j)| \leq \mu l_5(\mu) \|\phi_1 - \phi_2\| \quad (26)$$

где

$$l_5(\mu) = m_3(m_1(l_3(\mu) + l_1(\mu)(\beta - \alpha) + 1) + p l_3(\mu) \gamma(\mu) + \mu L) + l_4(\mu) \gamma(\mu).$$

Из (23) и (26) следует, что

$$|S_k(\eta_1, u_1, v_k^{(1)}, \mu) - S_k(\eta_2, u_2, v_k^{(2)}, \mu)| \leq \mu l^{(1)}(\mu) \|\phi_1 - \phi_2\|,$$

где $l^{(1)}(\mu) = m_1(p + 1)(l_5(\mu) + l_3(\mu) \mu l_4(\mu) \gamma(\mu))$ ограниченная функция. Так как есть конечное число S_j , то существует $L_1(\mu)$ такое, что

$$|S_i(\eta_1, u_1, v_i^{(1)}, \mu) - S_i(\eta_2, u_2, v_i^{(2)}, \mu)| \leq \mu L^{(1)}(\mu) \|\phi_1 - \phi_2\| \quad (27)$$

для $i = 1, 2, \dots, p$.

Аналогично для F_i , используя (14), можно показать, что существует функция $L^{(2)}(\mu)$, ограниченная, и такая, что

$$|\sum_{\alpha < \theta_i < t} F_i(t, \eta_2, u_2, \mu) - \sum_{\alpha < \theta_i < t} F_i(t, \eta_1, u_1, \mu)| \leq \mu L^{(2)}(\mu) \|\phi_1 - \phi_2\| \quad (28)$$

равномерно для $t \in [\alpha, \beta]$.

Пусть $m_4 = \max_{t,s} |H(t, s)|$. Используя (27) и (28) получим, что

$$|\psi(t, \phi_1, \mu) - \psi(t, \phi_2, \mu)| \leq p m_4 L^{(1)}(\mu) \|\phi_1 - \phi_2\|$$

и

$$|\alpha(t, \phi_1, \mu) - \alpha(t, \phi_2, \mu)| \leq (\beta - \alpha) m_4 L^{(2)}(\mu) \|\phi_1 - \phi_2\|.$$

Таким образом,

$$|\alpha(t, \phi_1, \mu) - \alpha(t, \phi_2, \mu)| + |\psi(t, \phi_1, \mu) - \psi(t, \phi_2, \mu)| \leq \mu L(\mu) \|\phi_1 - \phi_2\|,$$

где $L(\mu) = (\beta - \alpha)L^{(2)}(\mu) + p m_4 L^{(1)}(\mu)$.

Наконец, обозначая

$$m_5 = \max\{\max_t \|\Psi(t)\Psi^{-1}(\beta)\|, \max_t \|E^T(t)\Psi^{-1}(\beta)\|, \max_i \|P_i^T \Psi^{-1}(\beta)\|\},$$

можно получить, что

$$\|E(\phi_1, \mu) - E(\phi_2, \mu)\| \leq 2L(\mu)(1 + 6m_5)\|\phi_1 - \phi_2\|.$$

Следовательно, если $\mu_0 \leq \mu_4$ и μ_0 достаточно мало, так что для всех $\mu < \mu_0$ выполняется неравенство $2\mu L(\mu)(1 + 6m_5) < 1$, то оператор μE является сжимающим.

Таким образом, можно заключить, что оператор $\phi_0 + \mu E(\phi, \mu)$ имеет неподвижную точку ϕ^0 .

Нетрудно проверить, что ϕ^0 есть решение задачи λ_μ . Лемма доказана.

Теорема 1. Пусть для системы (1) выполнены все перечисленные выше условия и матрица $\Psi(\beta)$ невырожденная. Тогда задача управления γ_μ разрешима.

Доказательство. Согласно лемме 3 существует управление $\{\hat{u}, \hat{v}\}$, при котором система (8) имеет решение $y(t)$, такое, что $y(\alpha) = a, y(\beta) = b$. Так как (1) и (8) В-эквивалентны, то уравнение (1) с управлением $\{\hat{u}, \hat{v}\}$ допускает решение $x(t)$ такое, что $x(t) = y(t)$ для всех t , исключая $t \in [\theta_i, \zeta_i], i = \overline{1, p}$. Так как α и β не включены в $[\theta_i, \zeta_i], i = \overline{1, p}$, то $x(\alpha) = x(\beta)$. Значит, краевое условие (2) выполнено и теорема доказана.

Заключение

В данной работе исследована задача управляемости для квазилинейных интегро-дифференциальных систем с импульсным воздействием. Основное внимание уделено случаям, когда моменты импульсов зависят от состояния системы, что приводит к необходимости разработки новых аналитических методов. Показано, что применение метода сведения к уравнениям с фиксированными моментами импульсного воздействия позволяет существенно упростить анализ системы и сформулировать условия управляемости. Разработан метод построения разрешающего управления, учитывающий как непрерывные, так и импульсные компоненты воздействия. Продемонстрировано, что предложенные алгоритмы позволяют не только гарантировать достижение требуемого состояния системы, но и минимизировать функционал затрат, что обеспечивает оптимальность управления в среднем.

Перспективным направлением дальнейших исследований является разработка более общих методов управления для нелинейных систем с произвольной структурой импульсов, а также учет неопределенностей и стохастических факторов, влияющих на динамику процессов.

ЛИТЕРАТУРА

- 1 Abbasov, E.M., Agaeva, N.A., and Imamaliyev, S.A. Modeling of hydrodynamics of liquid motion in complex profile pipeline. *Journal of Engineering Thermophysics*, 29 (3), 2020.
- 2 Kler, A., Apanovich, D., and Maximov, A. An effective method for calculating the elements of thermal power plants, which are reduced to solving systems of partial differential equations. *E3S Web of Conferences*, 209, 03029 (2020). <https://doi.org/10.1051/e3sconf/202020903029>
- 3 Rodríguez, F., López, J.C.C., and Castro, M.A. *Models of Delay Differential Equations* (Basel: MDPI, 2021).
- 4 Tinyukova, T.S., and Chuburin, Yu.P. Majorana states near an impurity in the Kitaev infinite and semi-infinite model. *Theoretical and Mathematical Physics*, 200 (1), 1043–1052 (2019).
- 5 Samoilenko, A.M., and Perestyuk, N.A. *Differentsial'nye uravneniya s impul'snym vozdeistviem* [Differential Equations with Impulsive Action] (Kiev: Vishcha Shkola, 1987), 287 p. (In Russian).
- 6 Yuldashev, T.K., Odinaev, R.N., and Zarifzoda, S.K. On exact solutions of a class of singular partial integro-differential equations. *Lobachevskii Journal of Mathematics*, 42 (3), 676–684 (2021). <https://doi.org/10.1134/S1995080221030240>
- 7 Lakshmikantham, V., Bainov, D.D., and Simeonov, P.S. *Theory of Impulsive Differential Equations* (Singapore: World Scientific, 1989).
- 8 Kolmogorov, A.N., and Fomin, S.V. *Elementy teorii funktsii i funktsional'nogo analiza* [Elements of the Theory of Functions and Functional Analysis] (Moscow: Nauka, 1989), 623 p. (In Russian).
- 9 Akhmetov, M.U. Pochti periodicheskie resheniya integro-differentsial'nykh uravnenii s impul'snym vozdeistviem [Almost periodic solutions of integro-differential equations with impulsive action]. *Matematicheskaya Fizika i Nelineinye Kolebaniya*, 42, 5–6 (1987). (In Russian).
- 10 Lando, Yu.K. Ob upravlyaemykh integro-differentsial'nykh operatorakh [On controllable integro-differential operators]. *Differentsial'nye Uravneniya*, 9 (12), 2227–2230 (1973). (In Russian).
- 11 Aisagaliev, S.A. *Controllability Theory of the Dynamic Systems* (Almaty: Kazakh University, 2014), 158 p.
- 12 Barbu, V., Iannelli, M., and Martcheva, M. On the controllability of the Lotka-McKendrick model of population dynamics. *Journal of Mathematical Analysis and Applications*, 253 (1), 142–165 (2001).
- 13 Rama Mahana Rao, M., Srivastava, K.S., and Sivasundaram, S. Stability of linear delay impulsive differential equations with impulsive effect. *Journal of Mathematical Analysis and Applications*, 163, 47–59 (1992).
- 14 Anokhin, L., Berezansky, and Braverman, E. Stability of linear delay impulsive differential equations. *Dynamic Systems and Applications*, 4, 173–188 (1995).
- 15 Akhmet, M., Dauylbayev, M., and Mirzakulova, A. A singularly perturbed differential equation with piecewise constant argument of generalized type. *Turkish Journal of Mathematics*, 42 (4), 1680–1685 (2018).
- 16 Dzhumabaev, D.S. Solvability of a linear boundary value problem for a Fredholm integro-differential equation with impulsive inputs. *Differential Equations*, 51 (9), 1180–1196 (2015).
- 17 Akhmetov, M.U., and Perestyuk, N.A. O dvizhenii s impul'snym vozdeistviem na poverkhnostyakh [On motion with impulsive action on surfaces]. *Izvestiya AN KazSSR. Seriya fiziko-matematicheskaya*, 1, 11–14 (1988). (In Russian).
- 18 Akhmetov, M.U., and Perestyuk, N.A. O metode sravneniya dlya differentsial'nykh uravnenii s impul'snym vozdeistviem [On the comparison method for differential equations with impulsive action]. *Differentsial'nye Uravneniya*, 26 (9), 1475–1483 (1990). (In Russian).
- 19 Akhmet, M., Tleubergenova, M., Seilova, R., and Nugayeva, Z. Symmetrical impulsive inertial neural networks with unpredictable and Poisson-stable oscillations. *Symmetry*, 15 (10), 1812 (2023). <https://doi.org/10.3390/sym15101812>
- 20 Akhmet, M., Aviltay, N., Dauylbayev, M., and Seilova, R. A case of impulsive singularity. *Journal of Mathematics, Mechanics and Computer Science*, 117 (1) (2023). <https://doi.org/10.26577/JMMCS.2023.v117.i1.0>
- 21 Mynbayeva, S.T., and Tankeyeva, A.K. A computational method for solving a boundary value problem for impulsive integro-differential equation. *International Journal of Mathematics and Physics*, 14 (1), 45–52 (2023).
- 22 Mynbayeva, S.T., Assanova, A.T., and Tankeyeva, A.K. On the solvability of a quasilinear boundary value problem for an impulsive system. *Lobachevskii Journal of Mathematics*, 45 (10), 5146–5155 (2024).

¹Сейлова Р.Д.,

ф.-м.ғ.к., доцент, ORCID ID: 0000-0002-4666-8001,

e-mail: roza_seilova@mail.ru

¹Талипова М.Ж.,

ф.-м.ғ.к., доцент, ORCID ID: 0000-0001-9728-8378,

e-mail: mira_talipova@mail.ru

¹Бекбауова А.У.,

ф.-м.ғ.к., доцент, ORCID ID: 0000-0002-5847-9881,

e-mail: mirra478@mail.ru

^{1*}Еримбетова Г.Г.,

докторант, ORCID ID: 0000-0002-7352-5997,

*e-mail: gulmira.yerimbetova1982@gmail.com

¹К. Жұбанов атындағы Ақтөбе өңірлік университетінің Математика кафедрасы,
Ақтөбе қ., Қазақстан

КВАЗИСЫЗЫҚТЫ ИНТЕГРО-ДИФФЕРЕНЦИАЛЫҚ ЖҮЙЕЛЕРДІ ИМПУЛЬСТІК ӘСЕР ЕТУ ЖАҒДАЙЫНДАҒЫ БАСҚАРУ

Аңдатпа

Мақалада импульстік әсер ету жағдайындағы квазисызықты интегро-дифференциалдық жүйелерді басқару мәселесі қарастырылады. Импульстік әсер ету сәттерін анықтайтын теңдеулерге енгізілген әлсіз сызықты емес ауытқулардың ықпалы зерттеледі, бұл өзгермелі импульстік сәттері бар жүйелерді талдау қажеттілігіне әкеледі. Импульстік әсер ету сәттерін бекітілген теңдеулерге келтіру әдісі қолданылады, бұл классикалық басқару әдістерін пайдалануға мүмкіндік береді. Жүйені бастапқы күйден берілген соңғы күйге ауыстыруды қамтамасыз ететін басқару әдістері әзірленді. Интегро-дифференциалдық теңдеулер шешімдерінің болуы мен бірегейлігі шарттары, сондай-ақ белгілі бір шығын функционалын минимизациялауға бағытталған басқаруды орташа оңтайландыру мәселесі қарастырылды. Алынған нәтижелер автоматты реттеу, дискретті және үздіксіз әсер ететін динамикалық жүйелерді модельдеу, сондай-ақ технологиялық және экономикалық процестерді басқару міндеттерінде қолданылуы мүмкін. Бұл жұмыс дифференциалдық теңдеулер, басқару теориясы және қолданбалы математика салаларындағы мамандар үшін қызықты болады.

Түйін сөздер: басқарылу, квазисызықты интегро-дифференциалдық теңдеулер, импульстік әсер ету, оңтайлы басқару әдістері, шекаралық басқару есебі, әлсіз сызықты емес ауытқулар.

¹Seilova R.D.,

PhD, Associate Professor, ORCID ID 0000-0002-4666-8001,
e-mail: roza_seilova@mail.ru

¹Talipova M.Zh.

PhD, Associate Professor, ORCID ID 0000-0001-9728-8378,
e-mail: mira_talipova@mail.ru

¹Bekbauova A.U.,

PhD, Associate Professor, ORCID ID: 0000-0002-5847-9881,
e-mail: mirra478@mail.ru

^{1*}Yerimbetova G.G.,

PhD student, ORCID ID 0000-0002-7352-5997,
*e-mail: gulmira.yerimbetova1982@gmail.com

¹K. Zhubanov Aktobe Regional University, Aktobe, Kazakhstan

CONTROL OF QUASILINEAR INTEGRO-DIFFERENTIAL SYSTEMS WITH IMPULSE EFFECTS

Abstract

This paper examines the controllability problem of quasilinear integro-differential systems with impulse effects. The influence of weak nonlinear perturbations included in the equations that determine the moments of impulse effects is studied, leading to the necessity of analyzing systems with variable impulse moments. The method of reduction to equations with fixed impulse moments is applied, allowing for the use of classical control approaches. Methods for constructing admissible control that ensures the transition of the system from an initial state to a given final state are developed. The conditions for the existence and uniqueness of solutions, as well as the optimization of control in the mean, aimed at minimizing a given cost functional, are considered. The obtained results can be applied in automatic regulation tasks, modeling of dynamic systems with discrete and continuous influences, as well as in the control of technological and economic processes. This work will be of interest to specialists in differential equations, control theory, and applied mathematics.

Keywords: controllability, quasilinear integro-differential equations, impulse effects, optimal control methods, boundary control problem, weak nonlinear perturbations.

Received February 2, 2026; revised March 2, 2026; accepted March 6, 2026.

UDC 517.957
IRSTI 27.31.15, 27.31.21

<https://doi.org/10.55452/1998-6688-2026-23-2-26-34>

Shaikhova G.N.,

PhD, Associate Professor, ORCID ID: 0000-0002-0819-5338,

e-mail: g.shaikhova@gmail.com

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

ANALYTICAL SOLUTIONS OF THE (2+1)-DIMENSIONAL GENERALIZED BENJAMIN-ONO EQUATION

Abstract

The (2+1)-dimensional generalized Benjamin-Ono equation models the propagation of small-amplitude, long-wavelength waves on the surface of shallow water. Constructing explicit solutions of the (2+1)-dimensional generalized Benjamin-Ono equation not only provides theoretical support for experimental investigations but also offers a rigorous basis for addressing applied problems arising in nonlinear wave dynamics. In this work, we investigate wave propagation governed by the (2+1)-dimensional generalized Benjamin-Ono equation in nonlinear media, accounting for dispersive effects. To this end, the sine-cosine function method and the hyperbolic tangent method are employed as analytical tools for deriving explicit solutions. The methods prove effective for a broad class of nonlinear equations encountered in mathematical physics. Using these approaches, periodic-wave solutions and solitary wave solutions are obtained, and to illustrate the obtained results, we plot 3D and 2D plots by setting suitable values of the involved parameters.

Keywords: generalized Benjamin-Ono equation, sine-cosine method, hyperbolic tangent method, periodic wave solution, solitary wave solution.

Received January 1, 2026; revised March 6, 2026; accepted April 8, 2026.

Introduction

Nonlinear partial differential equations describe a broad spectrum of physical phenomena. Since nonlinear effects drive many processes in nature – from gravitational interactions to the dynamics of fluids and plasmas, there is strong motivation to derive explicit solutions for these equations. Explicit solutions provide reference results for validating numerical schemes, elucidating qualitative behaviors (e.g., wave propagation, stability, and interactions), and clarifying the influence of model parameters [1–5].

To this end, a variety of analytical techniques have been developed for deriving exact solutions of nonlinear partial differential equations, including the Darboux transformation method, the Hirota bilinear method, the first integral method, the Lie symmetry method, the Kudryashov method, the sine-cosine method, and the tanh method [6–16].

Motivated by these applications, the present study investigates the (2+1)-dimensional generalized Benjamin-Ono equation [17, 18], which is expressed as follows

$$u_{xxxx} + c_1 u_{tt} + c_2 u_{xt} + c_3 u_{xy} + c_4 (u^2)_{xx} = 0, \quad (1)$$

where c_1 , c_2 , c_3 , and c_4 are arbitrary constants. The equation (1) was studied by Jacobi elliptic function expansion method by the improved tanh-coth and tan-cot methods [17, 18].

Considered the wave transformation as

$$u(x, y, t) = u(\xi), \text{ where } \xi = ax + by - kt. \quad (2)$$

Taking the derivatives and submit into the Eq. (1) and we get the ODE as

$$a^4 u^{(4)} + u''(k^2 c_1 - c_2 ka + c_3 ab) + 2a^2 c_4 (uu'' + u'^2) = 0. \quad (3)$$

By taking integration of Eq. (3), we have

$$a^4 u'' + u(k^2 c_1 - c_2 ka + c_3 ab) + a^2 c_4 u^2 = 0. \quad (4)$$

Material and methods

To construct new analytical solutions, we applied the sine-cosine method and the hyperbolic tangent method. These approaches yielded several distinct families of explicit solutions.

The sine-cosine method

In this section, the sine-cosine method is described [1, 11–13]. Within this approach, a suitable traveling-wave transformation is introduced

$$u(x, y, t) = u(\xi), \quad \xi = (x + y - ct), \quad (5)$$

the partial differential equation (PDE)

$$E_1(u_t, u_x, u_y, u_{xx}, u_{yy}, u_{xxx}, u_{yyy} \dots) = 0, \quad (6)$$

can be rewritten as an ordinary differential equation (ODE)

$$E_2(u, u', u'', u''', \dots) = 0. \quad (7)$$

In Eq. (6) $u(x, y, t)$ is an unknown function, E_1 is a polynomial in $u(x, y, t)$ and its partial derivatives in which the highest order derivatives and the nonlinear terms are involved. In Eq. (7), $u(\xi)$ denotes the unknown function, and E_2 represents a polynomial expression in $u(\xi)$ and its derivatives, including the nonlinear terms.

Equation (7) is then integrated, noting that all terms are derivative terms; hence, the integration constants are taken to be zero. Consequently, the solutions of ODE (7) may be represented in the form

$$u(x, y, t) = \alpha \sin^\beta(\mu\xi), \quad (8)$$

and cosine solution

$$u(x, y, t) = \alpha \cos^\beta(\mu\xi), \quad (9)$$

where $\xi = x - ct$. The parameters μ , α , and β are to be determined; here μ denotes the wave number and c represents the wave speed [1]. The derivatives of (8) are given by

$$(u(\mu\xi))' = \beta\mu\alpha \sin^{\beta-1}(\mu\xi) \cos(\mu\xi), \quad (10)$$

$$(u(\mu\xi))'' = -\mu^2 \beta^2 \alpha \sin^\beta(\mu\xi) + \mu^2 \alpha \beta (\beta - 1) \sin^{\beta-2}(\mu\xi), \quad (11)$$

and the derivatives of (9) have next forms

$$(u(\mu\xi))' = -\beta\mu\alpha \cos^{\beta-1}(\mu\xi) \sin(\mu\xi), \quad (12)$$

$$(u(\mu\xi))'' = -\mu^2 \beta^2 \alpha \cos^\beta(\mu\xi) + \mu^2 \alpha \beta (\beta - 1) \cos^{\beta-2}(\mu\xi), \quad (13)$$

and so on for the other derivatives.

Substituting Eqs. (8)–(13) into the reduced ODE (7) yields a trigonometric identity involving powers of $\cos^r(\mu\xi)$ and/or $\sin^r(\mu\xi)$. The unknown parameters are then obtained as follows: first, we balance the exponents of the cosine (or sine) terms to determine β . Next, we group like powers of $\cos^r(\mu\xi)$ or $\sin^r(\mu\xi)$ and set the corresponding coefficients equal to zero. This procedure produces a system of algebraic equations for α and μ , from which the required coefficients can be determined.

The hyperbolic tangent method

In this subsection, we present the hyperbolic tangent method [14–16]. Within this framework, a suitable traveling-wave transformation is introduced to

$$u(x, y, t) = u(\xi), \quad \xi = (x + y - ct), \quad (14)$$

the partial differential equation (PDE)

$$E_1(u_t, u_x, u_y, u_{xx}, u_{yy}, u_{xxx}, u_{yyy} \dots) = 0, \quad (15)$$

can be rewritten as an ordinary differential equation (ODE)

$$E_2(u, u', u'', u''', \dots) = 0. \quad (16)$$

In Eq. (15) $u(x, y, t)$ is an unknown function, E_1 is a polynomial in $u(x, y, t)$ and its partial derivatives in which the highest order derivatives and the nonlinear terms are involved. In Eq. (16), $u(\xi)$ denotes the unknown function, and E_2 represents a polynomial expression in $u(\xi)$ and its derivatives, including the nonlinear terms.

The hyperbolic tangent method admits the use of the finite expansion

$$u(\mu\xi) = S(Y) = \sum_{k=0}^M a_k Y^k, \quad \text{with } Y = \tanh(\mu\xi), \quad (17)$$

where M is a positive integer that is generally determined as part of the method. To identify M , we balance the highest-order linear term in the resulting equation with the highest-order nonlinear term, following the procedure outlined above. Next, we group terms with the same power of Y and set their coefficients to zero. This leads to a system of algebraic equations for the unknown parameters a_k , μ , and c . Once these parameters are found, a closed-form analytical solution $u(x, y, t)$ can be constructed. The obtained solutions may represent solitons (typically expressed via sech^2), kink-type waves (in terms of $\tanh$), and in some cases periodic wave structures as well.

Results and discussion

The sine solution

Theorem 1 Assume that $c_1 \neq 0$, $c_2 \neq 0$, $c_3 \neq 0$, $c_4 \neq 0$ and let $\xi = ax + by - kt$, where a, b , and k are arbitrary real parameters with $a \neq 0$. Then the (2+1)-dimensional generalized Benjamin-Ono equation (1) has an explicit traveling wave solution of the form

$$u(x, y, t) = \alpha \sin^\beta(\mu(ax + by - kt)),$$

where the constants α, β and μ are given in terms of the parameters by

$$\alpha = -\frac{3k^2c_1 - c_2ka + c_3ab}{c_4a^2}, \quad \beta = -2, \quad \mu = \pm \frac{\sqrt{k^2c_1 - c_2ka + c_3ab}}{2a^2}.$$

Proof:

According to method the solution of the (4) can be found by transformation

$$u(\mu\xi) = \alpha \sin^\beta(\mu\xi). \quad (18)$$

For the sine-form solution, we use the ansatz (18) and the derivative relations (10)-(11). Inserting (18) and (10)–(11) into (4) gives

$$-a^4 \mu^2 \beta^2 \alpha \sin^\beta(\mu\xi) + \mu^2 \alpha \beta (\beta - 1) \sin^{\beta-2}(\mu\xi) + \alpha \sin^\beta(\mu\xi) (k^2 c_1 - c_2 ka + c_3 ab) + a^2 c_4 \alpha \sin^{2\beta}(\mu\xi) = 0. \quad (19)$$

By applying the balance method to Eq. (19) and matching the exponents of the $\sin^j$ terms, the parameter β is determined as

$$\beta - 1 \neq 0, \quad \beta - 2 = 2\beta \Rightarrow \beta = -2. \quad (20)$$

Inserting (20) into (19) yields the next relation:

$$-a^4 \mu^2 \beta^2 \alpha \sin^{-2}(\mu\xi) + a^4 \mu^2 \alpha \beta (\beta - 1) \sin^{-4}(\mu\xi) + \alpha \sin^{-2}(\mu\xi) (k^2 c_1 - c_2 ka + c_3 ab) + a^2 c_4 \alpha \sin^{-4}(\mu\xi) = 0. \quad (21)$$

Comparing the coefficients of the corresponding sine functions leads to the following system of algebraic equations:

$$\sin^{-2}(\mu\xi) : \quad -4a^4 \mu^2 \alpha + \alpha (k^2 c_1 - c_2 ka + c_3 ab) = 0, \quad (22)$$

$$\sin^{-4}(\mu\xi) : \quad 6a^4 \mu^2 \alpha - a^2 c_4 \alpha = 0. \quad (23)$$

From (22)–(23) we have

$$\mu = \pm \frac{\sqrt{k^2 c_1 - c_2 ka + c_3 ab}}{2a^2}, \quad \alpha = -\frac{3}{2} \frac{k^2 c_1 - c_2 ka + c_3 ab}{c_4 a^2}. \quad (24)$$

Substituting (24) into (18) we have the sine solution of the (2+1)-dimensional generalized BO equation (1)

$$u_{11}(x, y, t) = -\frac{3}{2} \frac{k^2 c_1 - c_2 ka + c_3 ab}{c_4 a^2} \csc^2 \left(\frac{\sqrt{k^2 c_1 - c_2 ka + c_3 ab}}{2a^2} (ax + by - kt) \right), \quad (25)$$

with $k^2 c_1 - c_2 ka + c_3 ab > 0$.

$$u_{12}(x, y, t) = -\frac{3}{2} \frac{k^2 c_1 - c_2 ka + c_3 ab}{c_4 a^2} \operatorname{csch}^2 \left(\frac{\sqrt{k^2 c_1 - c_2 ka + c_3 ab}}{2a^2} (ax + by - kt) \right), \quad (26)$$

with $k^2 c_1 - c_2 ka + c_3 ab < 0$.

In Figure 1, we show the plots of solutions (25) and (26):

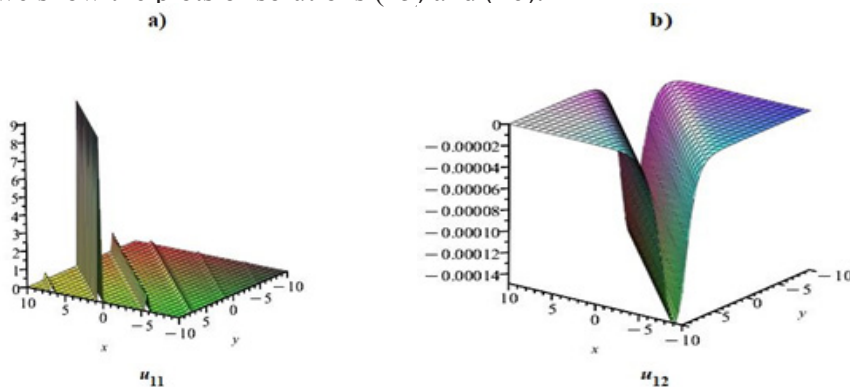


Figure 1 – a) The periodic solutions u_{11} with parameters $a = 0.01, b = 0.01, k = 0.01, c_1 = 1, c_2 = 1, c_3 = 1, c_4 = -1, t = 1$; b) The solitary solutions u_{12}

with parameters $a = 0.01, b = 0.01, k = 0.01, c_1 = -1, c_2 = 1, c_3 = 1, c_4 = -1, t = 1$.
The cosine solution

Theorem 2 Assume that $c_1 \neq 0, c_2 \neq 0, c_3 \neq 0, c_4 \neq 0$ and let $\xi = ax + by - kt$, where a, b , and k are arbitrary real parameters with $a \neq 0$. Then the (2+1)-dimensional generalized Benjamin–Ono equation (1) has an explicit traveling-wave solution of the form

$$u(x, y, t) = \alpha \cos^\beta(\mu(ax + by - kt)).$$

where the constants α, β and μ are given in terms of the parameters by

$$\alpha = -\frac{3k^2c_1 - c_2ka + c_3ab}{2c_4a^2}, \quad \beta = -2, \quad \mu = \pm \frac{\sqrt{k^2c_1 - c_2ka + c_3ab}}{2a^2}.$$

Proof:

To find cosine solution we use

$$u(\mu\xi) = \alpha \cos^\beta(\mu\xi), \quad (27)$$

and its derivatives (12)–(13). Inserting (27) and the derivatives (12)–(13) into (4) leads to

$$-a^4\mu^2\beta^2\alpha\cos^\beta(\mu\xi) + \mu^2\alpha\beta(\beta - 1)\cos^{\beta-2}(\mu\xi) + \alpha\cos^\beta(\mu\xi)(k^2c_1 - c_2ka + c_3ab) + a^2c_4\alpha\cos^{2\beta}(\mu\xi) = 0. \quad (28)$$

Applying the balance method to Eq. (28) and matching the exponents of $\cos^j$ terms, we obtain β as:

$$\beta - 1 \neq 0, \quad \beta - 2 = 2\beta \Rightarrow \beta = -2. \quad (29)$$

Substituting Eq. (29) into Eq. (28), we obtain the following equation:

$$-a^4\mu^2\beta^2\alpha\cos^{-2}(\mu\xi) + a^4\mu^2\alpha\beta(\beta - 1)\cos^{-4}(\mu\xi) + \alpha\cos^{-2}(\mu\xi)(k^2c_1 - c_2ka + c_3ab) + a^2c_4\alpha\cos^{-4}(\mu\xi) = 0. \quad (30)$$

By equating the coefficients of corresponding cosine terms, we obtain the following system of algebraic equations:

$$\cos^{-2}(\mu\xi): \quad -4a^4\mu^2\alpha + \alpha(k^2c_1 - c_2ka + c_3ab) = 0, \quad (31)$$

$$\cos^{-4}(\mu\xi): \quad 6a^4\mu^2\alpha - a^2c_4\alpha = 0. \quad (32)$$

From (31)–(32) we have

$$\mu = \pm \frac{\sqrt{k^2c_1 - c_2ka + c_3ab}}{2a^2}, \quad \alpha = -\frac{3k^2c_1 - c_2ka + c_3ab}{2c_4a^2}. \quad (33)$$

Substituting (33) into (27) we have the cosine solution of the (2+1)-dimensional generalized BO equation (1)

$$u_{21}(x, y, t) = -\frac{3k^2c_1 - c_2ka + c_3ab}{2c_4a^2} \sec^2\left(\frac{\sqrt{k^2c_1 - c_2ka + c_3ab}}{2a^2}(ax + by - kt)\right). \quad (34)$$

with $k^2c_1 - c_2ka + c_3ab > 0$.

$$u_{22}(x, y, t) = -\frac{3k^2c_1 - c_2ka + c_3ab}{2c_4a^2} \operatorname{sech}^2\left(\frac{\sqrt{k^2c_1 - c_2ka + c_3ab}}{2a^2}(ax + by - kt)\right). \quad (35)$$

with $k^2c_1 - c_2ka + c_3ab < 0$.

The plots of solutions (34) and (35) are presented in Figure 2.

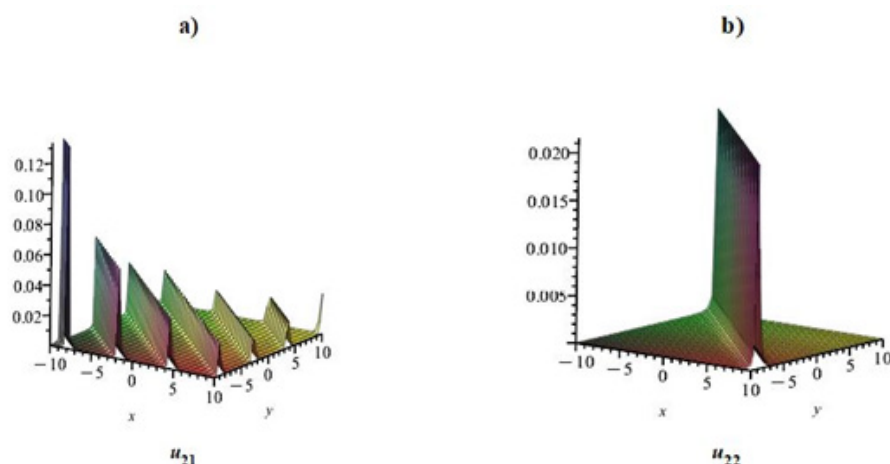


Figure 2 – a) The periodic solutions u_{21} with parameters $a = 0.01, b = 0.01, k = 0.01, c_1 = 1, c_2 = 1, c_3 = 1, c_4 = -1, t = 1$; b) The solitary solutions u_{22} with parameters $a = 0.01, b = 0.01, k = 0.01, c_1 = -1, c_2 = 1, c_3 = 1, c_4 = -1, t = 1$.

The Hyperbolic Tangent Method

Theorem 3 Assume that $c_1 \neq 0, c_2 \neq 0, c_3 \neq 0, c_4 \neq 0$ and let $\xi = ax + by - kt$, where a, b , and k are arbitrary real parameters with $a \neq 0$. Then the (2+1)-dimensional generalized Benjamin–Ono equation (1) admits an explicit traveling wave solution of the form

$$u(x, y, t) = a_0 + a_1 Y + a_2 Y^2,$$

where $Y = \tanh(\mu(ax + by - kt))$ and the constants a_0, a_1, a_2, μ are given in terms of the parameters by

$$a_0 = \frac{abc_3 - ac_2k + c_1k^2}{2a^2c_4}, a_1 = 0, a_2 = -\frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4}, \mu = \pm \frac{\sqrt{abc_3 - ac_2k + c_1k^2}}{2a^2}.$$

or

$$a_0 = -\frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4}, a_1 = 0, a_2 = \frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4}, \mu = \pm \frac{\sqrt{-abc_3 + ac_2k - c_1k^2}}{2a^2}.$$

Proof:

Balancing in Eq. (4) the nonlinear term u^2 , whose highest power is $2M$, with the highest-order derivative u''' , which contributes a term of order $M + 2$, gives $M + 2 = 2M$ and hence $M = 2$. Therefore, within the hyperbolic tangent method we may adopt the following substitution:

$$u(\xi) = a_0 + a_1 Y + a_2 Y^2, \quad (36)$$

where

$$Y = \tanh(\mu\xi), \quad \xi = ax + by - kt.$$

After substituting Eq. (36) into Eq. (4) and collecting terms with the same powers of Y , we obtain a system of algebraic equations for a_0, a_1, a_2 , and μ . Solving this system using Maple yields the following results:

$$a_0 = \frac{abc_3 - ac_2k + c_1k^2}{2a^2c_4}, a_1 = 0, a_2 = -\frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4}, \mu = \pm \frac{\sqrt{abc_3 - ac_2k + c_1k^2}}{2a^2}. \quad (37)$$

Case 2:

$$a_0 = -\frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4}, a_1 = 0, a_2 = \frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4}, \mu = \pm \frac{\sqrt{-abc_3 + ac_2k - c_1k^2}}{2a^2}. \quad (38)$$

Substituting Eqs. (37)–(38) into Eq. (36), and then inserting the resulting expressions into Eq. (2), we derive analytical solutions of the (2+1)-dimensional generalized Benjamin-Ono equation (1) in the following forms:

Case 1:

$$u_{31}(x, y, t) = \frac{abc_3 - ac_2k + c_1k^2}{2a^2c_4} - \frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4} \times \tanh^2 \left(\frac{\sqrt{abc_3 - ac_2k + c_1k^2}}{2a^2} (ax + by - kt) \right), \quad (38)$$

with $k^2c_1 - c_2ka + c_3ab > 0$.

Case 2:

$$u_{32}(x, y, t) = -\frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4} + \frac{3(abc_3 - ac_2k + c_1k^2)}{2a^2c_4} \tanh^2 \left(\frac{\sqrt{-abc_3 + ac_2k - c_1k^2}}{2a^2} (ax + by - kt) \right), \quad (39)$$

with $-k^2c_1 + c_2ka - c_3ab > 0$.

The graphs of solution (38) is presented in Figure 3.

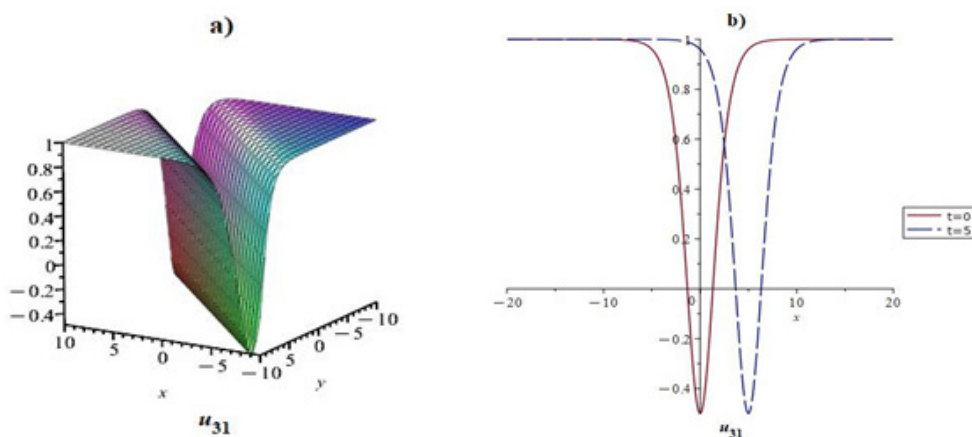


Figure 3 – a) The solitary wave solutions of the u_{31} with parameters

$a = 0.01, b = 0.01, k = 0.01, c_1 = 1, c_2 = 1, c_3 = 1, c_4 = -1, t = 1$;

b) Dynamics of the the solitary wave solutions u_{31} with parameters

$a = 0.01, b = 0.01, k = 0.01, c_1 = -1, c_2 = 1, c_3 = 1, c_4 = -1, t = 0, \text{ and } t = 5$.

Conclusion

In this article, the sine-cosine method and the hyperbolic tangent method are employed to derive exact solutions of the (2+1)-dimensional generalized Benjamin-Ono equation. By introducing an appropriate traveling-wave transformation, the nonlinear partial differential equation is reduced to a nonlinear ordinary differential equation, which is then treated via the proposed approaches. As a result, families of exact solutions, such as periodic wave solutions and solitary wave solution are constructed. To demonstrate the behavior of the derived solutions, 2D and 3D profiles are plotted for representative parameter settings. Overall, the outcomes indicate that the sine-cosine method and the hyperbolic tangent method are straightforward to implement and provide an efficient and effective framework for analyzing a broad class of nonlinear partial differential equations.

Acknowledgment

This work is supported by the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant AP26194665)

REFERENCES

- 1 Wazwaz, A.M., Partial Differential Equations and Solitary Waves Theory (Springer-Verlag, Berlin Heidelberg, 2009), 681 p.
- 2 Burdik, C., Shaikhova, G., Rakhimzhanov, B. Soliton solutions and travelling wave solutions for the two-dimensional generalized nonlinear Schrödinger equations, European Physical Journal Plus 136(10), 1095 (2021).
- 3 Manikandan, K., Serikbayev, N., Manigandan, M., Sabareeshwaran, M. Dynamical evolutions of optical smooth positons in variable coefficient nonlinear Schrödinger equation with external potentials. Optik 288, 171203 (2023).
- 4 Verhulst F. Nonlinear Differential Equations and Dynamical Systems (Springer, Berlin, 1996).
- 5 Shaikhova, G., Kutum, B., Myrzakulov, R.: Periodic traveling wave, bright and dark soliton solutions of the (2+1)-dimensional complex modified Korteweg-de Vries system of equations by using three different methods. AIMS Mathematics, 7(10), 18948–18970 (2022).
- 6 Chen, H., Zheng, Sh. Darboux transformation for nonlinear Schrödinger type hierarchies. Physica D: Nonlinear Phenomena, 454, 1338636 (2023).
- 7 Liu, J.G., Zhu, W.H. Multiple rogue waves, breather wave and interaction solutions of a generalized (3+1)-dimensional variable-coefficient nonlinear wave equation. Nonlinear Dynamics, 103, 1841–1850 (2021).
- 8 Dehghan, M., Hamidi, A., Shakourifar, M. The solution of coupled Burgers, equations using Adomian-Pade technique. Applied Mathematics and Computation. 189(2), 1034–1047 (2007).
- 9 Cherniha, R., Muzyka, L. Lie symmetries of the Shigesada-Kawasaki-Teramoto system. Communication Nonlinear Science and Numerical Simulation, 45, 81–92 (2017).
- 10 Hosseini, K., Mayeli, P., Kumar, D. New exact solutions of the coupled sine-Gordon equations in nonlinear optics using the modified Kudryashov method. Journal Modern Optics, 65(3), 361–364 (2018).
- 11 Yao Sh.-W, Behera. S., Inc, M., Rezazadeh, H., Viridi, J.P.S., Mahmoud, W., Arqu, O.A., Osman, M.S. Analytical solutions of conformable Drinfel'd-Sokolov-Wilson and Boiti Leon Pempinelli equations via sine-cosine method. Results in Physics, 42, 105990 (2022).
- 12 Wazwaz, A.M., The sine-cosine method for obtaining solutions with compact and noncompact structures, Applied Mathematics and Computation, 159(2), 559–576 (2004).
- 13 Ala, V., Shaikhova, G. Analytical Solutions of Nonlinear Beta Fractional Schrödinger Equation Via Sine-Cosine Method, Lobachevskii Journal of Mathematics, 43(11), 3033–3038 (2022).
- 14 Malfliet, W. Solitary wave solutions of nonlinear wave equations, American Journal of Physics., 60 (7), 650–654 (1992).
- 15 Malfliet, W. The tanh method: I. Exact solutions of nonlinear evolution and wave equations. Physica Scripta, 54, 563–568 (1996).
- 16 Wazwaz, A.M. The Tanh and the Sine-Cosine Methods for the Complex Modified KdV and the Generalized KdV Equations. Computers and Mathematics with Applications, 49, 1101–1112 (2005)
- 17 Hassaballa, A., Baber, M.Z., Butt, A., Yasin, M.W., Safdar, R., Bittaye, E. Dynamical description and analytical study of traveling wave solutions for generalized Benjamin-Ono equation. Scientific Reports, 15, 33923 (2025).
- 18 Ma, H., ·Yue, Sh., ·Gao, Y., Deng, A. Lump solution, breather soliton and more soliton solutions for a (2+1)-dimensional generalized Benjamin-Ono equation. Qualitative Theory of Dynamical Systems, 22, 72 (2023).

Шайхова Г.Н.,
PhD, доцент, ORCID ID: 0000-0002-0819-5338,
e-mail: g.shaikhova@gmail.com

Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

(2+1)-ӨЛШЕМДІ ЖАЛПЫЛАНҒАН БЕНДЖАМИН-ОНО ТЕҢДЕУІНІҢ АНАЛИТИКАЛЫҚ ШЕШІМДЕРІ

Аңдатпа

(2+1)-өлшемді жалпыланған Бенджамин-Оно теңдеуі таяз су бетінде таралатын кіші амплитудалы, ұзын толқынды толқындардың таралуын модельдейді. (2+1)-өлшемді жалпыланған Бенджамин-Оно теңдеуінің нақты шешімдерін құрастыру эксперименттік зерттеулерді теориялық тұрғыдан негіздеумен қатар, сызықтық емес толқындар динамикасында туындайтын қолданбалы мәселелерді шешуге берік әрі қатаң негіз береді. Мақалада дисперсиялық әсерлерді ескере отырып, сызықтық емес ортада (2+1)-өлшемді жалпыланған Бенджамин-Оно теңдеуімен сипатталатын толқындардың таралуын зерттейміз. Осы мақсатта айқын шешімдер алудың аналитикалық құралдары ретінде синус-косинус функциялары әдісі және гиперболалық тангенс әдісі қолданылады. Бұл әдістер математикалық физикада кездесетін кең ауқымды сызықтық емес теңдеулер класын шешуде тиімді екенін көрсетеді. Аталған тәсілдер арқылы периодты толқындық шешімдер және жекеленген толқын шешімдер алынды; алынған нәтижелерді көрнекі көрсету үшін параметрлердің тиісті мәндерін таңдап, 3D және 2D графиктер жасалды.

Түйін сөздер: жалпыланған Бенджамин-Оно теңдеуі, синус-косинус әдісі, гиперболалық тангенс әдісі, периодты толқындық шешім, жекеленген толқындық шешім.

Шайхова Г.Н.,
PhD, ассоциированный профессор, ORCID ID: 0000-0002-0819-5338,
e-mail: g.shaikhova@gmail.com

Евразийский национальный университет им. Л.Н. Гумилева, г. Астана, Казахстан

АНАЛИТИЧЕСКИЕ РЕШЕНИЯ (2+1)-МЕРНОГО ОБОБЩЕННОГО УРАВНЕНИЯ БЕНДЖАМИНА-ОНО

Аннотация

(2+1)-мерное обобщенное уравнение Бенджамина-Оно моделирует распространение длинноволновых волн малой амплитуды на поверхности мелкой воды. Построение явных решений (2+1)-мерного обобщенного уравнения Бенджамина-Оно не только обеспечивает теоретическую поддержку экспериментальным исследованиям, но и дает строгую основу для решения прикладных задач, возникающих в нелинейной динамике волн. В данной работе исследовано распространение волн, описываемое (2+1)-мерным обобщенным уравнением Бенджамина-Оно в нелинейных средах с учетом дисперсионных эффектов. Для этого в качестве аналитических инструментов получения явных решений используются метод функций синуса-косинуса и метод гиперболического тангенса. Показано, что эти методы эффективны для широкого класса нелинейных уравнений математической физики. С их помощью получены решения в виде периодических волн и уединенных волн; для графического представления результатов построены трехмерные и двумерные графики при выборе подходящих значений параметров модели.

Ключевые слова: обобщенное уравнение Бенджамина-Оно, метод синусов и косинусов, метод гиперболического тангенса, решение в виде периодической волны, решение в виде солитонной волны.

UDC 517.95
IRSTI 27.33.19

<https://doi.org/10.55452/1998-6688-2026-23-2-35-45>

¹**Abdikalikova G.A.,**

Cand. Phys.-Math. Sc., Professor, ORCID ID: 0000-0001-6280-4168,
e-mail: abdikalikova.galliya@gmail.com

¹**Sartabanov Zh. A.,**

Dr. Phys.-Math. Sc., Professor, ORCID ID: 0000-0003-2601-2678,
e-mail: sartabanov42@mail.ru

^{2*}**Aitenova G.M.,**

PhD, Associate Professor (Docent), ORCID ID: 0000-0002-4572-8252,
*e-mail: gulsezimaitenova@gmail.com

^{1,3}**Zhumagaziyev A.Kh.,**

PhD, ORCID ID: 0000-0002-6007-3311,
e-mail: azhumagaziyev@zhubanov.edu.kz

¹K. Zhubanov Aktobe Regional University, Aktobe, Kazakhstan

²M. Utemisov West Kazakhstan University, Uralsk, Kazakhstan

³Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan

INVESTIGATION OF A MULTIPERIODIC PROBLEM FOR A FINITE- HEREDITARY SYSTEM OF INTEGRO-DIFFERENTIAL EQUATIONS

Abstract

Numerous problems arising in various fields of natural sciences and engineering lead to the study of integro-differential equations. The investigation of such equations, which account for the historical behavior of processes or phenomena, originates from the pioneering works of V. Volterra, where the role of the integral term was emphasized. Various approaches have since been developed to describe hereditary effects and aftereffects in these equations. It should be noted that the existence of multiperiodic and quasiperiodic solutions for systems of Volterra-type integro-differential equations does not always guarantee their uniqueness. In this paper, a system of partial integro-differential equations with a special differentiation operator is considered. Using the method of periodic characteristics, integral representations of the solution manifold for such systems with aftereffects are constructed. The properties of iterated kernels and resolvents are studied, and corresponding estimates are obtained. Conditions for the existence of a Green-type matrix function for the multiperiodic problem are established, along with integral representations with suitable bounds. Finally, sufficient conditions for the existence and uniqueness of a multiperiodic solution, interpreted as an integral manifold of the integro-differential system with finite hereditary effects, are derived.

Keywords: matricant, kernel, resolving operator, dichotomy, Green type function.

Received January 1, 2026; revised January 21, 2026, April 23, 2026; accepted April 25, 2026.

Introduction

Investigation of some processes in natural sciences and technology often necessitates the study of integro-differential equations.

Many problems in natural sciences and various applied fields are effectively modeled through mathematical frameworks described by integro-differential equations.

It is well established that certain problems of mathematical biology, biomechanics, control theory, population dynamics, elasticity theory, etc. lead to the study of partial integro-differential equations [1–8].

Of particular interest are processes where the dynamics depend not only on their current state but also on their previous states. The influence of the integral term on certain classes of differential equations and its impact on the nature of their solutions has been explored in [9–12]. The study of such processes with aftereffects is intricately linked to finite-period heredity. Notably, these processes and phenomena induce temporal oscillations [13–14].

Research efforts focusing on multiperiodic and almost periodic solutions of specific classes of partial integro-differential equations provide a detailed review and analysis of these problems, considering the role of finite hereditary effects [15–18].

In this context, developing methods to extend the oscillatory solutions of differential equation systems to integro-differential systems with specialized differentiation operators in multidimensional time and along vector field directions is of significant interest.

To study the problem of integrating multiperiodic functions along the periodic characteristics of the differentiation operator with respect to (τ, t) , an approach is proposed [19–21]. By employing the linear transformation $t = t_0 + \beta^*(\tau - \tau_0) \equiv \beta(\tau, \tau_0, t_0)$, the θ – periodic β – helical characteristic of the operator D_e is constructed, which describe translational and rotational motion, and the θ – periodic by τ characteristic $\eta = \beta(\xi, \tau, \tilde{t})$ is determined.

By linear substitution $t_j = c_j \tilde{t}_j$, $j = \overline{1, m}$ or $t = c\tilde{t}$, the differentiation operator D_e is reduced to operator D_c with m -dimensional vector $e = (1, \dots, 1)$, where $c = \text{diag}[c_1, \dots, c_m]$ is a matrix, $\tilde{t}_j = (\tilde{t}_1, \dots, \tilde{t}_m)$.

Therefore, operator D_c has a θ – periodic characteristic

$$\eta = c\beta(\xi, \tau, c^{-1}t) \equiv h(\xi, \tau, t).$$

The vector-function $h(\xi, \tau, t)$ has the properties:

$$i) h(\xi + \theta, \tau, t) = h(\xi, \tau + \theta, t) = h(\xi, \tau, t),$$

$$ii) h(\xi, \tau, t + \omega) = h(\xi, \tau, t) + \omega,$$

it is periodic by t of the second kind with a period $\omega = (\omega_1, \dots, \omega_m)$.

For a specific class of partial integro-differential equations with a specialized differentiation operator, integral representations of the solution manifold are derived. The representation of solutions in terms of the kernel resolvent is analyzed, highlighting the critical role of these equations in modeling phenomena and processes characterized by aftereffects.

In contrast to the existing results on almost periodic and multiperiodic solutions of integro-differential equations, the present paper considers a system with a special differentiation operator D_c acting along periodic characteristics in multidimensional time [15–21].

The novelty of the approach consists in combining the method of periodic characteristics with the construction of a resolving operator based on the matricant and the kernel resolvent for systems with finite heredity. This makes it possible to obtain integral representations of solutions and to establish sufficient conditions for the existence and uniqueness of multiperiodic solutions in a more general setting.

Materials and methods

I. Consider a vector integro-differential equation of the form

$$D_c u(\tau, t) = A(\tau, t)u(\tau, t) + \int_{\tau}^{\tau+\theta} K(\xi, \tau, t, h(\xi, \tau, t))u(\xi, h(\xi, \tau, t))d\xi, \quad (1)$$

where $D_c = \frac{\partial}{\partial \tau} + \left\langle c, \frac{\partial}{\partial t} \right\rangle$ is differentiation operator; $h(\xi, \tau, t)$ is periodic characteristic of the operator D_c ; $A(\tau, t) \in C_{\tau, t}^{(0, \theta)}(R \times R^m)$ and $K(\xi, \tau, t, h) \in C_{\xi, \tau, t, h}^{(0, 0, \theta, \theta)}(R \times R \times R^m \times R^m)$ are $n \times n$ – matrices; e and $\tilde{e}$ are vectors with single components. We assume that the conditions (P) are met if:

$$A(\tau + \theta, t + \omega) = A(\tau, t); \quad (2)$$

$$K(\xi + \theta, \tau, t, h(\xi + \theta, \tau, t)) = K(\xi, \tau, t, h(\xi, \tau, t));$$

$$K(\xi, \tau + \theta, t + \omega, h(\xi, \tau + \theta, t + \omega)) = K(\xi, \tau, t, h(\xi, \tau, t)), \quad (3)$$

where $\omega_0 = \theta$, $\omega = (\omega_1, \dots, \omega_m)$ is rationally incommensurable positive constants.

The conditions (P) express the multiperiodicity, smoothness, and boundedness of the matrix-function $A(\tau, t)$ and the kernel $K(\xi, \tau, t, h)$. In particular, relations (2)-(3) ensure the invariance of the system with respect to shifts by the periods θ and ω , which is necessary for the construction of multiperiodic solutions along the periodic characteristics of the operator D_c .

Moreover, these conditions guarantee the boundedness of the corresponding operators and allow the application of the method of periodic characteristics and the construction of the resolving operator.

Similarly, condition (P₁) means that the nonhomogeneous term $f(\tau, t)$ possesses the same multiperiodicity properties and belongs to the same smoothness class as the coefficients of the system. This assumption is essential for preserving the multiperiodic structure of the solution.

Building upon the approach outlined in for constructing the matrix associated with the equation $D_c v = A(\tau, t)v$, we define the matrix $V(\tau_0, \tau, t)$ which satisfies the matrix equation and reduces to the identity $n \times n$ -matrix [17]. Utilizing the existence of the inverse operator D_c^{-1} , we research the matricant as a solution to the matrix equation expressed as a series, with its terms determined through recurrence relations. The absolute and uniform convergence of this series can be rigorously verified. The main properties of the matricant are discussed [17], [21]. Notably, one of the key attributes of the matricant $V(\tau_0, \tau, t)$ is its multiperiodicity with respect to τ_0 , τ and t :

$$V(\tau_0 + \theta, \tau + \theta, t + \omega) - V(\tau_0, \tau, t) = 0.$$

Using the conditions of multiperiodicity (2) of the matrix $A(\tau, t)$ and its smoothness with unknown vector-functions and the kernel $K(\xi, \tau, t, h(\xi, \tau, t))$ of the hereditary term, the integro-differential system (1) by linear transformation

$$u(\tau, t) = V(\tau_0, \tau, t)w(\tau, t), \quad (4)$$

is reduced to a system of equations of the form

$$D_c w(\tau_0, \tau, t) = \int_{\tau}^{\tau+\theta} Q(\xi, \tau_0, \tau, t, h(\xi, \tau, t))w(\tau_0, h(\xi, \tau, t))d\xi. \quad (5)$$

Here $Q(\xi, \tau_0, \tau, t, h(\xi, \tau, t)) = V^{-1}(\tau_0, \tau, t)K(\xi, \tau, t, h(\xi, \tau, t))V(\tau_0, \xi, h(\xi, \tau, t))$ is a kernel. Note that $D_c V^{-1}(\tau_0, \tau, t) = -V^{-1}(\tau_0, \tau, t)A(\tau, t)$, then $V^{-1}(\tau_0, \tau, t)$ also has the properties of smoothness, multiperiodicity and boundedness. Due to the properties of matrices $V(\tau_0, \tau, t)$ and $V^{-1}(\tau_0, \tau, t)$, as well as condition (3), the kernel has the property

$$\begin{aligned} Q(\xi + \theta, \tau_0 + \theta, \tau + \theta, t + \omega, h(\xi + \theta, \tau + \theta, t + \omega)) &= \\ &= Q(\xi, \tau_0, \tau, t, h(\xi, \tau, t)). \end{aligned}$$

The matricant $W(\tau_0, \tau, t)$ of the equation (5) is defined as the solution of the matrix integral equation

$$\begin{aligned} W(\tau_0, \tau, t) &= E + \\ &+ \int_{\tau_0}^{\tau} \int_{\eta}^{\eta+\theta} Q(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))W(\xi, s, h(\xi, \tau, t))d\xi d\eta. \end{aligned} \quad (6)$$

Using the iterative method, its resolvent $R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))$ is constructed, and satisfying the integral equation

$$\begin{aligned} R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t)) &= Q(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t)) + \\ &+ \int_s^{\tau} \int_{\eta}^{\eta+\theta} Q(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))d\xi d\eta \end{aligned} \quad (7)$$

Thus, the resolvent (7) is well-defined and satisfies the corresponding integral equation in the class of bounded matrix-functions. It is worth noting that, based on the product of matrices and through calculations involving changes in the order of integration while considering the group property,

iterated kernels $Q_m(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))$ are constructed. These kernels are determined through recurrence relations, and their corresponding estimates are established

$$\|Q_m(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))\| \leq Q_0^m \frac{\theta^{m-1}(\tau-\xi)^{m-1}}{(m-1)!}. \quad (8)$$

The integral equation (7) is obtained by successive approximations. Substituting the expression for R into the right-hand side and iterating, one obtains a series representation in terms of iterated kernels. The convergence of this series follows from estimate (8) and the boundedness of the kernel Q .

Taking into account (8), for the resolving kernel holds

$$\|R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))\| \leq Q_0 e^{Q_0 \theta(\tau-s)},$$

where $\|Q(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))\| \leq Q_0$, $Q_0 = \text{const}$; s and η changes between τ_0 and τ , $|\tau - \tau_0| < \Delta$, $\Delta - \text{const} > 0$.

Note that the resolvent $R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t))$ has the property of multiperiodicity:

$$\begin{aligned} R(\xi, s + \theta, \eta, h(\eta, \tau + \theta, t + \omega), h(\xi, \tau + \theta, t + \omega)) = \\ = R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t)). \end{aligned}$$

Next, the matrix (6) is represented by the resolvent

$$\begin{aligned} \mathbb{W}(\tau_0, \tau, t) = E + \\ + \int_{\tau_0}^{\tau} \int_{\eta}^{\eta+\theta} R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t)) d\xi d\eta. \end{aligned} \quad (9)$$

Thus, taking into account the transformation (4) and the resolvents, we have

$$\Phi(\tau_0, \tau, t) = \mathbb{V}(\tau_0, \tau, t) \mathbb{W}(\tau_0, \tau, t). \quad (10)$$

A matrix of the form (10) is called the resolving operator of the integro-differential equation (1).

This operator $\Phi(\tau_0, \tau, t)$ has the following properties:

- $i_1) D_c \Phi(\tau_0, \tau, t) = A(\tau, t) \Phi(\tau_0, \tau, t) +$
 $+ \int_{\tau}^{\tau+\theta} K(\xi, \tau, t, h(\xi, \tau, t)) \Phi(\xi, \tau_0, h(\xi, \tau, t)) d\xi;$
- $i_2) \Phi(\tau_0, \tau + \theta, t + \omega) = \Phi(\tau_0, \tau, t);$
- $i_3) \|\Phi(\tau_0, \tau, t)\| \leq d e^{-\alpha(\tau-\tau_0)}, \tau \geq \tau_0, d \geq 1, \alpha > 0.$

The set of conditions $i_1) - i_3)$ is called conditions (i).

For clarity, we note that all vector-functions are considered in the space $C_{\tau, t}^{(k, \theta)}(R \times R^m)$, and their dependence on the variables (τ, t) or (ξ, τ, t, h) is explicitly indicated whenever necessary.

Throughout the paper, the matrices $A(\tau, t)$, $K(\xi, \tau, t, h)$, $Q(\xi, \tau_0, \tau, t, h)$, as well as the operators Φ, V, W , and R are assumed to be defined for all admissible values of their arguments, and their dependence on these variables will be omitted only when it does not lead to ambiguity.

Construction of a multiperiod solution

Consider the linear inhomogeneous integro-differential equation

$$\begin{aligned} D_c u(\tau, t) = A(\tau, t) u(\tau, t) + \\ + \int_{\tau}^{\tau+\theta} K(\xi, \tau, t, h(\xi, \tau, t)) u(\xi, h(\xi, \tau, t)) d\xi + f(\tau, t). \end{aligned} \quad (11)$$

Assume that the conditions (P_1) are met if

$$f(\tau + \theta, t + \omega) = f(\tau, t) \in C_{\tau, t}^{(0, \theta)}(R \times R^m). \quad (12)$$

We will assume that equation (11) is uncritical (non-critical) for any vector-function $f(\tau, t)$ of $C_{\tau, t}^{(0, s)}(R \times R^m)$.

Now let's start considering the problem for the system of integro-differential equations (11) with the initial condition

$$u|_{\tau=\tau_0} = \varphi(t), \quad (13)$$

where $\varphi(t + \omega) = \varphi(t) \in C_t^{(s)}(R^m)$ is a well-known vector-function.

The function $u(\tau, t)$ belonging to class $C_{\tau, t}^{(1, s)}(R \times R^m)$ satisfying the integro-differential equation (11) and for $\tau = \tau_0$ – the initial condition (13) is called the solution of the Cauchy problem for the inhomogeneous integro-differential equation (11).

Lemma 1. If the conditions (P), $\varphi(t) \in C_t^{(s)}(R^m)$ are met, then there is a solution to the initial problem (1), (13) and it is given by the formula

$$\hat{u}(\tau, t) = \Phi(\tau_0, \tau, t) \varphi(h(\tau_0, \tau, t)), \quad \tau_0 \in R. \quad (14)$$

Proof. Based on (10), the formula (14) is represented as

$$\begin{aligned} \hat{u}(\tau, t) &= V(\tau_0, \tau, t) W(\tau_0, \tau, t) \varphi(h(\tau_0, \tau, t)) = V(\tau_0, \tau, t) [\varphi(h(\tau_0, \tau, t)) + \\ &+ \int_{\tau_0}^{\tau} \int_{\eta}^{\eta+\theta} R(\xi, s, \eta, h(\eta, \tau, t), h(\xi, \tau, t)) \varphi(h(\tau_0, \tau, t)) d\eta d\xi]. \end{aligned} \quad (15)$$

In (15), we used representation (9) along the way. Taking into account the properties of the operator D_c, i_1 , $D_c h(\tau_0, \tau, t) = 0$ and applying D_c to the relation (15), we have that $\hat{u}(\tau, t)$ is the solution of the problem under consideration.

It follows from (14) that for $\tau = \tau_0$, condition (13) holds.

Q.E.D.

Lemma 2. If conditions (P) and (P₁) are satisfied, then equation (11) admits a solution

$$\tilde{u}(\tau, t) = \int_{\tau_0}^{\tau} \Phi(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi, \quad (16)$$

satisfying condition

$$\tilde{u}|_{\tau=\tau_0} = 0. \quad (17)$$

Proof. We have

$$\begin{aligned} &= f(\tau, t) + \int_{\tau_0}^{\tau} [D_c \Phi(\xi, \tau, t) + A(\tau, t) \Phi(\xi, \tau, t)] f(\xi, h(\xi, \tau, t)) d\xi = \\ &= f(\tau, t) + \int_{\tau_0}^{\tau} T \Phi(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi = f(\tau, t) + T \tilde{u}(\tau, t), \end{aligned} \quad (18)$$

where $(Tu)(\tau, t) = \int_{\tau}^{\tau+\theta} K(\xi, \tau, t, h(\xi, \tau, t)) u(\xi, h(\xi, \tau, t)) d\xi$.

Consequently, from (18) we have that the vector function $\tilde{u}(\tau, t)$ satisfies equation (11).

The satisfiability (feasibility) of condition (17) is obvious.

It follows that $\tilde{u}(\tau, t)$ is the solution of problem (11), (17).

Q.E.D.

Note that we can directly write the integral representation (16) in terms of the resolvent as follows

$$\begin{aligned} \tilde{u}(\tau, t) &= \int_{\tau_0}^{\tau} V(\xi, \tau, t) [f(\xi, h(\xi, \tau, t)) + \\ &+ \int_{\xi}^{\xi} \int_{\eta}^{\eta+\theta} R(\xi', s, \eta, h(\eta, \tau, t), h(\xi, \tau, t)) f(\xi, h(\xi, \tau, t)) d\eta d\xi'] d\xi. \end{aligned} \quad (19)$$

Theorem 1. If the conditions (P), $\varphi(t) \in C_t^{(s)}(R^m)$ are met, then there is a unique solution $u(\tau, t)$ of the problem (11), (13) represented as

$$u(\tau, t) = \hat{u}(\tau, t) + \tilde{u}(\tau, t), \quad (20)$$

where $\hat{u}(\tau, t)$ satisfies the integro-differential equation (1) and condition (13), $\tilde{u}(\tau, t)$ – satisfies (11) and (17).

Proof. The proof of the theorem follows from Lemma 1 and Lemma 2.

Here the solutions $\hat{u}(\tau, t)$ and $\tilde{u}(\tau, t)$ are expressed in terms of (15) and (19), respectively. The uniqueness of the solution of the problem under consideration is proved by the method of the opposite.

The solution (20) of the Cauchy problem (11), (13) is represented as

$$u(\tau, t) = \Phi(\tau_0, \tau, t) \varphi(h(\tau_0, \tau, t)) + \int_{\tau_0}^{\tau} \Phi(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi. \quad (21)$$

Using the necessary and sufficient periodicity condition

$$u(\tau_0 + \theta, t) = u(\tau_0, t) C_{\tau_0, \tau}^{(0, \theta)}(R \times R^m)$$

the method of successive approximations, and the method of complete mathematical induction, we obtain a function

$$\varphi^*(h(\tau_0, \tau, t)) = \int_{-\infty}^{\tau_0} \Phi(\xi, \tau_0, h(\tau_0, \tau, t)) f(\xi, h(\xi, \tau_0, t)) d\xi, \quad (22)$$

that satisfies the requirements of the initial condition (13).

Substituting (22) into (21), we obtain the corresponding solution to the Cauchy problem

$$u^*(\tau, t) = \int_{-\infty}^{\tau} \Phi(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi. \quad (23)$$

Here we have used the group property. The convergence of the integral in (23) is provided by i_3) and the estimate

$$\|u^*(\tau, t)\| \leq \int_{-\infty}^{\tau} \|\Phi(\xi, \tau, t)\| \|f(\xi, h(\xi, \tau, t))\| d\xi \leq \frac{d}{\alpha} \|f\|.$$

Problem. To establish the existence of a unique (θ, ω) – periodic solution of the integro-differential equation (11).

It is assumed that the resolving operator $\Phi(\xi, \tau, t)$ is represented as

$$\Phi(\xi, \tau, t) = \Phi_{-}(\xi, \tau, t) + \Phi_{+}(\xi, \tau, t). \quad (24)$$

Here $\Phi_{-}(\xi, \tau, t)$ and $\Phi_{+}(\xi, \tau, t)$ are partial solutions of the vector-matrix integro-differential equation (1) with initial conditions

$$\begin{aligned} \Phi_{-}(\xi, \tau, t) &= E_{-}, \\ \Phi_{+}(\xi, \tau, t) &= E_{+}, \\ E &= E_{-} + E_{+} \end{aligned} \quad (25)$$

and with corresponding properties (i), as well as satisfying the estimates:

$$|\Phi_{-}(\xi, \tau, t)| \leq d e^{-\alpha(\tau-\xi)} \quad \text{при } \tau > \xi, \quad (26)$$

$$|\Phi_{+}(\xi, \tau, t)| \leq d e^{\alpha(\tau-\xi)}, \quad \tau < \xi, \quad (27)$$

where $d \geq 1, \alpha = \text{const} > 0$.

Thus, splitting the resolving operator into the sum of matrices with properties (i), as well as satisfying conditions (26) and (27), leads to a system of integro-differential equations with the property of exponential dichotomy.

Let's start constructing a matrix function of the Green type.

$$\tilde{\Phi}(s, \tau, t) = \begin{cases} \Phi_-(s, \tau, t), & \tau > s, \\ -\Phi_+(s, \tau, t), & \tau < s. \end{cases} \quad (28)$$

Here are some properties of $\tilde{\Phi}(s, \tau, t)$:

$$i_4) D_\tau \tilde{\Phi}(s, \tau, t) = A(\tau, t) \tilde{\Phi}(s, \tau, t) + \int_\tau^{\tau+\theta} K(\xi, \tau, t, h(\xi, \tau, t)) \tilde{\Phi}(s, \tau, h(\xi, \tau, t)) d\xi, \quad \tau \neq s.$$

Due to the fact that $\Phi_-(s, \tau, t)$ and $\Phi_+(s, \tau, t)$ are matrix solutions of the integro-differential equation (1), the satisfiability (feasibility) of this condition follows.

$$i_5) \tilde{\Phi}(s - 0, \tau, t) - \tilde{\Phi}(s + 0, \tau, t) = E.$$

From (25), one can directly verify the feasibility of this property.

$$i_6) \tilde{\Phi}(s + \theta, \tau + \theta, t + \omega) = \tilde{\Phi}(s, \tau, t).$$

Note that E_- , E_+ and E do not depend on the variable t , as well as $\Phi_-(s, \tau, t)$, $\Phi_+(s, \tau, t)$ and $\Phi(s, \tau, t)$ are multiperiodic matrix solutions, then the property $i_6)$ follows from the above.

$$i_7) |\tilde{\Phi}(s, \tau, t)| \leq d e^{-\tilde{\alpha}|\tau-s|}, \quad d \geq 1, \quad \tilde{\alpha} > 0.$$

From estimates (26) and (27) we obtain the property $i_7)$.

It should be noted that the assumptions on boundedness, smoothness, and multiperiodicity of the matrices and kernels are sufficient for the application of the proposed method. These conditions ensure the existence of the matricant, the convergence of the resolvent series, and the boundedness of the resolving operator.

At the same time, these assumptions are not necessarily minimal. Their possible weakening, in particular with respect to smoothness or periodicity, requires additional investigation and may be considered in further studies.

Theorem 2. If the conditions (P), (P₋₁), $k_0\theta + \sum_{j=1}^m k_j c_j \omega_j \neq 0$, $\sum_{j=0}^m |k_j| \neq 0$, $k_j \in \mathbb{Z}$, (24) and (25) are met, then the linear integro-differential equation (11) has a unique (θ, ω) -periodic solution representable in the form

$$u^{**}(\tau, t) = \int_{-\infty}^{+\infty} \tilde{\Phi}(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi. \quad (29)$$

Proof. Using (28), we write (29) as follows

$$\begin{aligned} u^{**}(\tau, t) &= \int_{-\infty}^{\tau} \tilde{\Phi}(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi + \\ &+ \int_{\tau}^{+\infty} \tilde{\Phi}(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi = \\ &= \int_{-\infty}^{\tau} \Phi_-(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi + \\ &+ \int_{\tau}^{+\infty} \Phi_+(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi. \end{aligned}$$

Acting with the D_c operator, using the properties $i_4)$ and $i_5)$ we have

$$\begin{aligned} D_c u^{**}(\tau, t) &= [\tilde{\Phi}(\xi - 0, \tau, t) - \tilde{\Phi}(\xi + 0, \tau, t)] f(\tau, t) + \\ &+ \int_{-\infty}^{+\infty} D_c \tilde{\Phi}(\xi, \tau, t) f(\xi, h(\xi, \tau, t)) d\xi + \\ &+ \int_{-\infty}^{+\infty} \tilde{\Phi}(\xi, \tau, t) D_c f(\xi, h(\xi, \tau, t)) d\xi \end{aligned} \quad (30)$$

Further in (30) taking into account that for $\tau \neq \xi$

$$\begin{aligned} D_c \Phi(\xi, \tau, t) &= [A(\tau, t)I + T]\Phi(\xi, \tau, t), \\ D_c f(\xi, h(\xi, \tau, t)) &= 0, \end{aligned} \quad (31)$$

where I – single operator;

$$(T\Phi)(\xi, \tau, t) = \int_{\tau}^{\tau+\theta} K(\xi, \tau, t, h(\xi, \tau, t))\Phi(\xi, \tau, t)d\xi$$

we have

$$D_c u^{**}(\tau, t) = f(\tau, t) + [A(\tau, t)I + T]u^{**}(\tau, t).$$

Relations (31) are used in deriving equation (30). Thus, $u^{**}(\tau, t)$ satisfies equation (11) and is its solution.

The convergence of improper integrals is ensured by i_7) and the limitation of the function $f(\tau, t)$, and the convergence is uniform with respect to the vector t and the resolving operator $\Phi(\xi, \tau, t)$ has the property of ω -periodicity by t .

The (θ, ω) -periodicity of the solution (29) with respect to independent variables is derived from the multiperiodic property of the characteristic $h(\xi, \tau, t)$ and the property (12) of the vector function $f(\tau, t)$. By shifting τ by the period θ , substituting $\xi = s + \theta$, and accounting for the periodicity of both $f(\tau, t)$ and $h(\xi, \tau, t)$, the main requirement for the vector-function $u^{**}(\tau, t)$ is satisfied: (θ, ω) -periodicity with respect to (τ, t) . The uniqueness of the solution is proved by the method of the opposite.

Q.E.D.

Note some properties of the function $u^{**}(\tau, t)$.

1. The function $u^{**}(\tau, t)$ satisfies equation (11).

2. It is a (θ, ω) -periodic function.

3. $\|u^{**}(\tau, t)\| \leq \int_{-\infty}^{+\infty} \|\Phi(\xi, \tau, h(\xi, \tau, t))\| \|f(\xi, h(\xi, \tau, t))\| d\xi \leq \frac{\tilde{\alpha}}{\alpha} \|f\|$,

where $\|f\| = \sup |f(\tau, t)|$.

4. The solution $u^{**}(\tau, t)$ is unique.

Remark. The obtained results can be applied to mathematical models describing processes with aftereffects, for example, in control theory and population dynamics, where the state of the system depends on its past values over a finite interval. In such models, the existence of multiperiodic solutions corresponds to stable oscillatory regimes determined by the periodic structure of the coefficients and the hereditary term.

Results and discussion

In this paper, for the integro-differential system (1) with finite hereditary effects and the operator D_c , an integral representation of the solution manifold has been constructed using the method of periodic characteristics.

By means of the transformation (4), the original system is reduced to equation (5) with kernel Q , which preserves the properties of smoothness, boundedness, and multiperiodicity. Based on this, the matricant $W(\tau_0, \tau, t)W(\tau_0, \tau, t)W(\tau_0, \tau, t)$ is defined as the solution of the integral equation (6), and its resolvent R is constructed via (7). The estimates obtained for the iterated kernels (8) ensure the convergence of the corresponding series and lead to the exponential bound for the resolvent. This, in turn, guarantees the boundedness of the resolving operator

$$\begin{aligned} \Phi(\tau_0, \tau, t) &= V(\tau_0, \tau, t)W(\tau_0, \tau, t), \quad \Phi(\tau_0, \tau, t) = V(\tau_0, \tau, t)W(\tau_0, \tau, t), \quad \Phi(\tau_0, \tau, t) \\ &= V(\tau_0, \tau, t)W(\tau_0, \tau, t), \end{aligned}$$

which satisfies conditions (i).

For the nonhomogeneous equation (11), the solution of the Cauchy problem (11), (13) is obtained in the form (21), combining the homogeneous component and the particular solution. This representation is expressed in terms of the resolving operator and reflects the influence of both the initial function and the right-hand side.

To construct a (θ, ω) -periodic solution, the resolving operator is decomposed as in (24) into Φ_- and Φ_+ , satisfying the exponential estimates (26), (27), which correspond to an exponential dichotomy. On this basis, a Green-type matrix function $\Phi(s, \tau, t)$ is introduced by (28) and its main properties are established.

Using this function, it is proved that under conditions (P), (P_1) and the non-resonance condition, equation (11) admits a unique (θ, ω) -periodic solution represented by (29). The convergence of the improper integral in (29) follows from the exponential estimate of Φ , while multiperiodicity is ensured by the properties of the characteristic $h(\xi, \tau, t)$ and the function $f(\tau, t)$.

Thus, the obtained results show that the existence of a resolving operator with properties (i) and its decomposition with exponential dichotomy provide sufficient conditions for the existence and uniqueness of a bounded multiperiodic solution of the integro-differential system with finite heredity.

Conclusion

The significant role of the equations under consideration in modeling phenomena and processes associated with aftereffects is highlighted. For a partial integro-differential equation with a specialized differentiation operator, integral representations of the solution manifold have been derived. Several properties of iterated kernels and resolvents have been established, along with the corresponding estimates. Utilizing the necessary and sufficient conditions for periodicity, sufficient conditions for the existence of a bounded and unique multiperiodic solution across all independent variables on the periodic characteristics of a system of integro-differential equations with a specialized differentiation operator have been determined.

Acknowledgments. This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19676629).

REFERENCES

- 1 Volterra, V. Theory of functionals, integral and integro-differential equations (Moscow: Dover Publications, 2005), 226.
- 2 De Bonis, M.C., Occorsio, D. Quadrature methods for integro-differential equations of Prandtl's type in weighted spaces of continuous functions. Applied Mathematics and Computation, 393, 125721 (2021). <https://doi.org/10.1016/j.amc.2020.125721>
- 3 Xie, Ch., Zhu, L., An, Ch., Shao, Y. A rapid steady/unsteady aerodynamic analysis method based on panel method coupled boundary layer theory. Aerospace Science and Technology, 165, 110494 (2025). DOI.10.1016/j.ast.2025.110494.
- 4 Doicu, A., Mishchenko, M., Efremenko, D., Trautmann, T. Spectral Spherical Harmonics Discrete Ordinate Method. Journal of Quantitative Spectroscopy & Radiative Transfer. 1–17 (2020). <https://doi.org/10.1016/j.jqsrt.2020.107386>
- 5 Yonghong, Y., Shoude, H. Almost periodic solution of delayed systems (Lotka–Volterra type). International Journal of Control, 2492–2506 (2016). <https://doi.org/10.1080/00207179.2016.1168526>.
- 6 Ros-Oton, X., Serra, J., Valdinoci, E. Pohozaev identities for anisotropic integro-differential operators. Communications in Partial Differential Equations, 1–32 (2016). <https://doi.org/10.48550/arXiv.1502.01431>.
- 7 Yuldashev, T.K. On a boundary value problem for Boussinesq type nonlinear partial integro - differential equation with reflecting argument. Lobachevskii J. Math., 41, 111–123 (2020).
- 8 Assanova, A.T. A two-point boundary value problem for a fourth order partial integro-differential equation. Lobachevskii J. Math., 42, 526–535 (2021).
- 9 Yuldashev, T.K., Kadirkulov, B.J. Boundary Value Problem for Weak Nonlinear Partial Differential Equations of Mixed Type with Fractional Hilfer Operator. Axioms, 9(2), 68 (2020). <https://doi.org/10.3390/axioms9020068>.

- 10 Bykov, Ya.V. O nekotorykh voprosakh kachestvennoi teorii integro-differentsialnykh uravnenii [On some questions of the qualitative theory of integro-differential equations]. Research on integrodifferential equations in Kyrgyzstan.-Frunze, 3–54 (1961). (in Russian)
- 11 Imanaliev, M.I. Nelineinye integro-differentsialnye uravneniia s chastnymi proizvodnymi [Nonlinear integro-differential equations with partial derivatives] (Bishkek: Ilim, 1992), 112. (in Russian)
- 12 Timoshin, S.A., Wang, Y. Periodic solutions of a population dynamics model with hysteresis. Nonlinear Analysis: Real World Applications, 77, 104050 (2024). <https://doi.org/10.1016/j.nonrwa.2023.104050>.
- 13 Veivoda, O. et al. Partial differential equations: Time – periodic solutions (Alphen aan den Rijn. Sijthoff: Noordhoff, 1981), 358.
- 14 Kadry, S., Alferov, G., Ivanov, G., Sharlay, A. Almost periodic solutions of first-order ordinary differential equations. Mathematics (MDPI), 6(9), 171 (2018). <https://doi.org/10.3390/math6090171>
- 15 Kharasakhal, V.Kh. Pochti-periodicheskie resheniia obyknovennykh differentsialnykh uravnenii [Almost periodic solutions of ordinary differential equations] (Alma-Ata: Nauka, 1970), 200. (in Russian)
- 16 Umbetzhano, D.U. Pochti mnogoperiodicheskie resheniia differentsialnykh uravnenii v chastnykh proizvodnykh [Almost multiperiodic solution of partial Differential Equations] (Alma-Ata: Nauka, 1979), 211. (in Russian)
- 17 Umbetzhano, D.U. Pochti periodicheskie resheniia evoliutsionnykh uravnenii [Almost periodic solutions of evolutionary equations] (Alma-Ata: Nauka, 1990), 184. (in Russian)
- 18 Sartabanov, Zh., Abdikalikova, G.A., Aitenova, G.M., Omarova, B.Zh., Zhumagazyev, A.Kh. Implementation of the Small Parameter Method for Studying Multiperiodic Solutions of Systems with a Diagonal Differentiation Operator. Lobachevskii journal of mathematics, 45(12), 6594-6609 (2024).
- 19 Sartabanov, Zh.A. Implementation of small parameter method for the study of multiperiodic solutions of systems with a diagonal differentiation operator. VII Congress TWMS, Turkestan, 133 (2023).
- 20 Sartabanov, Zh.A. Periodichnost kharakteristik operatora differentsirovaniia po diagonalii [Periodicity of characteristics of the diagonal differentiation operator]. Bulletin of KazNPU Abay, (2), 40–53 (2023). (in Russian)
- 21 Sartabanov, Zh.A., Aitenova, G.M., Abdikalikova, G.A. Multiperiodic solutions of quasilinear systems of integrodifferential equations with Dc-operator and ε -period of heredity. Eurasian Mathematical Journal, 13(1), 86–100 (2022).

¹Абдикаликова Г.А.,

ф.-м.ф.к., профессор, ORCID ID: 0000-0001-6280-4168

e-mail: abdikalikova.galliya@gmail.com

¹Сартабанов Ж.А.,

ф.-м.ф.д., профессор, ORCID ID: 0000-0003-2601-2678

e-mail: sartabanov42@mail.ru

^{2*}Айтенова Г.М.,

PhD, қауымдастырылған профессор, ORCID ID: 0000-0002-4572-8252

*e-mail: gulsezimaitenova@gmail.com

^{1,3}Жумагазиев А.Х.,

PhD, ORCID ID: 0000-0002-6007-3311

e-mail: azhumagazyev@zhubanov.edu.kz

¹Қ. Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе қ., Қазақстан

²М. Өтемісов атындағы Батыс Қазақстан университеті, Орал қ., Қазақстан

³Х. Досмұхамедов атындағы Атырау университеті, Атырау қ., Қазақстан

АҚЫРЛЫ-ЭРЕДИТАРЛЫ ИНТЕГРАЛДЫ-ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕР ЖҮЙЕСІ ҮШІН КӨППЕРИОДТЫ ЕСЕПТІҢ ШЕШІМІН ЗЕРТТЕУ

Аңдатпа

Интегралды-дифференциалдық теңдеулерді зерттеу қажеттілігі жаратылыстану мен техниканың бір-қатар салаларындағы көптеген қолданбалы есептермен негізделеді. Процестің немесе құбылыстың тарихын

сипаттайтын интегралды-дифференциалдық теңдеулерді зерттеу, интегралдық мүшенің қатысуын ескере отырып, В. Вольтерраның еңбектерінен бастау алады. «Тұқым қуалаушылықты», «салдарды» есепке алудың әртүрлі тәсілдері белгілі. Вольтерра типті интегралды-дифференциалдық теңдеулер жүйесінің көппериодты және квазипериодты шешімдерінің бар болуын анықтау әрдайым бірімәнді шешіле бермейтінін атап өтеміз. Арнайы дифференциалдау операторы бар дербес туындылы интегралды-дифференциалдық теңдеулер жүйесі қарастырылады. Периодтық сипаттамалар әдісін қолдану арқылы мұндай жүйелер шешімдерінің көпбейнесінің интегралдық көріністері салдарларымен бірге құрылды. Итерацияланған ядролар мен шешімдердің кейбір қасиеттері анықталып, сонымен қатар бағалаулар алынды. Көппериодты есептің Грин типті матрицалық функциясының бар болу шарттары және сәйкес бағалаулары бар интегралдық көрінісі табылды. Ақырлы эредитарлы интегралды-дифференциалдық жүйенің интегралдық көпбейнесі болатын көппериодты шешімнің бар болуы мен жалғыздығының жеткілікті шарттары алынды.

Түйін сөздер: матрицант, ядро, шешуші оператор, дихотомиялық, Грин типті функция.

¹**Абдикаликова Г.А.,**
к.ф.-м.н., профессор,
ORCID ID: 0000-0001-6280-4168
e-mail: abdikalikova.galliya@gmail.com

¹**Сартабанов Ж.А.,**
д.ф.-м.н., профессор,
ORCID ID: 0000-0003-2601-2678
e-mail: sartabanov42@mail.ru

^{2*}**Айтенова Г.М.,**
PhD, ассоциированный профессор,
ORCID ID: 0000-0002-4572-8252
*e-mail: gulsezimaitenova@gmail.com

^{1,3}**Жумагазиев А.Х.,**
PhD, ORCID ID: 0000-0002-6007-3311
e-mail: azhumagaziyev@zhubanov.edu.kz

¹Актюбинский региональный университет им. К. Жубанова, г. Актобе, Казахстан

²Западно-Казахстанский университет им. М. Утемисова, г. Уральск, Казахстан

³Атырауский университет им. Х. Досмухамедова, г. Атырау, Казахстан

ИССЛЕДОВАНИЕ МНОГОПЕРИОДИЧЕСКОЙ ЗАДАЧИ ДЛЯ КОНЕЧНО-ЭРЕДИТАРНОЙ СИСТЕМЫ ИНТЕГРО-ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Аннотация

К проблеме необходимости исследования интегро-дифференциальных уравнений приводят многочисленные задачи приложения ряда областей естествознания и техники. Исследование интегро-дифференциальных уравнений, описывающие предысторию процесса или явления, начались с работ В. Вольтерра с учетом влияния интегрального члена. Известны различные подходы учета «наследственности», «последствия». Отметим, что установление существования многопериодических и квазипериодических решений системы интегро-дифференциальных уравнений типа Вольтерра не всегда однозначно разрешимо. Рассматривается система интегро-дифференциальных уравнений в частных производных со специальным оператором дифференцирования. Используя метод периодических характеристик, построены интегральные представления многообразия решений таких систем с последствием. Установлены некоторые свойства итерированных ядер и резольвенты, а также получены оценки. Найдены условия существования матричной функции типа Грина многопериодической задачи и интегральное представление с соответствующей оценкой. Получены достаточные условия существования и единственности многопериодического решения, которые являются интегральным многообразием интегро-дифференциальной системы с конечной эредитарностью.

Ключевые слова: матрицант, ядро, разрешающий оператор, дихотомичность, функция типа Грина.

UDC 510.5

IRSTI 27.03.45

<https://doi.org/10.55452/1998-6688-2026-23-2-46-52>¹**Kalmurzayev B.S.,**

PhD, Associate Professor, ORCID ID: 0000-0002-4386-5915,

e-mail: birzhan.kalmurzayev@gmail.com

¹**Nurlanbek D.D.,**

PhD student, ORCID ID: 0000-0002-1275-1413,

e-mail: nurlanbek.dias21@gmail.com

²**Bazhenov N.A.**

PhD in Mathematical Logic, Algebra and Number Theory, ORCID ID: 0000-0002-5834-2770,

e-mail: nickbazh@yandex.ru

¹Kazakh-British Technical University, Almaty, Kazakhstan²Sobolev Institute of Mathematics, Novosibirsk, Russia**ON UNIVERSAL NUMBERINGS
OF TWO-ELEMENT FAMILIES IN THE ERSHOV HIERARCHY****Abstract**

The study of local and global invariants of the Rogers semilattice is an important and fundamental problem in numbering theory and computability theory. Global invariants include properties such as an existence of a universal numbering, the number of minimal numberings, the cardinality of the entire semilattice, and a criterion for determining whether a semilattice is a lattice. Local invariants, in turn, describe structures, such as initial segments or intervals within the semilattice. We say that a numbering $\nu \in \text{Com}(S)$ is universal if any other numbering $\mu \in \text{Com}(S)$ reduces to ν . The study of universal numberings is important for understanding the structure of semilattices and their classification. In this paper, an existence of universal numberings is considered for finite families of computably enumerable sets located at finite levels of the Ershov hierarchy. The main result is that for any two-element family of computably enumerable sets S , its Rogers semilattice, considered at the third level of the Ershov hierarchy, has universal numberings.

Keywords: computable numberings, universal numberings, Rogers semilattice, Ershov hierarchy.*Received March 17, 2026; revised March 30, April 28, 2026; accepted April 30, 2026.***Introduction**

The study of local and global invariants of the Rogers semilattice is one of the central problems of numbering theory. Among the global invariants are an existence of a universal numbering, the number of minimal numberings, the cardinality of the semilattice, and its lattice property. Local invariants characterize the structure of initial segments and intervals, in particular, their finiteness or density.

One of the important global invariants is an existence of a universal numbering. Recall that a numbering is called universal if any other numbering in the same family can be reduced to it via a computable function. In other words, there exists an algorithm that effectively reproduces all other algorithms that enumerate a given family. For families of computably enumerable (c.e.) sets, an existence of a universal numbering was established in the work of Yu. L. Ershov [1].

Algorithmic enumerations of elements of Σ_1^{-1} -sets are monotonic: if an element is enumerated into a set at some step of the computation, it remains in this set thereafter. Therefore, the growth of the set being enumerated is monotonic. The same thing happens in enumerations of Σ_n^0 -sets for any level n ; the only difference from Σ_1^{-1} -enumerations is that suitable oracles are used. In algorithms for enumerating Σ_n^{-1} -sets for $n > 1$, monotonicity is absent: initially, no number is present in the set; at some step of the algorithm, a number may be enumerated into the set; this is the first change

in the number's status; later, this number may be extracted from the set; this is the second change; then, possibly, the number is enumerated into the set again, this is the third change; and so on. Sets in the Ershov hierarchy are classified by the number of "mind changes" of the procedures of inclusion in the set or exclusion from the set of each of the elements of the set [2]. For all numbers in the set from the class Σ_n^{-1} , this number of mind changes is bounded by the number n . Thus, the class Σ_1^{-1} consists of computably enumerable sets, and the class Σ_2^{-1} consists of differences of computably enumerable sets. Even the first studies of computable numberings of families of sets from the class Σ_2^{-1} showed the difference between the fundamental invariants of their Rogers semilattices compared to the Rogers semilattices of families of Σ_1^{-1} sets.

By S. Badaev and Zh. Talasbayeva, a family was constructed consisting of one computably enumerable set and one Σ_2^{-1} -set, related by an inclusion relation, which had a one-element Rogers semilattice [3]. This is impossible in the classical case: the Rogers semilattice of any finite family of Σ_1^{-1} -sets containing at least one pair of nested sets is infinite and contains both the greatest and the least elements. At the same time, there exists a finite family of Σ_2^{-1} -sets whose Rogers semilattice does not have a greatest element constructed by K. Abeshev [4].

The main result of the paper is the following statement. For any two-element family of computably enumerable sets belonging to an odd level of the Ershov hierarchy, there exists a universal numbering.

The paper consists of an introduction, preliminaries and one chapter, which provides a complete proof of an existence of a universal numbering for the two-element family of computably enumerable sets on the third level of the Ershov hierarchy.

Materials and methods

In this paper we will follow to the notations and terminology adopted in [1], [5]. We define the pair $\langle \cdot, \cdot \rangle$ as a computable bijection from ω^2 to ω that is called Cantor's pairing function

$$\langle x, y \rangle = \frac{(x + y)^2 + 3x + y}{2}$$

For any $x, y \in \omega$ we can define the functions l and r , such that $x = \langle l(x), r(x) \rangle$, $l(\langle x, y \rangle) = x$ and $r(\langle x, y \rangle) = y$.

Let n be a non-zero natural number. The class Σ_n^{-1} of the Ershov hierarchy contains all n -c.e. sets. A set $B \subseteq \omega$ is n -c.e. (or n -computably enumerable) if $B = \lim_s A_s$ for a uniformly computable sequence of sets A_s such that $A_0 = \emptyset$, and for each $x \in \omega$, there are at most n -many s satisfying $A_s(x) \neq A_{s+1}(x)$. Here we say that the function $f(x, s) = A_s(x)$ is an n -approximation of the n -c.e. set B .

A numbering ν of some family $\mathcal{F} \in \Sigma_n^{-1}$ is called a Σ_n^{-1} -computable numbering if the set

$$\{(x, k) : x \in \nu(k), k \in \omega\}$$

is a n -c.e. set. We say that a numbering ν is reducible to a numbering μ (denoted as $\nu \leq \mu$) if there is a computable function f such that $\nu = \mu \circ f$. We say that ν and μ are equivalent ($\nu \equiv \mu$) if $\nu \leq \mu$ and $\mu \leq \nu$. $\text{Com}_n^{-1}(\mathcal{S})$ denotes the set of Σ_n^{-1} -computable numberings of the family $\mathcal{S}$. For numberings ν and μ their join $(\nu \oplus \mu)$ is defined as

$$(\nu \oplus \mu)(2k) = \nu(k) \text{ and } (\nu \oplus \mu)(2k + 1) = \mu(k).$$

The quotient structure $\mathcal{R}_n^{-1}(\mathcal{S}) = \{\text{Com}_n^{-1} \setminus \equiv, \leq, \oplus\}$ is called a Rogers semilattice of the Σ_n^{-1} -computable family $\mathcal{S}$.

A numbering ν is called Σ_n^{-1} -universal numbering if $\nu \in \text{Com}_n^{-1}(\mathcal{S})$ and any numbering $\mu \in \text{Com}_n^{-1}(\mathcal{S})$ is reduced to ν .

Results and discussion

In this section we will consider two-element family of c.e. sets $\mathcal{S}$.

Theorem 1. For any two-element family of c.e. sets $\mathcal{S} = \{A, B\}$, the Rogers semilattice $\mathcal{R}_3^{-1}(\mathcal{S})$ has a universal numbering.

Proof. We divide the proof into two cases:

Case 1: one of the sets is a subset of the other. Without loss of generality, we can assume that $A \subseteq B$. In this case, we fix some element $b \in B \setminus A$ and computable approximations $(A_s)_{s \in \omega}$ and $(B_s)_{s \in \omega}$ for c.e. sets A and B , respectively. Moreover, in this case, we can assume that $A_s \subseteq B_s$ for all $s \in \omega$. Then we define the numbering $v(x) := \lim_s V_{x,s}$ where

$$V_{x,s} = \begin{cases} A_s, & \text{if } b \notin \pi_s(x), \\ B_s, & \text{if } b \in \pi_s(x). \end{cases}$$

by $(\pi_s(x))_{s \in \omega}$ we define approximation of a universal numbering of all 3-c.e. sets.

Lemma 2. The constructed numbering $v \in \text{Com}_3^{-1}(\mathcal{S})$.

Proof. For any $x \in \omega$, the sequence $V_{x,s}$ coincides with either A_s or B_s at each step. Therefore, three options are possible:

$$\lim_s V_{x,s} = A;$$

$$\lim_s V_{x,s} = B;$$

the limit does not exist.

The third option is impossible, since it would imply an infinite number of changes in the status of the element b in the set $\pi(x)$, which contradicts the 3-computability of $\pi(x)$.

This means we have only two options: - $v(x) = A$ if $b \notin \pi(x)$ and - $v(x) = B$ if $b \in \pi(x)$. Therefore, v is the numbering of the family $\mathcal{S} = \{A, B\}$.

Now we prove that for a fixed x the limit $\lim_s V_{x,s}$ is a 3-c.e. set. Let z be an arbitrary element. Consider the cases.

1. $z \notin A$ and $z \notin B$. In this case z will not be enumerated into $V_{x,s}$. No mind-changes.

2. $z \in B \setminus A$.

a. $\lim_s V_{x,s} = A$. The element b either never appears in $\pi(x)$ or appears at some step s_0 and then will be permanently removed on step s_1 . If b never appears, then $V_{x,s} = A_s$ at all steps, and z never appears. If b appears at step s_0 and is permanently removed at step s_1 , then z may appear before s_1 (if it is already listed in B) and will be permanently removed at step s_1 . In this case, there are at most two changes.

b. $\lim_s V_{x,s} = B$. There exist steps $s_0 < s_1 < s_2$ such that b first enters π at step s_0 , is then removed at step s_1 , and finally enters at step s_2 . If z already belongs to B before s_1 , changes are possible at steps s_1 and s_2 . If z enters B after s_1 , it is added once and is not removed again. In all cases, there are no more than three changes.

$z \in A$ and $z \in B$.

a. $\lim_s V_{x,s} = A$. If b never occurs in $\pi(x)$, then $V_{x,s} = A_s$, and z appears once. If b occurs at step s_0 and is removed at step s_1 , then:

♦ if z enters A before s_0 , it can be temporarily removed and then added again;

♦ if z first enters B , it can be removed at s_1 and then added again later. In any case, there are at most three changes.

b. $\lim_s V_{x,s} = B$. Again, there are steps $s_0 < s_1 < s_2$. If z enters A before the corresponding switches, then, since $A_s \subseteq B_s$, it is no longer removed. If z first enters B , changes are possible at steps s_0, s_1, s_2 . The number of changes does not exceed three.

Since $V_{x,s}$ is constructed uniformly in x, s , the numbering v is 3-computable.

Lemma 3. The numbering v is a universal numbering in $\mathcal{R}_3^{-1}(\mathcal{S})$.

Proof. Since π is universal, for every computable numbering $\mu \in \text{Com}_3^{-1}(\mathcal{S})$ there exists a computable function g such that

$$\mu(x) = \pi(g(x)).$$

By construction, $v(g(x)) = \pi(g(x))$ and $v(g(x)) = \mu(x)$ for all x . Hence $\mu \leq v$, and therefore v is universal in $\mathcal{R}_3^{-1}(\mathcal{S})$. $\square$

Case 2. Suppose that A is not a subset of B , and vice versa. Fix elements $a \in A \setminus B$ and $b \in B \setminus A$.

Let π_e be a fixed universal computable numbering of all 3-c.e. families. For all e we will build the numbering v_e and function f_e .

Construction.

Stage 0. For all x we set $v_{e,0}(x) = \emptyset$ and $f_{e,0}(x) \uparrow$.

Stage $s + 1$. Let $l(s + 1) = y$ at each step. Check is $f_{e,s}(y) \downarrow$?

1. If $f_{e,s}(y)$ is already defined, let $f_{e,s}(y) = i$. Then we check what stage the module for (e, y) is at, and continue its execution.

2. If $f_{e,s}(y)$ is undefined, then we check whether stage s marks the first occurrence of one of the markers a or b in $\pi_e(y)$, that is:

- ♦ either $a \in \pi_{e,s}(y)$ and previously $a \notin \pi_{e,t}(y)$ and $b \notin \pi_{e,t}(y)$ for all $t < s$,
- ♦ or $b \in \pi_{e,s}(y)$ and previously $b \notin \pi_{e,t}(y)$ and $a \notin \pi_{e,t}(y)$ for all $t < s$.

If one of them occurs for the first time, then we fix a new index i , define $f_{e,s+1}(y) = i$ and run the module for the pair (e, y) with this index i .

3. If none of these events occurred, then we set $v_{e,s+1}(y) = A_{s+1}$ and keep the function $f_{e,s+1}(y) \uparrow$.

Description of the module for (e, y) with index i .

1. Wait for a stage s_0 such that either $a \in \pi_{e,s_0}(y)$ or $b \in \pi_{e,s_0}(y)$. Until such a stage appears, keep $v_{e,s}(i) = A_s$.

2. Suppose $a \in \pi_{e,s_0}(y)$. Set $v_{e,s_0}(i) = A_{s_0}$. Wait for a stage $s_1 > s_0$ such that $a \notin \pi_{e,s_1}(y)$ and $b \in \pi_{e,s_1}(y)$.

3. At stage s_1 , redefine

$$v_{e,s_1}(i) = B_{s_1},$$

that is, remove all elements of $A_{s_1} \setminus B_{s_1}$. Then wait for a stage $s_2 > s_1$ such that $b \notin \pi_{e,s_2}(y)$ and $a \in \pi_{e,s_2}(y)$.

4. At stage s_2 , redefine

$$v_{e,s_2}(i) = A_{s_2},$$

that is, remove all elements of $B_{s_2} \setminus A_{s_2}$, and for all steps $t > s_2$ $v_{e,t}(i) = A_t$.

If at stage s_0 we instead have $b \in \pi_{e,s_0}(y)$, the construction is defined symmetrically, interchanging the roles of a and b , and of A and B .

We now analyze the possible outcomes.

1. If the construction waits forever at step (1), then neither a nor b ever enumerates in $\pi_e(y)$. In this case $v_e(i) = A$.

2. If the construction waits forever at step (2), then $a \in \pi_e(y)$ and $b \notin \pi_e(y)$. Hence $\pi_e(y) = A$, and $v_e(i) = A$.

3. If the construction waits forever at step (3), then $b \in \pi_e(y)$ and $a \notin \pi_e(y)$. Hence $\pi_e(y) = B$, and $v_e(i) = B$.

4. If the construction reaches step (4), then the last redefinition yields $v_e(i) = A$. In this case $\pi_e(y) = B$, but according to the strategy the numbering $\pi_e(y)$ will switch back to A .

The numbering v we will define as follows: $v((e, x)) = v_e(x)$. $\square$

Lemma 4. The constructed numbering $v \in \text{Com}_3^{-1}(\mathcal{S})$.

Proof. For any $x \in \omega$, the sequence $v_{e,s}(x)$ coincides with either A_s or B_s at each step. Therefore, three options are possible:

1. $\lim_s v_{e,s}(x) = A$;
2. $\lim_s v_{e,s}(x) = B$;
3. the limit does not exist.

The third option is impossible, since it would imply an infinite number of changes in the status of the elements a and b in the set $\pi(x)$, which contradicts the 3-computability of $\pi(x)$.

This means we have only two options: - $v(x) = A$ if $a \in \pi(x)$ and - $v(x) = B$ if $b \in \pi(x)$, it is impossible to have both $a \in \pi(x)$ and $b \in \pi(x)$ because these elements distinguishes the sets A and B . Therefore, v is a numbering of the family $S = \{A, B\}$.

Now we prove that for a fixed x the limit $\lim_s v_s(x)$ is a 3-c.e. set. Let z be an arbitrary element. Consider the cases.

1. $z \notin A$ and $z \notin B$. In this case z will not be enumerated into $v_s(x)$. No mind-changes.
2. $z \in B \setminus A$.
 - a. $\lim_s v_{e,s}(x) = A$. The element b either never appears in $\pi(x)$ or appears at some step s_0 and then will be permanently removed on step s_1 and the element a can be enumerated on step s_3 , removed on s_4 and enumerated forever on some s_5 . If b never appears, then $v_s(x) = A_s$ at all steps, and z never appears. If b appears at step s_0 and is permanently removed at step s_1 , then z may appear before s_1 (if it is already listed in B), and will be permanently removed at step s_1 . Or it can be enumerated after step s_1 , then it will not appear in v . In this case, there are at most two changes.
 - b. $\lim_s v_{e,s}(x) = B$. There exist steps $s_0 < s_1 < s_2$ such that b first enters π at step s_0 , is then removed at step s_1 , and finally enters at step s_2 . And there exist steps $s_3 < s_4 < s_5$ such that a first enters π at step s_3 , is then removed at step s_4 , and finally enters at step s_5 . If z already belongs to B before s_1 , changes are possible at steps s_1 and s_2 . If z enters B after s_1 , it is added once and is not removed again. In all cases, there are no more than three changes.
3. $z \in A \setminus B$. Same as the previous one.
4. $z \in A$ and $z \in B$.
 - a. $\lim_s v_{e,s}(x) = A$. If b never occurs in $\pi(x)$, then $v_{e,s}(x) = A_s$, and z appears at step s_3 can be removed on s_4 and will be enumerated forever at s_5 , so at most three mind-changes. If b occurs at step s_0 and is removed at step s_1 , then:
 - ♦ if z enters A before s_0 , it can be temporarily removed and then added again;
 - ♦ if z first enters B , it can be removed at s_1 and then added again later, since it's in intersection of sets, it will never be removed when we see $z \in A$ and $z \in B$. In any case, there are at most three changes.
 - b. $\lim_s v_{e,s}(x) = B$. If a never occurs in $\pi(x)$, then $v_{e,s}(x) = B_s$, and z appears at step s_0 can be removed on s_1 and will be enumerated forever at s_2 , so at most three mind-changes. If a occurs at step s_3 and is removed at step s_4 , then:
 - ♦ if z enters B before s_3 , it can be temporarily removed and then added again;
 - ♦ if z first enters A , it can be removed at s_4 and then added again later, since it's in intersection of sets, it will never be removed when we see $z \in A$ and $z \in B$. In any case, there are at most three changes.

Lemma 5. If the numbering $\pi_e \in \text{Com}_3^{-1}(\mathcal{S})$, then the function f_e is total.

Proof. Let π_e enumerates the family $\mathcal{S}$. By construction, for each x , there exists a step s at which either the set A or the set B corresponding to the index x appears in π_e . At this point, the module (e, x) is activated, which determines the value of $f_e(x)$.

Therefore, for any x , the value of $f_e(x)$ is determined at some finite step, meaning that the function f_e is defined everywhere. Thus, f_e is a total function. $\square$

Lemma 6. If the numbering $\pi_e \in \text{Com}_3^{-1}(\mathcal{S})$, then $\forall x \pi_e(x) = v_e(f_e(x))$.

Proof. Consider the module (e, x) (see module analysis). By construction, if $\pi_e(x)$ enumerates the set A , then the module sets the value of $f_e(x)$ in such a way that $v_e(f_e(x))$ also enumerates A . Similarly, if $\pi_e(x)$ enumerates the set B , then $v_e(f_e(x))$ enumerates B .

Thus, in any case, we obtain the equality $\pi_e(x) = v_e(f_e(x))$. $\square$

Lemma 7. The numbering v is a universal numbering in $\mathcal{R}_3^{-1}(\mathcal{S})$.

Proof. Let $\pi_{\mathcal{S}} \in \text{Com}_3^{-1}(\mathcal{S})$. Then by Lemma 6 there exists a total function $f_{\mathcal{S}}$ such that for all x $\pi_{\mathcal{S}}(x) = v_{\mathcal{S}}(f_{\mathcal{S}}(x))$.

Consequently, $\pi_{\mathcal{S}}(x) = v(\langle e, f_{\mathcal{S}}(x) \rangle)$.

Since $f_{\mathcal{S}}$ is a total function, $v(\langle e, f_{\mathcal{S}} \rangle)$ defines a computable reducibility of $\pi_{\mathcal{S}}$ to v . Therefore, v is a universal numbering in $\mathcal{R}_3^{-1}(\mathcal{S})$. $\square$

Using the same module construction and the same reduction scheme, the result generalizes to all $2n + 1$ levels of the Ershov hierarchy: at each odd level, a universal numbering of the corresponding class is constructed by means of a similar parametrization and uniform reduction.

Corollary 8. For any family $\mathcal{S} = \{A, B\}$, where A, B are c.e. sets, $\mathcal{R}_{2n+1}^{-1}(\mathcal{S})$ has a universal numbering for every n .

Conclusion

This paper shows an existence of a universal numberings for the Rogers semilattice of two-element families of c.e. sets on odd levels of the Ershov hierarchy.

REFERENCES

- 1 Ershov, Y.L. Teoriya numeratsii [The Theory of Numberings] (Moscow: Nauka, 1977). (In Russian).
- 2 Ash, C.J., and Knight, J.F. Computable Structures and the Hyperarithmetical Hierarchy (Elsevier Science, 2000).
- 3 Badaev, S.A., and Talasbaeva, Zh.T. Computable numberings in the hierarchy of Ershov. In: Mathematical Logic in Asia, edited by S.S. Goncharov et al. (Singapore: World Scientific, 2006), pp. 17–30.
- 4 Abeshev, K.Sh. On the existence of universal numberings for finite families of d.c.e. sets. Mathematical Logic Quarterly, 60 (3), 161–167 (2014).
- 5 Soare, R.I. Recursively Enumerable Sets and Degrees: A Study of Computable Functions and Computably Generated Sets (Berlin: Springer Science & Business Media, 1999).

¹Калмурзаев Б.С.,

PhD, қауымдастырылған профессор, ORCID ID: 0000-0002-4386-5915,
e-mail: birzhan.kalmurzayev@gmail.com

¹Нұрланбек Д.Д.,

докторант, ORCID ID: 0000-0002-1275-1413,
e-mail: nurlanbek.dias21@gmail.com

²Баженев Н.А.,

физика-математика ғылымдарының кандидаты, ORCID ID: 0000-0002-5834-2770,
e-mail: nickbzh@yandex.ru

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

²Соболев атындағы математика институты, Новосибирск қ., Ресей

ЕРШОВ ИЕРАРХИЯСЫНДАҒЫ ЕКІ ЭЛЕМЕНТТІ ЖИЫНДАР ҮЙІРЛЕРІ ҮШІН УНИВЕРСАЛ НӨМІРЛЕУЛЕР ТУРАЛЫ

Аңдатпа

Роджерс жарты торының локалды және глобалды инварианттарын зерттеу – нөмірлеу теориясындағы маңызды әрі іргелі мәселелердің бірі. Глобалды инварианттарға универсал нөмірлеудің болуы, минималды нөмірлеулер саны, бүкіл жарты тордың қуаттылығы және жарты тордың тор болу критерийі сияқты қасиеттер жатады. Ал локалды инварианттар

жарты тор ішіндегі бастапқы сегменттер мен аралықтар тәрізді құрылымдарды сипаттайды. Егер кез келген $\mu \in Com(S)$ нөмірлеуі $\nu \in Com(S)$ нөмірлеуіне көшірілсе, онда ν нөмірлеуі универсал балады. Универсал нөмірлеуді зерттеу жарты торлардың құрылымын және олардың жіктелуін түсіну үшін маңызды. Мақалада Ершов иерархиясының шекті деңгейлерінде орналасқан есептелімді жиындардың шекті үйірлері үшін универсал нөмірлеулердің болуы қарастырылады. Негізгі нәтиже – есептелімді жиындардың кез келген екі элементті S үйірі үшін Ершов иерархиясының үшінші деңгейінде қарастырылатын Роджерс жарты торында универсал нөмірлеудің бар екендігі дәлелденеді.

Түйін сөздер: есептелімді нөмірлеулер, универсал нөмірлеулер, Роджерс жарты торы, Ершов иерархиясы.

¹Калмурзаев Б.С.,

PhD, ассоциированный профессор, ORCID ID: 0000-0002-4386-5915,
e-mail: birzhan.kalmurzayev@gmail.com

¹Нұрланбек Д.Д.,

докторант, ORCID ID: 0000-0002-1275-1413,
e-mail: nurlanbek.dias21@gmail.com

²Баженов Н.А.,

кандидат физико-математических наук, ORCID ID: 0000-0002-5834-2770,
e-mail: nickbazh@yandex.ru

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

²Институт математики им. Соболева, г. Новосибирск, Россия

ОБ УНИВЕРСАЛЬНЫХ НУМЕРАЦИЯХ ДЛЯ ДВУХЭЛЕМЕНТНЫХ СЕМЕЙСТВ В ИЕРАХИИ ЕРШОВА

Аннотация

Изучение локальных и глобальных инвариантов полурешетки Роджерса является важной и фундаментальной задачей в теории нумерации и теории вычислимости. Глобальные инварианты включают такие свойства, как существование универсальной нумерации, число минимальных нумераций, мощность всей полурешетки и критерий определения того, является ли полурешетка решеткой. Локальные инварианты, в свою очередь, описывают структуры, такие как начальные сегменты или интервалы внутри полурешетки. Мы говорим, что нумерация $\nu \in Com(S)$ является универсальной, если любая другая нумерация $\mu \in Com(S)$ сводится к ν . Изучение универсальных нумераций важно для понимания структуры полурешеток и их классификации. В данной работе рассматривается существование универсальных нумераций для конечных семейств вычислимо перечислимых множеств, расположенных на конечных уровнях иерархии Ершова. Основной результат заключается в том, что для любого семейства вычислимо перечислимых множеств S , состоящего из двух элементов, его полурешетка Роджерса, рассматриваемая на третьем уровне иерархии Ершова, имеет универсальную нумерацию.

Ключевые слова: вычислимые нумерации, универсальные нумерации, полурешетка Роджерса, иерархия Ершова.

ӘОЖ 517.927
FTAXP 27.31.15

<https://doi.org/10.55452/1998-6688-2026-23-2-53-60>

^{1,2*}**Иманбаев Н.С.,**

физ.-мат.ғ.к., профессор, ORCID ID: 0000-0002-5220-9899,

*e-mail: imanbaevnur@mail.ru

^{2,3}**Сайрам Н.Н.,**

магистр, докторант, ORCID ID: 0009-0004-9766-6219,

e-mail: nurgul.sairam02@mail.ru

¹Ж.А. Ташенев атындағы университет, Шымкент қ., Қазақстан

²Математика және математикалық модельдеу институты, Алматы қ., Қазақстан

³Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

ЛОКАЛДЫ ЕМЕС «СЕРПІМДІ» ШЕТТІК ШАРТТАРМЕН БЕРІЛГЕН ЕСЕЛІ ДИФФЕРЕНЦИАЛДАУ ШТУРМ-ЛИУВИЛЛ ОПЕРАТОРЫНЫҢ МЕНШІКТІ ЖӘНЕ ҚОСАРЛАНҒАН ФУНКЦИЯЛАР ЖҮЙЕСІНІҢ БАЗИСТІЛІГІ ТУРАЛЫ

Андатпа

Мақала сегментте анықталған локалды емес интегралдық «серпімді» шеттік шарттары бар Штурм–Лиувилл еселі дифференциалдау операторы үшін қойылған спектралдық есепті зерттеуге арналған. Қарастырылатын шеттік шарттар регулярлы болғанымен, күшейтілмеген регулярлы класқа жататындықтан, мұндай есептерді зерттеу спектралдық теория тұрғысынан елеулі теориялық қиындықтар туындатады және классикалық әдістерді тікелей қолдануға мүмкіндік бермейді. Зерттеудің негізгі мақсаты квадраттары қосындыланатын функциялар кеңістігінде меншікті және қосарланған функциялар жүйесінің базистік қасиеттерін анықтау, сондай-ақ олардың орнықтылығы мен орнықсыздығын жан-жақты талдау болып табылады. Жұмыста интегралдық «серпімді» Самарский–Ионкин типіндегі есеп дербес жағдай ретінде қарастырылып, оның меншікті және қосарланған функциялар жүйесінің Рисс базисін құрайтындығы дәлелденеді. Сонымен қатар, шеттік шарттардағы интегралдық ядроның шамалы өзгеруі кезінде базистік қасиеттердің сақталу немесе бұзылу шарттары айқындалады. Рисс базисін құрайтын түбірлік векторлар жүйесі мен жәй базисте құрмайтын меншікті және қосарланған функциялар жүйелерінің жиындарының L_1 кеңістігіндегі тығыздығы дәлелденеді. Алынған нәтижелер локалды емес шеттік шарттары бар дифференциалдық операторлардың спектралдық теориясын дамытуда, сондай-ақ локалды емес әсерлерді ескеретін қолданбалы есептерді зерттеуде пайдаланылуы мүмкін.

Түйін сөздер: Штурм–Лиувилл операторы, бейлокалды шеттік шарттар, интегралдық серпімді шарттар, Самарский–Ионкин есебі, Рисс базис, меншікті және қосарланған функциялар, орнықтылық.

Кіріспе

Локалды емес шеттік шарттары бар спектралдық есептер қазіргі заманғы дифференциалдық операторлар теориясының маңызды бағыттарының бірі болып табылады. Мұндай есептер әртүрлі физикалық, инженерлік және қолданбалы модельдерде кездесетіндіктен, олардың спектралдық қасиеттерін зерттеу теориялық тұрғыдан да, практикалық тұрғыдан да өзекті [1, 7]. Алайда локалды емес шеттік шарттардың болуы оператордың өзіне-өзі түйіндес еместігіне алып келеді, ал бұл жағдай классикалық спектралдық теория әдістерін тікелей қолдануға мүмкіндік бермейді [1].

Күшейтілмеген регулярлы шеттік шарттар жағдайында, тіпті екінші ретті жәй дифференциалдық теңдеулер үшін де меншікті және қосарланған функциялар жүйесінің базистік қасиеттері толық сипатталмаған. Бұл мәселенің күрделілігі шеттік шарттар мен теңдеу коэффициенттерінің өзара байланысына тәуелді екені белгілі. Атап айтқанда, шеттік

шарттардың коэффициенттері арқылы түбірлік функциялар жүйесінің базис құруы жалпы жағдайда мүмкін еместігі көрсетілген [1]. Алайда тендеу коэффициенттерінің әсерін ескере отырып, дербес жағдайларда оң нәтижелер алынған.

Коэффициенттері бар екінші ретті дифференциалдық тендеулер үшін күшейтілмеген регулярлы шеттік шарттар қойылған есептерде меншікті және қосарланған функциялар жүйесінің Рисс базисін құрайтын арнайы кластары бөлініп алынған [2]. Ал қарапайым жағдайда, яғни қосымша коэффициенттерсіз берілген екінші ретті дифференциалдық операторлар үшін күшейтілмеген регулярлы шеттік шарттармен берілген спектралдық есептердің базистік қасиеттері толық зерттелген [3].

Сонымен қатар, жалпыланған шеттік шарттары бар Лаплас операторы үшін периодтық және Самарский–Ионкин типіндегі шеттік шарттарды қамтитын спектралдық есептер бір-қатар еңбектерде қарастырылған [4, 5]. Жұлдыз тәрізді графта анықталған Штурм–Лиувилл операторының спектралдық сипаттамалары да жеке зерттеу объектісі ретінде белгілі [6]. Бұл нәтижелер локалды емес интегралдық шеттік шарттары бар операторлардың спектралдық теориясын одан әрі дамыту қажеттігін көрсетеді.

Осы мақалада регулярлы, бірақ күшейтілмеген регулярлы интегралдық «серпімді» шеттік шарттармен берілген Штурм–Лиувилл еселі дифференциалдау операторы үшін қойылған спектралдық есеп қарастырылады. Бұл жұмыста меншікті және қосарланған функциялар жүйесінің базистік қасиеттеріне, сондай-ақ олардың орнықтылығы мен орнықсыздығына баса назар аударылады.

Материалдар мен әдістер

$$Ly = -y''(x) + q(x)y(x) = \lambda y(x), \quad 0 < x < 1, \quad q(x) \in C[0,1] \quad (1.1)$$

дифференциалдық амалынан туындайтын

$$Y_1(y) \equiv a_{11}y'(0) + a_{12}y'(1) + a_{13}y(0) + a_{14}y(1) = 0, \quad (1.2)$$

$$Y_2(y) \equiv a_{23}y(0) + a_{24}y(1) = \int_0^1 \overline{p(x)}y(x)dx, \quad p(x) \in L_2(0,1)$$

регулярлы, бірақ күшейтілмеген регулярлы [7], интегралдық «серпімді» шеттік шарттармен берілген Штурм–Лиувилл еселі дифференциалдау операторын қарастырамыз. Мұндағы $Y_1(y)$, $Y_2(y)$ – күшейтілмеген регулярлық

$$|a_{11}| + |a_{12}| \neq 0, \quad a_{11}a_{24} + a_{12}a_{23} = \pm[a_{11}a_{23} + a_{12}a_{24}] \neq 0$$

шарттарды қанағаттандыратын, комплекс мәнді тұрақты коэффициенттермен берілген сызықтық тәуелсіз формалар.

[8] – мақалада жәй дифференциалдық оператор үшін кез-келген интегралдық «серпімді» шеттік шарттарда түбірлік векторлар жүйесінің жақшалы Рисс базисін құрайтындығы, ал күшейтілген регулярлы шеттік шарттар жағдайында меншікті және қосарланған функциялар жүйесінің Рисс базистілігі дәлелденген болатын. Дербес жағдайлардағы еселі дифференциалдау операторының интегралдық «серпімді» периодтық шеттік шарттармен берілген спектралдық есебі үшін меншікті және қосарланған функциялар жүйесінің базистілік қасиеттерінің орнықсыздығы [2]-ші мақалада зерттелген. [9; 10] біздің зерттеулерімізде $p(0)v(0)$ түріндегі жүктелген бірінші және екінші ретті дифференциалдық оператордың түбірлік векторларының базистілік қасиеттерінің орнықтылығы қарастырылды.

Нәтижелер мен талқылау

Теорема 1. Айталық $(p(x) \equiv 0)$ болған жағдайдағы «серпімсіз» L_0 операторының λ_k^0 – меншікті мәндері болсын және сәйкес меншікті-қосарланған функциялар жүйесі (МҚФ) $L_2(0, 1)$ кеңістігінде Рисс базисін құрсын делік. Онда (2) – «серпімді» регулярлы шеттік шарттармен берілген L операторының сипаттаушы (характеристикалық) анықтаушы келесі түрде болады:

$$\Delta(\lambda) = \Delta_0(\lambda) \cdot \left[1 - \sum_{k=1}^{\infty} \overline{a_{k_0}} \left(\frac{V_3(v_{k_0})}{\lambda - \lambda_k^0} - \frac{\sqrt{\lambda_k^0 V_3(v_{k_1})}}{(\lambda - \lambda_k^0)^2} \right) - \sum_{k=1}^{\infty} \overline{a_{k_1}} \left(\frac{V_3(v_{k_0}^0)}{\lambda - \lambda_k^0} \right) \right]$$

мұндағы $\Delta_0(\lambda) - L_0$ «серпімсіз» оператордың сипаттаушы (характеристикалық) анықтаушы; $V_3(v)$ – сызықтық біртекті формалар, яғни «серпімсіз» түйіндес L_0^* операторының шеттік шарттарын құру барысында пайда болады;

$\{v_{k_1}^0, v_{k_0}^0\}$ – «серпімсіз» түйіндес есептің меншікті және қосарланған функцияларының жүйесі; a_{k_1}, a_{k_0} – $\overline{p(x)}$ функциясының

$$\overline{p(x)} = \sum_{k=1}^{\infty} [\overline{a_{k_1}} v_{k_1}^0 + \overline{a_{k_0}} v_{k_0}^0] \quad (2.2)$$

жүйесі бойынша биортогоналды жіктелуінің Фурье коэффициенттері.

(1.1)–(1.2) есебінің сипаттаушы (характеристикалық) анықтаушының (2.1) формуласының негізінде әртүрлі күшейтілмеген регулярлы шеттік шарттардың жағдайы үшін меншікті мәндерінің асимптотикасын аламыз және меншікті және қосарланған функциялар жүйесінің базистілік қасиеттерінің орнықтылығы мен орнықсыздығы жайында қорытынды жасай аламыз.

Еселі дифференциалдау операторы үшін «серпімді» Самарский-Ионкин есебі, $q(x) \equiv 0$ жағдайында мысал бола алады:

$$y'(0) = y'(1), \quad y'(0) = \int_0^1 \overline{p(x)} y(x) dx, \quad (2.3)$$

$$y'(0) = y'(1) + \int_0^1 \overline{p(x)} y(x) dx, \quad y(0) = 0, \quad (2.4)$$

мұндағы $p(x) \in L_2(0, 1)$.

Қарастырылып отырған (1.1)–(2.3), (1.1)–(2.4) есептерінің ерекшелігі, бұл есепке түйіндес есептердің жүктелген екінші ретті дифференциалдық теңдеулер болуында:

$$l^*(v) = -v''(x) + p(x)v'(0) = \bar{\lambda}v(x), \quad V_1(v) = v'(1) = 0, \quad V_2(v) = v(0) - v(1) = 0, \quad (2.5)$$

$$l^*(v) = -v''(x) + p(x)v(0) = \bar{\lambda}v(x), \quad V_1(v) = v'(1) = 0, \quad V_2(v) = v(0) - v(1) = 0. \quad (2.6)$$

(1.1)–(2.3) спектралдық есебінің сипаттаушы (характеристикалық) анықтаушы келесі түрдегі көрініске ие болады:

$$\Delta_5(\lambda) = \Delta_0(\lambda) \left[1 + 4\sqrt{2}\pi \sum_{k=1}^{\infty} \overline{a_{k_0}} \frac{k}{\lambda - (2k\pi)^2} \right], \quad (2.7)$$

ал (1.1)–(2.4) есебінің сипаттаушы (характеристикалық) анықтаушы мына түрде:

$$\Delta_6(\lambda) = \Delta_0(\lambda) \left[1 + 4\sqrt{2} \sum_{k=1}^{\infty} \left(\pi \overline{a_{k_0}} \frac{k}{(\lambda - (2k\pi)^2)^2} + \overline{a_{k_1}} \frac{k}{\lambda - (2k\pi)^2} \right) \right], \quad (2.8)$$

Мұндағы $\Delta_0(\lambda) = 1 - \cos\sqrt{\lambda}$,

$$\overline{p(x)} = \sum_{k=1}^{\infty} \overline{a_{k_0}} \cdot 2\sqrt{2}(1-x) \cdot \sin 2k\pi x + \sum_{k=1}^{\infty} \overline{a_{k_1}} \cdot 4\sqrt{2} \cdot \cos 2k\pi x. \quad (2.9)$$

$\Delta_5(\lambda), \Delta_6(\lambda)$, – функциялары λ айнымалысы бойынша бүтін аналитикалық функциялар.

(2.7) формуласынан (1.1)–(2.3) «серпімді» есебінің меншікті мәндерінің екі сериясы шығады:

$$\lambda_k^{(1)} = \lambda_k^0 = (2k\pi)^2;$$

$$\lambda_k^{(2)} = \left[2k\pi + \overline{a_{k_0}} \left(\sqrt{2} + O\left(\frac{1}{\sqrt{k}}\right) \right) \right]^2$$

(1.1)–(2.3), (1.1)–(2.4) «серпімді» есептердің меншікті-қосарланған функциялар жүйесі мен «серпімсіз» (Рисс базисін құратын) Самарский-Ионкин есебінің түбірлік векторларының жүйесі тек қана алғашқы мүшелерінің ақырлы санымен ғана айырмашылығы болатындығын байқаймыз.

Олай болса, $L_2(0, 1)$ кеңістігінде (1.1)–(2.3), (1.1)–(2.4) «серпімді» есептердің меншікті-қосарланған функцияларының жүйелері де Рисс базисін құрайтындығы шығады.

(2.9) түріндегі $p(x)$ функцияларының жиыны $L_2(0, 1)$ кеңістігінде тығыз болады.

Теорема 2. «Серпімді» (1.1)–(2.3), (1.1)–(2.4) есептерінің $L_2(0, 1)$ кеңістігінде Рисс базисін құратын меншікті-қосарланған функцияларының жүйелерінің $p(x)$ функцияларының жиыны $L_2(0, 1)$ кеңістігінде тығыз болады.

(2.3) шеттік шартының шамалы ғана интегралдық «серпімді» болғандағы (1.1)–(2.3), (1.1)–(2.4) «серпімді» есептердің меншікті және қосарланған функцияларының жүйелерінің базистілік қасиетінің орнықсыздығын көрсету үшін (1.1)–(2.3) есебінің меншікті функцияларын табамыз, ал (2.4) шеттік шарты үшін (2.6) түйіндес есебінің $\lambda_j^0 = (2j\pi)^2$ меншікті мәндеріне сәйкес меншікті функцияларын табамыз, яғни

$$u_j^1(x) = C_1 \left\{ \cos(2j\pi x) + \frac{\sqrt{2}}{\overline{a_{j_0}}} \left[1 - \frac{\sqrt{2}}{4\pi} \cdot \frac{\overline{a_{j_0}}}{j} + \frac{\sqrt{2}}{\pi} \sum_{k \neq j}^{\infty} \overline{a_{k_0}} \frac{k}{j^2 - k^2} \right] \sin(2j\pi x) \right\}, \quad (2.10)$$

$$u_{j_0}(x) = C_1 \left\{ \cos(2j\pi x) - \left[2\sqrt{2} \cdot \overline{a_{j_1}} + O\left(\frac{1}{j^2}\right) \right] \left[\sin(2j\pi x) + \int_0^x p(\xi) \frac{\sin(2j\pi(x-\xi))d\xi}{2j\pi} \right] \right\} \quad (2.11)$$

(2.10) меншікті функциялары үшін C_1 тұрақтыны $(u_j^1(x), v_j^1(x)) = 1$ биортогоналдық шартынан таңдаймыз, ал (2.11) меншікті функциялары үшін $(u_{j_0}(x), v_{j_0}(x)) = 1$ шартынан таңдаймыз. Онда Юнг теоремасынан [11], (1.1)–(2.3) есебі үшін $\lim_{j \rightarrow \infty} \|u_j^1\| \cdot \|v_j^1\| = \infty$, ал (1.1)–(2.4) есебі үшін $\lim_{j \rightarrow \infty} \|u_{j_0}\| \cdot \|v_{j_0}\| = \infty$ болатындығы шығады.

Демек, меншікті және қосарланған функциялар жүйесінің бірқалыпты минималдық [12] шарты орындалмайды, олай болса $L_2(0, 1)$ кеңістігінде жәй базисте құрмайды. Бұл зерттеліп отырған есептің көкейкесті мәселесі, яғни келесі теорема осы нәтижені дәріптейді:

Теорема 3. (1.1)–(2.3), (1.1)–(2.4) «серпімді» есептердің $L_2(0, 1)$ кеңістігінде жәй базисте құрмайтын $p(x) \in L_2(0, 1)$ меншікті-қосарланған функциялар жүйелерінің жиыны $L_2(0, 1)$ кеңістігінде тығыз болады.

Ескерту. Түбірлік функциялардың Рисс базистілігі қасиеттеріне түйіндес операторлар бір мерзімде ие болады.

Қорытынды

Мақалада локалды емес интегралдық «серпімді» шеттік шарттармен берілген Штурм–Лиувилл еселі дифференциалдау операторы үшін қойылған спектралдық есептердің базистік қасиеттері зерттелді. Алынған нәтижелер күшейтілмеген регулярлы шеттік шарттар жағдайында меншікті және қосарланған функциялар жүйесінің құрылымы классикалық есептерден айтарлықтай ерекшеленетінін көрсетеді [1, 7]. Интегралдық «серпімді»

Самарский–Ионкин типіндегі есеп үшін меншікті және қосарланған функциялар жүйесінің квадраттары қосындыланатын функциялар кеңістігінде Рисс базисін құрайтыны дәлелденді. Сонымен қатар, шеттік шарттардағы интегралдық ядроның шамалы өзгеруі кезінде базистік қасиеттердің орнықсыздығы дәлелденді, бұл бұрын алынған нәтижелермен үйлеседі [2]. Рисс базисін құрайтын түбірлік векторлар жүйесінің жиыны мен жәй базисте құрмайтын меншікті және қосарланған функциялар жүйелерінің жиындары L_1 функционалдық кеңістігінде тығыз болатыны дәлелденді. Бұл нәтиже локалды емес шеттік шарттармен берілген операторлар үшін базистік мәселесінің күрделілігін және көпқырлы сипатын айқындайды [8, 11]. Алынған нәтижелер авторлардың алдыңғы зерттеулерін толықтырып, кеңейтеді [9, 10].

Сөз соңында, авторлар ф.-м.ғ.д., профессор, ҰҒА корр.-мүшесі М.А. Садыбековке берген ғылыми кеңесі үшін алғыстарын білдіреді.

Қаржыландыру туралы ақпарат

Бұл мақала Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитетінің гранты аясында (грант №AP23485279) қаржыландырылды.

ӘДЕБИЕТТЕР

- 1 Ильин В.А. О связи между видом краевых условий и свойствами базисности и равносходимости с тригонометрическим рядом разложений по корневым функциям несамосопряженного дифференциального оператора // Дифференциальные уравнения. – 1994. – Т. 30, № 9. – С. 1516–1529. URL: <https://www.mathnet.ru/rus/de9437>
- 2 Макин А.С. О нелокальном возмущении периодической задачи на собственные значения // Дифференциальные уравнения. – 2006. – Т. 42, № 4. – С. 560–562. <https://doi.org/10.1134/S0012266106040185>
- 3 Lang P., Locker J. Spectral theory of two-point differential operators determined by $D_{2n}D^{2n}$ // Journal of Mathematical Analysis and Applications. – 1990. – Vol. 146, No. 1. – P. 148–191. [https://doi.org/10.1016/0022-247X\(90\)90339-H](https://doi.org/10.1016/0022-247X(90)90339-H)
- 4 Dukenbaeva A., Sadybekov M.A. On a spectral problem for the Laplace operator with more general boundary conditions // Herald of the Kazakh-British Technical University. – 2024. – Vol. 21, No. 4. – P. 146–152. <https://doi.org/10.55452/1998-6688-2024-21-4-146-152>
- 5 Sadybekov M.A., Imanbaev N.S. Characteristic determinant of a boundary value problem, which does not have the basis property // Eurasian Math. J. – 2017. – Vol. 8, No. 2. – P. 148–191. URL: <https://www.mathnet.ru/rus/emj/v8/i2/p40>
- 6 Сатпаева З.З., Кангужин Б.Е. Регуляризованный след оператора Штурма–Лиувилля на графе-звезде // Вестник Казахстанско-Британского технического университета. – 2025. – Т. 22, № 1. – С. 229–238. <https://doi.org/10.55452/1998-6688-2025-22-1-229-238>
- 7 Наймарк М.А. Линейные дифференциальные операторы. – М.: Наука, 1969. – 352 с.
- 8 Шкаликов А.А. О базисности собственных функций обыкновенных дифференциальных операторов с интегральными краевыми условиями // Вестник МГУ. Серия: Математика. Механика. – 1982. – № 6. – С. 12–21. URL: <https://www.mathnet.ru/vmumm3582>
- 9 Imanbaev N.S., Sairam N.N. On the quadratic closeness of eigenfunctions of «unperturbed» and «perturbed» differentiation operators on an interval // Herald to National Engineering Academy of the Republic of Kazakhstan – 2024. – № 3(93). – С. 313–319. <https://doi.org/10.47533/2024.1606-146X.70>
- 10 Imanbaev N.S. On basic properties of eigenfunctions and associated functions of one loaded operator of multiple differentiation // Lobachevskii J. Math. – 2022. – Т. 43, № 3. – С. 749–754. URL: <https://www.mathnet.ru/vmumm3582>
- 11 Hardy G.H., Littlewood J.E., Pólya G. Inequalities. – Cambridge: Cambridge University Press, 1948. – 456 p.
- 12 Функциональный анализ / Под ред. С.Г. Крейна. – М.: Мир, 1972.

REFERENCES

- 1 Ilyin, V.A. O svyazi mezhdru vidom kraevykh uslovii i svoistvami bazisnosti i ravnoskhodimosti s trigonometricheskimi ryadami razlozhenii po kornevym funktsiyam nesamosopryazhennogo differentsial'nogo operatora [On the relation between the type of boundary conditions and the basis and equiconvergence properties with trigonometric series expansions in root functions of a non-selfadjoint differential operator]. *Differentsial'nye uravneniya*, 30 (9), 1516–1529 (1994). (In Russian). URL: <https://www.mathnet.ru/rus/de9437>
- 2 Makin, A.S. O nelokal'nom vozmushchenii periodicheskoi zadachi na sobstvennye znacheniya [On a nonlocal perturbation of a periodic eigenvalue problem]. *Differentsial'nye uravneniya*, 42 (4), 560–562 (2006). (In Russian). <https://doi.org/10.1134/S0012266106040185>
- 3 Lang, P., and Locker, J. Spectral theory of two-point differential operators determined by (D^{2n}) . *Journal of Mathematical Analysis and Applications*, 146 (1), 148–191 (1990). [https://doi.org/10.1016/0022-247X\(90\)90339-H](https://doi.org/10.1016/0022-247X(90)90339-H)
- 4 Dukenbaeva, A., and Sadybekov, M.A. On a spectral problem for the Laplace operator with more general boundary conditions. *Herald of the Kazakh-British Technical University*, 21 (4), 146–152 (2024). <https://doi.org/10.55452/1998-6688-2024-21-4-146-152>
- 5 Sadybekov, M.A., and Imanbaev, N.S. Characteristic determinant of a boundary value problem, which does not have the basis property. *Eurasian Mathematical Journal*, 8 (2), 148–191 (2017). URL: <https://www.mathnet.ru/rus/emj/v8/i2/p40>
- 6 Satpaeva, Z.Z., and Kanguzhin, B.E. Regularizovannyi sled operatora Shturma–Liuvillya na grafe-zvezde [Regularized trace of the Sturm–Liouville operator on a star graph]. *Vestnik Kazakhstanskogo-Britanskogo tekhnicheskogo universiteta*, 22 (1), 229–238 (2025). (In Russian). <https://doi.org/10.55452/1998-6688-2025-22-1-229-238>
- 7 Naimark, M.A. *Lineinye differentsial'nye operatory* [Linear Differential Operators] (Moscow: Nauka, 1969), 352 p. (In Russian).
- 8 Shkalikov, A.A. O bazisnosti sobstvennykh funktsii obyknovennykh differentsial'nykh operatorov s integral'nymi kraevymi usloviyami [On the basis property of eigenfunctions of ordinary differential operators with integral boundary conditions]. *Vestnik MGU. Seriya: Matematika. Mekhanika*, 6, 12–21 (1982). (In Russian). URL: <https://www.mathnet.ru/vmumm3582>
- 9 Imanbaev, N.S., and Sairam, N.N. On the quadratic closeness of eigenfunctions of “unperturbed” and “perturbed” differentiation operators on an interval. *Herald to National Engineering Academy of the Republic of Kazakhstan*, 3 (93), 313–319 (2024). <https://doi.org/10.47533/2024.1606-146X.70>
- 10 Imanbaev, N.S. On basic properties of eigenfunctions and associated functions of one loaded operator of multiple differentiation. *Lobachevskii Journal of Mathematics*, 43 (3), 749–754 (2022). URL: <https://www.mathnet.ru/vmumm3582>
- 11 Hardy, G.H., Littlewood, J.E., and Pólya, G. *Inequalities* (Cambridge: Cambridge University Press, 1948), 456 p.
- 12 *Funktsional'nyi analiz* [Functional Analysis]. Ed. by S.G. Krein (Moscow: Mir, 1972). (In Russian).

^{1,2*}**Иманбаев Н.С.,**

к.физ.-мат.н., профессор, ORCID ID: 0000-0002-5220-9899,

*e-mail: imanbaevnur@mail.ru

^{2,3}**Сайрам Н.Н.,**

магистр, докторант, ORCID ID: 0009-0004-9766-6219,

e-mail: nurgul.sairam02@mail.ru

¹Университет имени Ж.А. Ташенева, г. Шымкент, Казахстан

²Институт математики и математического моделирования, г. Алматы, Казахстан

³Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

О БАЗИСНОСТИ СИСТЕМ СОБСТВЕННЫХ И ПРИСОЕДИНЕННЫХ ФУНКЦИЙ ОПЕРАТОРА КРАТНОГО ДИФФЕРЕНЦИРОВАНИЯ ШТУРМА-ЛИУВИЛЛЯ С НЕЛОКАЛЬНЫМИ «ВОЗМУЩЕННЫМИ» КРАЕВЫМИ УСЛОВИЯМИ

Аннотация

В статье исследуется спектральная задача для оператора кратного дифференцирования Штурма–Лиувилля, заданного на отрезке с нелокальными интегральными «возмущенными» граничными условиями. Рассматриваемые граничные условия являются регулярными, но не усиленно регулярными, что существенно осложняет анализ спектральных свойств оператора и изучение базисных свойств соответствующих систем корневых функций, а также ограничивает применение классических методов спектральной теории. Основной целью работы является исследование свойств базисности систем собственных и присоединенных функций в пространстве квадратично суммируемых функций, а также анализ их устойчивости и неустойчивости при малых изменениях параметров задачи и интегральных ядер в граничных условиях. В качестве частного случая подробно рассматривается задача Самарского–Ионкина с интегральными «возмущенными» граничными условиями. Доказано, что система собственных и присоединенных функций данной задачи образует базис Рисса в пространстве квадратично суммируемых функций на отрезке. Показано, что при малых изменениях интегрального ядра в граничных условиях свойства базисности могут быть неустойчивыми. Установлено, что множества систем корневых функций, образующих базис Рисса, и систем корневых функций, не образующих обычного базиса, являются плотными в рассматриваемом функциональном пространстве L_1 . Полученные результаты представляют интерес для дальнейшего развития спектральной теории дифференциальных операторов на отрезке и ее приложений.

Ключевые слова: оператор Штурма–Лиувилля, нелокальные граничные условия, интегральное возмущение, задача Самарского–Ионкина, базис Рисса, собственные и присоединенные функции, устойчивость.

^{1,2*}**Imanbaev N.S.,**

Cand. Sci. (Phys.-Math.), Professor, ORCID ID: 0000-0002-5220-9899,

*e-mail: imanbaevnur@mail.ru

^{2,3}**Sairam N.N.,**

Master, PhD doctoral student,

ORCID ID: 0009-0004-9766-6219, e-mail: nurgul.sairam02@mail.ru

¹Z.A. Tashenev University, Shymkent, Kazakhstan

²Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan

³Al-Farabi Kazakh National University, Almaty, Kazakhstan

BASIS PROPERTY OF SYSTEMS OF EIGEN AND ASSOCIATED FUNCTIONS OF A SHTURM-LIOUVILLE MULTIPLE DIFFERENTIATION OPERATOR WITH NONLOCAL «PERTURBED» BOUNDARY CONDITIONS

Abstract

This paper is devoted to the study of a spectral problem for a Sturm–Liouville multiple differentiation operator defined on an interval with nonlocal integrally “perturbed” boundary conditions. The considered boundary conditions

are regular but not strongly regular, which leads to essential difficulties in the analysis of the spectral properties and the basis behavior of the corresponding systems of functions. The main objective of the study is to investigate the basis properties of systems of eigen and associated functions in the space of square-summable functions and to analyze their stability and instability under small perturbations of the boundary conditions. As a particular case, the Samarskii–Ionkin problem with integrally perturbed boundary conditions is examined in detail. It is proved that the system of eigen and associated functions of this problem forms a Riesz basis in the L_2 space on the interval. The paper also shows that small changes in the integral kernel of the boundary conditions may lead either to the preservation or to the loss of the basis property. Moreover, it is established that the sets of root function systems forming a Riesz basis and the sets of eigen and associated function systems that do not form an ordinary basis are dense in the corresponding functional space L_1 . The obtained results contribute to the development of the spectral theory of differential operators with nonlocal boundary conditions.

Keywords: Sturm–Liouville operator, nonlocal boundary conditions, integral perturbation, Samarskii–Ionkin problem, Riesz basis, eigen and associated functions, stability.

Received January 17, 2026; revised April 22, 2026; accepted May 4, 2026.

ӘОЖ 517.9
ҒТАХР 27.41.19

<https://doi.org/10.55452/1998-6688-2026-23-2-61-74>

^{1,3*}**Молыбайқызы А.,**

ғылыми қызметкер, аға оқытушы, ORCID ID: 0009-0008-2452-5932,
e-mail: altynai_bm@mail.ru

^{1,2}**Кабдрахова С.С.,**

ф.-м.ғ.к., бас ғылыми қызметкер, ORCID ID: 0000-0003-0247-5985,
e-mail: Symbat2909.sks@gmail.com

¹**Искакова Н.Б.,**

ф.-м.ғ.к., бас ғылыми қызметкер, ORCID ID: 0000-0002-0680-4099,
e-mail: narkesh77@gmail.com

^{1,3}**Минглибаева Б.Б.,**

ф.-м.ғ.к., ғылыми қызметкер, аға оқытушы, ORCID ID: 0000-0002-7195-4480,
e-mail: minglibayeva.bayan@gmail.com

¹Математика және математикалық моделдеу институты, Алматы қ., Қазақстан

²Әл-Фараби атындағы қазақ ұлттық университеті, Алматы қ., Қазақстан

³Қазақ ұлттық қыздар педагогикалық университеті, Алматы қ., Қазақстан

БӨЛІКТІ ТҰРАҚТЫ АРГУМЕНТТІ ИМПУЛЬСТІ ГИПЕРБОЛАЛЫҚ ТЕҢДЕУ ҮШІН ШЕТТІК ЕСЕПТІҢ ШЕШІМІ ТУРАЛЫ

Аңдатпа

Мақалада бөлікті тұрақты аргументі бар импульсті гиперболаалық теңдеу үшін шеттік есеп қарастырылады. Бөлікті тұрақты аргументі бар импульсті гиперболаалық теңдеулер нейрондық желілердегі, динамикалық жүйелердегі, гибридік жүйелердегі және т.б. физикалық процестерді сипаттаудың математикалық моделі ретінде пайда болады. Осындай теңдеулер үшін шеттік есептердің бар болуы және шешімдерін құру сұрақтары қазіргі таңда маңызды мәселелердің бірі болып қала береді. Бұл есептің шешілімділік шарттарын алу үшін Джумабаевтың параметрлеу әдісі қолданылады және жұмыста шешімнің жуық мәнін табуға арналған итерациялық алгоритм жасалды. Итерациялық процестің әрбір қадамы үшін функционалдық параметрлер арасындағы байланысты сипаттайтын $Q(x)$ матрицасы арқылы өрнектелетін интегралдық формулалар алынды. Егер осы матрицаның керісі бар болса, онда есептің параметрлік және бастапқы түрлері үшін шешімнің бар болуы мен жалғыздығы дәлелденеді. Ұсынылған әдіс есептің тек теориялық шешілуін дәлелдеумен шектелмей, шешімді табудың нақты құрылымдық процедурасын ұсынады. Бұл әрі қарайғы сандық жүзеге асырулар мен шешімдердің орнықтылығын талдауда маңызды рөл атқарады. Сонымен қатар, ұсынылған тәсілді бөлікті тұрақты аргументі бар басқа типтегі есептерге, соның ішінде импульстік шарттары бар жүйелерге, жады әсері бар нейрондық желілерге және бейсызық гибридік модельдерге қолдануға болады.

Түйін сөздер: импульсті гиперболаалық теңдеу, бөлікті тұрақты аргумент, шеттік есеп, алгоритм.

Кіріспе

Соңғы жылдары аргументі үзілісті дифференциалдық теңдеулер саласындағы зерттеулер күрделі гибридік процестерді сипаттау қажеттілігімен тығыз байланысты дамып келеді. Мұндай процестердің қатарына импульстік жүйелер, жады әсері бар желілер және бөлікті үзіліссіз динамикалық модельдер жатады [1]–[2]. Солардың ішінде бөлікті тұрақты аргументі және дискретті жады әсері бар гиперболаалық теңдеулерді зерттеу ерекше маңызға ие, себебі олар классикалық гиперболаалық теңдеулер мен аргументі кешеуілдеген немесе ілгеріленген функционалдық-дифференциалдық жүйелердің қасиеттерін біріктіреді [1]. Осындай

модельдер нейрондық желілер теориясында, секіrmелі әсерлері бар механикалық жүйелерде, автоматика мен басқару теориясында және күрделі гибридітік процестерді математикалық модельдеуде жиі кездеседі [3]–[10].

Атап айтсақ [7]–[9], және басқа ғалымдардың еңбектерінде бөлікті тұрақты аргументі бар теңдеулердің жалпы теориясы жасалып, интегралдық көпбейнеліктер құру әдістері, Грин функциялары, тұрақтылық және периодтылық шарттары қарастырылған. Кейінгі зерттеулерде [3]–[5] бұл теория дискретті жады әсері бар импульстік гиперболалық теңдеулерге дейін кеңейтілді, мұнда жүйенің динамикасы тек үзіліссіз емес, сонымен бірге дискретті уақыт сәттерінде әсер ететін импульстар арқылы сипатталады.

Сондай-ақ, зерттелген есептердің шешімдерінің бар болуы және жалғыздығы туралы теоремалар орнатылғанына қарамастан, импульстік гиперболалық теңдеулерге арналған көптеген шеттік есептер конструктивті тұрғыдан толық қарастырылмаған. Әсіресе, бөлікті тұрақты аргументі мен дискретті жады әсері бар есептер үшін бірімді шешімділік шарттары, жуық шешімдерді табудың алгоритмдері мен жинақтылығы мәселелері ұзақ уақыт бойы белгісіз болып келді.

Осы мақалада бөлікті тұрақты аргументі бар импульстік гиперболалық теңдеу үшін қойылған шеттік есеп қарастырылады. Есептің шешілу шарттарын анықтау және шешімін құру мақсатында Джумабаевтың параметрлеу әдісі [4] қолданылады. Бұл әдіс шеттік есептерді шешудің тиімді құралы ретінде кеңінен танылған. Осы тәсіл негізінде жаңа функционалдық параметрлер енгізіліп, бастапқы есепті параметрлері бар Коши есептерінің әулетіне түрлендіретін эквивалентті дифференциалдық және интегралдық теңдеулер жүйесі құрылады.

Осылайша, мақалада алынған нәтижелер дискретті жады және импульстік әсерлері бар теңдеулер теориясын кеңейтіп, бөлікті тұрақты аргументі бар гиперболалық жүйелердің шеттік есептерін шешудің жаңа конструктивті әдісін ұсынады. Алынған нәтижелер [4], [6, 7], [8, 9] еңбектеріндегі тәсілдерді жалпылайды және дамытады. Болашақ зерттеулердің перспективалы бағыттарына әдісті кеңістіктік-уақыттық (көпөлшемді) есептерге, стохастикалық және сызықтық емес гиперболалық теңдеулерге бейімдеу, сондай-ақ мұндай жүйелер үшін тиімді сандық схемаларды жасау жатады.

Материалдар мен әдістер

Есептің қойылымы және шешу әдісі.

$\Omega = [0, T] \times [0, \omega]$ облысында бөлікті тұрақты аргументі бар импульстік гиперболалық теңдеу үшін қойылған шеттік есепті қарастырамыз

$$\frac{\partial^2 u}{\partial t \partial x} = A(t, x) \frac{\partial u(t, x)}{\partial x} + B(t, x) \frac{\partial u(t, x)}{\partial t} + C(t, x) u(t, x) + A_0(t, x) \frac{\partial u(\gamma(t), x)}{\partial x} + B_0(t, x) \frac{\partial u(\gamma(t), x)}{\partial t} + C_0(t, x) u(\gamma(t), x) + f(t, x), \quad t \neq \theta_j, \quad j = \overline{1, N}, \quad (1)$$

$$P_2(x) \frac{\partial u(0, x)}{\partial x} + P_1(x) \frac{\partial u(0, x)}{\partial t} + P_0(x) u(0, x) + S_2(x) \frac{\partial u(T, x)}{\partial x} + S_1(x) \frac{\partial u(T, x)}{\partial t} + S_0(x) u(T, x) = \varphi_0(x), \quad x \in [0, \omega], \quad (2)$$

$$\lim_{t \rightarrow \theta_p + 0} u(t, x) - \lim_{t \rightarrow \theta_p - 0} u(t, x) = \varphi_p(x), \quad x \in [0, \omega], \quad p = \overline{1, N-1}, \quad (3)$$

$$u(t, 0) = \psi(t), \quad t \in [0, T]. \quad (4)$$

Мұндағы $u(t, x)$ белгісіз функция, $A(t, x)$, $B(t, x)$, $C(t, x)$, $A_0(t, x)$, $B_0(t, x)$, $C_0(t, x)$ функциялары және $f(t, x)$ функциясы Ω облысында үзіліссіз;

$$\gamma(t) = \zeta_s, \quad \text{егер} \quad t \in [\theta_{s-1}, \theta_s), \quad s = \overline{1, N}.$$

$$\theta_{s-1} < \zeta_s < \theta_s, \quad s = 1, 2, \dots, N; \quad 0 = \theta_0 < \theta_1 < \dots < \theta_{N-1} < \theta_N = T;$$

$P_2(x), P_1(x), P_0(x), S_2(x), S_1(x), S_0(x)$ үзіліссіз, $\varphi_p(x), \varphi_0(x)$ функциялары $[0, \omega]$ кесіндісінде үзіліссіз дифференциалданады, $p = \overline{1, N-1}$; $\psi(t)$ функциясы $[0, T]$ кесіндісінде үзіліссіз дифференциалданады.

$\Delta_N(\omega)$ деп Ω облысын $t = \theta_s$ түзулері бойынша бөлуді белгілейміз. $\Omega_s = [\theta_{s-1}, \theta_s) \times [0, \omega]$, $s = \overline{1, N}$, $\Omega = \bigcup_{s=1}^N \Omega_s$.

$PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$ арқылы Ω облысында бөлікті үзіліссіз, $t = \theta_j$, $j = \overline{1, N-1}$ түзулерінде үзілістері болуы мүмкін $u(t, x)$ функцияларынан тұратын кеңістігін белгілейік және нормасы келесідей анықталсын:

$$\|u\|_1 = \max_{s=\overline{1, N}} \sup_{(t, x) \in \Omega_s} |u(t, x)|.$$

$u(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$ функциясы (1)–(4) есебінің шешімі болады, егер төмендегі шарттар орындалса:

I) $u(t, x)$ функциясының дербес туындылары бар болады:

$$\frac{\partial u(t, x)}{\partial x} \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R}), \quad \frac{\partial u(t, x)}{\partial t} \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R});$$

II) әрбір $(t, x) \in \Omega$ нүктесінде $\frac{\partial^2 u}{\partial t \partial x}$ аралас дербес туындысы бар, тек (θ_{s-1}, x) , $s = \overline{1, N}$, $x \in [0, \omega]$ нүктелерін қоспағанда, бірақ бұл нүктелерде бір жақты аралас дербес туындылары бар болады;

III) (1) гиперболалық теңдеуі әрбір $(\theta_{s-1}, \theta_s) \times [0, \omega]$, $s = \overline{1, N}$ ішкі облысында $u(t, x)$ үшін орындалады, сондай-ақ (θ_{s-1}, x) , $s = \overline{1, N}$, $x \in [0, \omega]$ нүктелеріндегі $u(t, x)$ функциясының оң жақты екінші ретті аралас дербес туындысы үшін де орынды;

(IV) $u(t, x)$ үшін (2) шеттік шарт және (4) бастапқы шарт, сәйкес, $t=0$, $t=T$ және $x=0$ орындалады;

(V) $u(t, x)$ функциясы үшін (3) импульстік әсер шарттары $t = \theta_p$, $p = \overline{1, N-1}$, $x \in [0, \omega]$ орындалады.

Жаңа функциялар енгізу және Джумабаевтың параметрлеу әдісінің сұлбасы.

Алдымен, $v(t, x) = \frac{\partial u(t, x)}{\partial x}$, $w(t, x) = \frac{\partial u(t, x)}{\partial t}$ жаңа функцияларын енгіземіз. (1)–(4) есебі бөлікті тұрақты аргументі бар импульсті дифференциалдық теңдеулер әулеті үшін шеттік есепке ауысады:

$$\frac{\partial v(t, x)}{\partial t} = A(t, x)v(t, x) + A_0(t, x)v(\gamma(t), x) + B(t, x)w(t, x) + B_0(t, x)w(\gamma(t), x) + C(t, x)u(t, x) + C_0(t, x)u(\gamma(t), x) + f(t, x), t \neq \theta_j, j = \overline{1, N-1}, \quad (5)$$

$$P_2(x)v(0, x) + P_1(x)w(0, x) + P_0(x)u(0, x) + S_2(x)v(T, x) + S_1(x)w(T, x) + S_0(x)u(T, x) = \varphi_0(x), \quad (6)$$

$$\lim_{t \rightarrow \theta_p+0} v(t, x) - \lim_{t \rightarrow \theta_p-0} v(t, x) = \varphi_p(x), \quad x \in [0, \omega], \quad p = \overline{1, N-1}, \quad (7)$$

$$u(t, x) = \psi(t) + \int_0^x v(t, \xi) d\xi, \quad w(t, x) = \psi(t) + \int_0^x \frac{\partial v(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}. \quad (8)$$

$\{v(t, x), u(t, x), w(t, x)\}$ функцияларының үштігі (5)–(8) бөлікті тұрақты аргументі бар импульсті дифференциалдық теңдеулер әулеті үшін есептің шешімі болады, егер:

I) $v(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$ функциясының дербес туындысы бар және

$$\frac{\partial v(t, x)}{\partial t} \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R}), \quad u(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R}), \quad w(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R});$$

II) (5) дифференциалдық теңдеулер әулеті әрбір $[\theta_{s-1}, \theta_s) \times [0, \omega]$, $s = \overline{1, N}$, ішкі облысында $v(t, x)$, $u(t, x)$ және $w(t, x)$ функциялары үшін орындалады және (θ_{s-1}, x) , $s = \overline{1, N}$, $x \in [0, \omega]$ нүктелеріндегі $v(t, x)$ функциясының t бойынша оң жақты дербес туындысы үшін де орынды;

III) (6) шеттік шарты $v(t, x)$ функциясының әрбір $t=0, t=T$ түзулерінде орындалсын;

(IV) $u(t, x)$ және $w(t, x)$ функциялары $v(t, x)$ және $\frac{\partial v(t, x)}{\partial t}$ функцияларымен (8) интегралдық теңдеулері арқылы байланысқан болса.

Біз $v_s(t, x)$ деп $v(t, x)$ функциясының Ω_s -ішкі облыстағы тарылуын белгілейміз, яғни: $v_s(t, x) = v(t, x)$, $(t, x) \in \Omega_s$, $s = \overline{1, N}$.

$C(\Omega, \Delta_N(\omega), \mathbb{R}^N)$ кеңістігі деп: $v([t], x) = (v_1(t, x), v_2(t, x), \dots, v_N(t, x))'$ функцияларынан тұратын, $v_s: \Omega_s \rightarrow \mathbb{R}$ және $\lim_{t \rightarrow \theta_s - 0} v_s(t, x)$ сол жақты шегі бар, нормасы төмендегідей анықталған:

$$\|v[\cdot]\|_2 = \max_{s=\overline{1, N}} \sup_{(t, x) \in \Omega_s} \|v_s(t, x)\|$$

кеңістікті айтамыз.

$v([t], x) = (v_1(t, x), v_2(t, x), \dots, v_N(t, x))'$ функциясы $C(\Omega, \Delta_N(\omega), \mathbb{R}^N)$ кеңістігінде жатады, және оның элементтері $v_s(t, x)$, $s = \overline{1, N}$, төмендегі дифференциалдық теңдеулер жүйесін қанағаттандырады:

$$\frac{\partial v_s}{\partial t} = A(t, x)v_s(t, x) + A_0(t, x)v_s(\zeta_s, x) + B(t, x)w(t, x) + B_0(t, x)w(\zeta_s, x) + C(t, x)u(t, x) + C_0(t, x)u(\zeta_s, x) + f(t, x), \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}, \quad (9)$$

және келесі шарттар орындалады:

$$P_2(x)v_1(0, x) + P_1(x)w(0, x) + P_0(x)u(0, x) + S_2(x) \lim_{t \rightarrow T-0} v_N(t, x) + S_0(x) \lim_{t \rightarrow T-0} u(t, x) = \varphi_0(x), \quad x \in [0, \omega], \quad (10)$$

$$v_{p+1}(\theta_p, x) - \lim_{t \rightarrow \theta_p - 0} v_p(t, x) = \phi_p(x), \quad x \in [0, \omega], \quad p = \overline{1, N-1}, \quad (11)$$

$$u(t, x) = \psi(t) + \int_0^x v_s(t, \xi) d\xi,$$

$$w(t, x) = \psi(t) + \int_0^x \frac{\partial v_s(t, \xi)}{\partial \tau} d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}, \quad (12)$$

(9) теңдеуде біз мынаны ескердік $\gamma(t) = \zeta_s$, $t \in [\theta_{s-1}, \theta_s)$, $s = \overline{1, N}$.

Біз келесі түрде функционалдық параметрлерді енгіземіз: $\mu_s(x) = v_s(\zeta_s, x)$, $s = \overline{1, N}$ және $x \in [0, \omega]$.

Келесі ауыстыруды $v_s(t, x) = \tilde{v}_s(t, x) + \mu_s(x)$, $(t, x) \in \Omega_s$, $s = \overline{1, N}$ жасай отырып, төмендегідей түрдегі параметрлі дифференциалдық теңдеулер әулеті үшін есеп аламыз:

$$\frac{\partial \tilde{v}_s}{\partial t} = A(t, x)(\tilde{v}_s(t, x) + \mu_s(x)) + A_0(t, x)\mu_s(x) + B(t, x)w(t, x) + B_0(t, x)w(\zeta_s, x) + C(t, x)u(t, x) + C_0(t, x)u(\zeta_s, x) + f(t, x), \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}, \quad (13)$$

бастапқы шарт

$$\tilde{v}_s(\zeta_s, x) = 0, \quad s = \overline{1, N} \text{ және } x \in [0, \omega] \quad (14)$$

шеттік шарт

$$P_2(x)(\tilde{v}_1(0, x) + \mu_1(x)) + P_1(x)w(0, x) + P_0(x)u(0, x) + S_2(x) \lim_{t \rightarrow T-0} [\tilde{v}_N(t, x) + \mu_N(x)] + S_1(x)w(T, x) + S_0(x)u(T, x) = \varphi_0(x), \quad x \in [0, \omega], \quad (15)$$

импульс шарты

$$\tilde{v}_{p+1}(\theta_p, x) + \mu_{p+1}(x) - \lim_{t \rightarrow \theta_p - 0} \tilde{v}_p(t, x) - \mu_p(x) = \phi_p(x), \quad x \in [0, \omega], \quad p = \overline{1, N-1}, \quad (16)$$

$$u(t, x) = \psi(t) + \int_0^x [\tilde{v}_s(t, \xi) + \mu_s(\xi)] d\xi,$$

$$w(t, x) = \psi(t) + \int_0^x \frac{\partial \tilde{v}_s(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}, \quad (17)$$

(13)–(17) параметрлері бар есептің шешімі ретінде $\{v([t], x), \mu(x), u(t, x), w(t, x)\}$, төрттігін қарастырамыз, мұнда элементтер $v_s(t, x) \in C(\Omega, \Delta_N(\omega), \mathbb{R})$ функцияларының уақыт бойынша туындысы $\frac{\partial v_s(t, x)}{\partial t} \in C(\Omega, \Delta_N(\omega), \mathbb{R})$ бар; функционалдық параметрлер $\mu_s(x) \in C([0, \omega], \mathbb{R})$, $s = \overline{1, N}$; ал функциялар $u(t, x), w(t, x) \in C(\Omega, \Delta_N(\omega), \mathbb{R})$ болып табылады. Сонымен бірге, бұл элементтер барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін (13) дифференциалдық теңдеулердің жүйесін, (14) бастапқы шарттарын, (15) шеттік шартты және (16) импульстік әсерлер шарттарын $[0, \omega]$ аралығында қанағаттандырады. $u(t, x), w(t, x)$ функциялары барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін $v_s(t, x)$ және оның уақыт бойынша туындысы $\frac{\partial v_s(t, x)}{\partial t}$ арқылы (17) интегралдық теңдеулер $\mu_s(x), u(t, x), w(t, x)$ арқылы байланысқан болса (13)–(14) есебі дифференциалдық теңдеулерге арналған Коши есептерінің әулетіне айналады. Бекітілген $\mu_s(x), u(t, x), w(t, x)$ үшін (13)–(14) Коши есептері әулетінің шешімі жалғыз және төмендегідей түрге ие:

$$\begin{aligned} \tilde{v}_s(t, x) = & e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} [A(\tau, x) + A_0(\tau, x)] d\tau \mu_s(x) \\ & + e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} [B(\tau, x) w(\tau, x) + C(\tau, x) u(\tau, x)] d\tau \\ & + e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} [B_0(\tau, x) w(\zeta_s, x) + C_0(\tau, x) u(\zeta_s, x)] d\tau \\ & + e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} f(\tau, x) d\tau, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}. \end{aligned} \quad (18)$$

Мұндағы $\alpha_s(t, x) = \int_{\zeta_s}^t A(\tau, x) d\tau$, $(t, x) \in \Omega_s$, $s = \overline{1, N}$.

Келесі белгілеулерді енгіземіз:

$$\begin{aligned} D_s(t, x) &= e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} [A(\tau, x) + A_0(\tau, x)] d\tau, \\ H_s(t, x, \omega, u) &= e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} [B(\tau, x) w(\tau, x) + C(\tau, x) u(\tau, x)] d\tau + \\ &+ e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} [B_0(\tau, x) w(\zeta_s, x) + C_0(\tau, x) u(\zeta_s, x)] d\tau, \\ F_s(t, x) &= e^{\alpha_s(t, x)} \int_{\zeta_s}^t e^{-\alpha_s(t, \tau)} f(\tau, x) d\tau, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}. \end{aligned}$$

Онда (18) өрнегін келесі түрде жазуға болады.

$$\tilde{v}_s(t, x) = D_s(t, x) \mu_s(x) + H_s(t, x, w, u) + F_s(t, x), \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}. \quad (19)$$

(19) теңдеуден біз келесіні табамыз:

$$\tilde{v}_1(0, x), \quad \lim_{t \rightarrow T-0} \tilde{v}_N(t, x), \quad \tilde{v}_{p+1}(\theta_p, x), \quad \lim_{t \rightarrow \theta_p-0} \tilde{v}_p(t, x), \quad p = \overline{1, N-1}.$$

Табылған өрнектерді (15) және (16) қатынастарына қойып, келесі өрнектерді аламыз:

$$\begin{aligned} P_2(x)[1 + D_1(0, x)]\mu_1(x) + S_2(x)[1 + D_N(T, x)]\mu_N(x) = & -P_2(x)[F_1(0, x) + H_1(0, x, w, u)] - \\ S_2(x) \lim_{t \rightarrow T-0} [H_N(t, x, w, u) + F_N(t, x)] - & P_1(x)\omega(0, x) - P_0(x)u(0, x) - S_1(x)\omega(T, x) - \\ S_0(x)u(T, x) + \varphi_0(x). \end{aligned} \quad (20)$$

$$[1 + D_{p+1}(\theta_p, x)]\mu_{p+1}(x) - [D_p(\theta_p, x) + 1]\mu_p(x) = \varphi_p(x) + F_{p+1}(\theta_p, x) - F_{p+1}(\theta_p, x) + H_p(\theta_p, x, w, u) - H_{p+1}(\theta_p, x, w, u), \quad p = \overline{1, N-1}. \quad (21)$$

(20), (21) теңдеулер жүйесінің сол жақ бөлігіндегі $\mu_s(x)$, $s = \overline{1, N}$ үшін коэффициенттерді пайдалана отырып, келесі түрде $N \times N$ өлшемді $Q(x)$ матрицасын құрамыз:

$$Q(x)\mu(x) = F_*(x) + H_*(x, w, u), \quad (22)$$

$$Q(x) = \begin{bmatrix} P_2(x)[1 + D_1(0, x)] & 0 & 0 & \dots & 0 & S_2(x)[1 + D_N(T, x)] \\ -[1 + D_1(\theta_1, x)] & 1 + D_2(\theta_1, x) & 0 & \dots & 0 & 0 \\ 0 & -[1 + D_2(\theta_2, x)] & 1 + D_3(\theta_2, x) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -[1 + D_{N-1}(\theta_{N-1}, x)] & 1 + D_N(\theta_{N-1}, x) \end{bmatrix}$$

$F_*(x)$ пен $H_*(x, u)$ векторлары төмендегідей етіп аламыз:

$$F_*(x) = \begin{bmatrix} -P_2(x)F_1(0, x) - S_2F_N(T, x) + \varphi_0(x) \\ -F_2(\theta_1, x) + F_2(\theta_2, x) + \varphi_1(x) \\ -F_3(\theta_2, x) + F_2(\theta_3, x) + \varphi_2(x) \\ \dots \\ -F_N(\theta_{N-1}, x) + F_N(\theta_N, x) + \varphi_{N-1}(x) \end{bmatrix}$$

$$H_*(x, w, u) = \begin{bmatrix} -P_2(x)H_1(0, x, w, u) - S_2(H_N(t, x, w, u) - P_1(x)w(0, x) - P_0(x)u(0, x) - S_1(x)w(T, x) - S_0(x)u(T, x)) \\ H_2(\theta_1, x, w, u) + (H_1(\theta_1, x, w, u)) \\ H_3(\theta_2, x, w, u) + (H_2(\theta_2, x, w, u)) \\ \dots \\ H_N(\theta_{N-1}, x, w, u) + (H_{N-1}(\theta_{N-1}, x, w, u)) \end{bmatrix}$$

Нәтижелер мен талқылау

Параметрлеу әдісінің алгоритмі.

Егер $w(t, x)$ және $u(t, x)$ функциялары барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін белгілі болса, онда (22) функционалдық теңдеулер жүйесінен $\mu(x)$ функциясын табамыз, оның компоненттері $\mu_s(x) \in C([0, \omega], \mathbb{R})$, $s = \overline{1, N}$. Содан кейін (19) интегралдық өрнек және (13) дифференциалдық теңдеулер арқылы $\tilde{v}_s(t, x)$ функциясын және оның туындысын $\frac{\partial \tilde{v}_s}{\partial t}$ барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін анықтаймыз. Табылған $\tilde{v}_s(t, x)$ функциясын және оның туындысы $\frac{\partial \tilde{v}_s}{\partial t}$, $(t, x) \in \Omega_s$, $s = \overline{1, N}$, және функционалдық параметрлер $\mu(x)$ көмегімен $w(t, x)$ және $u(t, x)$ функцияларын анықтаймыз. Бұл жерде (13)–(17) есептер әулетінің шешімін табу үшін итерациялық әдісті қолданамыз. Біз $\{\tilde{v}^*([t], x), \mu^*(x), u^*(t, x), w^*(t, x)\}$, төрттіктер $\{\tilde{v}^{(k)}([t], x), \mu^{(k)}(x), u^{(k)}(t, x), w^{(k)}(t, x)\}$ төрттіктер тізбегінің шегі ретінде анықтаймыз.

0-қадам.

Барлық $x \in [0, \omega]$ үшін $(N \times N)$ өлшемді $Q(x)$ матрицасының керісі бар деп ұйғарамыз. (22) жүйесінің оң жағындағы $u(x) = \psi(t)$, $w(t, x) = \psi(t)$ деп алып, функционалдық параметрдің бастапқы жуықтауын $\mu^{(0)}(x) = (\mu_1^{(0)}(x), \mu_2^{(0)}(x), \dots, \mu_N^{(0)}(x))$ функционалдық теңдеулер жүйесінен табылады:

$$Q(x)\mu(x) = F_*(x) + H_*(x, \psi, \psi),$$

$$\mu^{(0)}(x) = [Q(x)]^{-1}F_*(x) + [Q(x)]^{-1}H_*(x, \psi, \psi), \quad x \in [0, \omega].$$

мұндағы компоненттер $\mu_s(x) \in C([0, \omega], \mathbb{R})$, $s = \overline{1, N}$.

(13) дифференциалдық теңдеулердің оң жағын $u(x) = \psi(t)$, $w(t, x) = \psi(t)$, $\mu_s(x) = \mu_s^{(0)}(x)$, $s = \overline{1, N}$ деп алып және (13)-(14) Коши есебі әулетінен $\tilde{v}_s^{(0)}(t, x)$ табамыз.

$$\tilde{v}_s^{(0)}(t, x) = D_s(t, x)\mu_s^{(0)}(x) + H_s(t, x, \psi, \psi) + F_s(t, x).$$

Және оның туындысын төмендегідей анықтаймыз:

$$\begin{aligned} \frac{\partial \tilde{v}_s^{(0)}(t, x)}{\partial t} = & A(t, x)\tilde{v}_s^{(0)}(t, x) + (A(t, x) + A_0(t, x))\mu_s^{(0)}(x) + B(t, x)\psi(t) + B_0(t, x)\psi(\zeta_s) \\ & + C(t, x)\psi(t) + C_0(t, x)\psi(\zeta_s) + f(t, x), \\ & (t, x) \in \Omega_s, \quad s = \overline{1, N}. \end{aligned}$$

Содан соң (17) интегралдық теңдеуден $u^{(0)}(t, x)$ функцияларды анықтаймыз.

$$u^{(0)}(t, x) = \psi(t) + \int_0^x (\tilde{v}_s^{(0)}(t, \xi) + \mu_s^{(0)}(\xi)) d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}.$$

$$w^{(0)}(t, x) = \psi(t) + \int_0^x \frac{\partial \tilde{v}_s^{(0)}(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N},$$

1-қадам. (22) жүйенің оң жағына $u(t, x) = u^{(0)}(t, x)$, $\omega(t, x) = \omega^{(0)}(t, x)$ қойып,

$\mu^{(1)}(x) = (\mu_1^{(1)}(x), \mu_2^{(1)}(x), \dots, \mu_N^{(1)}(x))$ функционалдык параметрінің

$$Q(x)\mu(x) = F_*(x) + H_*(x, w^{(0)}, u^{(0)}), \quad x \in [0, \omega]$$

$$\mu^{(1)}(x) = Q(x)^{-1}(F_*(x) + H_*(x, w^{(0)}, u^{(0)})), \quad x \in [0, \omega]$$

теңдеуінен бірінші жуықтауын анықтаймыз. (13) дифференциалдық теңдеулердің оң жағын $u(t, x) = u^{(0)}(t, x)$, $w(t, x) = w^{(0)}(t, x)$, $\mu_s(x) = \mu_s^{(1)}(x)$, $s = \overline{1, N}$ қойып, (13)-(14) Коши есебі әулетінен $\tilde{v}_s^{(1)}(t, x)$ табамыз.

$$\tilde{v}_s^{(1)}(t, x) = D_s(t, x)\mu_s^{(1)}(x) + H_s(t, x, w^{(0)}, u^{(0)}) + F_s(t, x).$$

$$\begin{aligned} \frac{\partial \tilde{v}_s^{(1)}(t, x)}{\partial t} = & A(t, x)\tilde{v}_s^{(1)}(t, x) + (A(t, x) + A_0(t, x))\mu_s^{(1)}(x) + B(t, x)w^{(0)}(t, x) + B_0(t, x)\psi(\zeta_s) + \\ & C(t, x)u^{(0)}(t, x) + C_0(t, x)\psi(\zeta_s) + f(t, x), \quad (t, x) \in \Omega_s, \end{aligned}$$

туындысын табамыз. (17) интегралдық теңдеуден төмендегі функцияларды анықтаймыз.

$$u^{(1)}(t, x) = \psi(t) + \int_0^x (\tilde{v}_s^{(1)}(t, \xi) + \mu_s^{(1)}(\xi)) d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}.$$

$$w^{(1)}(t, x) = \psi(t) + \int_0^x \frac{\partial \tilde{v}_s^{(1)}(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N},$$

Дәл осылай k -қадамға жалғастырамыз. (22) жүйенің оң жағына $u(t, x) = u^{(k-1)}(t, x)$, $w(t, x) = w^{(k-1)}(t, x)$ қойып, $\mu^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \dots, \mu_N^{(k)}(x))$ төмендегі теңдеуден анықтаймыз.

$$Q(x)\mu(x) = F_*(x) + H_*(x, w^{(k-1)}, u^{(k-1)}),$$

$$\mu^{(k)}(x) = Q(x)^{-1}(F_*(x) + H_*(x, w^{(k-1)}, u^{(k-1)})), \quad x \in [0, \omega].$$

Осыдан

$$\tilde{v}_s^{(k)}(t, x) = D_s(t, x)\mu_s^{(k)}(x) + H_s(t, x, w^{(k-1)}, u^{(k-1)}) + F_s(t, x).$$

$$\begin{aligned} \frac{\partial \tilde{v}_s^{(k)}(t, x)}{\partial t} = & A(t, x)\tilde{v}_s^{(k)}(t, x) + (A(t, x) + A_0(t, x))\mu_s^{(k)}(x) + B(t, x)w^{(k-1)}(t, x) + \\ & B_0(t, x)\psi(\zeta_s) + C(t, x)u^{(k-1)}(t, x) + C_0(t, x)\psi(\zeta_s) + f(t, x), \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}, \end{aligned}$$

$$u^{(k)}(t, x) = \psi(t) + \int_0^x (\tilde{v}_s^{(k)}(t, \xi) + \mu_s^{(k)}(x)) d\xi,$$

$$w^{(k)}(t, x) = \psi'(t) + \int_0^x \frac{\partial \tilde{v}_s^{(k)}(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = \overline{1, N}$$

табылады.

Мұнда $k=1, 2, \dots$

Әдіс белгісіз функцияларды үш кезеңде табуға мүмкіндік береді:

1. (22) функционалдық теңдеулер жүйесінен $\mu_s(x)$, $s = \overline{1, N}$ енгізілген функционалдық параметрлерді барлық $x \in [0, \omega]$ үшін анықтаймыз.

2. Коши есептерінің (13), (14) әулетінен $v_s(t, x)$ функцияларын барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін және оның туындысын $\frac{\partial v_s(t, x)}{\partial t}$ барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін табамыз.

3. (17) интегралдық теңдеулерден барлық $(t, x) \in \Omega_s$, $s = \overline{1, N}$ үшін $u(t, x)$ және $w(t, x)$ функцияларын анықтаймыз.

Бастапқы есепке қойылған шарттар орындалғанда, әрбір көмекші есептің жалғыз шешімі бар. Алгоритмді жүзеге асыру үшін (13)–(17) параметрлері бар есептің жуық шешімдерінің дәл шешімге жақындауын орнату қажет. Келесі теорема ұсынылған алгоритмнің жинақталу шарттарын және (13)–(17) параметрлері бар есептің жалғыз шешімінің бар болу шарттарын анықтайды.

Теорема 1.

$N \times N$ өлшемді $Q(x)$ матрицасы барлық $x \in [0, \omega]$ үшін қайтарымды болса және

I) $v(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$ функциясының дербес туындысы бар және $\frac{\partial v(t, x)}{\partial t} \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$, $u(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$, $w(t, x) \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R})$;

II) (5) дифференциалдық теңдеулер әулеті әрбір $[\theta_{s-1}, \theta_s) \times [0, \omega]$, $s = \overline{1, N}$, ішкі облысында $v(t, x)$, $u(t, x)$ және $w(t, x)$ функциялары үшін орындалады және (θ_{s-1}, x) , $s = \overline{1, N}$, $x \in [0, \omega]$ нүктелеріндегі $v(t, x)$ функциясының t бойынша оң жақты дербес туындысы үшін де орынды;

III) (6) шеттік шарты $v(t, x)$ функциясының әрбір $t=0$, $t=T$ түзулерінде орындалсын;

(IV) $u(t, x)$ және $w(t, x)$ функциялары $v(t, x)$ және $\frac{\partial v(t, x)}{\partial t}$ функцияларымен (8) интегралдық теңдеулері арқылы байланысқан болса, онда (13)–(17) параметрлері бар есептің жалғыз шешімі бар.

Негізгі нәтиже және мысал.

(1)–(4) және (13)–(17) пара-пар есептерден келесіні аламыз.

Теорема 2. Егер $N \times N$ өлшемді $Q(x)$ матрицасы барлық $x \in [0, \omega]$ үшін қайтарымды болса және төмендегі шарттар орындалса:

I) $u(t, x)$ функциясының дербес туындылары бар:

$$\frac{\partial u(t, x)}{\partial x} \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R}), \quad \frac{\partial u(t, x)}{\partial t} \in PC(\Omega, \{\theta_j\}_{j=1}^{N-1}, \mathbb{R});$$

II) $\frac{\partial^2 u}{\partial t \partial x}$ аралас дербес туындысы бар және әрбір $(t, x) \in \Omega$ нүктесінде, барлық $x \in [0, \omega]$ үшін (θ_{s-1}, x) , $s = \overline{1, N}$ нүктелерін қоспағанда, бір жақты аралас дербес туындылары бар;

III) (1) гиперболалық теңдеу әрбір $[\theta_{s-1}, \theta_s) \times [0, \omega]$, $s = \overline{1, N}$ үшін ішкі облыста $u(t, x)$ функциясы бойынша орындалады және $u(t, x)$ функциясының екінші ретті аралас дербес туындысы оң жақтан $[\theta_{s-1}, \theta_s) \times [0, \omega]$, $s = \overline{1, N}$, $x \in [0, \omega]$ нүктелерінде бар;

(IV) $u(t, x)$ үшін (2) шеттік шарт және (4) бастапқы шарт, сәйкес, $t=0$, $t=T$ және $x=0$ орындалады;

(V) $u(t, x)$ функциясы үшін (3) импульстік әсер шарттары $t = \theta_p$, $p = \overline{1, N-1}$, $x \in [0, \omega]$ орындалса онда бөлікті тұрақты импульстік гиперболалық теңдеу үшін шеттік (1)–(4) есептің жалғыз шешімі бар.

1 және 2 теоремалардың дәлелдеу схемасы [5] мақаланың теоремаларының дәлелдеу схемасы бойынша жүзеге асырылады.

Мысал. Алынған нәтижелердің иллюстрациясы ретінде келесі мысалды қарастырайық:

$$\frac{\partial^2 u}{\partial t \partial x} = (x + t) \frac{\partial u(t, x)}{\partial x} + (x^2 + t^2) \frac{\partial u(t, x)}{\partial t} - 2(x + t)u(t, x) + 3(x + t)u(\gamma(t), x) + f(t, x), \quad t \neq 0, 5; \quad (23)$$

$$3 \frac{\partial u(0, x)}{\partial x} + x^2 u(0, x) - 2xu(1, x) = x^4 - 3x^3 + 4x, \quad x \in [0, 1], \quad (24)$$

$$\frac{\partial^2 u}{\partial t \partial x} = A(t, x) \frac{\partial u(t, x)}{\partial x} + B(t, x) \frac{\partial u(t, x)}{\partial t} + C(t, x)u(t, x) + A_0(t, x) \frac{\partial u(\gamma(t), x)}{\partial x} + B_0(t, x) \frac{\partial u(\gamma(t), x)}{\partial t} + C_0(t, x)u(\gamma(t), x) + f(t, x),$$

$$\lim_{t \rightarrow 0,5+0} u(t, x) - \lim_{t \rightarrow 0,5-0} u(t, x) = 2x^2 + 3x + 5, \quad x \in [0, 1], \quad (25)$$

$$u(t, 0) = t^2. \quad (26)$$

Мұндағы $u(t, x)$ белгісіз функция және

$$A(t, x) = x + t, \quad B(t, x) = x^2 + t^2, \quad C = -2(x + 1), \quad C_0(t, x) = 3(x + t), \quad A_0(t, x) = B_0(t, x) = 0, \quad P_0(x) = x^2, \quad P_2(x) = 3, \quad S_0(x) = -2x, \quad P_1(x) = S_1(x) = S_2(x) = 0,$$

$\psi(t) = t^2$, $\varphi(x) = 2x^2 + 3x + 5$, $\varphi_0(x) = x^4 - 3x^3 + 4x$ функциялары Ω облысында үзіліссіз.

$$\gamma_1(t) = \zeta_1 = \frac{1}{4}, \quad \text{егер } t \in [0, 0.5), \quad \gamma_2(t) = \zeta_2 = \frac{3}{4} \quad \text{егер } t \in [0.5, 1].$$

Алдымен, $v(t, x) = \frac{\partial u(t, x)}{\partial x}$, $w(t, x) = \frac{\partial u(t, x)}{\partial t}$ жаңа функцияларын енгіземіз. (23)–(26) есебі бөлікті тұрақты аргументі бар импульсті дифференциалдық теңдеулер әулетіне үшін шеттік есепке ауысады.

$$\frac{\partial v(t, x)}{\partial t} = (x + t)v(t, x) + (x^2 + t^2)w(t, x) - 2(x + t)u(t, x) + 3(x + t)u(\gamma(t), x) + f(t, x), \quad t \neq 0, 5;$$

$$3v(0, x) + x^2 u(0, x) - 2xu(1, x) = x^4 - 3x^3 + 4x, \quad x \in [0, 1],$$

$$\lim_{t \rightarrow 0,5+0} v(t, x) - \lim_{t \rightarrow 0,5-0} v(t, x) = 4x + 3, \quad x \in [0, 1],$$

$$u(t, x) = t^2 + \int_0^x v(t, \xi) d\xi, \quad w(t, x) = 2t + \int_0^x \frac{\partial v(t, \xi)}{\partial t} d\xi.$$

$$\theta_{s-1} < \zeta_s < \theta_s, \quad s = 1, 2; \quad 0 < 0.5 < 1;$$

$$u(t, 0) = t^2, \quad x \in [0, 1]$$

$v([t], x)$ функциясы $C(\Omega, \Delta_2(\omega), \mathbb{R}^2)$ кеңістігінде жатады, және оның элементтері $v_s(t, x)$, $s = 1, 2$, төмендегі дифференциалдық теңдеулер жүйесін қанағаттандырады:

$$\frac{\partial v_1(t, x)}{\partial t} = (x + t)v_1(t, x) + (x^2 + t^2)w(t, x) - 2(x + t)u(t, x) + 3(x + t)u(0.25, x) + 2x^2 + 2xt + x^3 - 2xt^2 + \frac{3x}{16} + \frac{3t}{16} + 3tx^2,$$

$$(t, x) \in \Omega_1 = [0, 0.5) \times [0, 1];$$

$$\frac{\partial v_2(t, x)}{\partial t} = (x + t)v_2(t, x) + (x^2 + t^2)w(t, x) - 2(x + t)u(t, x) + 3(x + t)u(0.75, x) + 2x^2 + 2xt + x^3 - 2xt^2 + \frac{27x}{16} + \frac{27t}{16} + 3tx^2,$$

$$(t, x) \in \Omega_1 = [0.5, 1) \times [0, 1];$$

$$3v(0, x) + x^2 u(0, x) - 2xu(1, x) = x^4 - 3x^3 + 4x, \quad x \in [0, 1],$$

$$\lim_{t \rightarrow 0,5+0} v(t, x) - \lim_{t \rightarrow 0,5-0} v(t, x) = 4x + 3, \quad x \in [0, 1], \quad u(t, x) = t^2 + \int_0^x v(t, \xi) d\xi,$$

$$w(t, x) = 2t + \int_0^x \frac{\partial v(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = 1, 2.$$

$$u(t, 0) = t^2, \quad x \in [0, 1]$$

Біз келесі түрде функционалдық параметрлерді енгіземіз:

$$\mu_1(x) = v_1(\zeta_s, x), \quad \mu_2(x) = v_2(\zeta_s, x).$$

Келесі ауыстыруды жасай отырып: $v_s(t, x) = \tilde{v}_s(t, x) + \mu_s(x)$, $(t, x) \in \Omega_s$, $s = \overline{1, 2}$, төмендегідей түрдегі параметрлі дифференциалдық теңдеулер әулеті үшін есеп аламыз:

$$\frac{\partial v_1(t, x)}{\partial t} = (x + t)(\tilde{v}_1(t, x) + \mu_1(x)) + (x^2 + t^2)w(t, x) - 2(x + t)u(t, x) + 3(x + t)u(0.25, x) + 2x^2 + 2xt + x^3 - 2xt^2 + \frac{3x}{16} + \frac{3t}{16} + 3tx^2,$$

$$(t, x) \in \Omega_1 = [0, 0.5] \times [0, 1];$$

$$\frac{\partial v_2(t, x)}{\partial t} = (x + t)(\tilde{v}_2(t, x) + \mu_2(x)) + (x^2 + t^2)w(t, x) - 2(x + t)u(t, x) + 3(x + t)u(0.75, x)2x^2 + 2xt + x^3 - 2xt^2 + \frac{27x}{16} + \frac{27t}{16} + 3tx^2,$$

$$(t, x) \in \Omega_2 = [0.5, 1] \times [0, 1]; \quad (27)$$

$$3[\tilde{v}_s(0, x) + \mu_s(x)] + x^2u(0, x) - 2xu(1, x) = x^4 - 3x^3 + 4x, \quad x \in [0, 1], \quad (28)$$

$$\lim_{t \rightarrow 0.5+0} [\tilde{v}_2(t, x) + \mu_2(x)] - \lim_{t \rightarrow 0.5-0} [\tilde{v}_1(t, x) + \mu_1(x)] = 4x + 3, \quad x \in [0, 1], \quad (29)$$

$$u(t, x) = t^2 + \int_0^x [\tilde{v}_s(t, \xi) + \mu_s(\xi)] d\xi,$$

$$w(t, x) = 2t + \int_0^x \frac{\partial \tilde{v}_s(t, \xi)}{\partial t} d\xi, \quad (t, x) \in \Omega_s, \quad s = 1, 2. \quad (30)$$

Мынадай белгілеу енгізейік:

$$\alpha_1(t, x) = \int_{\frac{1}{4}}^t (x + t) d\tau = xt + \frac{t^2}{2} - \frac{x}{4} - \frac{1}{32}, \quad (t, x) \in \Omega_1;$$

$$\alpha_2(t, x) = \int_{\frac{5}{4}}^t (x + t) d\tau = xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}, \quad (t, x) \in \Omega_2.$$

Бекітілген $\mu_s(x)$, $u(t, x)$, $w(t, x)$ үшін (27)–(30) Коши есептері әулетінің шешімі жалғыз және келесі түрге ие:

$$\begin{aligned} \tilde{v}_1(t, x) = & e^{xt + \frac{t^2}{2} - \frac{x}{4} - \frac{1}{32}} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} (x + \tau) d\tau \mu_1(x) \\ & + e^{xt + \frac{t^2}{2} - \frac{x}{4} - \frac{1}{32}} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} [(x^2 + \tau^2)w(\tau, x) - 2(x + \tau)u(\tau, x)] d\tau \\ & + e^{xt + \frac{t^2}{2} - \frac{x}{4} - \frac{1}{32}} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} [3(x + \tau)u(0.25, x)] d\tau \\ & + e^{xt + \frac{t^2}{2} - \frac{x}{4} - \frac{1}{32}} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} (2x^2 + 2x\tau + x^3 - 2x\tau^2 + \frac{3x}{16} + \frac{3\tau}{16} + 3\tau x^2) d\tau, \end{aligned}$$

$$\begin{aligned} \tilde{v}_2(t, x) = & e^{xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} (x + \tau) d\tau \mu_2(x) \\ & + e^{xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} [(x^2 + \tau^2)w(\tau, x) - 2(x + \tau)u(\tau, x)] d\tau \\ & + e^{xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} [3(x + \tau)u(0.75, x)] d\tau \\ & + e^{xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} \left(2xt + x^3 - 2xt^2 + \frac{27x}{16} + \frac{27t}{16} + 3tx^2 \right) d\tau, \\ (t, x) \in \Omega_s, \quad s = \overline{1, N}. \end{aligned}$$

Келесі белгілеуді енгіземіз:

$$\begin{aligned} D_1(t, x) &= e^{xt + \frac{t^2}{2} - \frac{x}{4} - 1/32} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} (x + \tau) d\tau = -1 + e^{xt + \frac{t^2}{2} - \frac{x}{4} - 1/32}, \\ D_2(t, x) &= e^{xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} (x + \tau) d\tau = -1 + e^{xt + \frac{t^2}{2} - \frac{3x}{4} - 9/32}, \\ H_1(t, x, w, u) &= e^{xt + \frac{t^2}{2} - \frac{x}{4} - 1/32} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} [(x^2 + \tau^2)w(\tau, x) - 2(x + \tau)u(\tau, x)] d\tau \\ &+ e^{xt + \frac{t^2}{2} - \frac{x}{4} - 1/32} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} [3(x + \tau)u(0.25, x)] d\tau \\ H_2(t, x, w, u) &= e^{xt + \frac{t^2}{2} - \frac{3x}{4} - 9/32} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} [(x^2 + \tau^2)w(\tau, x) - 2(x + \tau)u(\tau, x)] d\tau \\ &+ e^{xt + \frac{t^2}{2} - \frac{x}{4} - 1/32} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} [3(x + \tau)u(0.75, x)] d\tau \\ F_1(t, x) &= e^{xt + \frac{t^2}{2} - \frac{x}{4} - \frac{1}{32}} \int_{1/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{x}{4} - \frac{1}{32})} (2x^2 + 2x\tau + x^3 - 2x\tau^2 + \frac{3x}{16} + \frac{3\tau}{16} + 3\tau x^2) d\tau, \\ F_2(t, x) &= e^{xt + \frac{t^2}{2} - \frac{3x}{4} - \frac{9}{32}} \int_{3/4}^t e^{-(x\tau + \frac{\tau^2}{2} - \frac{3x}{4} - \frac{9}{32})} \left(2xt + x^3 - 2xt^2 + \frac{27x}{16} + \frac{27t}{16} + 3tx^2 \right) d\tau. \end{aligned}$$

Табылған өрнектерді (20) және (21) қатынастарына қойып, келесі шектер мен мәндерді табамыз:

$$3e^{-\frac{x}{4} - 1/32} \mu_1(x) - 0 * e^{-\frac{x}{4} - \frac{1}{32}} \mu_2(x) = -3[F_1(0, x) + H_1(0, x, w, u)] - x^2 u(0, x) - 2xu(T, x) + \varphi_0(x). \quad (50)$$

$$\begin{aligned} e^{-\frac{x}{4} - \frac{1}{32}} \mu_1(x) + e^{-\frac{3x}{4} - \frac{9}{32}} \mu_2(x) &= \phi_p(x) + F_1(\theta_p, x) - F_2(\theta_p, x) + \\ &H_1(\theta_p, x, w, u) - H_2(\theta_p, x, w, u), \quad p = \overline{1, N-1}. \end{aligned}$$

$$Q(x) = \begin{bmatrix} 3e^{-\frac{x}{4} - 1/32} & 0 * e^{-\frac{x}{4} - 1/32} \\ -e^{-\frac{x}{4} - 1/32} & e^{-\frac{3x}{4} - 9/32} \end{bmatrix}, \quad \det Q(x) = 3e^{-x - 5/16} \neq 0.$$

$Q(x)$ матрицасының анықтаушы нолден өзгеше болатыны анық. Демек, $Q(x)$ матрицасының керісі бар және жоғарыдағы келтірілген теорема бойынша қарастырылып отырған есептің шешімі бар және ол жалғыз келесі түрде

$$u^*(t, x) = x^2 + t^2$$

өрнектеуге болады.

Сонымен қатар, мысалда берілген (23)-(26) тендеуде $A_0(t, x)$, $B_0(t, x)$ функцияларын нөлден өзгеше болатындай $A_0(t, x) = x - t$, $B_0(t, x) = 3x - 2t$ деп алып, дербес жағдайын қарастыруға болады.

$$\begin{aligned} \frac{\partial^2 u}{\partial t \partial x} &= (x+t) \frac{\partial u(t, x)}{\partial x} + (x^2 + t^2) \frac{\partial u(t, x)}{\partial t} - 2(x+t)u(t, x) + (x-t) \frac{\partial u(\gamma(t), x)}{\partial x} \\ &\quad + (3x-2t) \frac{\partial u(\gamma(t), x)}{\partial t} + 3(x+t)u(\gamma(t), x) + f(t, x), \quad t \neq 0,5; \\ 3 \frac{\partial u(0, x)}{\partial x} + x^2 u(0, x) - 2xu(1, x) &= x^4 - 3x^3 + 4x, \quad x \in [0,1], \\ \lim_{t \rightarrow 0,5+0} u(t, x) - \lim_{t \rightarrow 0,5-0} u(t, x) &= 2x^2 + 3x + 5, \quad x \in [0,1], \\ u(t, 0) &= t^2. \end{aligned}$$

Мұндағы $u(t, x)$ белгісіз функция және

$$A(t, x) = x + t, \quad B(t, x) = x^2 + t^2, \quad C = -2(x+1), \quad A_0(t, x) = x - t, \quad B_0(t, x) = 3x - 2t, \\ C_0(t, x) = 3(x+t),$$

$P_0(x) = x^2$, $P_2(x) = 3$, $S_0(x) = -2x$, $P_1(x) = S_1(x) = S_2(x) = 0$,
 $\psi(t) = t^2$, $\varphi(x) = 2x^2 + 3x + 5$, $\varphi_0(x) = x^4 - 3x^3 + 4x$ функциялары Ω облысында үзіліссіз.

$$\gamma_1(t) = \zeta_1 = \frac{1}{2}, \text{ егер } t \in [0, 0.5), \quad \gamma_2(t) = \zeta_2 = \frac{3}{2}, \text{ егер } t \in [0.5, 1].$$

$$D_1(t, x) = e^{xt + \frac{t^2}{2} \frac{x}{4} \frac{1}{32}} \int_{\frac{1}{2}}^t e^{-\left(x\tau + \frac{\tau^2}{2} \frac{x}{4} \frac{1}{32}\right)} (x + \tau + x - \tau) d\tau$$

$$D_2(t, x) = e^{xt + \frac{t^2}{2} \frac{3x}{4} \frac{9}{32}} \int_{\frac{3}{2}}^t e^{-\left(x\tau + \frac{\tau^2}{2} \frac{3x}{4} \frac{9}{32}\right)} (x + \tau + x - \tau) d\tau.$$

$$Q(x) = \begin{bmatrix} 3P_2(x)[1 + D_1(0, x)] & 0 * S_2(x)[1 + D_N(T, x)] \\ -[1 + D_1(\theta_1, x)] & 1 + D_2(\theta_1, x) \end{bmatrix}, \quad \det Q(x) \neq 0.$$

Бұл қойылған есеп үшін де $Q(x)$ матрицасы нөлден өзгеше екенін көреміз. Демек, $Q(x)$ матрицасының қайтарымы бар және мақалада ұсынылған алгоритм бойынша есептің шешімін бар екендігіне көз жеткізуге болады.

Қорытынды

Мақалада бөлікті тұрақты аргументі бар импульстік гиперболалық тендеу үшін қойылған шеттік есеп қарастырылған. Джумабаевтың параметрлеу әдісін қолдану арқылы есептің шешімінің бар болуы мен бірімәнділігінің жеткілікті шарттары алынды. Бастапқы шеттік есеп параметрлері бар Коши есептерінің эквивалентті әулетіне келтірілді. Шешімнің жуық мәнін анықтауға арналған итерациялық алгоритм ұсынылды. $Q(x)$ матрицасының кері матрицасы бар болу шартында есептің параметрлік және бастапқы түрлері үшін шешімнің бірімәнділігі қамтамасыз етілетіні көрсетілді. Алынған нәтижелер бөлікті тұрақты аргументі және импульстік әсерлері бар гиперболалық тендеулер үшін шеттік есептер теориясын толықтырады және оларды сандық тұрғыдан жүзеге асыруға негіз болады.

Қаржыландыру туралы ақпарат. Бұл зерттеуді Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитеті (грант № AP26194447) қаржыландырады.

ӘДЕБИЕТТЕР

1 Wiener, J. Generalized Solutions of Functional Differential Equations (Singapore: World Scientific, 1993).

- 2 Samoilenko, A.M., and Perestyuk, N.A. Impulsive Differential Equations (Singapore: World Scientific, 1995).
- 3 Imanchiyev, A.E., Assanova, A.T., and Molybaikyzy, A. Properties of a nonlocal problem for hyperbolic equations with impulse discrete memory. Lobachevskii Journal of Mathematics, 44 (10), 4299–4309 (2023). <https://doi.org/10.1134/S1995080223100177>
- 4 Dzhumabayev, D.S. Criteria for the unique solvability of a linear boundary-value problem for an ordinary differential equation. U.S.S.R. Computational Mathematics and Mathematical Physics, 29 (1), 34–46 (1989).
- 5 Assanova, A.T. Hyperbolic equation with piecewise-constant argument of generalized type and solving boundary value problems for it. Lobachevskii Journal of Mathematics, 42, 3584–3593 (2021).
- 6 Assanova, A.T., and Molybaikyzy, A. Solution to the periodic problem for the impulsive hyperbolic equation with discrete memory. Kazakh Mathematical Journal, 25 (1), 16–27 (2025).
- 7 Akhmet, M.U. Integral manifolds of differential equations with piecewise constant argument of generalized type. Nonlinear Analysis, 66 (3), 367–383 (2007). <https://doi.org/10.1016/j.na.2005.11.032>
- 8 Akhmet, M.U. Almost periodic solutions of differential equations with piecewise constant argument of generalized type. Journal of Mathematical Analysis and Applications, 336 (4), 646–663 (2007). <https://doi.org/10.1016/j.jmaa.2007.03.010>
- 9 Nieto, J.J., and Rodriguez-Lopez, R. Green's function for second order periodic BVPs with piecewise constant argument. Journal of Mathematical Analysis and Applications, 304 (1), 33–57 (2005). <https://doi.org/10.1016/j.jmaa.2004.09.023>
- 10 Aviltay, N., Akhmet, T., and Zhamanshin, A. Asymptotic solutions of differential equations with singular impulses. Carpathian Journal of Mathematics, 40 (3), 581–598 (2024).

^{1,3*}**Молыбайқызы А.,**

научный сотрудник, ст. преподаватель, ORCID ID: 0009-0008-2452-5932,
e-mail: altynaimolybai@gmail.com

^{1,2}**Кабдрахова С.С.,**

к. ф.-м.н., главный научный сотрудник, ORCID ID: 0000-0003-0247-5985,
e-mail: Symbat2909.sks@gmail.com

¹**Искакова Н.Б.,**

к. ф.-м.н., главный научный сотрудник, ORCID ID: 0000-0002-8697-8920,
e-mail: narkesh77@gmail.com

^{1,3}**Минглибаева Б.Б.,**

к. ф.-м.н., научный сотрудник, ст. преподаватель, ORCID ID: 0000-0002-7195-4480,
e-mail: minglibayeva.bayan@gmail.com

¹Институт математики и математического моделирования, г. Алматы, Казахстан

²Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

³Казахский национальный женский педагогический университет, г. Алматы, Казахстан

О РЕШЕНИИ КРАЕВОЙ ЗАДАЧИ ДЛЯ ИМПУЛЬСНОГО ГИПЕРБОЛИЧЕСКОГО УРАВНЕНИЯ С КУСОЧНО-ПОСТОЯННЫМ АРГУМЕНТОМ

Аннотация

В данной статье рассматривается краевая задача для импульсного гиперболического уравнения с кусочно-постоянным аргументом. Импульсные гиперболические уравнения с кусочно-постоянным аргументом возникают как математические модели физических процессов в нейронных сетях, непрерывных динамических системах, гибридных системах и других областях. Вопросы существования краевых задач и построения их решений для таких уравнений в настоящее время остаются одними из актуальных проблем. Для получения условий разрешимости данной задачи используется метод параметризации Джумабаева, а также в работе разработан итерационный алгоритм нахождения приближенного решения. Для каждого шага итерационного процесса получены интегральные формулы, выражаемые через матрицу $Q(x)$, описывающую

связь между функциональными параметрами. Если данная матрица является обратимой, то доказываются существование и единственность решения как для параметрической, так и для исходной задачи. Предложенный метод не ограничивается лишь доказательством теоретической разрешимости задачи, а предлагает конкретную конструктивную процедуру нахождения решения. Это имеет важное значение для последующих численных реализаций и анализа устойчивости решений. Кроме того, предложенный подход может быть применен к другим типам задач с кусочно-постоянным аргументом, включая системы с импульсными условиями, нейронные сети с эффектом памяти и нелинейные гибридные модели.

Ключевые слова: импульсное гиперболическое уравнение, кусочно-постоянный аргумент, краевая задача, алгоритм.

^{1,3*}**Molybaikyzy A.,**

Researcher, Senior Lecturer, ORCID ID: 0009-0008-2452-5932,
e-mail: altynaimolybai@gmail.com

^{1,2}**Kabdrakhova S.S.,**

Cand.Phys.Math.Sc., Chief Researcher, ORCID ID: 0000-0003-0247-5985,
e-mail: Symbat2909.sks@gmail.com

¹**Iskakova N.B.,**

Cand.Phys.Math.Sc., Chief Researcher, ORCID ID: 0000-0002-8697-8920,
e-mail: narkesh77@gmail.com

^{1,3}**Minglibayeva B.B.,**

Cand.Phys.Math.Sc., Researcher, Senior Lecturer, ORCID ID: 0000-0002-7195-4480,
e-mail: minglibayeva.bayan@gmail.com

¹Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan

²Al-Farabi Kazakh National University, Almaty, Kazakhstan

³Kazakh National Women's Teacher Training University, Almaty, Kazakhstan

ON THE SOLUTION OF A BOUNDARY VALUE PROBLEM FOR AN IMPULSIVE HYPERBOLIC EQUATION WITH A PIECEWISE-CONSTANT ARGUMENT

Abstract

This article investigates a boundary value problem for an impulsive hyperbolic equation with a piecewise constant argument. Impulsive hyperbolic equations with a piecewise constant argument arise as mathematical models of physical processes in neural networks, continuous dynamical systems, hybrid systems, and other fields. Issues related to the existence of boundary value problems and the construction of their solutions for such equations remain among the most relevant and challenging problems at present. To obtain solvability conditions for the considered problem, the Dzhumabaev parametrization method is employed, and an iterative algorithm for constructing an approximate solution is developed. For each step of the iterative process, integral formulas are derived and expressed in terms of the matrix $Q(x)$, which describes the relationship between the functional parameters. If this matrix is invertible, the existence and uniqueness of the solution are proved both for the parametric and the original problems. The proposed method is not limited to proving theoretical solvability; it also provides a concrete constructive procedure for obtaining the solution. This is of significant importance for subsequent numerical implementations and for the analysis of solution stability. Moreover, the proposed approach can be applied to other types of problems with a piecewise constant argument, including systems with impulsive conditions, neural networks with memory effects, and nonlinear hybrid models.

Keywords: impulsive hyperbolic equation, piecewise constant argument, boundary value problem, algorithm.

Received February 5, 2026; revised April 12, April 20, 2026; accepted April 24, 2026

UDC 512.57
IRSTI 27.17.23

<https://doi.org/10.55452/1998-6688-2026-23-2-75-82>

^{1*}**Yersaliyeva A.,**

Master's student, ORCID ID: 0009-0007-0952-706X,

*e-mail: airisays@gmail.com

¹SDU University, Kaskelen, Kazakhstan

LIE POLYNOMIALS IN FREE SPECIAL TORTKEN ALGEBRAS

Abstract

This paper studies Lie elements and symmetric (Tortken) elements in a free Novikov algebra and examines whether a nonzero multilinear element can belong to both classes simultaneously. We use the Euler operator and the null Lagrangian criterion to test membership in the symmetric subspace for elements represented in the standard multilinear Lie basis. For the Lie component, we employ left-normed commutators with a fixed first variable, which form a convenient basis of the multilinear part. The case $n = 3$ is worked out explicitly by expanding the commutators in the Novikov product and applying the Euler operator. For degrees $n = 4, 5, 6, 7$, the corresponding linear systems are obtained and solved computationally in Wolfram Mathematica and Albert. The computations show that the intersection of the multilinear Lie subspace with the subspace of symmetric elements is trivial for all $n \leq 7$. Thus, up to degree 7 there is no nonzero multilinear element in a free Novikov algebra that is simultaneously Lie and symmetric. These results provide a starting point for studying the problem in higher degrees.

Keywords: Lie polynomials, free special Tortken algebras, Lie elements, free Novikov algebras, multilinear Lie elements, Euler operator, symmetric elements.

Received March 25, 2026; accepted April 24, 2026.

Introduction

An algebra A over a field $\mathbb{K}$ is called a (right) Novikov algebra if it satisfies the following two identities:

$$(a, b, c) = (a, c, b) \quad (1)$$

$$a(bc) = b(ac) \quad (2)$$

for all $a, b, c \in A$, where $(a, b, c) = (ab)c - a(bc)$.

The identity (1) is called the right-symmetric identity, and any algebra satisfying this identity is called a right-symmetric algebra. The identity (2) is called the left-commutativity identity. The right-symmetric algebras are a well known class of algebra that appear in geometry and physics [4].

Novikov algebras form an important subclass of right-symmetric algebras. They first appeared in the study of Hamiltonian operators in the formal calculus of variations by Gel'fand and Dorfman [3], and later in the classification of linear Poisson brackets of hydrodynamic type by Balinskii and Novikov [2]. A standard method for constructing Novikov algebras is the Gel'fand-Dorfman construction, which uses a differential polynomial algebra with one derivation.

Let $\mathbb{K}[x]$ be the polynomial algebra in one variable x over a field $\mathbb{K}$, and let

$$D = \frac{d}{dx}$$

Define a bilinear multiplication $\cdot$ on $\mathbb{K}[x]$ by

$$f \cdot g := D(f)g = f'(x)g(x), f, g \in \mathbb{K}[x]. \quad (3)$$

Then it is easy to check that $(\mathbb{K}[x], \cdot)$ is a (right) Novikov algebra.

A vector space L over a field $\mathbb{K}$ equipped with a bilinear map

$$[\cdot, \cdot]: L \times L \rightarrow L$$

is called a Lie algebra if it satisfies the following two identities:

$$[a, b] = -[b, a] \quad (4)$$

$$[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0 \quad (5)$$

for all $a, b, c \in L$.

Since right-symmetric algebras are Lie-admissible, Novikov algebras are also Lie-admissible. Namely, let A be a (right) Novikov algebra with multiplication xy . Define the commutator (Lie bracket) on A by

$$[a, b] := ab - ba, \quad a, b \in A \quad (6)$$

Then $(A, [\cdot, \cdot])$ is a Lie algebra.

The Witt algebra W_1 of index 1 is the Lie algebra of all derivations of $\mathbb{K}[x]$ over:

$$W_1 = \{u\partial \mid u \in \mathbb{K}[x]\}$$

where $\partial = \partial/\partial x$. Define a product on W_1 by

$$u\partial \cdot v\partial = (v\partial(u))\partial$$

Then $(W_1, \cdot)$ is a Novikov algebra, called the Novikov-Witt algebra.

Define the symmetrized product (anticommutator) by

$$a \circ b := ab + ba.$$

The space of symmetric elements is the smallest linear subspace of A that contains X and is closed under the operation $\circ$. Elements of this subspace are called symmetric elements, or Tortken elements. Dzhumadil'daev [6] proved that the anticommutator algebra of any Novikov algebra satisfies the following identity:

$$(a \circ b) \circ (c \circ d) - (a \circ d) \circ (c \circ b) = (a, b, c)_\circ \circ d - (a, d, c)_\circ \circ b$$

where

$$(a, b, c)_\circ = a \circ (b \circ c) - (a \circ b) \circ c$$

A commutative algebra satisfying the identity above is called a Tortken algebra.

Later, he showed in [7] that Novikov algebras additionally satisfy an identity of degree 5:

$$\begin{aligned} & (((a \circ a) \circ a) \circ b) \circ b + (((a \circ b) \circ b) \circ a) \circ a + 2(((a \circ a) \circ b) \circ b) \circ a \\ & + 2(((a \circ b) \circ a) \circ a) \circ b - 3(((a \circ a) \circ b) \circ a) \circ b - 3(((a \circ b) \circ a) \circ b) \circ a = 0. \end{aligned}$$

This identity is called the besken identity, and it is the special identity satisfied by Novikov algebras under the anticommutator.

Recently, Dzhumadil'daev and Ismailov [9] studied Tortken elements in free Novikov algebras. They constructed a basis for the space of Tortken elements and obtained a criterion for determining them. In addition, the module structure of the space of Tortken elements over the symmetric group was completely described. The criterion was given in terms of Euler operators. However, a Lie criterion for Novikov polynomials is still an open question. This raises the essential question of whether there exists a Lie polynomial that is Tortken. In the classical case, every homogeneous Lie element of odd degree in a free associative algebra can be written in terms of anticommutator products of the given variables [10]. An analogue of this fact also holds for bicommutative algebras [8].

In this paper, we investigate whether Lie elements of odd degree in a free Novikov algebra are Tortken. We show that, in general, they are not. We verify this up to degree seven by means of computations carried out in Wolfram Mathematica. We further conjecture that the spaces of Lie elements and Tortken elements are disjoint in the space of Novikov polynomials. This work may be viewed as a step toward the study of Lie elements in free Novikov algebras.

Throughout the paper, all algebras are considered over a field of characteristic zero.

Materials and methods

Let $x = (x_1, \dots, x_m)$, and let $x_i^{(k)} = D^k(x_i)$, where D is the total derivative. The Euler operator [12] is the m -tuple $E = (E_1, \dots, E_m)$ with components

$$E_i = \sum_{k \geq 0} (-D)^k \frac{\partial}{\partial x_i^{(k)}}, \quad i = 1, \dots, m$$

In [9], Dzhumadil'daev and Ismailov proved the following statement. Suppose that $f = f(x_1, \dots, x_n)$ is a homogeneous polynomial in $\text{Nov}(X)$ such that $\deg_{x_i}(f) > 0$ for all i . Then the following conditions are equivalent:

- (1) f is a symmetric, or Tortken, element and $\deg(f) > 1$;
- (2) f is a null Lagrangian, that is, $E_i(f) = 0$ for all $i \in \{1, \dots, n\}$;
- (3) $E_i(f) = 0$ for some $i \in \{1, \dots, n\}$.

For example, consider $x_1(t), x_2(t)$ and $D = \frac{d}{dt}$. Let

$$F = x_1 x_2' + (x_1')^2.$$

Since F depends only on x_1, x_2 and their derivations, the Euler operator $E = (E_1, E_2)$ reduces to

$$E_i(F) = \frac{\partial F}{\partial x_i} - D \left(\frac{\partial F}{\partial x_i'} \right), \quad i = 1, 2.$$

Hence

$$E_1(F) = x_2' - D(2x_1') = x_2' - 2x_1'', \quad E_2(F) = 0 - D(x_1') = -x_1'.$$

Since we work over a field of characteristic zero, it is enough to perform all calculations for multilinear polynomials, that is, for the subspace in which each variable $x_1, \dots, x_n$ occurs exactly once. It is known that a multilinear basis of the free Lie algebra in degree n is given as follows [13]. To obtain canonical representatives, one often uses left-normed bracketing (all brackets associate to the left) and fixes the first variable; namely, one takes all elements of the form

$$\left[\left[\dots \left[[x_1, x_{\sigma(2)}], x_{\sigma(3)} \right], \dots \right], x_{\sigma(n)} \right],$$

where σ runs over all permutations of the set $\{2, 3, \dots, n\}$. This gives exactly $(n-1)!$ elements and yields a standard convenient basis of the multilinear component.

In particular, for $n = 3$ and $n = 4$ we obtain the following basis elements.

$$\begin{aligned} & [[x_1, x_2], x_3], [[x_1, x_3], x_2], \\ & [[[x_1, x_2], x_3], x_4], [[[x_1, x_2], x_4], x_3], [[[x_1, x_3], x_2], x_4] \\ & [[[x_1, x_3], x_4], x_2], [[[x_1, x_4], x_2], x_3], [[[x_1, x_4], x_3], x_2]. \end{aligned}$$

Now recall the basis of the free Novikov algebra in terms of differential polynomials, as well as the basis of the space of Tortken elements in the free Novikov algebra. A. S. Dzhumadil'daev and C. Löfwall [5] constructed the following basis.

Let $\mathcal{M}$ be the set of all differential monomials in $x_1, \dots, x_n$ with one derivation. Then the following holds.

Theorem 1. The set

$$\mathcal{N} = \{u \in \mathcal{M} \mid \deg(u) - d(u) = 1\}$$

is a basis of the free Novikov algebra $\text{Nov}(X)$.

A basis of the space of Tortken elements in a free Novikov algebra was constructed by Dzhumadil'daev and Ismailov [9]:

Theorem 2. The set

$$\overline{\mathcal{M}} := X \cup \{u' \mid u \in \mathcal{M}, \deg(u) - d(u) = 2\}$$

is a basis of the space of Tortken elements.

Results and discussion

The case $n = 3$. In this subsection, we present our calculations in degree three. More precisely, by performing explicit calculations, we show that a nonzero Lie element of degree 3 is not Tortken.

Let Y be a Lie polynomial in the free Novikov algebra generated by x_1, x_2, x_3 . We express it as a linear combination of the multilinear basis elements of the free Lie algebra, using the bases introduced above:

$$Y(x_1, x_2, x_3) = \lambda_1 [[x_1, x_2], x_3] + \lambda_2 [[x_1, x_3], x_2], \quad \lambda_1, \lambda_2 \in \mathbb{K}$$

Using (6), we expand each commutator explicitly in free Novikov algebra:

$$\begin{aligned} [[x_1, x_2], x_3] &= (x_1 x_2 - x_2 x_1) x_3 = (x_1 x_2 - x_2 x_1) x_3 - x_3 (x_1 x_2 - x_2 x_1) = \\ &= (x_1 x_2) x_3 - (x_2 x_1) x_3 - x_3 (x_1 x_2) + x_3 (x_2 x_1) \\ [[x_1, x_3], x_2] &= (x_1 x_3 - x_3 x_1) x_2 = (x_1 x_3 - x_3 x_1) x_2 - x_2 (x_1 x_3 - x_3 x_1) = \\ &= (x_1 x_3) x_2 - (x_3 x_1) x_2 - x_2 (x_1 x_3) + x_2 (x_3 x_1) \end{aligned}$$

Next, we rewrite the above expressions using the Novikov product (3):

$$\begin{aligned} [[x_1, x_2], x_3] &= x_1'' x_2 x_3 - x_1 x_2'' x_3 - x_1' x_2' x_3' + x_1 x_2' x_3' = f_1 \\ [[x_1, x_3], x_2] &= x_1'' x_2 x_3 - x_1 x_2 x_3'' - x_1' x_2' x_3' + x_1 x_2' x_3' = f_2 \end{aligned}$$

In the next step, we apply the Euler operator with respect to only variable x_1 :

$$\begin{aligned} E_{x_1}(f_1) &= (-x_2'' x_3 + x_2' x_3') - D(-x_2 x_3') + D^2(x_2 x_3) = 4x_2' x_3' + 2x_2 x_3'' \\ E_{x_1}(f_2) &= (-x_2 x_3'' + x_2' x_3') - D(-x_2' x_3) + D^2(x_2 x_3) = 4x_2' x_3' + 2x_2'' x_3 \end{aligned}$$

Substituting the obtained expressions into the original Y , we obtain:

$$\begin{aligned} Y(X) &= \lambda_1 [[x_1, x_2], x_3] + \lambda_2 [[x_1, x_3], x_2] = \lambda_1 (4x_2' x_3' + 2x_2 x_3'') + \lambda_2 (4x_2' x_3' + 2x_2'' x_3) \\ Y(X) &= 4(\lambda_1 + \lambda_2) x_2' x_3' + 2\lambda_1 x_2 x_3'' + 2\lambda_2 x_2'' x_3 \end{aligned}$$

We equate the coefficients of the linearly independent monomials and solve the resulting system of equations:

$$\begin{cases} 4(\lambda_1 + \lambda_2) = 0, \\ 2\lambda_1 = 0, \\ 2\lambda_2 = 0, \end{cases} \Rightarrow \lambda_1 = 0, \lambda_2 = 0$$

The verification of the case $n = 5$ showed that the corresponding system has only the trivial solution $\lambda_1 = \lambda_2 = 0$. Therefore, in degree 3, there are no nonzero elements that are simultaneously Lie and Tortken.

The cases $n = 4, 5, 6$, and 7. For the cases $n = 4, 5, 6, 7$, the calculations become substantially more involved. Therefore, we use code written in Wolfram Mathematica together with the software Albert [1]. In Wolfram Mathematica, we implemented an algorithm to construct the corresponding expressions and performed direct symbolic computations, including simplification, expansion in a chosen basis, and verification of identities. Starting from degree 5, there appears a new Lie identity satisfied by the commutators in every Novikov algebra, and the corresponding basis becomes more complicated:

$$\sum_{\sigma \in S_{\{2,3,4,5\}}} (-1)^\sigma \left[x_{\sigma(1)}, \left[x_{\sigma(2)}, \left[x_{\sigma(3)}, \left[x_{\sigma(4)}, x_5 \right] \right] \right] \right] = 0.$$

For this reason, we use Albert to generate the multilinear basis of the space of Lie polynomials in the free Novikov algebra in degrees 5, 6, 7, and then use it to represent elements and to control dimensions and linear independence.

First we write a code for the Euler operator.

Listing 1. Code implementing the Euler operator E

```
 $\delta_j[f\_ , n\_Integer] := \text{ExpandAll}[\text{Sum}[(-1)^i \text{Nest}[\text{Dr}, D[f, \text{Nest}[\text{Dr}, x, i]], i], \{i, 0, n\}]]$ 
```

Before presenting the following code, we introduce the necessary definitions. We observe that after applying the Euler operator E_i (or δ_i in our code) to our multilinear polynomial, the resulting expression does not contain the variable x_i . Therefore, it is important to know what polynomials span the image of E_i . The set of differential monomials of total order n in the variables $x_{i_1}, \dots, x_{i_m}$ is defined by

$$B_n(x_{i_1}, \dots, x_{i_m}) = \left\{ \prod_{j=1}^m Dr^{\alpha_j}(x_{i_j}) : \alpha_j \in \mathbb{Z}_{\geq 0}, \sum_{j=1}^m \alpha_j = n \right\},$$

where Dr denotes the derivation D in our code, $Dr^0(x) = x$, and for $k \geq 1$,

$$Dr^k(x) = \underbrace{Dr(Dr(\dots Dr(x) \dots))}_{k \text{ times}}.$$

In other words, $B_n(x_{i_1}, \dots, x_{i_m})$ consists of all products of derivatives whose total number of applications of the operator Dr equals n .

The cardinality (the number of elements) of $B_n(x_{i_1}, \dots, x_{i_m})$ is given by

$$|B_n(x_{i_1}, \dots, x_{i_m})| = \binom{n+m-1}{m-1}$$

For $m = 4$, for example, (the variables x_2, x_3, x_4, x_5) we have

$$B_n(x_2, x_3, x_4, x_5) = \{Dr^{\alpha_2}(x_2)Dr^{\alpha_3}(x_3)Dr^{\alpha_4}(x_4)Dr^{\alpha_5}(x_5) : \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 = n\}.$$

Let $l[\cdot, \cdot]$ be a binary operation defined by formula (6). For $n \geq 2$, define the left-normed n -ary expression $l^{(n)}[y_1, \dots, y_n]$ recursively by setting

$$l^{(2)}[y_1, y_2] = l[y_1, y_2]$$

and for $n \geq 3$,

$$l^{(n)}[y_1, \dots, y_n] = l[l^{(n-1)}[y_1, \dots, y_{n-1}], y_n]$$

Then for variables $x_1, \dots, x_n$ we set

$$L_n[x_1, \dots, x_n] := \sum_{\sigma \in S_{n-1}} \lambda_{\sigma} l^{(n)}[x_1, x_{\sigma(2)}, x_{\sigma(3)}, \dots, x_{\sigma(n)}]$$

where S_{n-1} is the set of all permutations of $\{2, 3, \dots, n\}$, and λ_{σ} are arbitrary scalars.

For instance, when $n = 4$, the expression L_4 takes the form

$$\begin{aligned} L_4[x_1, x_2, x_3, x_4] = & \lambda_1 l[l[l[x_1, x_2], x_3], x_4] + \lambda_2 l[l[l[x_1, x_3], x_2], x_4] \\ & + \lambda_3 l[l[l[x_1, x_2], x_4], x_3] + \lambda_4 l[l[l[x_1, x_3], x_4], x_2] \\ & + \lambda_5 l[l[l[x_1, x_4], x_2], x_3] + \lambda_6 l[l[l[x_1, x_4], x_3], x_2] \end{aligned}$$

After introducing the definitions above, we can present the Wolfram Mathematica code that we used to compute the coefficients.

Listing 2. Determination of the coefficients λ_k using the Euler operator.

```
TableForm[
First[Solve[
Map[S,Coefficient[
Collect[ $\delta_1[$ 
ExpandAll[
 $L_n, n], B_m]$ 
Array[ $\lambda[\#]&, (n-1)!]$ ]]]]]
```

For clarity, we present the full code for the case $n = 4$

Listing 3. The case $n = 4$.

```
TableForm[First[Solve[Map[S,Coefficient[Collect[ $\delta_1[\lambda_1 l[l[x1,x2],x3],x4]+\lambda_2$ 
 $l[l[l[x1,x3],x2],x4]+\lambda_3 l[l[l[x1,x2],x4],x3]+\lambda_4 l[l[l[x1,x3],x4],x2]+\lambda_5 l[l[l[x1,x4],x2],x3]$ 
 $+\lambda_6 l[l[l[x1,x4],x3],x2],$ 
6
],
{Dr[x2]Dr[x3]Dr[x4],Dr[Dr[x4]]Dr[x3]x2,Dr[Dr[x3]]Dr[x4]x2,
Dr[Dr[x4]]Dr[x2]x3,Dr[Dr[x2]]Dr[x4]x3,Dr[Dr[Dr[x4]]]x2x3,
Dr[Dr[x3]]Dr[x2]x4,Dr[Dr[x2]]Dr[x3]x4,Dr[Dr[Dr[x3]]]x2x4,
Dr[Dr[Dr[x2]]]x3x4}
],
{Dr[x2]Dr[x3]Dr[x4],Dr[Dr[x4]]Dr[x3]x2,Dr[Dr[x3]]Dr[x4]x2,
Dr[Dr[x4]]Dr[x2]x3,Dr[Dr[x2]]Dr[x4]x3,Dr[Dr[Dr[x4]]]x2x3,
Dr[Dr[x3]]Dr[x2]x4,Dr[Dr[x2]]Dr[x3]x4,Dr[Dr[Dr[x3]]]x2x4,
Dr[Dr[Dr[x2]]]x3x4}]],Array[ $\lambda[\#]&,6]]]$ 
```

Again, we find that the system has only the trivial solution. To continue our calculations, we need several further facts. Using Molev's result [11], one can determine the codimensions of the space of Lie elements in a free Novikov algebra. The dimensions of the space of Lie polynomials up to degree 7 are

$$L(n) = 1, 1, 2, 6, 20, 71, 259, \dots$$

The explicit bases elements are obtained by Albert. The computations performed in Wolfram Mathematica show that for $n = 3, 4, 5, 6, 7$ the resulting system of linear equations has only the trivial solution

$$\lambda_1 = \dots = \lambda_r = 0.$$

Consequently, $y = 0$, that is,

$$\text{Lie}_n \cap T_n = \{0\} \text{ for all } n \leq 7$$

where T_n is the space of multilinear Tortken (symmetric elements) in free Novikov algebras.

In other words, up to degree 7 there is no nonzero element that is simultaneously a Lie element and a Tortken element.

Conclusion

The present work provides a partial answer to the question of the intersection of the spaces of Lie and symmetric elements in a free Novikov algebra. We have shown that for $n \leq 7$, the intersection $L_n \cap T_n$ is trivial, in contrast to the classical situation for associative algebras and to the case of bicommutative algebras. These results may serve as a starting point for further investigation of the problem in higher degrees. Conjecture 0.3. The intersection $L_n \cap T_n$ is trivial for any $n \geq 2$.

Acknowledgments. The author expresses her gratitude to her scientific supervisor, professor Nurlan Ismailov, for his guidance, helpful discussions, and support provided during the preparation of this work.

REFERENCES

- 1 Albert, A. Version 4.0M6. URL: <https://web.osu.cz/~Zusmanovich/soft/albert/> (accessed 2026).
- 2 Balinskii, A.A., and Novikov, S.P. Poisson brackets of hydrodynamic type, Frobenius algebras and Lie algebras. Soviet Mathematics Doklady, 32, 228–231 (1985). <https://web.archive.org/web/https://homepage.mi-ras.ru/~snovikov/95.pdf>

- 3 Gel'fand, I.M., and Dorfman, I.Ya. Hamiltonian operators and algebraic structures related to them. *Functional Analysis and Its Applications*, 13 (4), 248–262 (1979). <https://doi.org/10.1007/BF01078363>
- 4 Burde, D. Left-symmetric algebras, or pre-Lie algebras in geometry and physics. *Central European Journal of Mathematics*, 4 (3), 323–357 (2006). <https://doi.org/10.2478/s11533-006-0014-9>
- 5 Dzhumadil'daev, A., and Löfwall, C. Trees, free right-symmetric algebras, free Novikov algebras and identities. *Homology, Homotopy and Applications*, 4 (2), 165–190 (2002). <http://eudml.org/doc/50501>
- 6 Dzhumadil'daev, A.S. Novikov-Jordan algebras. *Communications in Algebra*, 30 (11), 5205–5240 (2002). <https://doi.org/10.1081/AGB-120015649>
- 7 Dzhumadil'daev, A.S. Special identity for Novikov-Jordan algebras. *Communications in Algebra*, 33 (5), 1279–1287 (2005). <https://doi.org/10.1081/AGB-200060504>
- 8 Dzhumadil'daev, A.S., and Ismailov, N.A. Polynomial identities of bicommutative algebras, Lie and Jordan elements. *Communications in Algebra*, 46 (12), 5242–5252 (2018). <https://doi.org/10.1080/00927872.2018.1461890>
- 9 Dzhumadil'daev, A.S., and Ismailov, N.A. Null Lagrangians in free Novikov algebras. *arXiv preprint* (2026). <https://arxiv.org/abs/2601.11168>
- 10 Jacobson, N. *Structure and Representations of Jordan Algebras* (Providence, RI: American Mathematical Society, 1968). <https://bookstore.ams.org/COLL/39>
- 11 Molev, A.I. On the algebraic structure of the Lie algebra of vector fields on the line. *Mathematics of the USSR-Sbornik*, 62 (1), 83–94 (1989). <https://www.mathnet.ru/eng/sm2650>
- 12 Olver, P.J. *Applications of Lie Groups to Differential Equations* (2nd ed.) (New York: Springer, 1993). <https://link.springer.com/book/10.1007/978-1-4612-4350-2>
- 13 Reutenauer, C. *Free Lie Algebras* (Oxford: Oxford University Press, 1993). <https://global.oup.com/academic/product/free-lie-algebras-9780198536796>

^{1*}**Ерсалиева А.,**
магистрант, ORCID ID: 0009-0007-0952-706X,
*e-mail: airisays@gmail.com

¹SDU University, Қаскелен қ., Қазақстан

ЕРКІН АРНАЙЫ ТӨРТКЕН АЛГЕБРАЛАРЫНДАҒЫ ЛИ КӨПМҮШЕЛЕРІ

Аңдатпа

Мақалада еркін Новиков алгебрасындағы Ли элементтері мен симметриялық (Төрткен) элементтер зерттеледі, сондай-ақ нөлден өзге мультилиниялы элементтің осы екі класқа бір мезгілде тиесілі болу мүмкіндігі қарастырылады. Стандартты мультилиниялы Ли базисінде берілген элементтердің симметриялық ішкікеңістікке тиесілігін тексеру үшін Эйлер операторы мен нөлдік лагранжиан критерийі қолданылады. Ли бөлігі үшін бірінші айнымалысы бекітілген сол жақтан нормаланған коммутаторлар пайдаланылады, олар мультилиниялы бөліктің ыңғайлы базисін құрайды. $n = 3$ жағдайы коммутаторларды Новиков көбейтіндісі арқылы жіктеу және Эйлер операторын қолдану арқылы айқын түрде қарастырылады. $n = 4, 5, 6, 7$ дәрежелері үшін сәйкес сызықтық жүйелер құрылып, Wolfram Mathematica және Albert бағдарламаларында есептеу жолымен шешіледі. Есептеулер барлық $n \leq 7$ үшін мультилиниялы Ли ішкікеңістігі мен симметриялық элементтер ішкікеңістігінің қиылысуы тривиалды екенін көрсетеді. Демек, 7-дәрежеге дейін еркін Новиков алгебрасында бір мезгілде әрі Ли элементі, әрі симметриялық элемент болатын нөлден өзге мультилиниялы элемент жоқ. Бұл нәтижелер мәселені жоғары дәрежелерде зерттеуге бастапқы негіз болады.

Түйін сөздер: Ли көпмүшелері, еркін арнайы Төрткен алгебралары, Ли элементтері, еркін Новиков алгебралары, мультисызықты Ли элементтері, Эйлер операторы, симметриялық элементтер.

¹*Ерсалиева А.

магистрант, ORCID ID: 0009-0007-0952-706X,

*e-mail: airisays@gmail.com

¹SDU University, г. Каскелен, Казахстан

ПОЛИНОМЫ ЛИ В СВОБОДНЫХ СПЕЦИАЛЬНЫХ ТОРТКЕН-АЛГЕБРАХ

Аннотация

В данной работе изучаются элементы Ли и симметрические элементы (Tortken) в свободной алгебре Новикова, а также исследуется вопрос о том, может ли ненулевой мультилинейный элемент одновременно принадлежать обоим классам. Для проверки принадлежности симметрическому подпространству элементов, представленных в стандартном мультилинейном базисе Ли, используются оператор Эйлера и критерий нулевого лагранжиана. Для части Ли применяются левонормированные коммутаторы с фиксированной первой переменной, образующие удобный базис мультилинейной компоненты. Случай $n = 3$ рассматривается явно путем разложения коммутаторов через произведение Новикова и применения оператора Эйлера. Для степеней $n = 4, 5, 6, 7$ соответствующие линейные системы строятся и решаются вычислительно в Wolfram Mathematica и Albert. Вычисления показывают, что пересечение мультилинейного подпространства Ли с подпространством симметрических элементов тривиально для всех $n \leq 7$. Таким образом, до степени 7 в свободной алгебре Новикова не существует ненулевого мультилинейного элемента, который был бы одновременно элементом Ли и симметрическим элементом. Эти результаты служат отправной точкой для дальнейшего изучения данной задачи в более высоких степенях.

Ключевые слова: Лиевы полиномы, свободные специальные алгебры Торткена, Лиевы элементы, свободные алгебры Новикова, мультилинейные Лиевы элементы, оператор Эйлера, симметрические элементы.

УДК 519.62

МРНТИ 27.35.21; 27.35.47; 27.41.19

<https://doi.org/10.55452/1998-6688-2026-23-2-83-107>

^{1*}**Исахов А.А.,**

профессор, ORCID ID: 0000-0002-1937-8615,

*e-mail: alibek.issakhov@gmail.com

²**Нигметова А.,**

докторант, ORCID ID: 0009-0004-5395-4569,

e-mail: a.nygmetova@kbtu.kz

¹**Абылкасымова А.Б.,**

ассоциированный профессор, ORCID ID: 0000-0002-5967-6959,

e-mail: abylkassymova.aizhan@gmail.com

¹**Аханова Н.Е.**

профессор, ORCID ID: 0000-0003-2767-8711,

e-mail: n.akhanova@kbtu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

²Международный университет информационных технологий, г. Алматы, Казахстан

ПРОГНОЗИРОВАНИЕ ЗАГРЯЗНЕНИЯ ВОЗДУШНОЙ СРЕДЫ ОТ ВЫБРОСОВ В ГОРОДСКИХ КАНЬОНАХ С ИСПОЛЬЗОВАНИЕМ ML-CFD МОДЕЛИРОВАНИЯ

Аннотация

В работе представлен гибридный подход для прогнозирования распространения загрязняющих веществ в городских уличных каньонах с учетом шумозащитных барьеров. Методология объединяет детальное CFD-моделирование и суррогатную модель на основе нейросетевой архитектуры BiLSTM с механизмом внимания. Исследованы конфигурации с барьерами высотой 0.1H, 0.2H и 0.3H. CFD-расчеты выявили нелинейное влияние высоты барьера на аэродинамику и формирование зон накопления примеси, причем наиболее сложный нестационарный режим наблюдается при высоте 0.2H. Суррогатная модель успешно прогнозирует эволюцию концентрации для случаев без барьера и с барьером 0.1H, демонстрируя среднюю абсолютную процентную ошибку менее 15%. Для барьера 0.2H точность снижается в зонах интенсивной турбулентности, что связано с высокой нестационарностью процесса. Подход позволяет значительно сократить вычислительные затраты при сохранении физической достоверности, что перспективно для систем поддержки решений в области городской экологии. Модель обеспечивает ускоренное прогнозирования по сравнению с CFD-расчетом на 7–8 порядков, так потраченное время инференса показало 1–5 мс, тогда как одна CFD-симуляция занимает около 54 часов на CPU.

Ключевые слова: загрязнение воздуха, уличные каньоны, шумозащитные барьеры, вычислительная гидродинамика (CFD), машинное обучение, BiLSTM, суррогатное моделирование, прогнозирование концентраций

Введение

Загрязнение атмосферного воздуха в городской среде остается одной из наиболее значимых угроз общественному здоровью и устойчивому развитию. Многочисленные эпидемиологические исследования указывают на устойчивую связь между повышенными концентрациями загрязняющих веществ в приземном слое атмосферы и ростом заболеваемости респираторными, сердечно-сосудистыми и онкологическими заболеваниями [1–4]. Наибольшую опасность представляют мелкодисперсные аэрозольные частицы PM_{2.5} и PM₁₀, способные глубоко проникать в дыхательные пути и накапливаться в тканях организма.

В условиях современных городов доминирующим источником загрязнения воздуха остается автотранспорт. Рост урбанизации, увеличение плотности застройки и транспортной нагрузки приводят к формированию локальных зон повышенной концентрации загрязняющих веществ, особенно в часы пик [5–8]. Расширение городских территорий и зависимость от личного транспорта усугубляют эту проблему, формируя устойчивые очаги загрязнения вблизи магистралей и жилых районов.

Ключевую роль в формировании пространственного распределения загрязняющих веществ играет архитектурная структура городской среды. Плотная и высотная застройка формирует так называемые уличные каньоны, в которых ограниченная вентиляция и сложная структура турбулентности способствуют накоплению загрязнений [9–11]. В таких условиях даже при умеренных скоростях ветра возникают застойные зоны и устойчивые вихревые структуры, приводящие к формированию длительных эпизодов смога, рассеивание которых возможно лишь при благоприятных метеорологических условиях [12].

Для анализа процессов переноса и дисперсии загрязняющих веществ в сложной городской геометрии широко применяются методы вычислительной гидродинамики. CFD-моделирование позволяет детально учитывать форму застройки, турбулентные эффекты и нестационарные граничные условия, обеспечивая высокую физическую достоверность результатов по сравнению с параметрическими моделями [13–16]. Численные исследования уличных каньонов показали, что CFD является эффективным инструментом для оценки распределения концентраций и анализа влияния архитектурных и инженерных элементов городской инфраструктуры.

Однако практическое применение CFD в задачах оперативного прогнозирования и сценарного анализа существенно ограничено высокой вычислительной стоимостью. Детализированное моделирование турбулентных потоков в городской среде требует значительных вычислительных ресурсов, что затрудняет проведение многовариантных расчетов и использование таких моделей в системах поддержки принятия решений [15,16]. Эти ограничения стимулировали развитие альтернативных и комплементарных подходов, направленных на ускорение расчетов при сохранении приемлемой точности.

В последние годы все более активно развиваются методы машинного обучения, применяемые для прогнозирования качества воздуха и анализа результатов численного моделирования. ML-подходы используются как для обработки данных наблюдений, так и для построения суррогатных моделей, аппроксимирующих результаты CFD-расчетов [17, 18]. Такие модели позволяют существенно сократить вычислительные затраты и обеспечить быстрый отклик при анализе различных сценариев. Особый интерес представляют гибридные CFD–ML подходы, в которых машинное обучение используется в качестве дополнения к физическим моделям, а не их замены, что повышает устойчивость и интерпретируемость прогнозов.

Отдельного внимания заслуживает влияние инженерных решений и элементов городской инфраструктуры на процессы дисперсии загрязняющих веществ. Шумозащитные барьеры, элементы застройки и планировочные решения способны существенно изменять структуру потока, как снижая концентрации в отдельных зонах, так и формируя локальные области повышенного загрязнения [19–22]. Анализ таких эффектов требует сочетания физически обоснованных численных моделей и инструментов быстрого анализа, что делает гибридные подходы особенно актуальными.

Таким образом, разработка и применение методов, объединяющих CFD-моделирование и машинное обучение, представляется перспективным направлением для исследования загрязнения воздуха в уличных каньонах. Такой подход позволяет совместить физическую достоверность численных моделей с вычислительной эффективностью ML-алгоритмов и создать инструменты, пригодные для анализа сценариев, оптимизации городской инфраструктуры и поддержки экологически ориентированных решений.

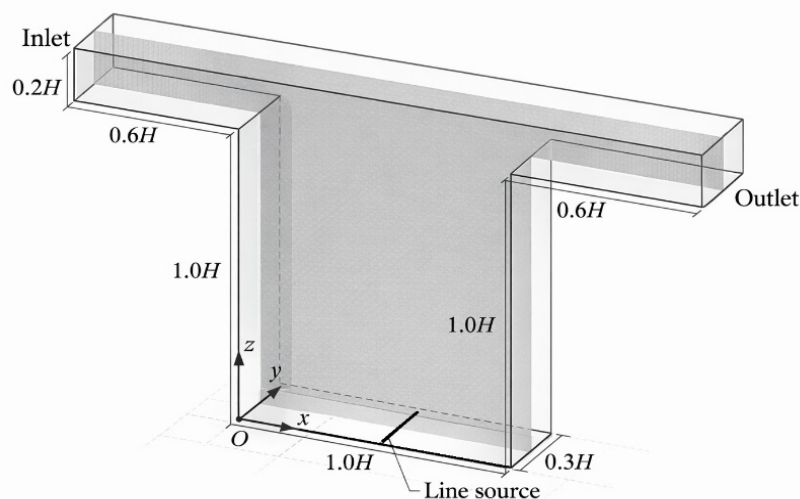
В рамках настоящего исследования рассматривается задача прогнозирования пространственного распределения загрязняющих веществ в идеализированных городских уличных каньонах с учетом влияния инженерных элементов городской инфраструктуры. Основное внимание уделяется разработке и валидации гибридного подхода, сочетающего численное моделирование на основе методов вычислительной гидродинамики и методы машинного обучения, ориентированные на построение вычислительно эффективных суррогатных моделей. CFD-расчеты используются для получения физически обоснованных данных о структуре потока и распределении концентраций при различных геометрических и аэродинамических конфигурациях, тогда как алгоритмы машинного обучения применяются для выявления скрытых закономерностей и аппроксимации результатов численного моделирования.

Материалы и методы

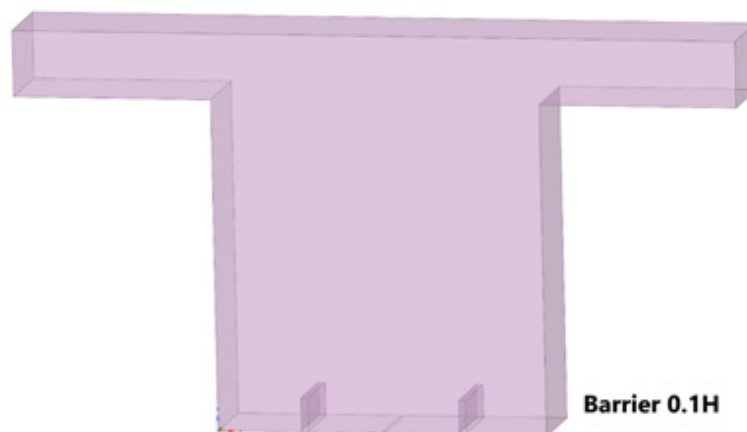
Постановка задачи и геометрия вычислительной области

Численное моделирование процессов переноса и дисперсии загрязняющих веществ в условиях городской застройки проводилось на основе методов вычислительной гидродинамики (CFD). В качестве объекта исследования рассматривался идеализированный уличный каньон, образованный двумя параллельными зданиями одинаковой высоты H (рис. 1а). Длина каньона значительно превышала его ширину и высоту, что позволяло пренебречь влиянием торцевых границ на установившийся режим течения в центральной части.

Границы вычислительной области, включая поверхности зданий, дорожное полотно и верхнюю границу, задавались в соответствии с условиями, имитирующими нейтральную стратификацию атмосферного пограничного слоя. Для анализа влияния инфраструктурных элементов вдоль линии, соответствующей расположению источника выбросов, были смоделированы шумозащитные барьеры различной высоты (рис. 1б). Высота барьера варьировалась в долях от высоты зданий и принимала значения $0.1H$, $0.2H$ и $0.3H$, что позволило исследовать зависимость структуры потока и распределения примеси от этого параметра. Толщина барьера принята $0.01H$.



a



b

Рисунок 1 – Геометрия вычислительной области:
а) без барьера, б) барьер с высотой 0.1H

Уравнения, описывающие процесс, и физические предположения

Моделирование источника загрязняющих веществ было реализовано в виде линейного источника, расположенного на уровне дорожного полотна и ориентированного вдоль оси каньона (рисунок 1). Такая постановка соответствует выбросам от непрерывного потока автотранспорта. В качестве модельного загрязнителя использовался этилен (C_2H_4). Вещество рассматривалось как пассивная примесь, не влияющая на плотность и динамику воздушного потока и не участвующая в химических реакциях.

Движение воздушной среды в уличном каньоне описывается осредненными по Рейнольдсу уравнениями Навье–Стокса для несжимаемого турбулентного течения. Такой подход является общепринятым при моделировании атмосферных потоков в городской среде и позволяет адекватно воспроизводить основные характеристики средних полей скорости и давления при умеренных вычислительных затратах. Плотность воздуха считалась постоянной, а влияние температурной стратификации и плавучести не учитывалось, что соответствует приближению нейтральной атмосферной устойчивости.

Система уравнений включает уравнение неразрывности, уравнения сохранения импульса и уравнение конвективно-диффузионного переноса пассивной примеси:

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0$$

(1)

$$\frac{\partial \rho \bar{u}_i}{\partial t} + \frac{\partial \rho \bar{u}_i \bar{u}_j}{\partial x_j} = - \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j} \left[(\mu + \mu_t) \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) \right] - \frac{2}{3} \delta_{ij} \rho k$$

(2)

$$\frac{\partial \rho \bar{C}}{\partial t} + \frac{\partial \rho \bar{C} \bar{u}_i}{\partial x_i} = \frac{\partial}{\partial x_i} \left[\rho \left(D + \frac{\nu_t}{Sc_t} \right) \frac{\partial \bar{C}}{\partial x_i} \right]$$

(3)

Здесь $\bar{u}_i$ – компоненты осредненной скорости, $\bar{p}$ – осредненное давление, $\bar{C}$ – осредненная концентрация примеси, μ – динамическая вязкость (для воздуха – 1.79×10^{-5} кг/(м·с), для этилена – 1.03×10^{-5} кг/(м·с)), μ_t – турбулентная вязкость, k – турбулентная кинетическая энергия, D – молекулярный коэффициент диффузии (для этилена – 2.88×10^{-5} м²/с), ν_t – турбулентная кинематическая вязкость ($\nu_t = \mu_t / \rho$), δ_{ij} – символ Кронекера, а Sc_t – турбулентное число Шмидта, принятое равным 0.7 [14].

Для замыкания системы использовалась модель турбулентности SST $k-\omega$. Эта двухпараметрическая модель сочетает точность стандартной $k-\omega$ модели в пристенной области с устойчивостью $k-\epsilon$ модели в удаленных от стен зонах, что делает ее эффективной для задач обтекания зданий, где существенны отрывные течения, сдвиговые слои и взаимодействие вихревых структур. Модель включает два транспортных уравнения – для турбулентной кинетической энергии k и удельной скорости диссипации ω , которые решались совместно с уравнениями (1)–(3).

Граничные условия были заданы следующим образом. На входной границе моделировался набегающий поток воздуха с постоянной скоростью $V=1\text{ м/с}$, направленной перпендикулярно оси каньона и линии источника. Такой сценарий соответствует наихудшим условиям вентиляции и способствует накоплению примеси, что важно для оценки защитной эффективности барьеров. Концентрация загрязняющего вещества на входе принималась равной нулю. Источник этилена задавался через эквивалентную скорость подачи, равную 0.01923 м/с . На выходной границе использовались условия свободного истечения для давления и нулевых градиентов для скорости и концентрации. Все твердые поверхности (здания, дорожное полотно, барьеры) моделировались как стенки с условием прилипания для скорости и нулевым нормальным диффузионным потоком для концентрации. Плотность воздуха принималась постоянной ($\rho = 1.225\text{ кг/м}^3$).

Численная дискретизация и расчетная сетка

Решение системы управляющих уравнений выполнялось методом конечных объемов. Конвективные члены в уравнениях переноса импульса и примеси аппроксимировались схемой второго порядка *upwind*, что позволило минимизировать численную диффузию и корректно описать перенос в зонах с высокой турбулентностью. Для давления применялась стандартная схема второго порядка. Связь полей скорости и давления осуществлялась с помощью алгоритма SIMPLE, обеспечивающего устойчивую итерационную процедуру для несжимаемых течений. Для проведения гидродинамических расчетов был использован ANSYS Fluent.

Дискретизация расчетной области выполнялась с использованием неструктурированной вычислительной сетки, сформированной из элементов додекаэдрической формы (рисунок 2). Повышенная плотность сетки задавалась в зонах, определяющих аэродинамику и перенос примеси: вблизи дорожного покрытия, линейного источника выбросов, поверхностей зданий и шумозащитного барьера. Общее количество ячеек в наиболее детализированной конфигурации составило порядка 1.2×10^6 , что обеспечило баланс между точностью решения и вычислительными затратами.

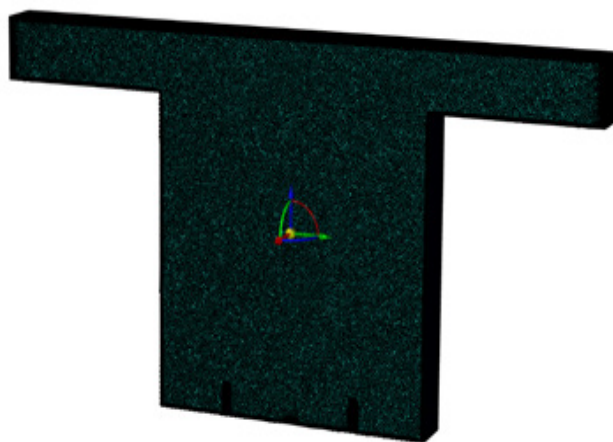


Рисунок 2 – Вычислительная сетка

Независимость численного решения от размера вычислительной сетки оценивалась путем сравнения результатов расчетов, выполненных на нескольких уровнях дискретизации (таблица 1) и экспериментальными данными Kikumoto & Ooka (2017) [14]. Анализ проводился на основе распределений продольной скорости и концентрации загрязняющего вещества в характерных сечениях уличного каньона. Различия между результатами для наиболее детализированных сеток находились в пределах численной погрешности (рисунок 3), что позволяет считать выбранную сетку достаточной для дальнейших расчетов.

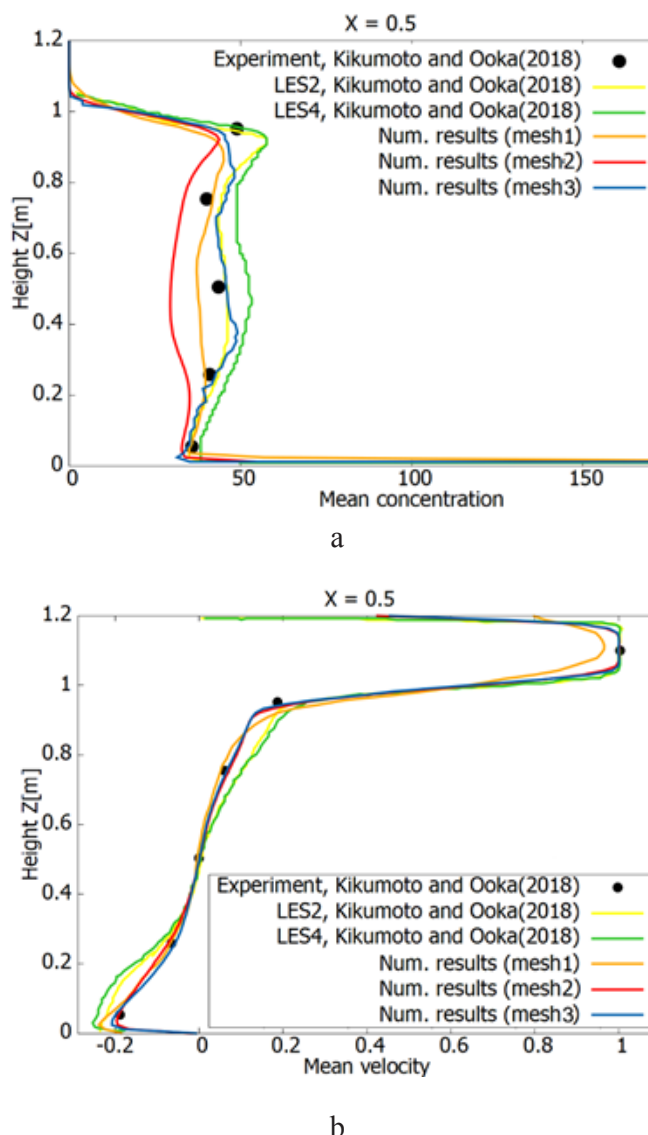


Рисунок 3 – Сравнение численных результатов с экспериментальными данными [14] для:
а) усредненная скорость; б) усредненная концентрация

Таблица 1 – Параметры вычислительной сетки

Сетка №	Минимальный размер элемента [м]	Максимальный размер элемента [м]	Количество элементов
1	0.0125	0.0625	8763
2	0.003125	0.015625	227030
3	0.001	0.01	1195547

Суррогатная модель на основе BiLSTM: данные, архитектура и обучение

Для построения суррогатной модели, предназначенной для прогнозирования временной эволюции концентрации загрязняющих веществ на основе данных CFD-моделирования, использовалась нейросетевая архитектура двунаправленной долгой краткосрочной памяти (BiLSTM), дополненная механизмом внимания. Выбор данной архитектуры обусловлен необходимостью учитывать временные зависимости различной протяженности, характерные для нестационарного переноса примеси в уличных каньонах, а также способностью механизма attention адаптивно выделять наиболее значимые участки временных последовательностей при формировании прогноза.

Формирование обучающего набора данных

Обучающая выборка формировалась на основе результатов нестационарных CFD-расчетов, выполненных для четырех геометрических конфигураций уличного каньона: без шумозащитного барьера и с барьерами относительной высоты $0.1H$, $0.2H$ и $0.3H$. Для каждой конфигурации регистрировались временные ряды концентрации этилена в пяти контрольных точках, расположенных в характерных зонах внутри каньона, включая области рециркуляции и основной поток (рисунок 4).

Продолжительность моделирования составляла 200 с. Шаг по времени принимался равным 0.25 с для всех конфигураций, что обеспечивало достаточное временное разрешение для анализа динамики концентрации при различных режимах обтекания. Сохранение численных результатов выполнялось с интервалом 0.5 с.

На основе исходных временных рядов формировались обучающие примеры в виде последовательностей типа «прошлое–будущее». Входные данные представляли собой фрагменты временного ряда фиксированной длины, соответствующие предыдущим значениям концентрации, тогда как выходными данными являлись значения концентрации в последующие моменты времени. Такая организация данных позволила сформулировать задачу как прогнозирование временной эволюции концентрации на основе ее предыстории.

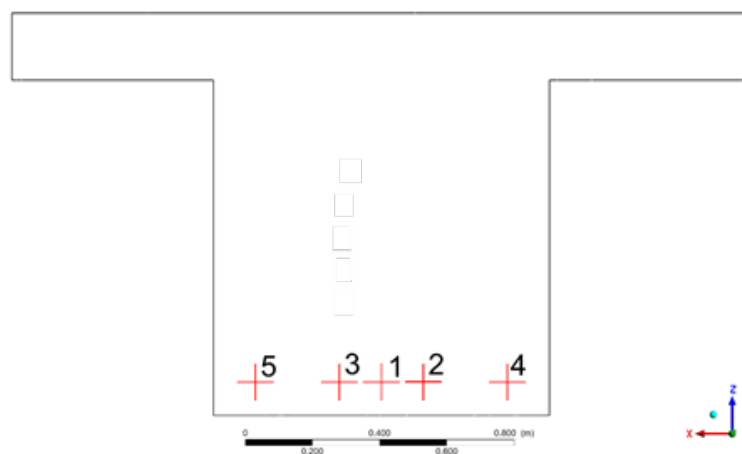


Рисунок 4 – Расположение контрольных вычислительных узлов в области вычисления

Сформированный набор данных (16 000 временных точек) был разделен на обучающую и тестовую выборки в соотношении 90/10 без перемешивания временных последовательностей. Формирование набора данных основывалась в таком виде (800 точек $\times$ 5 (контрольные точки) $\times$ 4 (варианта вычисления) = 16 000). Сохранение временного порядка обеспечивало корректную оценку прогностических свойств модели и исключало утечку информации между выборками.

В рамках данного исследования для каждой конфигурации (без барьера, $0.1H$, $0.2H$ и $0.3H$) обучалась отдельная модель машинного обучения. Такой подход обусловлен существенными

различиями в структуре потока и характере распределения загрязняющих веществ при изменении геометрии области. Обучение отдельных моделей позволяет избежать смещения различных аэродинамических режимов в одном наборе данных и обеспечивает более корректное воспроизведение зависимостей между входными параметрами и концентрацией загрязняющего вещества. Кроме того, раздельное обучение моделей повышает точность предсказаний, поскольку каждая модель адаптируется к специфике конкретной конфигурации потока. Сравнение качества предсказаний выполнялось независимо для каждой модели с использованием метрик MAE, MAPE и коэффициента детерминации (R^2).

Архитектура нейронной сети и обоснование ее проектирования

Суррогатная модель построена на основе двунаправленной рекуррентной архитектуры LSTM, предназначенной для аппроксимации нестационарной временной эволюции концентрации загрязняющего вещества. Выбор архитектуры LSTM обусловлен рядом особенностей рассматриваемой задачи. Нестационарный перенос загрязняющих веществ в уличном каньоне характеризуется сложными временными зависимостями различной протяженности: краткосрочными флуктуациями концентрации, вызванными турбулентными пульсациями, и долгосрочными тенденциями, связанными с эволюцией вихревых структур. Рекуррентные ячейки LSTM, в отличие от стандартных RNN, устойчиво улавливают такие зависимости благодаря механизму вентилей, предотвращающему затухание градиентов при обучении на длинных последовательностях [23].

Использование двунаправленной конфигурации дополнительно обосновано спецификой задачи суррогатного моделирования: в отличие от задач онлайн-прогнозирования, где будущие значения недоступны, суррогатная модель обучается на полных временных последовательностях CFD-расчетов, что делает двунаправленную обработку физически оправданной и позволяет модели формировать более полное внутреннее представление динамики процесса. Дополнительный механизм внимания обеспечивает адаптивное взвешивание временных шагов, что особенно важно в условиях нестационарного режима течения с резкими изменениями концентрации. Размерность скрытого состояния составляет 256 нейронов в каждом направлении, что обеспечивает достаточную выразительную способность модели без избыточного увеличения числа параметров.

Последовательность скрытых состояний, формируемая двунаправленным LSTM-слоем (рисунок 5), далее обрабатывается механизмом внимания, предназначенным для адаптивного перераспределения весов между временными шагами. Механизм внимания реализован как параметрическое отображение скрытых состояний во временные коэффициенты значимости с последующей нормализацией. Полученные коэффициенты используются для вычисления контекстного вектора в виде взвешенной суммы скрытых состояний, который аккумулирует наиболее информативные элементы временной последовательности. Контекстный вектор поступает на выходной полносвязный слой, формирующий прогноз концентрации загрязняющего вещества на заданном временном горизонте.

Данная архитектура обеспечивает баланс между аппроксимационной способностью и интерпретируемостью модели. Использование внимания позволяет не только повысить точность прогноза, но и проанализировать, какие участки временной эволюции оказывают наибольшее влияние на формирование результата, что представляет интерес с точки зрения физической интерпретации нестационарных процессов переноса примеси.

В качестве основной архитектуры использовалась двунаправленная рекуррентная нейронная сеть BiLSTM, дополненная механизмом внимания. В данной работе применен механизм additive attention (Bahdanau attention), который позволяет учитывать вклад скрытых состояний на каждом временном шаге при формировании итогового прогноза. Такой подход обеспечивает более гибкое взвешивание информации по временной последовательности по сравнению с простыми схемами усреднения.

Входные данные формировались в виде временных окон фиксированной длины. Размер входного окна (look-back) составлял ($L = 20$) временных шагов, что обеспечивало достаточный контекст для учета динамики изменения концентрации. Горизонт прогноза (forecast horizon) был выбран равным ($H = 5$) временным шагам, что соответствует краткосрочному прогнозированию концентрации загрязняющего вещества.

В качестве входного вектора использовались многомерные временные ряды, включающие значения концентрации во всех 9 контрольных точках одновременно. Таким образом, размерность входного вектора на каждом временном шаге составляла 9. Это позволило модели учитывать пространственные корреляции между различными точками наблюдения, что особенно важно в условиях сложной аэродинамики городского каньона.

Подбор гиперпараметров модели выполнялся с использованием метода grid search с последующей локальной донастройкой. В процессе оптимизации варьировались следующие параметры: количество слоев BiLSTM (1–3), размер скрытого слоя (32–128 нейронов), размер батча (16–64), скорость обучения ($10^{-4} - 10^{-2}$), а также параметры механизма внимания. Критерием выбора оптимальной конфигурации служило минимальное значение функции потерь на валидационной выборке (MSE). Дополнительно применялась ранняя остановка (early stopping) для предотвращения переобучения. Оптимальная конфигурация для расчетов была выбрана как 2 слоя BiLSTM, размер скрытого слоя 128 нейронов, размер батча 32, скорость обучения 10^{-3} . Финальная модель обучалась с указанными параметрами; ранняя остановка применялась при отсутствии улучшения MSE на валидации в течение 30 эпох.

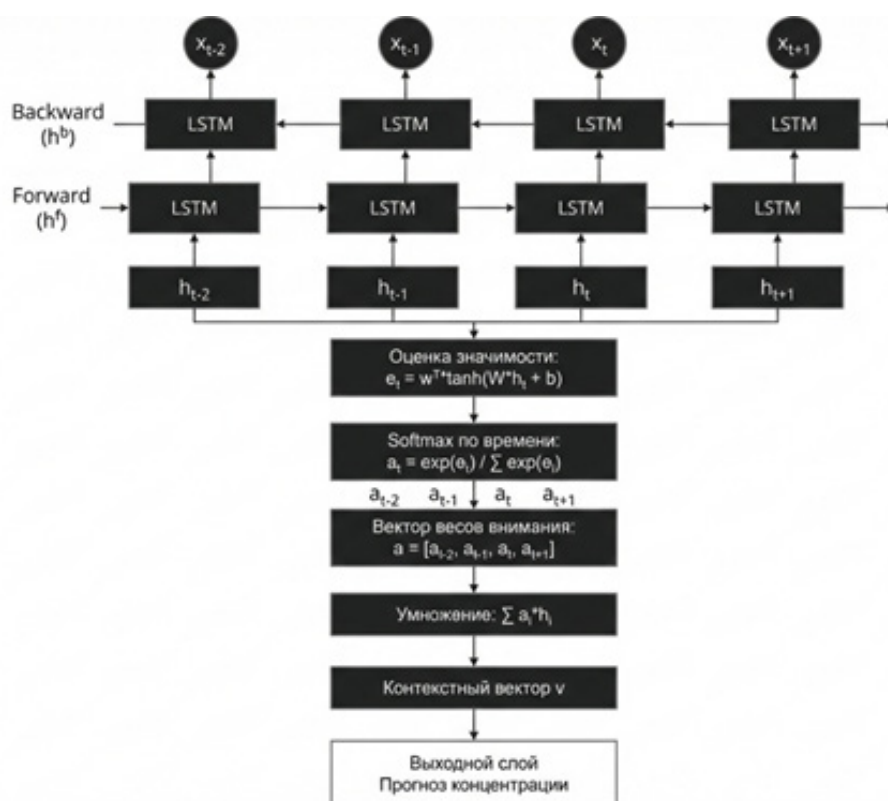


Рисунок 5 – Bi-LSTM модель с механизмом внимания

Обучение и регуляризация

Обучение нейросетевой модели выполнялось с использованием оптимизатора RAdam. Его применение позволило избежать неустойчивого поведения градиентов на начальной стадии обучения и обеспечить более предсказуемую сходимость без дополнительной подстройки

гиперпараметров. Для сглаживания траектории оптимизации и повышения устойчивости получаемого решения использовался алгоритм Lookahead, в рамках которого обновления параметров осуществлялись по схеме чередования быстрых и медленных весов.

В качестве функции потерь было выбрано среднеквадратичное отклонение с временными весами, смещающими акцент ошибки в сторону начальных шагов прогнозной последовательности. Такое взвешивание позволило улучшить качество краткосрочного прогноза, не влияя на структуру модели и не вводя дополнительных ограничений в процесс обучения.

Для снижения риска переобучения после рекуррентного слоя применялся dropout с коэффициентом 0.2. Обучение контролировалось с использованием механизма ранней остановки: процесс оптимизации прекращался при отсутствии улучшения значения функции потерь на валидационной выборке в течение 30 последовательных эпох. Этот критерий оказался достаточным для стабилизации параметров модели без преждевременного прекращения обучения. Все расчеты выполнялись с использованием фреймворка PyTorch. Обучение модели осуществлялось с аппаратным ускорением на графическом процессоре с поддержкой CUDA.

Результаты и обсуждение

Влияние высоты барьера на аэродинамику и распределение концентрации

Нестационарные CFD-расчеты позволили проследить временную эволюцию концентрации загрязняющего вещества в уличном каньоне при различных конфигурациях шумозащитного барьера. Во всех рассмотренных случаях формирование концентрационного поля носило выражено нестационарный характер и определялось взаимодействием основного поперечного потока с вихревыми структурами внутри каньона и в приповерхностной зоне.

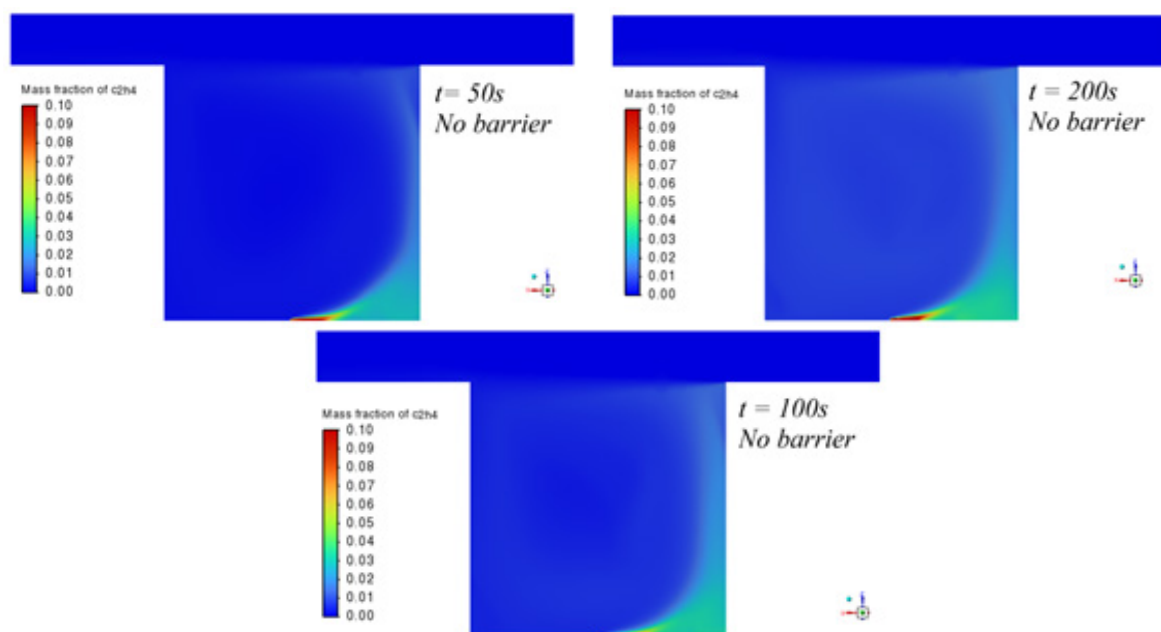
Введение барьера вдоль линии источника кардинально изменяет структуру течения внутри каньона. При отсутствии барьера (рисунок 6а) основной вихрь занимает почти весь объем каньона, обеспечивая относительно однородное перемешивание. Линейный источник создает шлейф примеси, который захватывается этим вихрем и распределяется по всей высоте подветренной зоны.

Барьер высотой 0.1H (рисунок 6б) не разрушает основной вихрь, но создает локальную зону застоя непосредственно перед препятствием. Поток разделяется, огибая барьер сверху, что приводит к формированию компактного вторичного вихря между барьером и подветренной стеной. Концентрация примеси резко возрастает в этой замкнутой зоне, создавая локальный максимум вблизи точки 5 (согласно нумерации на рис. 4). Напротив, за барьером (точка 3) концентрация остается низкой, что подтверждает его экранирующую функцию.

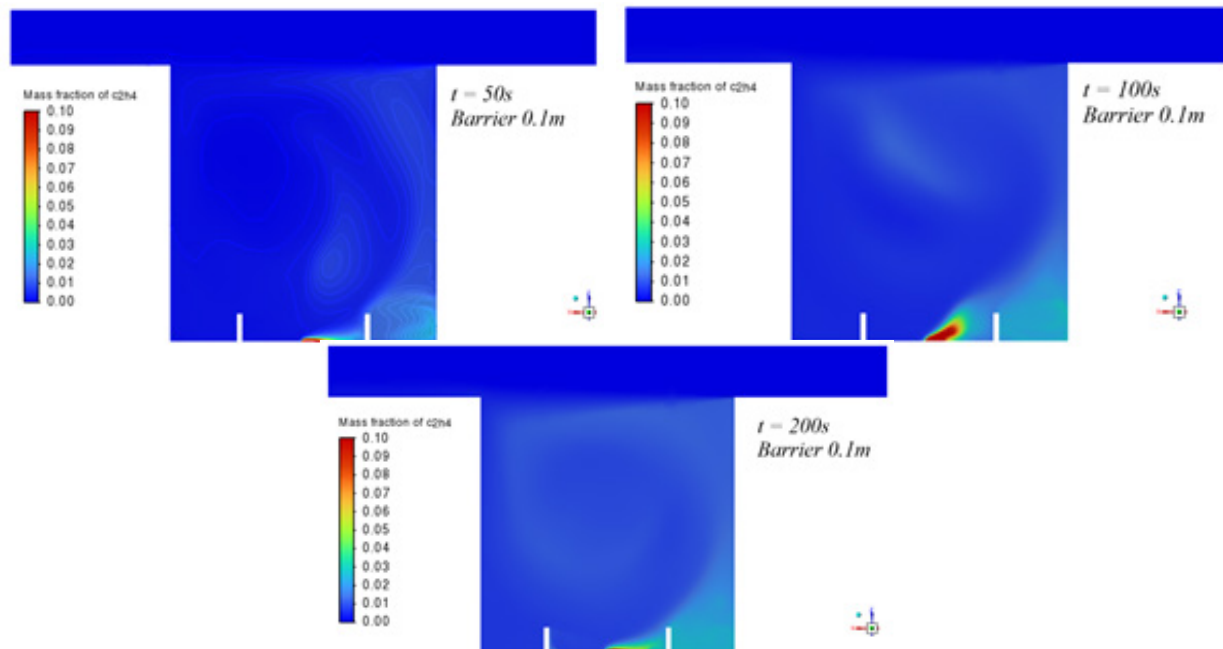
Увеличение высоты барьера до 0.2H (рис. 6в) усиливает этот эффект. Основной вихрь деформируется, его центр смещается вверх. Пространство между барьером и подветренной стеной становится областью сложной турбулентной структуры с несколькими мелкими вихрями. Динамика переноса примеси здесь становится существенно нестационарной, а распределение концентрации – менее предсказуемым. Пиковые значения концентрации в зоне между барьером и стеной возрастают.

При высоте барьера 0.3H (рисунок 6г) аэродинамическая картина снова упорядочивается. Барьер эффективно разделяет поток на два канала. Большая часть потока перенаправляется над барьером, создавая область пониженного давления за ним и устойчивую рециркуляцию. В этой более стабильной, хотя и интенсивной, вихревой зоне концентрация примеси остается высокой, но ее временная изменчивость снижается по сравнению со случаем 0.2H.

Полученные CFD-результаты сформировали основу для построения и валидации суррогатной модели машинного обучения, позволяя оценивать не только средние концентрации, но и временную динамику загрязняющего вещества в характерных зонах уличного каньона.



a



б

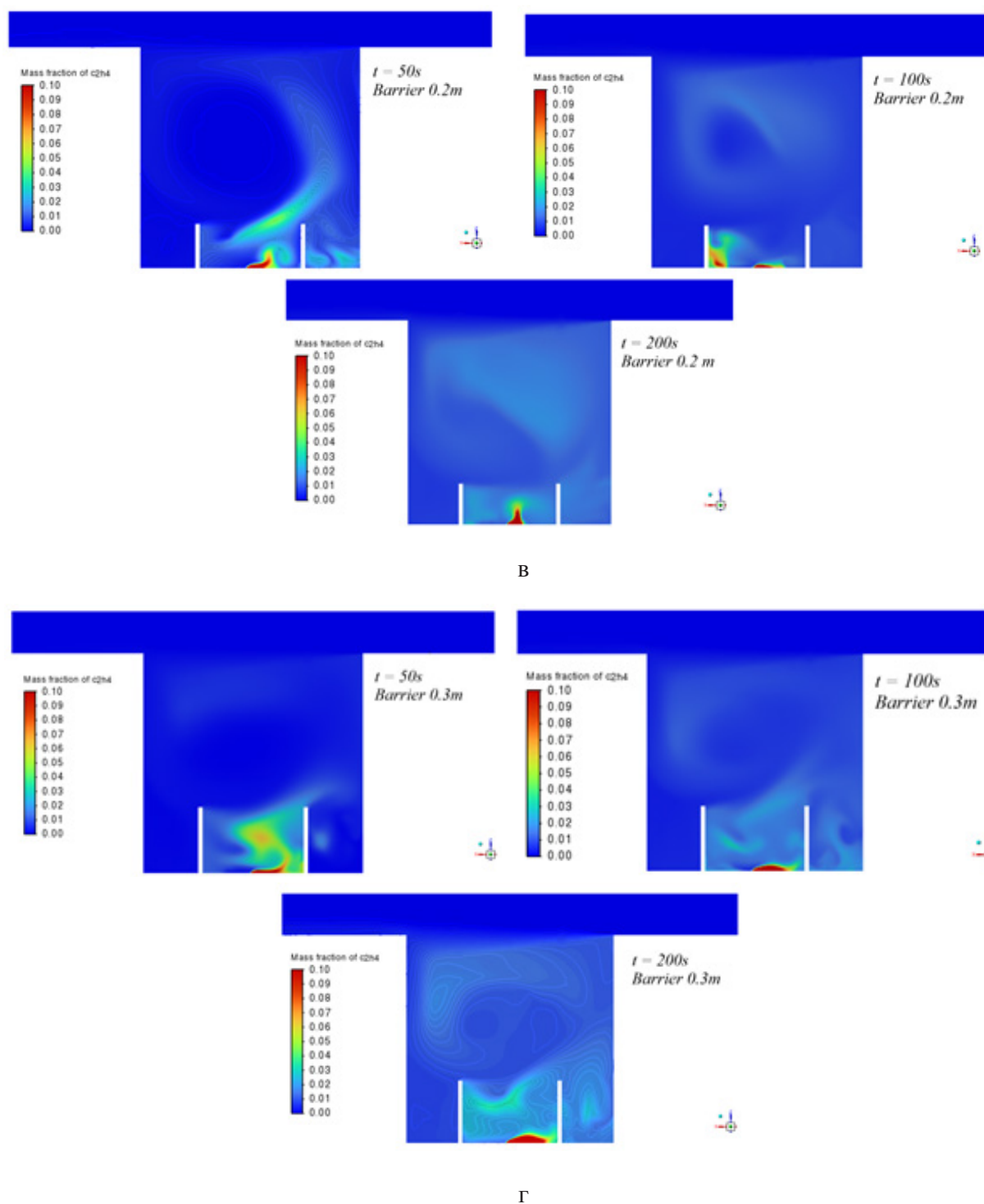


Рисунок 6 – Распределение поля концентрации вещества в различные моменты времени:
а) без барьера, б) барьер 0.1Н, в) барьер 0.2Н, г) барьер 0.3Н

Прогнозирование концентраций с помощью суррогатной модели BiLSTM

Качество работы суррогатной модели оценивалось путем прямого сопоставления прогнозируемых и расчетных по CFD временных рядов концентрации в контрольных точках для всех рассмотренных конфигураций. На рисунке 7 представлены типичные временные зависимости

концентрации, полученные из CFD-расчетов и предсказанные моделью машинного обучения, для различных положений внутри уличного каньона.

Во всех конфигурациях модель корректно воспроизводит общий уровень концентрации и характер временной изменчивости сигнала. Основные особенности нестационарной динамики, включая фазы нарастания концентрации и последующие квазипериодические колебания, совпадают с CFD-результатами по фазе и амплитуде.

Вблизи дорожного полотна и в областях, прилегающих к шумозащитному барьеру, расхождения между CFD и ML-прогнозами увеличиваются. В этих зонах модель демонстрирует сглаживание резких пиков концентрации, что особенно заметно на начальных этапах моделирования и при переходных режимах. Несмотря на это, средний уровень концентрации и общая форма временного сигнала сохраняются с приемлемой точностью.

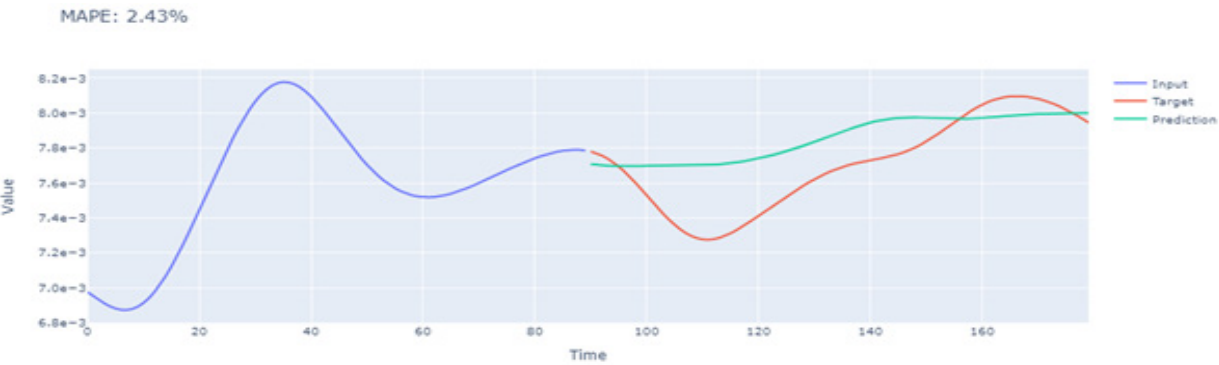
Для случая отсутствия барьера модель демонстрирует исключительно высокую точность. Временные ряды прогнозируемой концентрации практически идентичны исходным данным, визуальные расхождения минимальны. Это подтверждается низкими значениями метрик ошибок: средняя абсолютная процентная ошибка (MAPE) не превышает 6.5%, а коэффициент детерминации R^2 остается положительным для всех точек, достигая 0.6 в точках с более выраженной динамикой (таблица 2). Модель успешно улавливает как общий тренд, связанный с выходом на стационарный режим, так и мелкомасштабные флуктуации, вызванные турбулентностью.

Аналогичная ситуация наблюдается при введении низкого барьера (0.1H). Несмотря на усложнение структуры потока, модель обеспечивает стабильный и точный прогноз. Наибольшие ошибки, закономерно, фиксируются в точках 1 и 2, расположенных в непосредственной близости от барьера в зоне интенсивного вихреобразования (MAPE до 14.7%). Однако даже здесь модель корректно предсказывает момент резкого роста концентрации и ее последующее осциллирующее поведение.

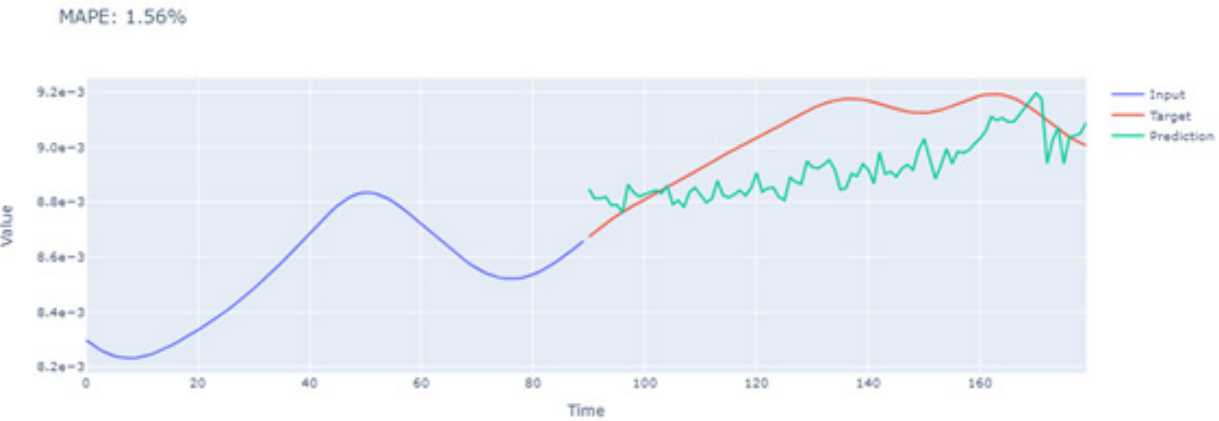
Качественный скачок в ошибке прогноза происходит при увеличении высоты барьера до 0.2H. На графиках для точек 1 и 2 (рисунок 7 (к) и рисунок 7 (л)) становятся заметны выраженные расхождения: модель запаздывает с реакцией на резкие изменения концентрации или генерирует сглаженный прогноз, не улавливая пиковых значений. Это напрямую связано с описанной выше сложной и нестационарной аэродинамикой данного случая. Процессы интенсивного перемешивания и быстрого изменения полей концентрации формируют временные ряды с высокочастотными составляющими, которые трудно предсказать на основе ограниченной истории, особенно когда модель обучена на данных, где такие экстремальные режимы представлены недостаточно полно.

При дальнейшем увеличении высоты барьера до 0.3H точность прогноза частично восстанавливается. Хотя абсолютный уровень концентраций в зоне за барьером высок, их временная изменчивость снижается благодаря стабилизации вихревой структуры. Это позволяет модели, обученной на осредненных закономерностях, давать более устойчивые прогнозы. Значения MAPE для точек 1 и 2 снижаются по сравнению со случаем 0.2H, а R^2 возвращается в положительную область.

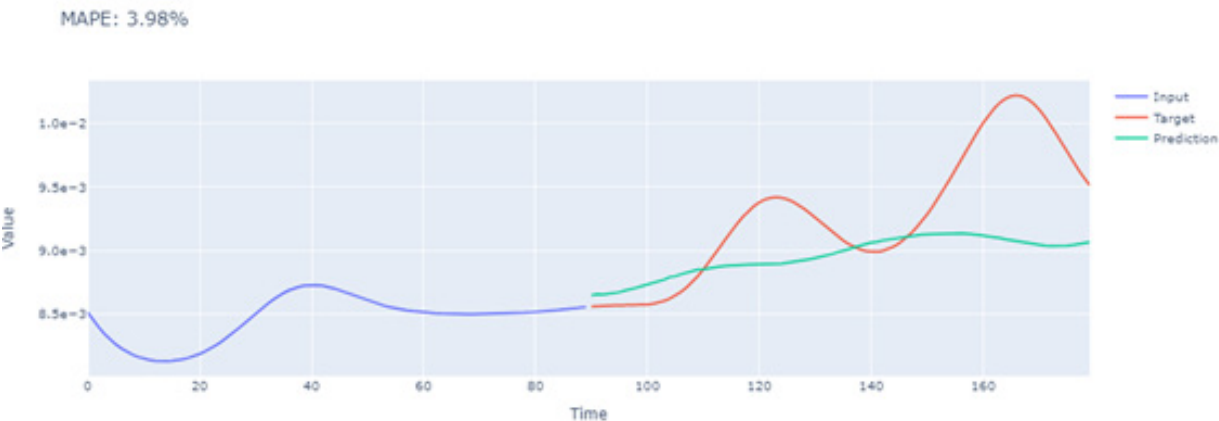
Модель обучалась на всех 5 точках одновременно – и она нашла компромиссное решение, которое хорошо работает для большинства точек (MAPE < 15%), но систематически ошибается именно в точке 2. Данное отклонение можно объяснить классической проблемой multi-output регрессии – оптимизация усредненной loss-функции, что не гарантирует приемлемого качества для каждого выхода по отдельности. Точка 2 при 0.2H – физически наиболее аномальная зона, и она «переголосована» остальными 4 точками при обучении.



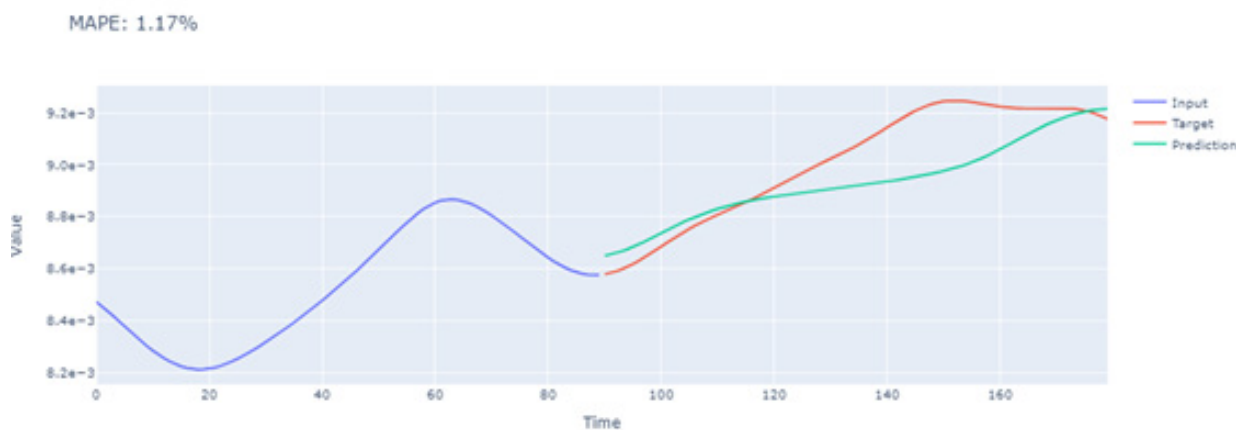
a) 0H barrier, point-1



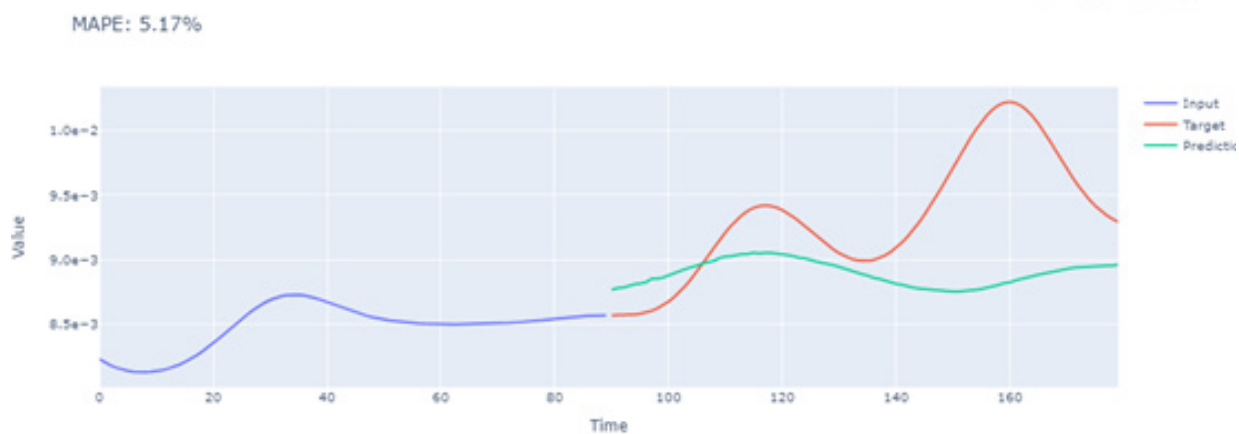
b) 0H barrier, point-2



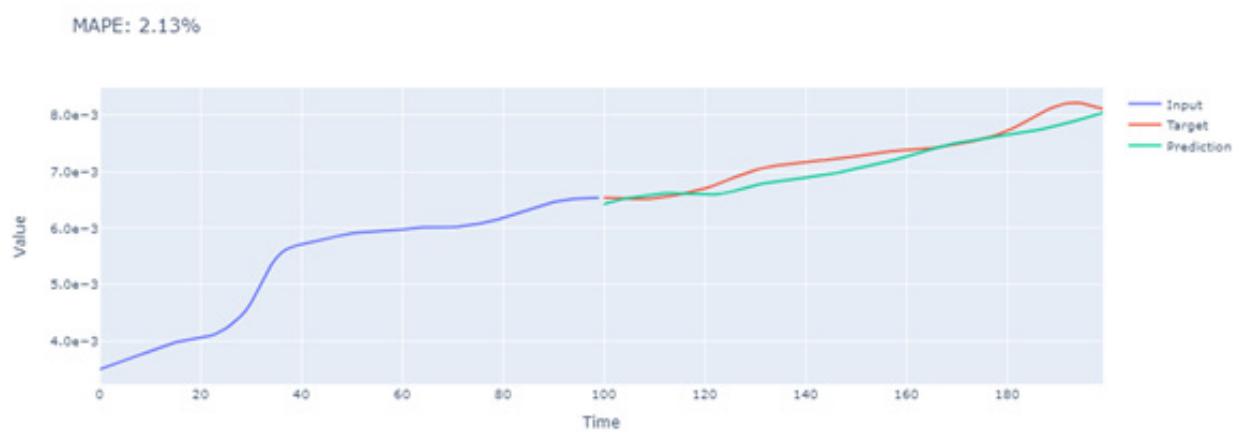
c) 0H barrier, point-3



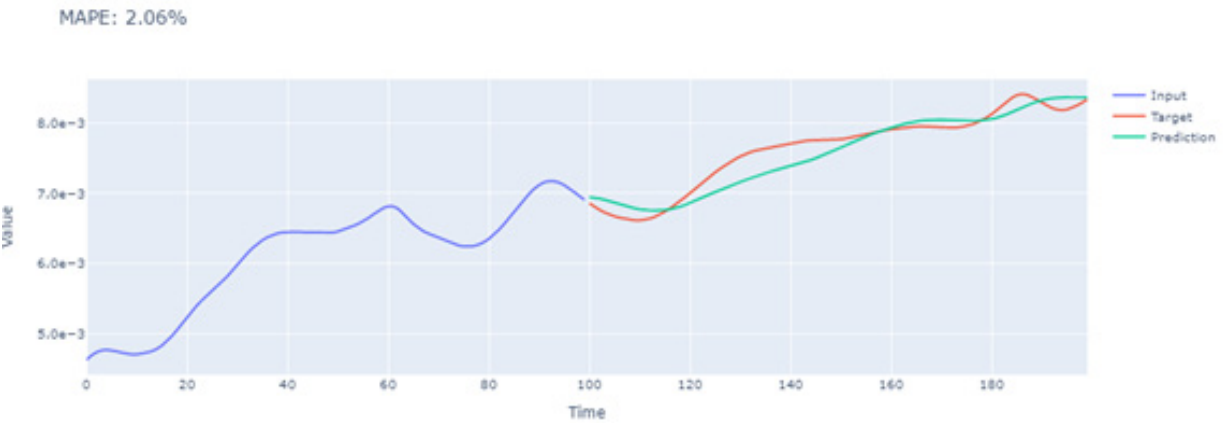
d) 0H barrier, point-4



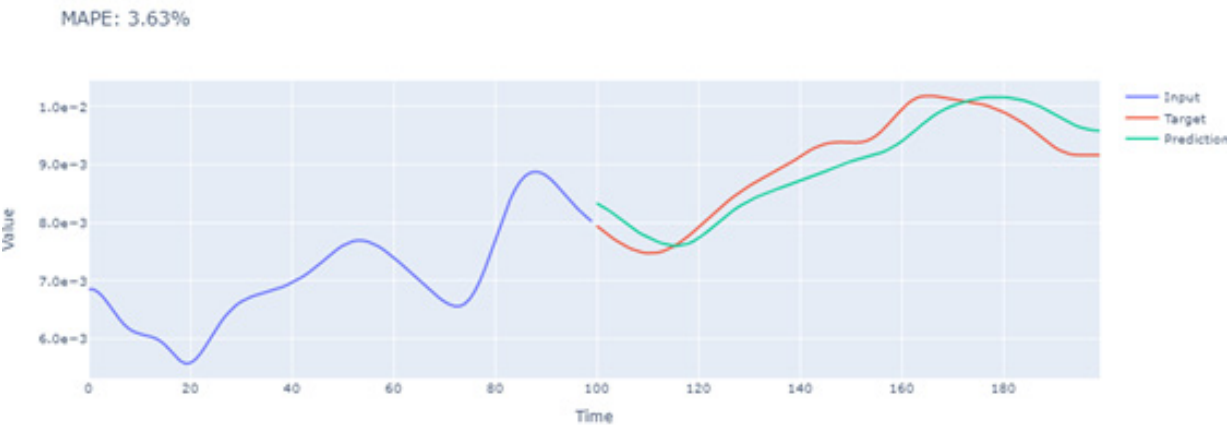
e) 0H barrier, point-5



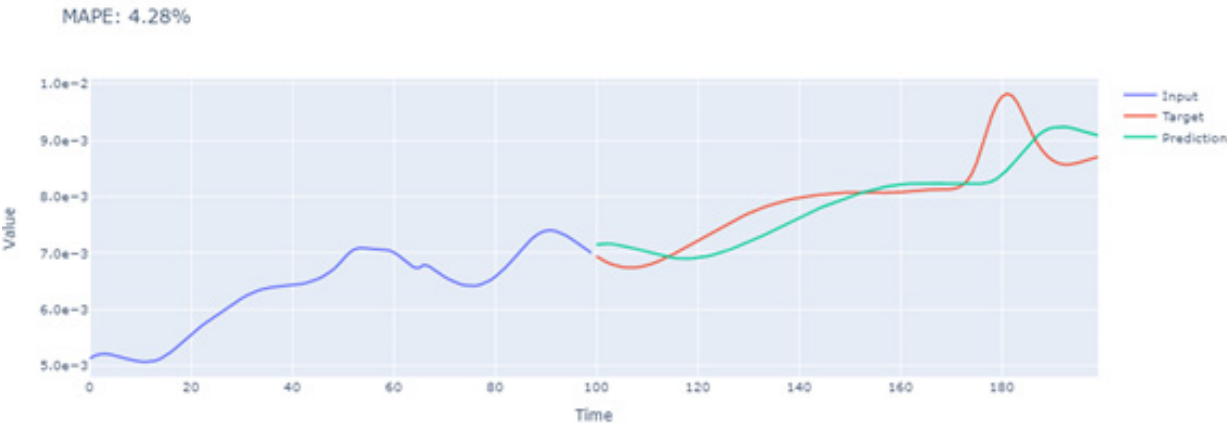
f) 0.1H barrier, point-1



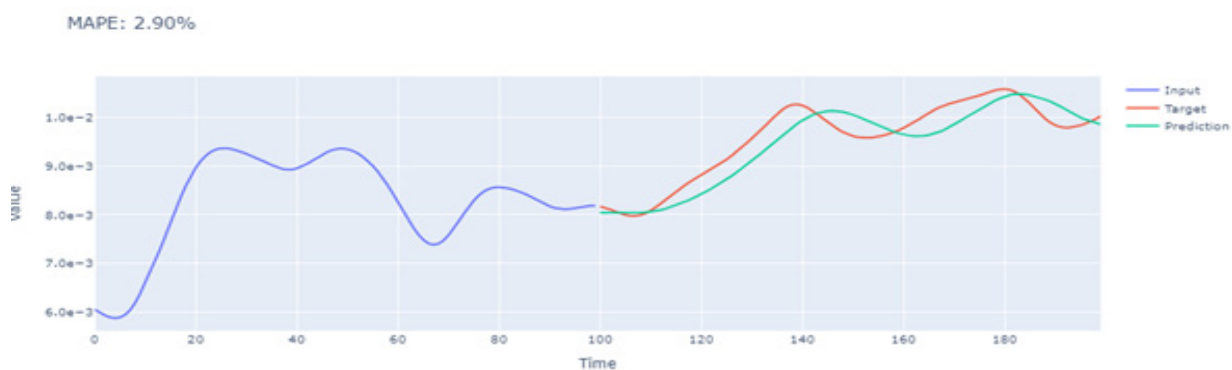
g) 0.1H barrier, point-2



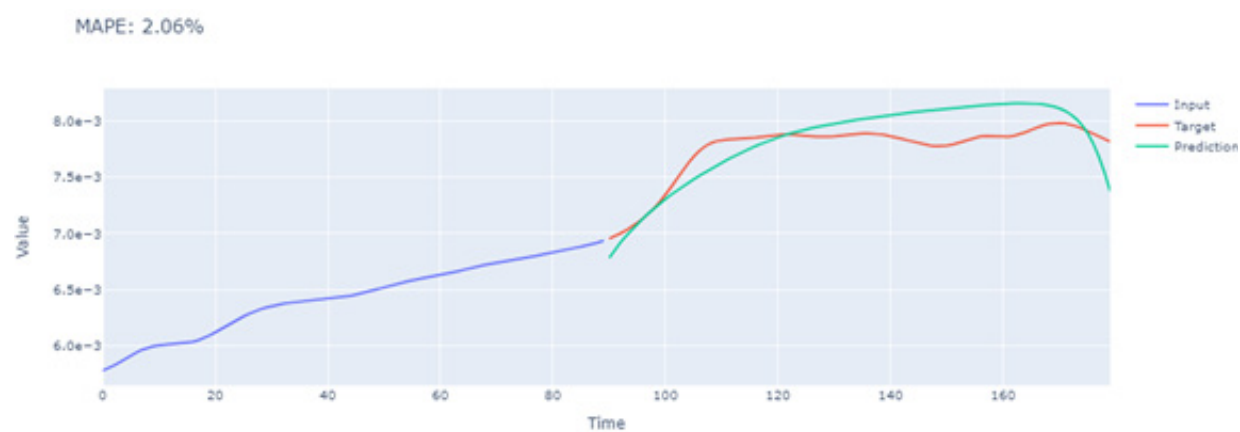
h) 0.1H barrier, point-3



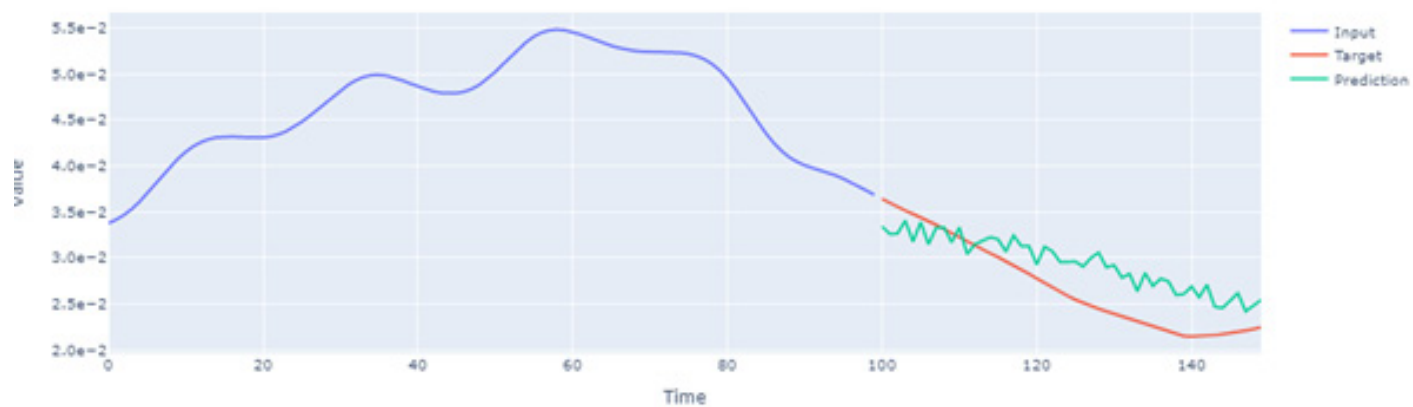
i) 0.1H barrier, point-4



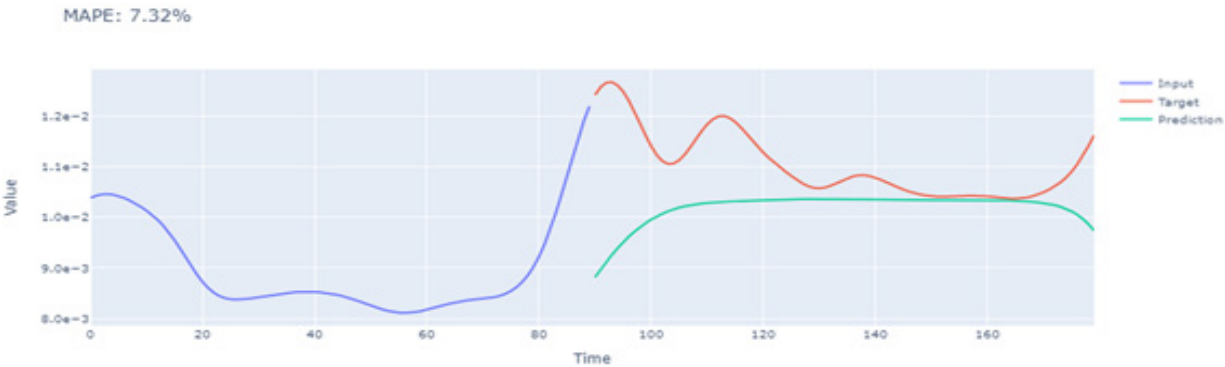
j) 0.1H barrier, point-5



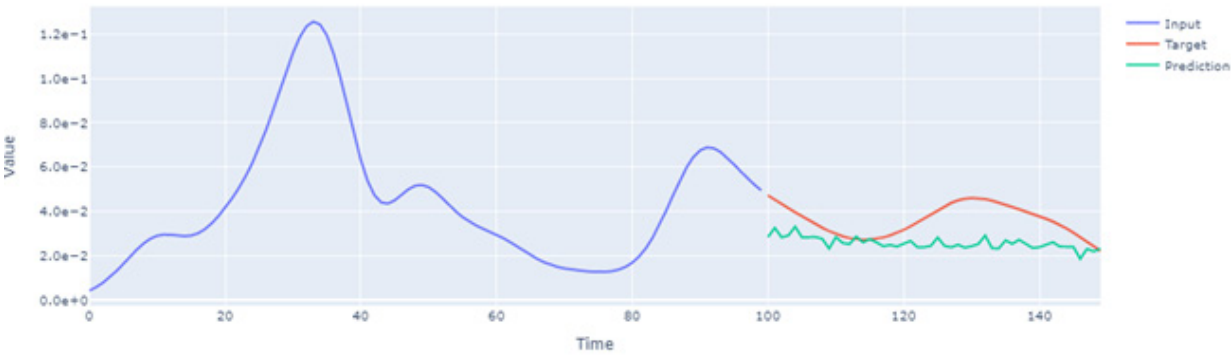
k) 0.2H barrier, point-1



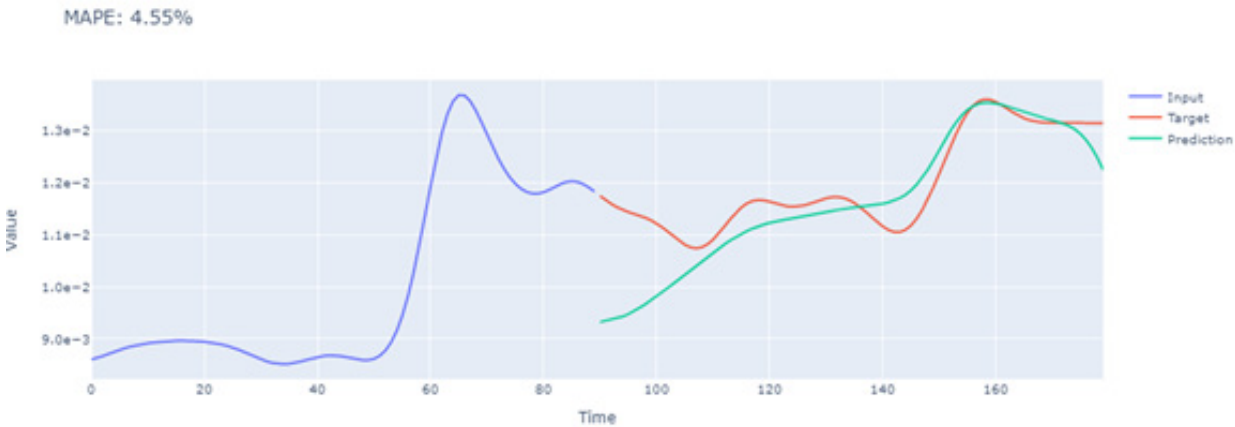
l) 0.2H barrier, point-2



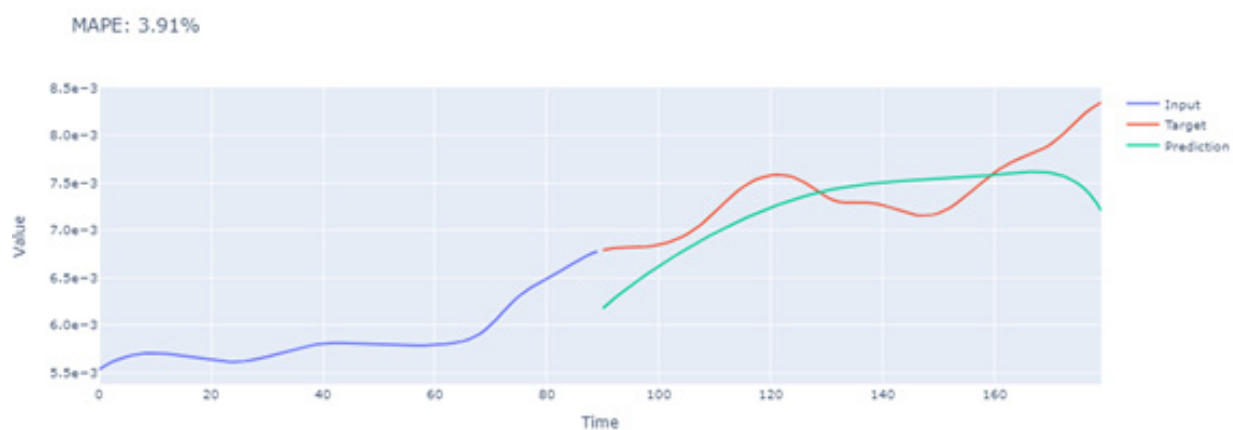
m) 0.2H barrier, point-3



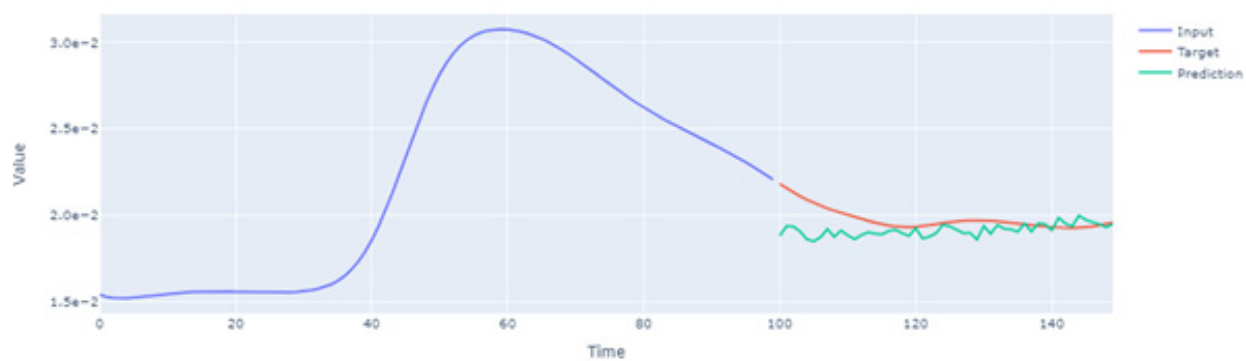
n) 0.2H barrier, point-4



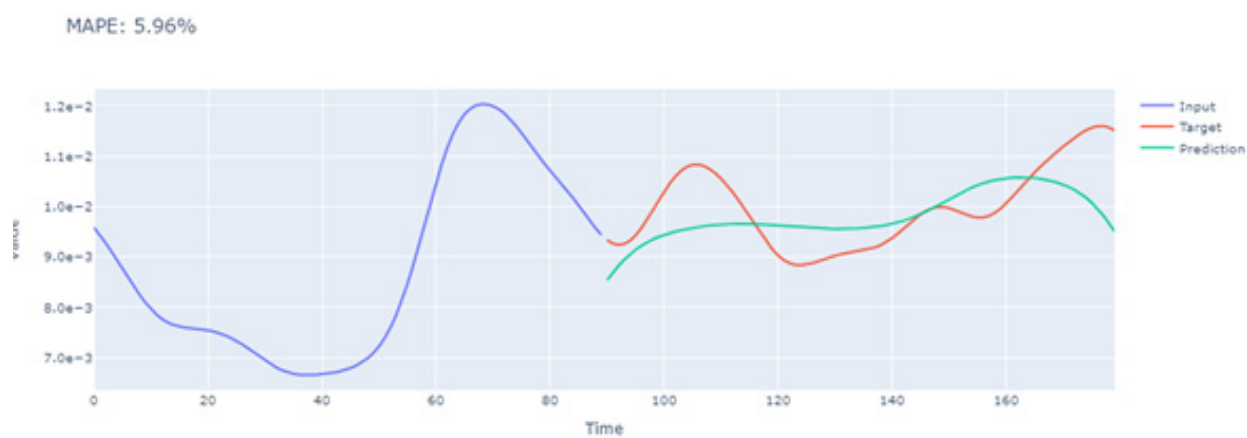
o) 0.2H barrier, point-5



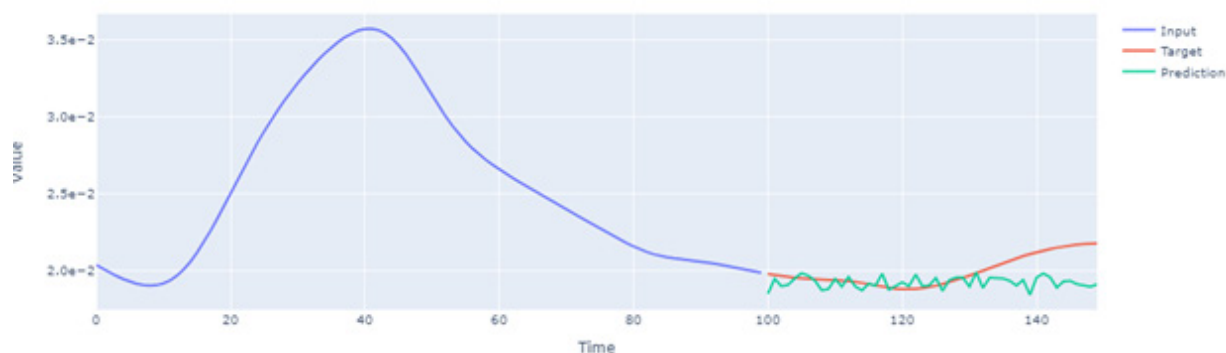
p) 0.3H barrier, point-1



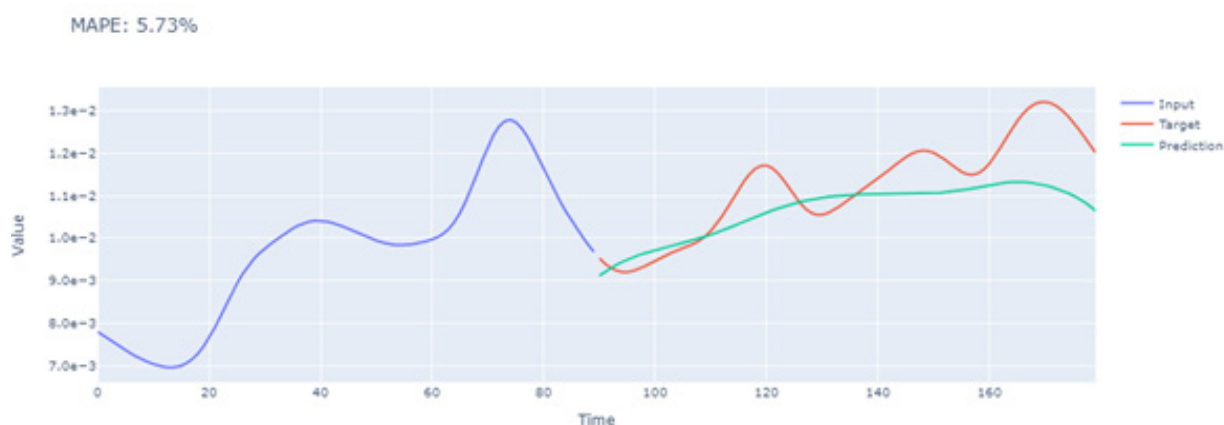
q) 0.3H barrier, point-2



r) 0.3H barrier, point-3



s) 0.3H barrier, point-4



t) 0.3H barrier, point-5

Рисунок 7 – Предсказания суррогатной модели для контрольных узлов сетки и различных высот барьера

Количественный анализ ошибок прогнозирования

Систематический анализ метрик MAE, MAPE и R^2 (таблица 2) позволил выявить четкие пространственные закономерности в точности модели. Наибольшая чувствительность к изменению высоты барьера наблюдается в точках, непосредственно находящихся в аэродинамической тени барьера и в зонах отрыва потока (точки 1, 2, 3). Для барьера высотой 0.2H в точке 2 фиксируется минимальное значение $R^2 = -4.5$. Отрицательный коэффициент детерминации указывает, что простая константная модель (среднее значение) в данном случае дает более точный прогноз, чем построенная BiLSTM. Это явный признак того, что модель не смогла выявить физически обоснованную зависимость в данных, что вызвано чрезвычайной сложностью и, вероятно, неполной представительностью обучающей выборки для данного специфического режима течения.

Следует отметить, что коэффициент детерминации R^2 для нестационарных временных рядов может давать нестабильные результаты. В контексте гидрологических и аэродинамических задач эквивалентной метрикой является коэффициент эффективности Нэша–Сатклиффа (NSE), который вычисляется по формуле, идентичной R^2 , и численно совпадает с приведенными в таблице 2 значениями.

Коэффициент эффективности Нэша–Сатклиффа (NSE), вычисляемый по формуле, математически эквивалентной R^2 , может принимать отрицательные значения, что указывает на то,

что простое использование среднего значения наблюдений обеспечивает более точный прогноз, чем рассматриваемая модель. Аналогичное поведение характерно для R^2 , вычисленного по той же формуле, в частности в библиотеке scikit-learn. Значение $NSE = -4.5$ в точке 2 для конфигурации барьера 0.2H свидетельствует о том, что модель не смогла выявить устойчивую зависимость в данных для указанного режима течения, что обусловлено экстремальной нестационарностью динамики концентрации в зоне аэродинамического влияния барьера.

В то же время точки, расположенные за барьером относительно направления ветра (точка 4), или на значительном удалении от зоны его непосредственного влияния (точка 5), демонстрируют гораздо более устойчивые результаты. Ошибки прогноза растут незначительно с увеличением высоты барьера, а значения (R^2) остаются в допустимом диапазоне. Это свидетельствует о том, что модель успешно экстраполирует изученные закономерности переноса для относительно спокойных зон потока, даже при изменении глобальной конфигурации.

Интересной находкой является нелинейная зависимость ошибки от высоты барьера. Наиболее проблемной с точки зрения прогнозирования оказалась не максимальная (0.3H), а средняя (0.2H) высота. Это подтверждает гипотезу о том, что ключевым фактором, лимитирующим точность суррогатной ML-модели в данной задаче, является не абсолютный уровень сложности физического явления, а степень его нестационарности и наличие в данных резких, плохо коррелирующих изменений.

Таблица 2 – Метрики ошибок для модели BiLSTM

Средняя абсолютная процентная ошибка (MAPE)				
№	No barrier	0.1 barrier	0.2 barrier	0.3 barrier
1	2.08%	5.91%	10.70%	6.65%
2	1.80%	14.70%	17.56%	14.15%
3	3.22%	4.00%	12.60%	3.07%
4	1.81%	4.59%	3.07%	7.62%
5	2.70%	3.47%	3.99%	7.10%
Средняя абсолютная ошибка (MAE)				
1	0.0001984	0.000285	0.01561	0.00237
2	0.0001942	0.002661	0.0267	0.00683
3	0.0002086	0.0002132	0.003559	0.000168
4	0.0009714	0.001019	0.0001686	0.00222
5	0.0002340	0.0001147	0.0001681	0.000335
R^2 – метрика				
1	0.45	0.47	-0.15	0.26
2	0.32	0.52	-4.5	0.29
3	0.32	0.71	0.51	0.63
4	0.46	0.61	0.71	0.23
5	0.60	0.72	0.63	0.61

Заключение

В ходе настоящего исследования была решена задача разработки и валидации гибридного подхода, сочетающего методы вычислительной гидродинамики (CFD) и машинного обучения (ML), для прогнозирования пространственно-временного распределения загрязняющих веществ в идеализированных городских уличных каньонах с учетом влияния инженерных элементов инфраструктуры – шумозащитных барьеров различной высоты.

CFD-моделирование позволило детально изучить аэродинамические особенности и процессы дисперсии примеси в различных геометрических конфигурациях. Установлено, что высота барьера оказывает существенное нелинейное влияние на структуру течения и характер

накопления загрязняющего вещества. В частности, барьер высотой $0.2H$ создает наиболее сложный и нестационарный режим течения с интенсивным вихреобразованием, что приводит к значительным пространственно-временным флуктуациям концентрации.

На основе данных нестационарных CFD-расчетов была разработана суррогатная модель на основе архитектуры BiLSTM с механизмом внимания. Модель продемонстрировала высокую эффективность в прогнозировании временной эволюции концентрации в большинстве контрольных точек, особенно для случаев отсутствия барьера и при наличии барьера малой высоты ($0.1H$). Средняя абсолютная процентная ошибка (MAPE) в этих конфигурациях не превышала 15%, а коэффициент детерминации R^2 оставался положительным, что свидетельствует о хорошей обобщающей способности модели.

Однако при увеличении высоты барьера до $0.2H$ точность прогноза снизилась, особенно в зонах непосредственного аэродинамического влияния барьера, где наблюдались резкие и слабо коррелированные изменения концентрации. В наиболее сложной точке (точка 2) модель показала отрицательное значение R^2 , что указывает на недостаточную репрезентативность обучающих данных для данного режима и ограниченную способность архитектуры улавливать экстремально нестационарные процессы.

Тем не менее, для барьера высотой $0.3H$ точность прогноза частично восстановилась благодаря стабилизации вихревой структуры и снижению временной изменчивости полей концентрации. Это подтверждает, что ключевым фактором, определяющим эффективность суррогатной ML-модели, является не абсолютная сложность физического явления, а степень его нестационарности и наличие в данных резких, плохо предсказуемых изменений.

Проведенное исследование подтвердило перспективность гибридного CFD–ML подхода для задач оценки и прогнозирования качества воздуха в городской среде. Предложенная методология позволяет сочетать физическую достоверность численного моделирования с вычислительной эффективностью машинного обучения, что открывает возможности для оперативного анализа сценариев, оптимизации городской инфраструктуры и поддержки принятия экологически ориентированных решений. Суррогатная модель обеспечивает ускорение прогнозирования по сравнению с CFD-расчетом на 7–8 порядков: время инференса составляет порядка 1–5 мс, тогда как одна CFD-симуляция занимает около 54 часов на CPU (32 потока, 128 GB RAM). Таким образом, предложенный гибридный подход позволяет проводить многовариантный анализ сценариев при вычислительных затратах, недостижимых для прямого численного моделирования.

В качестве направлений дальнейших исследований целесообразно расширение обучающей выборки за счет включения большего числа вариаций геометрических и метеорологических параметров, применение более сложных архитектур глубокого обучения, а также интеграция предложенного подхода в системы мониторинга и управления качеством городского воздуха в режиме, близком к реальному времени.

Информация о финансировании

Работа поддержана грантом Министерства науки и высшего образования Республики Казахстан (BR28712901).

REFERENCES

- 1 World Health Organization. Ambient (outdoor) air pollution. World Health Organization (2023). URL: [https://www.who.int/ru/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/ru/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- 2 Turner, M.C., Andersen, Z.J., Baccarelli, A., Diver, W.R., Gapstur, S.M., Pope, C.A., Prada, D., Samet, J., Thurston, G., and Cohen, A. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. CA: A Cancer Journal for Clinicians, 70 (6), 460–479 (2020). <https://doi.org/10.3322/caac.21632>

- 3 Singh, J., Payra, S., and Verma, S. Air quality and human health. In: *Geospatial Analytics for Environmental Pollution Modeling*, pp. 317–352 (Cham: Springer, 2023). https://doi.org/10.1007/978-3-031-45300-7_13
- 4 Hänninen, O., Lehtomäki, H., Korhonen, A., Kokkola, T., Hartikainen, A., Sippula, O., Haverinen-Shaughnessy, U., Leviäkangas, P., and Rumrich, I.K. Health risks related to air pollution by transport categories and vehicle types: Comparison by mortality indicators. *Environment International*, 202, 109657 (2025). <https://doi.org/10.1016/j.envint.2025.109657>
- 5 Lolli, S. Urban PM_{2.5} concentration monitoring: A review of recent advances in ground-based, satellite, model, and machine learning integration. *Urban Climate*, 63, 102566 (2025). <https://doi.org/10.1016/j.uclim.2025.102566>
- 6 Natterer, E.S., Rao, S.R., Tejada Lapuerta, A., Engelhardt, R., Hörl, S., and Bogenberger, K. Machine learning surrogates for agent-based models in transportation policy analysis. *Transportation Research Part C: Emerging Technologies*, 180, 105360 (2025). <https://doi.org/10.1016/j.trc.2025.105360>
- 7 Xia, F., Cheng, X., Lei, Z., Liu, Y., Wang, J., and Zhang, H. Heterogeneous impacts of local traffic congestion on local air pollution within a city: Utilizing taxi trajectory data. *Journal of Environmental Economics and Management*, 122, 102896 (2023). <https://doi.org/10.1016/j.jeem.2023.102896>
- 8 Santiago, J.L., Sanchez, B., Quaassdorff, C., de la Paz, D., Martilli, A., Martín, F., and Lechón, Y. Performance evaluation of a multiscale modelling system applied to particulate matter dispersion in a real traffic hot spot in Madrid (Spain). *Atmospheric Pollution Research*, 11 (1), 141–155 (2020). <https://doi.org/10.1016/j.apr.2019.10.001>
- 9 Lin, C., Wang, Y., Ooka, R., Flageul, C., Kim, Y., Kikumoto, H., Wang, Z., and Sartelet, K. Modeling of street-scale pollutant dispersion by coupled simulation of chemical reaction, aerosol dynamics, and CFD. *Atmospheric Chemistry and Physics*, 23, 1421–1436 (2023). <https://doi.org/10.5194/acp-23-1421-2023>
- 10 Vahidi, S., Yeganeh, M., and Ghasaban, M. The effect of the morphology of highly polluted urban areas neighborhood on suspended particles of PM_{2.5} and PM₁₀ air pollutants (Case study: Ahvaz, Iran). *Energy Nexus*, 20, 100527 (2025). <https://doi.org/10.1016/j.nexus.2025.100527>
- 11 Fellini, S., Majumdar, D., Salizzoni, P., and van Reeuwijk, M. Three-dimensional pollutant dispersion in tree-lined urban canyons: Combined wind-tunnel and LES analysis. *Atmospheric Environment*, 367, 121748 (2026). <https://doi.org/10.1016/j.atmosenv.2025.121748>
- 12 Mousavi, D., Ghaemian, M., and Brinkerhoff, J. Multiscale CFD analysis of urban air pollution dome and ventilation enhancement via an urban chimney. *Atmospheric Environment*, 337, 120783 (2024). <https://doi.org/10.1016/j.atmosenv.2024.120783>
- 13 Huertas, J.I., and Prato, D.F. CFD modeling of near-roadway air pollution. *Environmental Modeling & Assessment*, 25 (1), 129–145 (2020). <https://doi.org/10.1007/s10666-019-09666-w>
- 14 Kikumoto, H., and Ooka, R. Large-eddy simulation of pollutant dispersion in a cavity at fine grid resolutions. *Building and Environment*, 127, 127–137 (2017). <https://doi.org/10.1016/j.buildenv.2017.11.005>
- 15 Reiminger, N., Jurado, X., Vazquez, J., Wemmert, C., and Meunier, G. Methodologies to assess mean annual air pollution concentration combining numerical results and wind roses. *Sustainable Cities and Society*, 59, 102221 (2020). <https://doi.org/10.1016/j.scs.2020.102221>
- 16 Martín, F., Janssen, S., Rodrigues, V., Sousa, J., Santiago, J.L., Rivas, E., Stocker, J., Jackson, R., Russo, F., Villani, M.G., Tinarelli, G., Barbero, D., José, R.S., Pérez-Camanyo, J.L., Santos, G.S., Bartzis, J., Sakellaris, I., Horváth, Z., Környei, L., Cuvelier, C. Using dispersion models at microscale to assess long-term air pollution in urban hot spots: A FAIRMODE joint intercomparison exercise for a case study in Antwerp. *Science of the Total Environment*, 925, 171761 (2024). <https://doi.org/10.1016/j.scitotenv.2024.171761>
- 17 Bi, J., Knowland, K.E., Keller, C.A., and Liu, Y. Combining machine learning and numerical simulation for high-resolution PM_{2.5} concentration forecast. *Environmental Science & Technology*, 56 (3), 1544–1556 (2022). <https://doi.org/10.1021/acs.est.1c05578>
- 18 Jurado, X., Reiminger, N., Benmoussa, M., Vazquez, J., and Wemmert, C. Deep learning methods evaluation to predict air quality based on computational fluid dynamics. *Expert Systems with Applications*, 203, 117294 (2022). <https://doi.org/10.1016/j.eswa.2022.117294>
- 19 Schulte, N., Snyder, M., Isakov, V., Heist, D., and Venkatram, A. Effects of solid barriers on dispersion of roadway emissions. *Atmospheric Environment*, 97, 286–295 (2014). <https://doi.org/10.1016/j.atmosenv.2014.08.026>

20 Venkatram, A., Isakov, V., Deshmukh, P., and Baldauf, R. Modeling the impact of solid noise barriers on near-road air quality. *Atmospheric Environment*, 141, 462–469 (2016). <https://doi.org/10.1016/j.atmosenv.2016.07.005>

21 Li, B., Qiu, Z., and Zheng, J. Impacts of noise barriers on near-viaduct air quality in a city: A case study in Xi'an. *Building and Environment*, 196, 107751 (2021). <https://doi.org/10.1016/j.buildenv.2021.107751>

22 Wu, Y., and Chen, H. Optimizing block morphology for reducing traffic pollutant concentration in adjacent external spaces of street canyons: A machine learning approach. *Building and Environment*, 242, 110587 (2023). <https://doi.org/10.1016/j.buildenv.2023.110587>

23 Mora-Mariano, D., and Flores-Tlacuahuac, A. Bayesian LSTM framework for the surrogate modeling of process engineering systems. *Computers & Chemical Engineering*, 181, 108553 (2023). <https://doi.org/10.1016/j.compchemeng.2023.108553>

^{1*}**Исахов А.А.,**

профессор, ORCID ID: 0000-0002-1937-8615,

*e-mail: alibek.issakhov@gmail.com

²**Нығметова Ә.,**

докторант, ORCID ID: 0009-0004-5395-4569,

e-mail: a.nygmetova@kbtu.kz

¹**Абылкасымова А.Б.,**

қауымдастырылған профессор, ORCID ID: 0000-0002-5967-6959,

e-mail: abylkassymova.aizhan@gmail.com

¹**Аханова Н.Е.,**

профессор, ORCID ID: 0000-0003-2767-8711,

e-mail: n.akhanova@kbtu.kz

¹Қазақстан-Британ техникалық университет, Алматы қ., Қазақстан

²Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

ML-CFD МОДЕЛЬДЕУІН ҚОЛДАНА ОТЫРЫП, ҚАЛАЛЫҚ КАНЬОНДАРДАҒЫ ШЫҒАРЫНДЫЛАРДАН АУАНЫҢ ЛАСТАНУЫН БОЛЖАУ

Аңдатпа

Мақалада шу кедергілерін ескере отырып, қалалық көше каньондарындағы ластаушы заттардың таралуын болжаудың гибриді тәсілі ұсынылған. Әдістеме егжей-тегжейлі CFD модельдеуін және назар аудару механизмі бар BiLSTM нейрондық желі архитектурасына негізделген суррогат модельді біріктіреді. 0,1Н, 0,2Н және 0,3Н тосқауыл биіктігі бар конфигурациялар зерттелді. CFD есептеулері тосқауыл биіктігінің аэродинамикаға және ластаушы заттардың жиналу аймақтарының қалыптасуына сызықтық емес әсер ететінін көрсетті, ал ең күрделі стационарлық емес мінез-құлық 0,2Н биіктікте байқалды. Суррогат модель кедергісіз және 0,1Н тосқауыл жағдайларында концентрация эволюциясын сәтті болжады, орташа абсолюттік пайыздық қателіктің 15%-дан аз екенін көрсетті. Алайда 0,2Н тосқауыл үшін процестің жоғары дәрежедегі стационарлық емес сипатына байланысты қарқынды турбуленттілік аймақтарында дәлдік төмендеді. Бұл тәсіл физикалық дәлдікті сақтай отырып, есептеу шығындарын айтарлықтай азайтады, сондықтан қалалық экологиядағы шешім қабылдауды қолдау жүйелері үшін перспективалы болып табылады. Модель CFD есептеулерімен салыстырғанда болжауды 7–8 есе жылдам қамтамасыз етеді: қорытындылау уақыты 1–5 мс болса, бір CFD модельдеуін орындау CPU-де шамамен 54 сағатты алады.

Түйін сөздер: ауаның ластануы, көше каньондары, шу кедергілері, есептеу сұйықтық динамикасы (CFD), машиналық оқыту, BiLSTM, суррогат модельдеу, концентрацияны болжау.

^{1*}**Issakhov A.A.,**

Professor, ORCID ID: 0000-0002-1937-8615,

*e-mail: alibek.issakhov@gmail.com

²**Nygmetyova A.,**

PhD student, ORCID ID: 0009-0004-5395-4569,

e-mail: a.nygmetyova@kbtu.kz

¹**Abylkassymova A.B.,**

Associate Professor, ORCID ID: 0000-0002-5967-6959,

e-mail: abylkassymova.aizhan@gmail.com

¹**Akhanova N.Ye.**

Professor, ORCID ID: 0000-0003-2767-8711,

e-mail: n.akhanova@kbtu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

²International Information Technology University, Almaty, Kazakhstan

FORECASTING AIR POLLUTION FROM EMISSIONS IN URBAN CANYONS USING ML-CFD MODELING

Abstract

This paper presents a hybrid approach for predicting pollutant dispersion in urban street canyons, taking into account noise barriers. The methodology combines detailed CFD modeling and a surrogate model based on the BiLSTM neural network architecture with an attention mechanism. Configurations with barrier heights of 0.1H, 0.2H, and 0.3H were studied. CFD calculations revealed a nonlinear effect of barrier height on aerodynamics and the formation of pollutant accumulation zones, with the most complex non-stationary behavior observed at a height of 0.2H. The surrogate model successfully predicts concentration evolution for both the barrier-free and 0.1H barrier cases, demonstrating an average absolute percentage error of less than 15%. For a 0.2H barrier, accuracy decreases in zones of intense turbulence due to the highly non-stationary nature of the process. This approach significantly reduces computational costs while maintaining physical accuracy, which is promising for decision support systems in urban ecology. The model provides accelerated forecasting compared to CFD calculations by 7–8 orders of magnitude, so the inference time showed 1–5 ms, while one CFD simulation takes about 54 hours on the CPU.

Keywords: air pollution, street canyons, noise barriers, computational fluid dynamics (CFD), machine learning, BiLSTM, surrogate modeling, concentration prediction

Received February 12, 2026; revised April 16, June 1, 2026; accepted June 3, 2026

УДК 519.635.4
МПНТИ 27.35.16

<https://doi.org/10.55452/1998-6688-2026-23-2-108-123>

^{1*}**Sinita A.V.,**

PhD, Assistant Professor, ORCID ID: 0000-0002-5828-7820,

*e-mail: a.sinita@kbtu.kz

¹**Tskhay Yu.A.,**

MSc, Lecturer, ORCID ID: 0009-0003-0772-4319,

e-mail: y.tskhay@kbtu.kz.

¹**Shkorko A.K.,**

MSc, Lecturer, ORCID ID: 0009-0003-0402-4720,

e-mail: a.ukassova@kbtu.kz.

²**Karduck A.P.,**

PhD, Professor, ORCID ID: 0000-0002-8142-6710,

e-mail: achim.karduck@hs-furtwangen.de

¹Kazakh-British Technical University, Almaty, Kazakhstan

²Furtwangen University, Furtwangen, Germany

NUMERICAL SOLUTION OF THE INVERSE COEFFICIENT ACOUSTIC PROBLEM USING LAPLACE TRANSFORM AND DIFFERENTIAL EVOLUTION METHOD

Abstract

This paper addresses the solution of the inverse coefficient problem for the wave equation aimed at reconstructing the spatial distribution of the speed of sound in an inhomogeneous medium. The Laplace transform is applied to solve the direct problem, eliminating time dependence and reducing the problem to ordinary differential equations in the frequency domain, which significantly decreases computational costs. The inverse problem is formulated as an optimization task: minimizing the residual functional between calculated and measured acoustic pressure values using the stochastic global optimization method, Differential Evolution. Numerical experiments were conducted on a multilayer medium model (sand, soil, rock, water, air) using synthetic data with added random noise. An adaptive combined reconstruction method is proposed to reduce errors at medium boundaries. The results demonstrate high accuracy: the relative error of the sound speed profile reconstruction was approximately from 2.5 to 4.3%, confirming the approach's effectiveness for acoustic diagnostics and tomography applications.

Keywords: Inverse problem, acoustic pressure, speed of sound, Laplace transform, differential evolution, inhomogeneous medium, optimization.

Received January 30, 2026; revised May 26, June 3, 2026; accepted June 8, 2026.

Introduction

Mathematical modeling is a cornerstone in analyzing complex physical processes where direct observation is constrained or impossible. Among the various classes of mathematical problems, inverse problems hold a distinct significance. Unlike direct problems, where the system's output is determined from known parameters, inverse problems aim to reconstruct the unknown internal characteristics of a medium, such as the speed of sound, density, or source parameters, based on observed data [1, 2]. This class of problems is fundamental to acoustic tomography, geophysical exploration, and non-destructive testing [3, 4]. The solution of inverse problems is inherently ill-posed in the sense of Hadamard, meaning that minor perturbations in the measured data can lead to significant errors in the reconstructed parameters [5]. Consequently, the development of

stable numerical algorithms and regularization techniques remains a priority in computational mathematics. Significant contributions to the theory of ill-posed and inverse problems have been made by the global scientific community, including the fundamental works of A.N. Tikhonov and M.M. Lavrentiev. In Kazakhstan, the theory of inverse problems has been extensively developed by the scientific school of S.I. Kabanikhin and his colleagues. For instance, Kabanikhin et al. in [6] and [7] investigated optimization methods for hyperbolic inverse problems, providing a theoretical basis for gradient-based reconstruction. Furthermore, Bektemesov et al. [8] and Iskakov [9] proposed efficient numerical algorithms for solving coefficient inverse problems in layered media, which are highly relevant to geophysical applications.

Traditionally, wave propagation problems are solved using grid-based methods such as Finite Difference (FDM) or Finite Element Methods (FEM) in the time domain [10, 11]. However, these approaches often incur high computational costs and suffer from numerical dispersion, especially in high-frequency regimes and multiscale domains [12]. As noted by El Moçayd et al. [13], data-driven hybrid modeling can mitigate some of these issues, yet the demand for computationally efficient analytical or quasi-analytical methods persists.

To address the limitations of time-domain solvers, various transformations have been proposed. The Laplace transform is particularly effective as it eliminates the time variable, reducing the hyperbolic partial differential equation to a parameter-dependent ordinary differential equation [14]. Keyantuo [15] demonstrated the applicability of the ascent method for abstract wave equations using the Laplace transform. Recent studies have expanded this approach to more complex scenarios. For example, Li et al. [16] applied the Laplace transform to study stress wave propagation in multi-layer media, while Owolabi et al. [17] utilized it for fractional time-diffusive models. It is worth noting that the approach proposed in the present study is conceptually related to the work of Meirmanov et al. [18], who considered an inverse coefficient problem for seismic wave propagation in composite layered media and reduced the governing equations to the frequency domain via the Fourier transform. The key distinction of the present work lies in the use of the Laplace transform instead of the Fourier transform: the Laplace transform allows for the incorporation of non-zero initial conditions in a more natural manner and yields a closed-form analytical solution directly in the complex frequency domain, which is then used to construct the objective function for the Differential Evolution optimization. In contrast, the approach of [18] operates in the Fourier frequency domain and employs the Nelder-Mead algorithm for the minimization of the data misfit functional.

In the context of reconstruction, optimization algorithms play a crucial role. While gradient-based methods are powerful, they are often susceptible to local minima. Consequently, stochastic global optimization methods have gained traction. The Differential Evolution (DE) algorithm, discussed by Wojciechowski et al. [19] and Reda et al. [20], offers robustness against noisy data and does not require gradient information. Meanwhile, the integration of machine learning into inverse problems is becoming increasingly prominent. Qu et al. [21] and Sun et al. [22] explored deep-learning-based tomography reconstruction, showing promising results in accelerating the inversion process.

In this paper, we propose a method that combines the analytical advantages of the Laplace transform with the robustness of the Differential Evolution algorithm. We focus on reconstructing the sound speed profile in a one-dimensional enclosed medium. This work builds upon our previous research [23], extending the analysis to more complex stratified environments. By transforming the wave equation into the frequency domain, we derive an analytical solution for the direct problem, which is then used to construct an objective function for the inverse problem.

The scientific novelty of the present work is threefold. First, unlike prior studies that apply the Fourier transform to reduce the wave equation to the frequency domain, the present approach employs the Laplace transform, which accommodates non-zero initial conditions naturally and yields a closed-form analytical solution valid for arbitrary real-valued Laplace parameter s . Second, the resulting analytical solution is used directly as the forward model inside a Differential Evolution optimization loop – eliminating the need for a numerical PDE solver at each iteration, which reduces computational cost by approximately two orders of magnitude compared to finite-difference-

based inversion schemes. Third, a combined adaptive post-processing method is introduced that dynamically selects between raw and smoothed reconstructed values at each spatial node, preserving sharp acoustic impedance contrasts while suppressing stochastic noise in homogeneous layers. The combination of these three elements constitutes the original contribution of this study.

Let us consider the process of acoustic wave propagation in a one-dimensional enclosed domain $\mathbf{x} \in \Omega: [0, L_x] \times [0, L_y] \times [0, L_z]$ and $t \in [0, T_{\max}]$. The general three-dimensional acoustic wave equation governing the sound pressure $P(\mathbf{r}, t)$ in an inhomogeneous medium with spatially varying speed of sound $c(\mathbf{r})$ is:

$$\frac{\partial^2 P}{\partial t^2} = c^2(\mathbf{r}) \left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2} \right). \quad (1)$$

Where: $P(\mathbf{r}, t)$ is the sound pressure; $\mathbf{r} = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}$ – the position vector of domain point with standard basis set, defined by the unit vectors $\{\hat{i}, \hat{j}, \hat{k}\}$; c – speed of sound, the desired coefficient for the inverse problem.

Since the medium is assumed to be homogeneous in the y and z directions and the acoustic source varies only along x , the problem reduces to one spatial dimension. Setting $\frac{\partial}{\partial y} = \frac{\partial}{\partial z} = 0$ and writing $c(x)$ for the speed of sound profile, the governing equation becomes:

$$\frac{\partial^2 P(x, t)}{\partial t^2} = c^2(x) \cdot \frac{\partial^2 P(x, t)}{\partial x^2}, \quad x \in [0, L], \quad t \in [0, T_{\max}]. \quad (2)$$

Where $P(x, t)$ is the acoustic pressure [Pa]; $c(x) > 0$ is the local speed of sound [m/s], which is the unknown coefficient to be recovered by the inverse problem; L is the length of the one-dimensional domain [m]; and $T_{\max}$ is the total observation time [s].

The initial and conditions define the pressure distribution and its rate of change at the initial moment and certain positions. It is assumed that the initial pressure is equal to a constant P_0 , and the initial rate of change of pressure is zero, while the boundary conditions determine the interaction of the wave with the domain boundaries. A Dirichlet boundary condition is imposed at $x = 0$, modeling a constant acoustic pressure source. An erroneous or simplified specification of boundary behavior can lead to artificial reflections, resonance effects or energy loss not caused by the real properties of the medium. This is especially critical in inverse modeling problems, where the slightest distortions in the simulated acoustic field can significantly affect the accuracy of reconstructing parameters such as the sound velocity profile. For the boundary at $x = 0$, we set the condition $p = 1$, which corresponds to a fixed acoustic pressure on the given surface. This allows us to model the source of the sound field, since the pressure is constant and does not depend on time. Thus, we obtain the following:

$$\begin{cases} P|_{t=0} = P_0, \\ \frac{\partial P}{\partial t}|_{t=0} = 0, \\ \frac{\partial P}{\partial n}|_{x=L} = 0, \quad n \in \{x, y, z\}, \\ P|_{x=0} = 1. \end{cases} \quad (3)$$

On the remaining boundaries of the computational domain, we set the Neumann boundary state equal to the basis of what physically corresponds to an absolutely rigid boundary. This means that the derivative pressure on the normal boundaries can be the same, i.e. the sound pressure can change, but the normal component of the mass flow through the boundary is absent.

With given initial data ($c(x)$ – speed of sound and boundary conditions (provided above in equations (2) – (3)), compute $P(x, t)$ which satisfies the given equation above and the harmonic source sound $s(x, t) = \Re\{S(x)e^{-st}\}$, then in the frequency domain:

$$P(x, t) = \Re\{\hat{u}(x, \omega)e^{-i\omega t}\}. \quad (4)$$

Solution in the Frequency Domain via Laplace Transform

To solve the forward problem and eliminate the time variable, the Laplace transform is applied:

$$\tilde{P}(x, s) = \int_0^\infty P(x, t)e^{-st}dt, \quad s = \sigma + i\omega, \quad \sigma > 0. \quad (5)$$

Where $s = \sigma + i\omega$ is the complex Laplace variable, $\sigma > 0$ is the real part (convergence abscissa), and ω is the imaginary frequency component. In numerical implementation the values of s are taken as $s = \sigma + i\omega_k$, $k = 1, \dots, N$, with σ fixed and ω_k sampled over the frequency range of interest.

Applying this integral transform to the original wave equation, taking into account the initial conditions, leads to an ordinary differential equation of the second order with respect to the spatial variable x :

$$s^2\tilde{P}(x, s) - c^2\frac{d^2\tilde{P}(x, s)}{dx^2} = sp_0. \quad (6)$$

Or, in the standard form of the inhomogeneous Helmholtz equation:

$$\frac{d^2\tilde{P}(x, s)}{dx^2} - \frac{s^2}{c^2}\tilde{P}(x, s) = -\frac{sp_0}{c^2}. \quad (7)$$

The boundary conditions in the frequency domain take the following form:

$$\tilde{P}(0, s) = \frac{1}{s}, \quad \frac{d\tilde{P}}{dx}(L, s) = 0. \quad (8)$$

The general solution to this equation is represented as the sum of the homogeneous solution and a particular solution of the inhomogeneous part:

$$\tilde{P}(x, s) = Ae^{\frac{sx}{c}} + Be^{-\frac{sx}{c}} + \frac{p_0}{s}. \quad (9)$$

The constants **A** and **B** are determined by substituting the boundary conditions, resulting in a system of algebraic equations. Solving this system yields the exact analytical expression for pressure in the frequency domain:

$$\tilde{P}(x, s) = \frac{p_0}{s} + \frac{(1 - p_0)}{s} \cdot \frac{e^{-\frac{sx}{c}} + e^{-\frac{2sL}{c}}e^{-\frac{sx}{c}}}{1 + e^{-\frac{2sL}{c}}}. \quad (10)$$

The figure below shows a graph of the interpolated function, in the form of a surface in three-dimensional space, due to the dependence of the function on 2 independent variables.

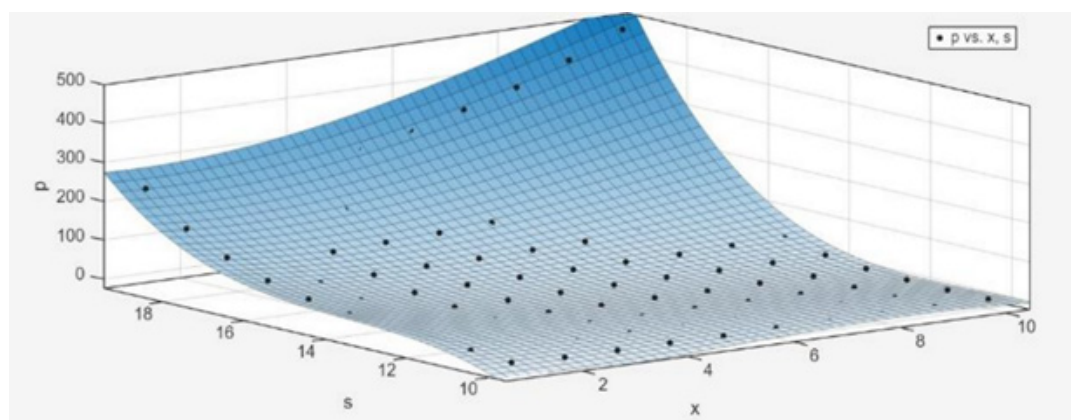


Figure 1 – Surface of function $P(x, s)$

Inverse Transformation and Optimization

To return to the time domain, a series expansion is used, allowing the solution to be interpreted as a superposition of time-delayed waves. Considering the properties of the Laplace transform for time-shifted functions, the analytical solution in the time domain is expressed using the Heaviside step function $H(t)$:

$$P(x, t) = p_0 + \sum_{n=0}^{\infty} (-1)^n \left[(1 - p_0) H\left(t - \frac{x+2nL}{c}\right) + (1 - p_0) H\left(t - \frac{2L+x+2nL}{c}\right) \right]. \quad (11)$$

To ensure numerical stability and differentiability during the optimization process, the discontinuous Heaviside function is replaced by its smooth approximation, the sigmoid function:

$$H(t) \approx \sigma(t; \alpha) = \frac{1}{1 + e^{-\alpha t}}. \quad (12)$$

The inverse problem is formulated as finding the sound speed distribution $c(x)$ that minimizes the residual functional between the calculated pressure P_{calc} and the measured pressure P_{meas} :

$$J(c) = \int_0^{T_{\text{max}}} [P_{\text{calc}}(x, t, c) - P_{\text{meas}}(x, t)]^2 dt \rightarrow \min. \quad (13)$$

To minimize this functional, the Differential Evolution algorithm is employed. This stochastic global optimization method is capable of finding the global minimum in a multidimensional parameter space without requiring gradient calculations, which is particularly crucial for problems involving noisy data. The spectral graph provides information about the resonance properties of the medium, its filtration behavior and frequency sensitivity, which is especially valuable when constructing tomographic or diagnostic algorithms.

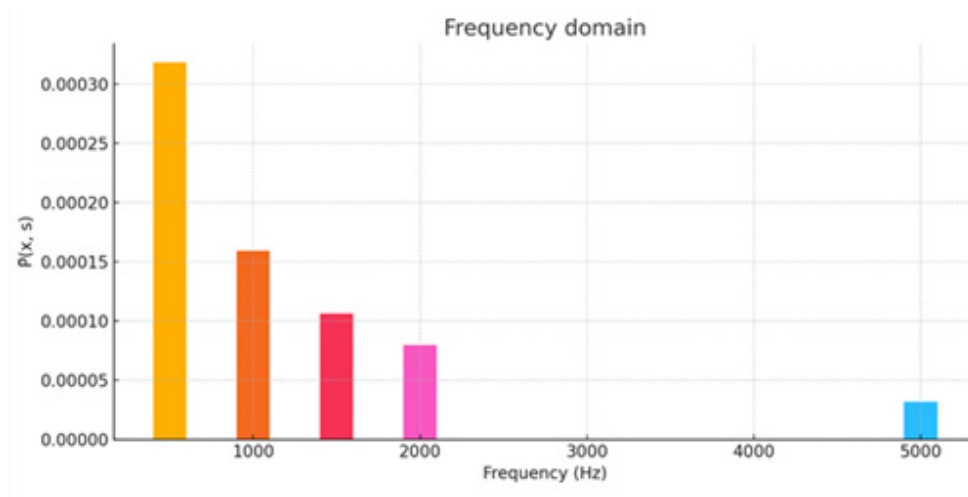


Figure 2 – Sound wave pressure with different frequency values (Hz),
sound pressure vs. frequency diagram

We will also plot the same waves, but in the time domain. Where we clearly see the difference in representations in the domains.

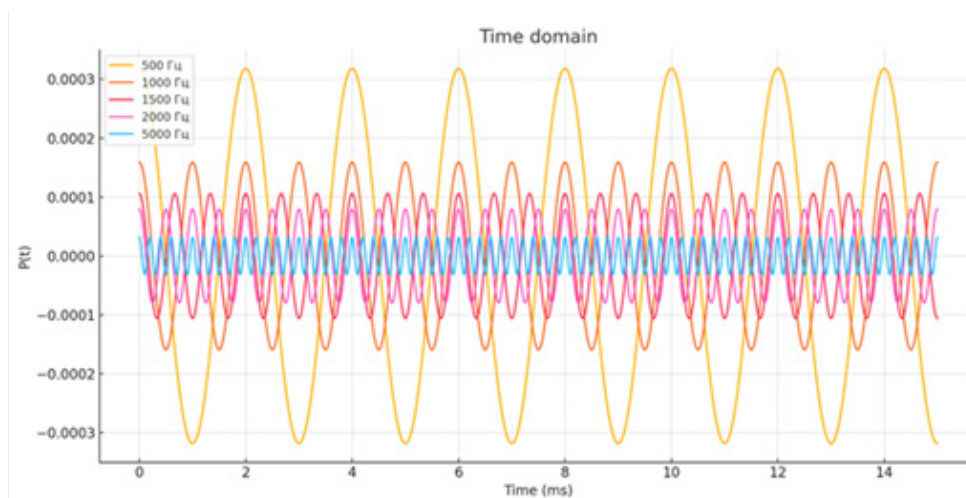


Figure 3 – Sound wave pressure with different frequency values (Hz),
sound pressure vs. time graph

Materials and methods

Numerical Implementation of the Inverse Solver

The core of the computational framework is the solution of the inverse coefficient problem via stochastic global optimization. The algorithm was implemented using the Python programming language, utilizing the `scipy.optimize` library.

The optimization process employs the Differential Evolution (DE) method. This algorithm was selected due to its robustness against local minima and its ability to explore the parameter space without requiring gradient information, which is computationally expensive for the defined objective function.

The inverse problem is treated as a parameter estimation task where the local speed of sound c_i at each spatial node x_i is determined by minimizing the misfit between the simulated pressure P_{calc} and the reference data. The Differential Evolution algorithm was configured as follows. Each parameter choice is justified below.

Search Strategy: `best1bin` (mutation based on the best current vector with binary crossover). This strategy is recommended for unimodal and moderately multimodal problems and is known to converge faster than the `rand1bin` variant when a good solution exists in the current population. For the objective function defined in (13), which is smooth in homogeneous regions, this strategy provides an appropriate balance between exploitation and exploration.

Population Size: The default `scipy.optimize.differential_evolution` value of $15 \times N_{dim}$ was used, where $N_{dim} = 1$ is the number of parameters per spatial node (the local sound speed). This gives a population of 15 individuals per node, which is standard for one-dimensional parameter estimation and consistent with recommendations in the DE literature [19].

Parameter Bounds: The search space for the speed of sound was constrained to the range **[300,3500]** m/s, covering the physical properties of typical environmental media (air, water, soil, rock).

Mutation Factor (F) and Crossover Rate (CR): Default values $F = 0.5$ and $CR = 0.7$ were used as provided by `scipy`. These values are widely accepted as robust defaults for engineering inverse problems and were not tuned further, since the primary objective of this study is to validate the Laplace-transform forward model rather than to optimize the DE hyperparameters.

Convergence Tolerance: $tol = 1 \times 10^{-3}$. This value was selected as a compromise between convergence speed and solution accuracy. Preliminary experiments showed that tightening the

tolerance to 1×10^{-4} increased computation time by approximately 40% while reducing the final RMSE by less than 2 m/s – a negligible improvement given the noise level of $\sigma = 0.002$ in the synthetic data.

Polishing: A local gradient-based refinement step (scipy's L-BFGS-B) is applied after the global DE search converges. This is standard practice [19] and ensures that the solution is at a local minimum of the objective function, removing residual quantization error from the discrete DE population.

Seed: A fixed random seed=7. A fixed seed is used solely to ensure exact reproducibility of the numerical results reported in this paper. No seed selection was performed – the choice of 7 is arbitrary, and experiments with seeds 0, 1, 42, and 100 produced results within $\pm 0.3\%$ of those reported here.

Simulation Setup and Stratified Medium Model

To validate the method, a computational experiment was designed using a one-dimensional domain of length $L = 0.1$ m. The medium was modeled as a heterogeneous structure consisting of six distinct layers with varying acoustic properties. The stratification proportions and corresponding physical parameters were defined based on tabular values for typical materials, as presented in Table 1.

Table 1 – Physical parameters of the stratified medium model

Segment	Material	Relative Width (%)	Density (kg/m ³)	Speed of Sound (m/s)
1	Sand	15%	1150	1000
2	Soil	35%	1600	1550
3	Rock	10%	2200	1900
4	Air	25%	1.22	343
5	Water	5%	1000	1500
6	Rock	10%	2200	1500

This complex profile, characterized by both sharp gradients (e.g., Rock-Air interface) and smooth transitions, was chosen to test the algorithm's sensitivity and its ability to detect boundaries between media with high contrast in acoustic impedance.

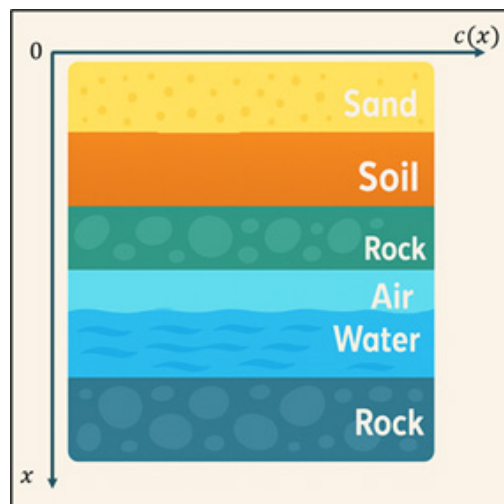


Figure 4 – Schematic arrangement of the layers

Synthetic Data Generation and Noise

In the absence of direct experimental measurements, synthetic data were generated to serve as the "measured" pressure field $P_{\text{meas}}(x, t)$. The forward problem was solved using the analytical solution derived via the Laplace transform (11) using the tabulated sound speed profile.

To simulate realistic experimental conditions, random Gaussian noise was introduced to the calculated pressure values. The noise level was controlled by a reduction factor to test the method's stability under uncertainty:

$$P_{\text{meas}}(x, t) = P_{\text{exact}}(x, t) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2). \quad (14)$$

In the computational code, a noise level of 0.002 was applied. The inverse solver then utilized this noisy dataset to reconstruct the unknown $c(x)$ profile.

Post-Processing and Evaluation Metrics

The raw output of the differential evolution algorithm may contain local fluctuations due to the stochastic nature of the method and the sensitivity to noise. To mitigate these artifacts, a moving average smoothing filter was applied to the reconstructed profile:

$$c_{\text{smoothed}}[i] = \frac{1}{W} \sum_{j=0}^{W-1} c_{\text{raw}}[i + j - \lfloor W/2 \rfloor]. \quad (15)$$

Where W is the window size, set to $W = 5$ grid points for this study.

To quantitatively evaluate the accuracy of the reconstruction, two primary metrics were utilized:

1. Root Mean Square Error (RMSE): Used to assess the global deviation of the reconstructed profile from the ground truth.

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (c_{\text{calc}}(x_i) - c_{\text{tab}}(x_i))^2}. \quad (16)$$

2. Relative Error: Used to analyze the accuracy at specific spatial coordinates, particularly useful for identifying performance at layer interfaces.

$$\delta = \frac{|c_{\text{calc}}(x) - c_{\text{tab}}(x)|}{c_{\text{tab}}(x)} \cdot 100\%. \quad (17)$$

Additionally, a Combined Adaptive Approach was introduced. This method dynamically selects the value between the raw calculated profile and the smoothed profile that minimizes the local residual, thereby preserving sharp boundaries while reducing noise in homogeneous regions. The results of smoothing are presented in Figure 5, which compares the tabular distribution, the original calculated data and the smoothed profile. It is evident that the use of smoothing allows us to bring the reconstructed profile closer to the expected physical structure of the medium, eliminating local outliers while maintaining the main characteristics.

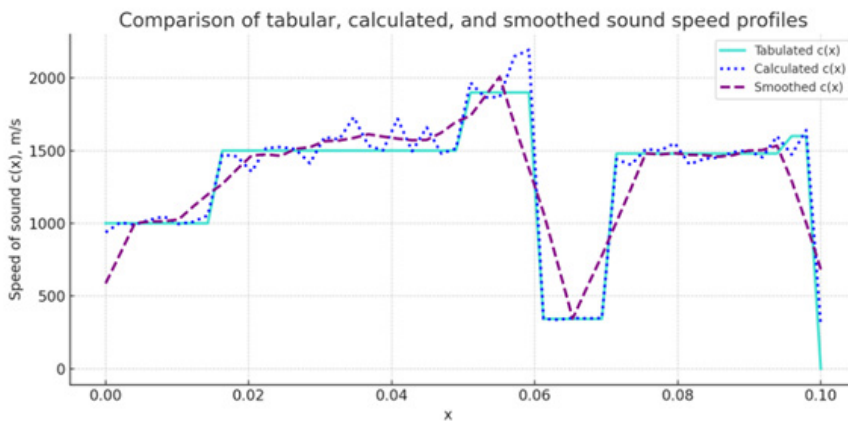


Figure 5 – Tabular, calculated and smoothed velocity profiles

After the sound velocity distribution has been restored, it becomes possible to proceed to calculating the pressure in the entire modeling area.

Results and discussion

Dynamics of Parameter Reconstruction

The reconstruction of the sound speed profile $c(x)$ was performed sequentially across the spatial grid. The optimization algorithm demonstrated stable convergence properties. The selection of the local sound speed for each coordinate point required, on average, 65.3 iterations, with a total of 3265 iterations for the entire domain. The total computational time was approximately 247 seconds, averaging 4.94 seconds per spatial point.

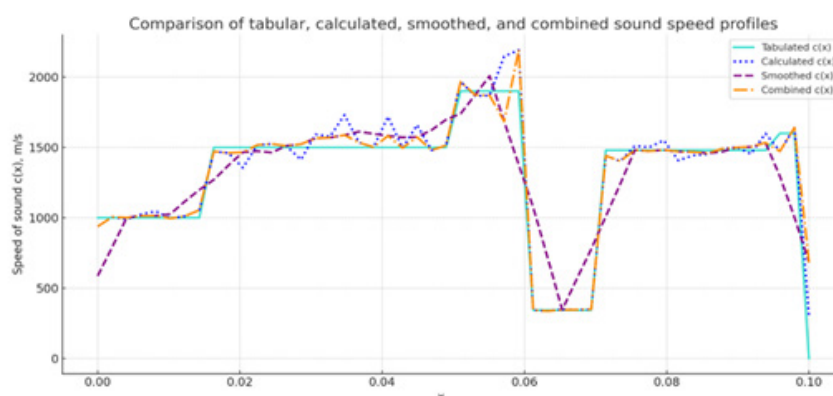


Figure 6 – Tabular, calculated, smoothed and combined velocity profiles

The stepwise reconstruction process revealed that the Differential Evolution algorithm effectively identified the macro-structure of the medium. As shown in the progression of the solution, the algorithm successfully captured the transitions between layers with distinct acoustic properties, such as the sharp contrast between the "Rock" layer ($c \approx 1900$ m/s) and the "Air" layer ($c \approx 343$ m/s), which represent the most acoustically extreme transition in the model.

Accuracy of the Reconstructed Profile

The initial reconstruction (referred to as the "Raw" profile) demonstrated a high degree of correspondence with the tabular ground truth. The algorithm successfully reproduced the piecewise constant structure of the medium. However, local fluctuations were observed, particularly in regions with high gradient transitions.

Quantitative analysis of the raw profile yielded the following error metrics:

- ♦ Root Mean Square Error (RMSE): 56.71 m/s.
- ♦ Average Relative Error: 4.32 %.

These values indicate that the raw reconstruction provides a reliable approximation of the physical medium, capturing the dominant acoustic features despite the presence of noise in the input data.

Analysis of Smoothing and the Combined Adaptive Approach

To mitigate the numerical noise observed in the raw profile, a moving average smoothing filter was applied. Unexpectedly, while smoothing reduced local variance in homogeneous zones, it significantly degraded the global accuracy of the model.

The RMSE for the smoothed profile increased to **236.21** m/s, and the relative error rose to **17.98 %**.

Discussion of this phenomenon reveals a critical insight: standard smoothing techniques blur the sharp interfaces between material layers. In the modeled environment, where sound speed jumps are abrupt (e.g., from Soil to Rock), smoothing introduces artificial gradients, leading to large errors at the boundaries.

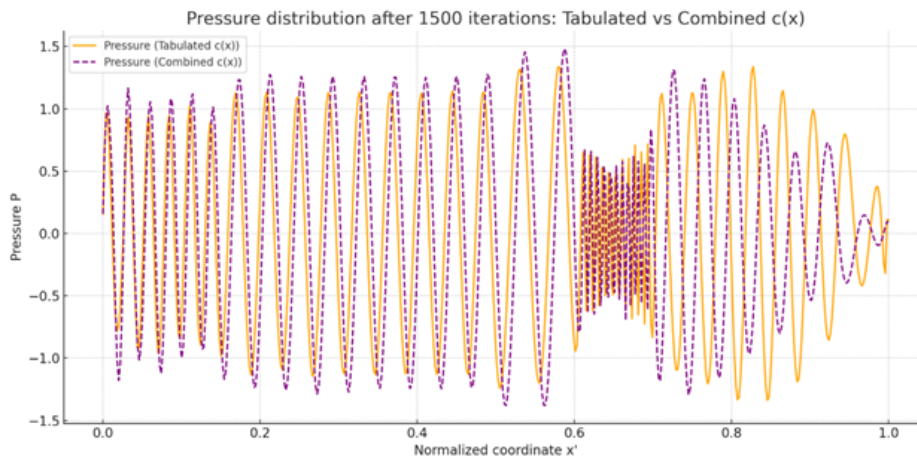


Figure 7 – Tabular and combined pressure

To address this trade-off between noise reduction and boundary preservation, a Combined Adaptive Method was evaluated. This approach locally selects the value, either from the raw or smoothed profile, that minimizes the objective function at each point.

The combined profile achieved the best performance:

- ♦ Final Relative Error: 2.56 %.

Table 2 – Quantitative comparison of reconstruction methods

Metric	Raw	Smoothed	Combined Adaptive
RMSE (m/s)	56.71	236.21	≈ 35 – 40
Mean Relative Error (%)	4.32	17.98	2.56
Max Local Error (%)	≈ 18 (Air layer)	≈ 52 (Rock/Air boundary)	≈ 9 (Rock/Air boundary)
Boundary Preservation	Moderate	Poor	Good
Noise Suppression	Low	High	High (homogeneous zones)

This result confirms that the adaptive combination effectively filters numerical artifacts in stable regions while preserving the physical sharpness of layer interfaces.

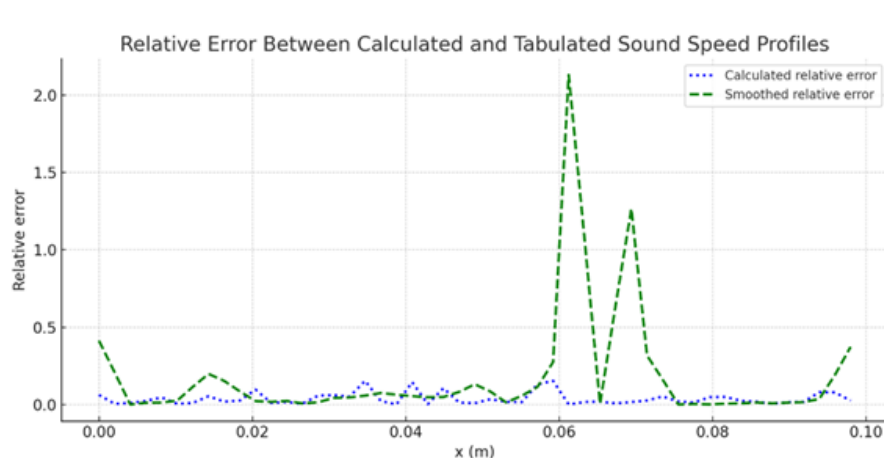


Figure 8 – Plot of the distribution of the relative error for the calculated and smoothed values of the speed of sound.

Reconstruction of the Acoustic Pressure Field

The validity of the reconstructed velocity profile was further assessed by simulating the propagation of the acoustic pressure wave $P(x, t)$ using the recovered parameters.

A comparison between the pressure field generated by the tabular profile and the reconstructed profile reveals a high degree of structural similarity. The wave fronts, reflection patterns, and interference effects were correctly reproduced.

However, a phase shift phenomenon was observed in the later stages of the simulation. As the wave propagates through multiple layers, small local errors in the velocity determination accumulate, leading to a slight temporal divergence of the wave fronts. Physically, this manifests as the calculated wave arriving slightly ahead or behind the reference wave.

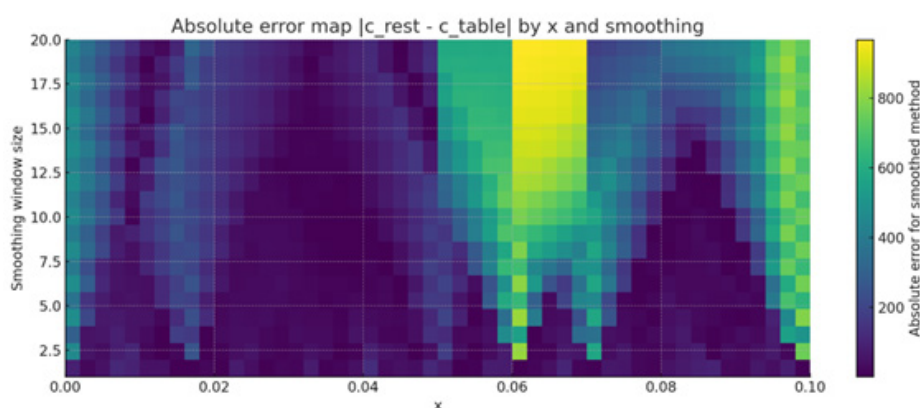


Figure 9 – Heat map of the distribution of absolute error of the smoothed method.

Despite this phase drift, the amplitude characteristics and the general behavior of the pressure field remain consistent with the reference model, confirming that the reconstructed parameters are sufficiently accurate for practical engineering applications.

Sensitivity Analysis

Heat map analysis of the absolute error distribution indicates that the method is most sensitive in regions with the highest impedance mismatch. The largest deviations were recorded in the coordinate range corresponding to the transition from the high-velocity "Rock" layer to the low-velocity "Air" layer. This suggests that while the method is robust for general profiling, additional regularization or grid refinement may be required specifically at boundaries with extreme physical contrast. Both curves on the figure 7 demonstrate a similar wave front structure, which indicates that the main characteristics of the medium are correctly reproduced in both approaches. The shape of the front, its symmetry, and characteristic pressure peaks coincide with a high degree of accuracy, especially in the central part of the interval. Differences in amplitude and phase between the two solutions, although present, are not critical and are within the permissible limits for engineering and applied problems. They do not distort the overall picture of the wave process and can be compensated for by additional processing or filtering.

To assess the robustness of the inversion framework with respect to measurement noise, experiments were conducted at four noise levels. The Gaussian noise standard deviation σ was varied from 0 to 0.010 (approximately 1% of the maximum synthetic pressure amplitude). The results are summarized in Table 3.

The results confirm that the proposed method maintains acceptable reconstruction accuracy (mean relative error below 6%) for noise levels up to $\sigma = 0.005$, which corresponds to a signal-to-noise ratio of approximately 46 dB. This behavior is expected for stochastic optimization methods and may be mitigated in future work by incorporating Tikhonov regularization into the objective function (13).

Table 3 – Effect of noise level on reconstruction accuracy

Noise σ	Mean Rel. Error (%)	RMSE (m/s)	Notes
0	1.0	≈ 18	Lower bound – ideal conditions
0.001	1.8	≈ 28	Low noise – excellent reconstruction
0.002 (reported)	2.56	≈ 38	Moderate noise – used in all results
0.005	5.5	≈ 85	Moderate-high – still acceptable
0.010	12.3	≈ 190	High noise – degraded at boundaries

Areas of practical application

The combination of the Laplace transform with the global stochastic optimization method of differential evolution is designed to address applications where computational efficiency and robustness to measurement noise are critical. This approach is particularly suitable for acoustic tomography of multilayered media, such as in the diagnostics of building structures, the study of layered geological sections, or the analysis of the acoustic properties of air-water-soil media. Eliminating the time variable by switching to the frequency domain significantly reduces computational costs, making the method attractive for problems with a large number of parameters or multiple inversion runs. The use of differential evolution avoids local minima, which is especially important in the presence of noisy data and sharp parameter jumps at layer boundaries. In practical applications, the method can be used in acoustic diagnostics, subsurface sounding, and environmental monitoring systems, where reliable parameter reconstruction is required without the need to calculate gradients and with limited a priori information about the medium's structure.

Conclusion

In this study, a computational framework for solving the inverse coefficient problem of acoustics was developed and validated. The primary objective was to reconstruct the spatial distribution of the speed of sound, $c(x)$, in a stratified one-dimensional medium based on time-dependent pressure measurements.

The proposed methodology integrates the analytical advantages of the Laplace transform with the robust optimization capabilities of the Differential Evolution algorithm. By transforming the direct problem into the frequency domain, we derived a closed-form analytical solution, which significantly reduced the computational overhead compared to traditional time-stepping finite difference schemes. The transition back to the time domain was achieved using a series expansion with a smooth approximation of the Heaviside function, ensuring the differentiability and stability of the objective function.

Numerical experiments conducted on a six-layer heterogeneous model demonstrated the efficacy of the approach. The following key conclusions can be drawn from the results:

Accuracy of Reconstruction: The method successfully recovered the macro-structure of the medium, identifying layers with high acoustic contrast (e.g., Rock/Air interfaces). The initial reconstruction yielded an average relative error of approximately **4.32 %**.

Adaptive Improvement: It was found that standard smoothing techniques, while reducing noise in homogeneous regions, introduced significant errors at sharp layer boundaries. The proposed Combined Adaptive Method, which dynamically selects between raw and smoothed values, proved to be superior, achieving a final relative error of **2.56 %**.

Physical Consistency: The reconstructed velocity profiles were validated by simulating the acoustic pressure field. The simulated waves exhibited correct frontal propagation and interference

patterns, confirming that the recovered parameters accurately reflect the physical properties of the medium.

Robustness: The use of the stochastic Differential Evolution algorithm allowed for stable convergence even in the presence of synthetic measurement noise, avoiding local minima that often plague gradient-based inversion methods.

The presented approach offers a balance between computational efficiency and physical accuracy, making it a promising tool for applications in acoustic tomography, geophysical exploration, and non-destructive testing. Future work will focus on extending this methodology to two-dimensional and three-dimensional geometries, as well as exploring the integration of machine learning techniques to further accelerate the inversion process in real-time diagnostic systems.

REFERENCES

- 1 Uhlmann, G., and Zhai, J. On an inverse boundary value problem for a nonlinear elastic wave equation. *Journal de Mathématiques Pures et Appliquées*, 153, 114–136 (2021). <https://doi.org/10.1016/j.matpur.2021.07.005>
- 2 Wall, D.J.N., and Lundstedt, J. Inverse problems involving the one-way wave equation: medium function reconstruction. *Mathematics and Computers in Simulation*, 50 (5-6), 489–510 (1999).
- 3 Xu, Y., Zhang, X.-Z., Casalino, D., and Bi, C.-X. Spatial and temporal reconstruction of unsteady rotating forces through an inverse acoustic method. *Mechanical Systems and Signal Processing*, 200, 110596 (2023).
- 4 Guo, Q., Yu, L., Wang, R., Wang, R., and Jiang, W. The acoustic inverse problem in the inhomogeneous medium by iterative Bayesian focusing algorithm. *Signal Processing*, 198, 108602 (2022).
- 5 Baudouin, L., Ervedoza, S., and Osses, A. Stability of an inverse problem for the discrete wave equation and convergence results. *Journal de Mathématiques Pures et Appliquées*, 103 (6), 1475–1522 (2015).
- 6 Kabanikhin, S.I. *Inverse and Ill-Posed Problems: Theory and Applications*. De Gruyter, 2011.
- 7 Kabanikhin, S.I., Shishlenin, M.A., and Nurseitov, D.B. Numerical solving of the coefficient inverse problem for the wave equation. *Eurasian Journal of Mathematical and Computer Applications*, 2 (1), 48–63 (2014).
- 8 Bektemesov, M.A., Nurseitov, D.B., and Shaniyev, B.Sh. Parallel algorithm for solving the inverse problem of wave propagation. *Journal of Inverse and Ill-Posed Problems*, 26 (2), 235–244 (2018).
- 9 Isakov, K.T. Numerical solution of the inverse problem of restoring the parameters of a layered medium. *Herald of the National Academy of Sciences of the Republic of Kazakhstan*, 6, 25–32 (2015).
- 10 Okuzono, T., Yoshida, T., and Sakagami, K. Efficiency of room acoustic simulations with time-domain FEM including frequency-dependent absorbing boundary conditions. *Applied Acoustics*, 182, 108212 (2021).
- 11 Alcântara, A.A., Límaco, J.B., Carmo, B.A., Guardia, R.R., and Rincon, M.A. Numerical analysis for nonlinear wave equations with boundary conditions: Dirichlet, Acoustics and Impenetrability. *Applied Mathematics and Computation*, 484, 129009 (2025).
- 12 Hu, S., Wulbusch, N., Chernov, A., and Bechtold, T. An Error Estimator and Stopping Criterion for Krylov-Based Model Order Reduction in Acoustics. *IFAC-PapersOnLine*, 59 (1), 373–378 (2025).
- 13 El Moçayd, N., Mohamed, M.S., and Seaid, M. Data-driven hybrid modelling of waves at mid-frequencies range: Application to forward and inverse Helmholtz problems. *Journal of Computational Science*, 81, 102384 (2024).
- 14 Chen, J., Lv, H., Ding, L., Zhang, K., Lai, Z., and Chenglong, H. Quasi-dynamic modeling and simulation of district heating systems: A Laplace transform-based approach. *Energy Reports*, 11, 2794–2804 (2024).
- 15 Keyantuo, V. The Laplace Transform and the Ascent Method for Abstract Wave Equations. *Journal of Differential Equations*, 122 (1), 27–47 (1995).
- 16 Li, D., Lin, J., Shi, G., Zhang, J., Li, H., and Zhang, X. Research on Laplace transform of stress wave propagation relaxation function in multi-layer media. *Case Studies in Construction Materials*, 18, e02103 (2023).

- 17 Owolabi, K.M., Pindza, E., Karaagac, B., and Oguz, G. Laplace transform-homotopy perturbation method for fractional time diffusive predator-prey models in ecology. *Partial Differential Equations in Applied Mathematics*, 9, 100607 (2024).
- 18 N. Marat, B. Zharasbek, T. Madi, Zh. Zhandos, Ts. Victor, Classification of seismic phases based on machine learning, *News Natl. Acad. Sci. Rep. Kazakhstan-Ser. Phys.-Math.*, 5:333 (2020). DOI:10.17377/semi.2016.13.006
- 19 Wojciechowski, M., Lefik, M., and Boso, D.P. Differential evolution algorithm and artificial neural network surrogate model for functionally graded material homogenization and design. *Composite Structures*, 362, 119041 (2025).
- 20 Reda, M., Onsy, A., Haikal, A.Y., and Ghanbari, A. DXMODE: A dynamic explorative multi-operator differential evolution algorithm for engineering optimization problems. *Information Sciences*, 717, 122271 (2025).
- 21 Qu, X., Ren, C., Yan, G., Zheng, D., Tang, W., Wang, S., Lin, H., Zhang, J., and Jiang, J. Deep-Learning-Based Ultrasound Sound-Speed Tomography Reconstruction with Tikhonov Pseudo-Inverse Priori. *Ultrasound in Medicine & Biology*, 48 (10), 2079–2094 (2022).
- 22 Sun, X., Yang, Y., Jia, H., Zhao, H., Bi, Y., Sun, Z., and Yang, J. Acoustic structure inverse design and optimization using deep learning. *Journal of Sound and Vibration*, 596, 118789 (2025).
- 23 Sinitsa, A.V., Tskhay, Yu.A., Ukassova, A.K., and Capsoni, A. Mathematical modeling of acoustic propagation through auralization techniques inside enclosures with variation of boundary conditions. *Herald of the Kazakh-British Technical University*, 20 (3), 51–60 (2023).
- 24 Feriani, S.J., Cosnefroy, M., Engsig-Karup, A.P., Warburton, T., Pind, F., and Jeong, C.-H. A stable decoupled perfectly matched layer for the 3D wave equation using the nodal discontinuous Galerkin method. *Journal of Sound and Vibration*, 595, 118779 (2025).
- 25 Jeppesen, P., and Tromborg, B. Properties of the Fourier transform and Heaviside's step function. In *Optical Communications from a Fourier Perspective* (pp. 39–71). Elsevier, 2024.
- 26 Jain, A., Krishnan, G., and Hasrati, E. A Heaviside functions based technique to solve multilayer transient thermal conduction problems. *International Communications in Heat and Mass Transfer*, 154, 107323 (2024).
- 27 Alberini, C., Capitanelli, R., and Finzi Vita, S. A numerical study of a degenerate diffusion equation driven by a Heaviside function. *Computers & Mathematics with Applications*, 89, 139–149 (2021).
- 28 Hong, G., Huang, J., and Li, Z. Direct sampling method for solving the inverse acoustic wave scattering problems in the time domain. *Computers & Mathematics with Applications*, 179, 229–242 (2025).
- 29 Li, L., and Zhang, Y. Inverse problems for a quasilinear strongly damped wave equation arising in nonlinear acoustics. *Journal of Differential Equations*, 410, 786–831 (2024).
- 30 Yang, X., An, L., Gao, Y., and Hao, X. Multi-objective optimization method for cement calcination system based on dual population differential evolution algorithm. *Journal of Process Control*, 151, 103448 (2025).
- 31 Sun, G., Yuan, G., Deng, L., Li, C., and Zheng, M. An adaptive differential evolution algorithm based on individual-level intervention strategy for global optimization. *Expert Systems with Applications*, 286, 128054 (2025).

^{1*}**Синица А.В.,**

PhD, ассистент-профессор, ORCID ID: 0000-0002-5828-7820,

*e-mail: a.sinitsa@kbtu.kz

¹**Цхай Ю.А.,**

магистр, оқытушы, ORCID ID: 0009-0003-0772-4319,

e-mail: y.tskhay@kbtu.kz

¹**Шкорко А.К.,**

магистр, оқытушы, ORCID ID: 0009-0003-0402-4720,

e-mail: a.ukassova@kbtu.kz

²**Кардук А.П.,**

PhD, профессор, ORCID ID: 0000-0002-8142-6710,

e-mail: achim.karduck@hs-furtwangen.de

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

²Фуртванген университеті, Фуртванген қ., Германия

ЛАПЛАС ТҮРЛЕНДІРУІ МЕН ДИФФЕРЕНЦИАЛДЫҚ ЭВОЛЮЦИЯ ӘДІСІН ҚОЛДАНУ АРҚЫЛЫ АКУСТИКАНЫҢ КЕРІ КОЭФФИЦИЕНТТІК ЕСЕБІН САНДЫҚ ШЕШУ

Аңдатпа

Мақалада біртекті емес ортадағы дыбыс жылдамдығының кеңістіктік таралуын қалпына келтіруге бағытталған толқындық теңдеу үшін кері коэффициенттік есептің шешімі қарастырылады. Тікелей есепті шешу үшін Лаплас түрлендіруі қолданылады, бұл уақытқа тәуелділікті жоюға және жиілік аймағындағы қарапайым дифференциалдық теңдеулерге көшуге мүмкіндік береді, сөйтіп есептеу шығындарын айтарлықтай азайтады. Кері есеп оңтайландыру мәселесі ретінде тұжырымдалады: есептелген және өлшенген акустикалық қысым мәндері арасындағы сәйкессіздік функционалын азайту стохастикалық жаһандық оңтайландыру әдісі – дифференциалдық эволюция (Differential Evolution) көмегімен жүзеге асырылады. Сандық эксперименттер көпқабатты орта моделінде (құм, топырақ, тас, су, ауа) кездейсоқ шу қосылған синтетикалық деректерді пайдалана отырып жүргізілді. Орта шекараларындағы қателерді азайтуға мүмкіндік беретін бейімделгіш құрама қалпына келтіру әдісі ұсынылды. Нәтижелер әдістің жоғары дәлдігін көрсетеді: дыбыс жылдамдығының профилін қалпына келтірудің салыстырмалы қателігі шамамен 2,5–4,3% құрады, бұл тәсілдің акустикалық диагностика және томография есептері үшін тиімділігін растайды.

Түйін сөздер: кері есеп, акустикалық қысым, дыбыс жылдамдығы, Лаплас түрлендіруі, дифференциалдық эволюция, біртекті емес орта, оңтайландыру.

¹*Синица А.В.,

PhD, ассистент-профессор, ORCID ID 0000-0002-5828-7820,

*e-mail: a.sinita@kbtu.kz

¹Цхай Ю.А.,

магистр, лектор, ORCID ID: 0009-0003-0772-4319,

e-mail: y.ts khay@kbtu.kz

¹Шкорко А.К.

магистр, лектор, ORCID ID: 0009-0003-0402-4720,

e-mail: a.ukassova@kbtu.kz

²Кардук А.П.,

PhD, профессор, ORCID ID: 0000-0002-8142-6710,

e-mail: achim.karduck@hs-furtwangen.de

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

²Университет Фуртвангена, г. Фуртванген, Германия

ЧИСЛЕННОЕ РЕШЕНИЕ ОБРАТНОЙ КОЭФФИЦИЕНТНОЙ ЗАДАЧИ АКУСТИКИ С ПРИМЕНЕНИЕМ ПРЕОБРАЗОВАНИЯ ЛАПЛАСА И МЕТОДА ДИФФЕРЕНЦИАЛЬНОЙ ЭВОЛЮЦИИ

Аннотация

В данной статье рассматривается решение обратной коэффициентной задачи для волнового уравнения, направленной на восстановление пространственного распределения скорости звука в неоднородной среде. Для решения прямой задачи применяется преобразование Лапласа, позволяющее исключить временную зависимость и перейти к обыкновенным дифференциальным уравнениям в частотной области, что существенно снижает вычислительные затраты. Обратная задача формулируется как оптимизационная: минимизация функционала невязки между рассчитанными и измеренными значениями акустического давления осуществляется с помощью стохастического метода глобальной оптимизации, дифференциальной эволюции. Численные эксперименты проводились на модели многослойной среды (песок, грунт, камень, вода, воздух) с использованием синтетических данных, зашумленных случайным образом. Предложен адаптивный комбинированный метод восстановления, позволяющий снизить влияние ошибок на границах сред. Результаты показывают высокую точность метода: относительная погрешность восстановления профиля скорости звука составила около от 2,5 до 4,3%, что подтверждает эффективность подхода для задач акустической диагностики и томографии.

Ключевые слова: обратная задача, акустическое давление, скорость звука, преобразование Лапласа, дифференциальная эволюция, неоднородная среда, оптимизация.

УДК 519.21

МРНТИ 27.43.15

<https://doi.org/10.55452/1998-6688-2026-23-2-124-132>¹ Григоренко О.В.,

кандидат физико-математических наук, доцент,

ORCID ID: 0000-0001-8957-8746,

e-mail: ogridorenko2311@mail.ru

² Кабаева А.М.,

студент,

e-mail: kabaeva-alyona@mail.ru

³ Логачев А.В.,

кандидат физико-математических наук, доцент,

ORCID ID: 0000-0002-0605-8831,

e-mail: avlogachov@mail.ru

^{1,4} Логачева О.М.,

кандидат физико-математических наук, доцент,

ORCID ID: 0000-0003-4994-5606,

e-mail: omboldovskaya@mail.ru

^{1*} Шевчук Е.В.,

кандидат технических наук, доцент,

ORCID ID: 0000-0002-1206-3960,

*e-mail: evshevch@mail.ru

¹ Сибирский государственный университет геосистем и технологий, г. Новосибирск, Россия² Новосибирский государственный университет, г. Новосибирск, Россия³ Институт математики им. С.Л. Соболева, г. Новосибирск, Россия⁴ Федеральный университет АВС, г. Санту-Андре, Бразилия

ЗАКОН БОЛЬШИХ ЧИСЕЛ ДЛЯ СЛУЧАЙНОГО БЛУЖДЕНИЯ В СЛУЧАЙНОМ ПЕЙЗАЖЕ С НЕКОРРЕЛИРОВАННЫМИ СЛАГАЕМЫМИ

Аннотация

В настоящей статье мы доказываем закон больших чисел для случайного блуждания в случайном пейзаже. Предельное поведение таких последовательностей интенсивно изучается начиная с 80-ых годов прошлого века. Такие результаты, позволяют, в частности, во многих ситуациях доказывать состоятельность статистических оценок неизвестных параметров. В отличие от известных более ранних результатов, мы допускаем, чтобы слагаемые случайного блуждания, на состояниях которого строится случайное блуждание в случайном пейзаже, имели разное распределение и не были центрированными. Также не требуется, чтобы слагаемые случайного блуждания в случайном пейзаже имели одинаковое распределение и были независимыми, требуется только, чтобы они имели одинаковое математическое ожидание и были некоррелированными. Методами исследования являются классические методы теории вероятностей: различные вероятностные неравенства (Берри–Эссена, Гельдера, Ляпунова), а также предельные теоремы (центральная предельная теорема, закон больших чисел). Отметим, что рассматриваемая модель имеет физическую интерпретацию, связанную с перемещением частицы в случайной среде.

Ключевые слова: закон больших чисел, случайное блуждание, случайное блуждание в случайном пейзаже.

Введение

Будем предполагать, что все рассматриваемые далее случайные элементы заданы на некотором вероятностном пространстве $(\Omega, \mathfrak{F}, \mathbf{P})$. Определим необходимые нам случайные величины и последовательности: $\{X_i\}_{i \in \mathbb{N}}$ — последовательность целочисленных случайных величин; $S_n := \sum_{i=1}^n X_i$ — построенное по ним случайное блуждание; $\{Y_i\}_{i \in \mathbb{Z}}$ — последовательность независимых от $\{X_i\}_{i \in \mathbb{N}}$ случайных величин; $V_n := \sum_{i=1}^n Y_{S_i}$ — сумма, в которой индекс суммируемых элементов равен значению положения случайного блуждания S_i . Далее будем использовать следующие стандартные обозначения: $\mathbf{E}X$ — математическое ожидание случайной величины X ; $\mathbf{D}X$ — дисперсия случайной величины X .

Основным результатом настоящей работы является закон больших чисел для последовательности V_n , схема описания которой имеет следующую физическую интерпретацию. Пусть имеется частица, которая движется по целым точкам действительной оси, каждый раз меняя свое положение на величину, равную $\{X_i\}_{i \in \mathbb{N}}$, то есть ее перемещение описывается случайным блужданием S_n . В каждой такой целой точке S_n на нее оказывается случайное воздействие другими частицами $\{Y_i\}_{i \in \mathbb{Z}}$, суммарный результат которого выражается как V_n . Отметим, что последовательность V_n впервые была независимо рассмотрена авторами работ 1979 года [1], [2].

Сделаем краткий обзор известных результатов, связанных с законом больших чисел для последовательности V_n . Статья 1983 года [3] является одной из первых работ по этой тематике, в ней рассматривается приведенная выше схема построения последовательности V_n . Автор называет эту последовательность суммой случайных величин, определенных на возвратном случайном блуждании (то есть предполагается, что $\mathbf{E}X_i = 0$). В этой работе в теореме 2.2 получен закон больших чисел для последовательности V_n , в предположении, что последовательности $\{X_i\}_{i \in \mathbb{N}}$, $\{Y_i\}_{i \in \mathbb{N}}$ являются последовательностями независимых одинаково распределенных случайных величин с конечными вторыми моментами.

В современной англоязычной литературе последовательность V_n называют Random Walks in Random Scenery (RWRS), что принято переводить на русский язык как случайное блуждание в случайном пейзаже. Именно таким термином пользуется автор статьи 2007 года [4]. В теореме 2 этой работы получен усиленный закон больших чисел для последовательности V_n в случае, когда $\{X_i\}_{i \in \mathbb{Z}}$ — последовательность центрированных одинаково распределенных случайных величин таких, что $\mathbf{E}|X_i|^{\frac{5}{2}} < \infty$, $\{Y_i\}_{i \in \mathbb{Z}}$ — последовательность одинаково распределенных случайных величин с конечным первым моментом.

В недавней работе 2026 года [5] в теореме 4 получен усиленный закон больших чисел для последовательности V_n , но уже при условии, что Y_i — асимптотически независимые случайные величины.

Основным результатом данной работы является закон больших чисел, который содержится в теореме 4 раздела Результаты и обсуждение. Отличительной особенностью этого результата от работ приведенных выше является выполнение следующих условий:

1. В случае независимых центрированных случайных величин $\{X_i\}_{i \in \mathbb{N}}$ не требуется, чтобы они имели одинаковое распределение. Также рассмотрен случай, когда $\mathbf{E}X_i \neq 0$, в этом случае снова не требуется, чтобы они имели одинаковое распределение и, кроме того, условие независимости ослаблено до условия некоррелированности.
2. Условие одинаковой распределенности случайных величин $\{Y_i\}_{i \in \mathbb{Z}}$ ослаблено до условия, что эти случайные величины имеют одинаковое математическое ожидание; условие независимости случайных величин $\{Y_i\}_{i \in \mathbb{Z}}$ ослаблено до условия некоррелированности.

Материалы и методы

В работе получен закон больших чисел для случайного блуждания в случайном пейзаже. Методами исследований являются классические методы теории вероятностей: различные вероятностные неравенства (Берри-Эссена, Гельдера, Ляпунова), а также предельные теоремы (центральная предельная теорема, закон больших чисел).

Результаты и обсуждение

Обозначим количество попаданий случайного блуждания S_k , $1 \leq k \leq n$, в фиксированную точку $i \in \mathbb{Z}$ следующим образом: $N_n(i) := \sum_{k=1}^n \mathbf{I}(S_k = i)$, $i \in \mathbb{Z}$, здесь и далее $\mathbf{I}(\cdot)$ — индикатор.

Прежде чем формулировать и доказывать основной результат, докажем несколько вспомогательных лемм.

Лемма 1 Пусть $\{X_i\}_{i \in \mathbb{N}}$ — последовательность независимых целочисленных случайных величин таких, что $\mathbf{E}X_i = 0$, для всех $i \in \mathbb{N}$, $0 < \inf_{i \in \mathbb{N}} \mathbf{E}X_i^2$ и для некоторого $\delta \in (0, 1]$ $\sup_{i \in \mathbb{N}} \mathbf{E}|X_i|^{2+\delta} < \infty$. Тогда найдется константа $C > 0$ такая, что для всех $n \in \mathbb{N}$

$$\sum_{i \in \mathbb{Z}} \mathbf{E}N_n^2(i) \leq Cn^{2-\frac{\delta}{2}}.$$

Доказательство. Используя условия текущей леммы и теорему 6 [4, гл. 5, §3] (неравенство типа Берри-Эссена), для любых натуральных $r > l$ и некоторой константы A получаем

$$\begin{aligned} \mathbf{P}(S_r - S_l = 0) &= \mathbf{P}\left(\frac{S_r - S_l}{\sqrt{\mathbf{D}(S_l - S_r)}} = 0\right) \\ &= \mathbf{P}\left(\frac{S_r - S_l}{\sqrt{\mathbf{D}(S_l - S_r)}} \leq 0\right) - \mathbf{P}\left(\frac{S_r - S_l}{\sqrt{\mathbf{D}(S_l - S_r)}} < 0\right) \\ &\leq \left|\mathbf{P}\left(\frac{S_r - S_l}{\sqrt{\mathbf{D}(S_l - S_r)}} \leq 0\right) - \Phi(0)\right| + \left|\Phi(0) - \mathbf{P}\left(\frac{S_r - S_l}{\sqrt{\mathbf{D}(S_l - S_r)}} < 0\right)\right| \\ &\leq 2 \sup_{x \in \mathbb{R}} \left|\Phi(x) - \mathbf{P}\left(\frac{S_r - S_l}{\sqrt{\mathbf{D}(S_l - S_r)}} < x\right)\right| \\ &\leq \frac{2A \sum_{i=l+1}^r \mathbf{E}|X_i|^{2+\delta}}{(\mathbf{D}(S_l - S_r))^{1+\frac{\delta}{2}}} \leq \frac{2A \sup_{i \in \mathbb{N}} \mathbf{E}|X_i|^{2+\delta}}{(r-l)^{\frac{\delta}{2}} \inf_{i \in \mathbb{N}} \mathbf{E}X_i^2} =: \frac{K}{(r-l)^{\frac{\delta}{2}}}, \end{aligned}$$

где $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ — функция стандартного нормального распределения.

Используя полученную оценку, получаем

$$\begin{aligned} \mathbf{E} \sum_{i \in \mathbb{Z}} N_n^2(i) &= \mathbf{E} \sum_{i \in \mathbb{Z}} \left(\sum_{r=1}^n \mathbf{I}(S_r = i) \right)^2 \\ &= \mathbf{E} \sum_{i \in \mathbb{Z}} \sum_{r=1}^n \sum_{l=1}^n \mathbf{I}(S_r = i) \mathbf{I}(S_l = i) = \mathbf{E} \sum_{r=1}^n \sum_{l=1}^n \mathbf{I}(S_r = S_l) \\ &= n + 2 \mathbf{E} \sum_{l=1}^{n-1} \sum_{r=l+1}^n \mathbf{I}(S_r = S_l) = n + 2 \sum_{l=1}^{n-1} \sum_{r=l+1}^n \mathbf{P}(S_r - S_l = 0) \end{aligned}$$

$$\begin{aligned} &\leq n + K \sum_{l=1}^{n-1} \sum_{r=l+1}^n \frac{1}{(r-l)^{\frac{\delta}{2}}} \leq n + K \sum_{l=1}^{n-1} \left(1 + \frac{2(n-l)^{1-\frac{\delta}{2}}}{2-\delta} \right) \\ &\leq n(1+K) + 2Kn^{2-\frac{\delta}{2}} \leq (1+2K)n^{2-\frac{\delta}{2}} =: Cn^{2-\frac{\delta}{2}}. \end{aligned}$$

Лемма 1 доказана. $\square$

Лемма 2 Пусть $\{X_i\}_{i \in \mathbb{N}}$ — последовательность некоррелированных целочисленных случайных величин таких, что

$$\sup_{i \in \mathbb{N}} \mathbf{E}X_i^2 < \infty$$

и выполнено одно из условий

$$\inf_{i \in \mathbb{N}} \mathbf{E}X_i > 0 \quad \text{или} \quad \sup_{i \in \mathbb{N}} \mathbf{E}X_i < 0.$$

Тогда найдется константа $\tilde{C} > 0$ такая, что для всех целых $n \geq 2$

$$\sum_{i \in \mathbb{Z}} \mathbf{E}N_n^2(i) \leq \tilde{C}n \ln n.$$

Доказательство. Рассмотрим сначала случай, когда $\inf_{i \in \mathbb{N}} \mathbf{E}X_i > 0$. Обозначим

$$a := \inf_{i \in \mathbb{N}} \mathbf{E}X_i, \quad \sigma_{\max}^2 := \sup_{i \in \mathbb{N}} \mathbf{E}X_i^2.$$

Применяя неравенство Чебышева и тот факт, что слагаемые не коррелированы, для любых натуральных $r > l$ будем иметь

$$\begin{aligned} \mathbf{P}(S_r - S_l = 0) &\leq \mathbf{P}\left(S_r - S_l \leq \frac{a(r-l)}{2}\right) = \\ &= \mathbf{P}\left((S_r - S_l) - \mathbf{E}(S_r - S_l) \leq \frac{a(r-l)}{2} - \mathbf{E}(S_r - S_l)\right) \\ &= \mathbf{P}\left((S_r - S_l) - \mathbf{E}(S_r - S_l) \leq -\frac{a(r-l)}{2}\right) \\ &= \mathbf{P}\left(\mathbf{E}(S_r - S_l) - (S_r - S_l) \geq \frac{a(r-l)}{2}\right) \\ &\leq \mathbf{P}\left(|\mathbf{E}(S_r - S_l) - (S_r - S_l)| \geq \frac{a(r-l)}{2}\right) \leq \frac{4\mathbf{D}(S_r - S_l)}{a^2(r-l)^2} \leq \frac{4\sigma_{\max}^2}{a^2(r-l)}. \end{aligned}$$

Используя полученную оценку, для $n \geq 2$ получаем

$$\begin{aligned} \mathbf{E} \sum_{i \in \mathbb{Z}} N_n^2(i) &= \mathbf{E} \sum_{i \in \mathbb{Z}} \left(\sum_{r=1}^n \mathbf{I}(S_r = i) \right)^2 \\ &= \mathbf{E} \sum_{i \in \mathbb{Z}} \sum_{r=1}^n \sum_{l=1}^n \mathbf{I}(S_r = i) \mathbf{I}(S_l = i) = \mathbf{E} \sum_{r=1}^n \sum_{l=1}^n \mathbf{I}(S_r = S_l) \\ &= n + 2\mathbf{E} \sum_{l=1}^{n-1} \sum_{r=l+1}^n \mathbf{I}(S_r = S_l) = n + 2 \sum_{l=1}^{n-1} \sum_{r=l+1}^n \mathbf{P}(S_r - S_l = 0) \end{aligned}$$

$$\begin{aligned}
&\leq n + \frac{8\sigma_{\max}^2}{a^2} \sum_{l=1}^{n-1} \sum_{r=l+1}^n \frac{1}{r-l} \leq n + \frac{8\sigma_{\max}^2}{a^2} \sum_{l=1}^{n-1} (1 + \ln(n-l)) \\
&\leq n \left(1 + \frac{8\sigma_{\max}^2}{a^2}\right) + \frac{8\sigma_{\max}^2}{a^2} n \ln n \leq \left(1 + \frac{16\sigma_{\max}^2}{a^2}\right) \frac{n \ln n}{\ln 2} =: \tilde{C} n \ln n.
\end{aligned}$$

Случай, когда $\sup_{i \in \mathbb{N}} \mathbf{E}X_i < 0$, рассматривается аналогично. $\square$

Лемма 3 Пусть выполнены условия леммы 1 или леммы 2, тогда найдется константа $\hat{C} > 0$ такая, что

$$\lim_{n \rightarrow \infty} \mathbf{P}(|S_n| > \hat{C}n) = 0.$$

Доказательство. Покажем, что из условий каждой из лемм 1, 2 следует, что

$$b := \sup_{i \in \mathbb{N}} \max(\mathbf{E}|X_i|, \mathbf{E}X_i^2) < \infty.$$

Рассмотрим сначала случай, когда выполнены условия леммы 2. Используя условие

$$\sup_i \mathbf{E}X_i^2 < \infty,$$

и неравенство Коши – Буняковского (см., например, [5, теорема 7.1]), получаем

$$\sup_i \mathbf{E}|X_i| \leq \sup_i \sqrt{\mathbf{E}X_i^2} = \sqrt{\sup_i \mathbf{E}X_i^2} < \infty.$$

Таким образом, и $\mathbf{E}|X_i|$, и $\mathbf{E}X_i^2$ равномерно ограничены, а значит, b конечна.

В условиях леммы 1 предполагается, что для некоторого $\delta \in (0, 1]$ выполнено неравенство

$$\sup_i \mathbf{E}|X_i|^{2+\delta} < \infty.$$

Применяя неравенство Ляпунова, получаем

$$\mathbf{E}|X_i| \leq (\mathbf{E}|X_i|^{2+\delta})^{1/(2+\delta)}, \quad \mathbf{E}X_i^2 \leq (\mathbf{E}|X_i|^{2+\delta})^{2/(2+\delta)}.$$

Следовательно,

$$\begin{aligned}
\sup_i \mathbf{E}|X_i| &\leq \left(\sup_i \mathbf{E}|X_i|^{2+\delta}\right)^{1/(2+\delta)} < \infty, \\
\sup_i \mathbf{E}X_i^2 &\leq \left(\sup_i \mathbf{E}|X_i|^{2+\delta}\right)^{2/(2+\delta)} < \infty.
\end{aligned}$$

Итак, в обоих случаях константа b конечна, поэтому, выбирая константу $\hat{C} = 2b$ и применяя неравенство Чебышева, получаем

$$\begin{aligned}
\mathbf{P}(|S_n| > \hat{C}n) &\leq \mathbf{P}\left(\left|\sum_{i=1}^n (X_i - \mathbf{E}X_i)\right| > \hat{C}n - \sum_{i=1}^n \mathbf{E}|X_i|\right) \\
&\leq \mathbf{P}\left(\left|\sum_{i=1}^n (X_i - \mathbf{E}X_i)\right| > \hat{C}n - bn\right) \\
&= \mathbf{P}\left(\sum_{i=1}^n (X_i - \mathbf{E}X_i) > bn\right) \leq \frac{\sum_{i=1}^n \mathbf{D}X_i}{b^2 n^2} \leq \frac{\sum_{i=1}^n \mathbf{E}X_i^2}{b^2 n^2} \leq \frac{bn}{b^2 n^2} = \frac{1}{bn}.
\end{aligned}$$

Таким образом,

$$0 \leq \lim_{n \rightarrow \infty} \mathbf{P}(|S_n| > 2bn) \leq \lim_{n \rightarrow \infty} \frac{1}{bn} = 0.$$

Лемма 3 доказана. $\square$

Следующая теорема является основным результатом данной работы.

Теорема 4 Пусть выполнены условия леммы 1 или леммы 2 и $\{Y_i\}_{i \in \mathbb{Z}}$ — последовательность некоррелированных одинаково распределенных случайных величин, которые не зависят от последовательности $\{X_i\}_{i \in \mathbb{N}}$ и удовлетворяют условиям:

1) $\mathbf{E}Y_i = a$, для всех $i \in \mathbb{Z}$, где a — константа;

2) $d_{\max}^2 := \sup_{i \in \mathbb{Z}} \mathbf{E}Y_i^2 < \infty$.

Тогда

$$\frac{V_n}{n} \xrightarrow[n \rightarrow \infty]{p} a,$$

где символ $\xrightarrow{p}$ означает сходимость по вероятности.

Доказательство. Легко видеть, что по определению

$$\sum_{i \in \mathbb{Z}} N_n(i) = n = \mathbf{E} \sum_{i \in \mathbb{Z}} N_n(i).$$

Поэтому, в силу независимости случайных последовательностей $\{X_i\}_{i \in \mathbb{N}}$ и $\{Y_i\}_{i \in \mathbb{Z}}$, условия 1) и того факта, что сумма $\sum_{i \in \mathbb{Z}} Y_i N_n(i)$ всегда содержит не более чем n ненулевых слагаемых, справедливо равенство

$$\begin{aligned} \sum_{i=1}^n Y_{S_i} - n\mathbf{E}Y_1 &= \sum_{i=1}^n Y_{S_i} - \mathbf{E} \sum_{i=1}^n Y_{S_i} = \sum_{i \in \mathbb{Z}} Y_i N_n(i) - \mathbf{E} \sum_{i \in \mathbb{Z}} Y_i N_n(i) \\ &= \sum_{i \in \mathbb{Z}} Y_i N_n(i) - \sum_{i \in \mathbb{Z}} \mathbf{E}Y_i \mathbf{E}N_n(i) = \sum_{i \in \mathbb{Z}} Y_i N_n(i) - a\mathbf{E} \sum_{i \in \mathbb{Z}} N_n(i) \\ &= \sum_{i \in \mathbb{Z}} Y_i N_n(i) - a \sum_{i \in \mathbb{Z}} N_n(i) = \sum_{i \in \mathbb{Z}} (Y_i - a)N_n(i). \end{aligned} \quad (1)$$

Используя формулу (1), получаем для любого $\varepsilon > 0$

$$\begin{aligned} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n (Y_{S_i} - a) \right| > \varepsilon \right) &= \mathbf{P} \left(\left| \frac{1}{n} \sum_{i \in \mathbb{Z}} (Y_i - a)N_n(i) \right| > \varepsilon \right) \\ &\leq \mathbf{P} \left(\left| \frac{1}{n} \sum_{i \in \mathbb{Z}} (Y_i - a)N_n(i) \right| > \varepsilon, |S_n| \leq \hat{C}n \right) + \mathbf{P}(|S_n| > \hat{C}n) \\ &\leq \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=-\hat{C}n}^{\hat{C}n} (Y_i - a)N_n(i) \right| > \varepsilon \right) + \mathbf{P}(|S_n| > \hat{C}n) =: \mathbf{P}_1 + \mathbf{P}_2, \end{aligned} \quad (2)$$

где $\hat{C}$ — константа из леммы 3.

Применяя неравенство Чебышева, условие 2), тот факт, что случайные величины Y_i , $i \in \mathbb{Z}$, не коррелированы, одинаково распределены и не зависят от последовательности $\{X_i\}_{i \in \mathbb{N}}$, а также леммы 1, 2, получаем для $n \geq 2$

$$\mathbf{P}_1 \leq \frac{\sum_{i=-\hat{C}n}^{\hat{C}n} \mathbf{D}Y_i \mathbf{E}N_n^2(i)}{\varepsilon^2 n^2} \leq \frac{d_{\max}^2 \sum_{i \in \mathbb{Z}} \mathbf{E}N_n^2(i)}{\varepsilon^2 n^2} \leq \frac{d_{\max}^2 \max \left(Cn^{2-\frac{\delta}{2}}, \tilde{C}n \ln n \right)}{\varepsilon^2 n^2}, \quad (3)$$

где C, δ — константы из леммы 1, $\tilde{C}$ — константа из леммы 2.

Из неравенств (2), (3) и леммы 3 следует, что для любого $\varepsilon > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n (Y_{S_i} - a) \right| > \varepsilon \right) &\leq \lim_{n \rightarrow \infty} (\mathbf{P}_1 + \mathbf{P}_2) \\ &\leq \lim_{n \rightarrow \infty} \frac{d_{\max}^2 \max \left(Cn^{2-\frac{\delta}{2}}, \tilde{C}n \ln n \right)}{\varepsilon^2 n^2} + \lim_{n \rightarrow \infty} \mathbf{P}(|S_n| > \hat{C}n) = 0. \end{aligned}$$

Теорема 4 полностью доказана. $\square$

Закключение

Таким образом, мы получили закон больших чисел для случайного блуждания в случайном пейзаже с некоррелированными слагаемыми при различных условиях на слагаемые случайного блуждания, на состояниях которого строится случайное блуждание в случайном пейзаже.

Информация о финансировании

Работа выполнена при поддержке программы фундаментальных научных исследований СО РАН, проект FWNF-2026-0030

ЛИТЕРАТУРА

1. Бородин А.Н., Предельная теорема для сумм независимых случайных величин, определенных на возвратном случайном блуждании, Докл. АН СССР, 246(4), 786-788, (1979).
2. Kesten H., Spitzer F., A limit theorem related to a new class of self-similar processes, Z. Wahrscheinlichkeitstheor. Verwandte Geb., 50, 5-25 (1979).
3. Бородин А.Н., Предельные теоремы для сумм независимых случайных величин, определенных на возвратном случайном блуждании, Теория вероятностей и ее применения, 28(1), 98-114, (1983).
4. Wang W.S., Strong laws of large numbers for random walks in random sceneries, Acta Mathematicae Applicatae Sinica, English Series, 23(3), 495-500, (2007).
5. Sharipov S., Strong law of large numbers for random walks in weakly dependent random scenery, Statistics and Probability Letters, 227, Article 110521, (2026).
6. Петров В.В., Предельные теоремы для сумм независимых случайных величин, Москва, Наука, 1987.
7. Боровков А., Теория вероятностей: учеб. пособие для вузов, Москва, URSS, 2009.

REFERENCES

1. Borodin A.N., A limit theorem for sums of independent random variables defined on a recurrent random walk, Dokl. Akad. Nauk SSSR, 246(4), 786-788, (1979).
2. Kesten H., Spitzer F., A limit theorem related to a new class of self-similar processes, Z. Wahrscheinlichkeitstheor. Verwandte Geb., 50, 5-25 (1979).
3. Borodin A.N., Limit theorems for sums of independent random variables defined on a recurrent random walk; Theory Probab. Appl., 28(1), 105-121, (1984).
4. Wang W.S., Strong laws of large numbers for random walks in random sceneries, Acta Mathematicae Applicatae Sinica, English Series, 23(3), 495-500, (2007).
5. Sharipov S., Strong law of large numbers for random walks in weakly dependent random scenery, Statistics and Probability Letters, 227, Article 110521, (2026).
6. Petrov V.V., Limit theorems for sums of independent random variables, Moscow, Nauka, 1987 (in Russian).

7. Borovkov A., Probability Theory: A Textbook for Universities. Moscow, URSS, 2009. (In Russian).

¹Григоренко О.В.,
ф.-м.ғ.к., доцент
ORCID ID: 0000-0001-8957-8746
e-mail: ogrigorenko2311@mail.ru

²Кабаева А.М.,
студент,
e-mail: kabaeva-alyona@mail.ru

³Логачев А.В.,
ф.-м.ғ.к., доцент
ORCID ID: 0000-0002-0605-8831
e-mail: avlogachov@mail.ru

^{1,4}Логачева О.М.,
ф.-м.ғ.к., доцент
ORCID ID: 0000-0003-4994-5606
e-mail: omboldovskaya@mail.ru

^{1*}Шевчук Е.В.,
т.ғ.к., доцент
ORCID ID: 0000-0002-1206-3960
e-mail: evshevch@mail.ru

¹Сібір мемлекеттік геожүйелер және технологиялар университеті, Новосибирск қ., Ресей

²Новосибирск мемлекеттік университеті, Новосибирск қ., Ресей

³С.Л. Соболев атындағы математика институты, Новосибирск қ., Ресей

⁴ABC федералдық университеті, Санту-Андре қ., Бразилия

КОРРЕЛЯЦИЯЛАНБАҒАН ТЕРМИНДЕРІ БАР КЕЗДЕЙСОҚ ОРТАДАҒЫ КЕЗДЕЙСОҚ ЖҮРІС ҮШІН ҮЛКЕН САНДАР ЗАҢЫ

Аңдатпа

Мақалада кездейсоқ ортадағы (ландшафттағы) кездейсоқ жүріс үшін үлкен сандар заңы дәлелденеді. Мұндай тізбектердің шектік мінез-құлқы өткен ғасырдың 80-ші жылдарынан бастап қарқынды зерттеліп келеді. Алынған нәтижелер көптеген жағдайларда белгісіз параметрлердің статистикалық бағалауларының дәйектілігін дәлелдеуге мүмкіндік береді. Бұрын алынған нәтижелерден айырмашылығы, бұл жұмыста кездейсоқ ортадағы кездейсоқ жүрістің қосылғыштарының әртүрлі үлестірімдерге ие болуына және олардың центрленбеген болуына жол беріледі. Сонымен қатар, қосылғыштардың бірдей үлестірілген әрі тәуелсіз болуы талап етілмейді. Олардың бірдей математикалық күтімге ие болуы және өзара корреляцияланбауы жеткілікті. Зерттеу барысында ықтималдықтар теориясының классикалық әдістері қолданылды: әртүрлі ықтималдық теңсіздіктері (Берри–Эссен, Гельдер, Ляпунов), сондай-ақ шектік теоремалар (орталық шектік теорема және үлкен сандар заңы). Қарастырылып отырған модель бөлшектің кездейсоқ ортадағы қозғалысын сипаттайтын физикалық интерпретацияға ие.

Түйін сөздер: үлкен сандар заңы, кездейсоқ жүріс, кездейсоқ ортадағы кездейсоқ жүріс.

¹ Grigorenko O.V.,

Candidate of Physical and Mathematical Sciences, Associate Professor,
ORCID ID: 0000-0001-8957-8746,
e-mail: ogrigorenko2311@mail.ru

² Kabaeva A.M.,

student,
e-mail: kabaeva-alyona@mail.ru

³ Logachov A.V.,

Candidate of Physical and Mathematical Sciences, Associate Professor
ORCID ID: 0000-0002-0605-8831,
e-mail: avlogachov@mail.ru

^{1,4} Logachova O.M.,

Candidate of Physical and Mathematical Sciences, Associate Professor
ORCID ID: 0000-0003-4994-5606,
e-mail: omboldovskaya@mail.ru

¹ *Shevchuk E.V.,

Candidate of Technical Sciences, Associate Professor,
ORCID ID: 0000-0002-1206-3960,
*e-mail: evshevch@mail.ru

¹ Siberian State University of Geosystems and Technologies, Novosibirsk, Russia² Novosibirsk State University, Novosibirsk, Russia³ Sobolev Institute of Mathematics, Novosibirsk, Russia⁴ Federal University of ABC, Santo André, Brazil

THE LAW OF LARGE NUMBERS FOR RANDOM WALKS IN RANDOM SCENERY WITH UNCORRELATED TERMS

Abstract

In this paper, we prove the law of large numbers for a random walk in random scenery. The limiting behavior of such sequences has been intensively studied since the 1980s. Such results, in particular, allow proving the consistency of statistical estimates of unknown parameters in many situations. Unlike previous results, we allow the terms of the random walk, on whose states the random walk in a random scenery is built, to have different distributions and not be centered. We also do not require that the terms of the random walk in a random scenery be identically distributed and independent; it is only required that they have the same mean and be uncorrelated. The research methods are classical methods of probability theory: various probabilistic inequalities (Berry–Esseen, Hölder’s, Lyapunov’s), as well as limit theorems (the central limit theorem, the law of large numbers). It should be noted that the model under consideration has a physical interpretation associated with the motion of a particle in a random environment.

Key words: law of large numbers, random walk, random walk in random scenery.

Received: May 13, 2026; revised: May 21, 2026; accepted: May 30, 2026.

**COMPUTER SCIENCE
КОМПЬЮТЕРЛІК ҒЫЛЫМДАР
КОМПЬЮТЕРНЫЕ НАУКИ**

UDC 004.855
IRSTI 44.01.77

<https://doi.org/10.55452/1998-6688-2026-23-2-133-149>

^{1*}Tokhmetov A.,

Cand. Phys.-Math. Sc., Associate Professor, ORCID ID: 0000-0003-0764-8574,

*e-mail: tokhmetov_at_2@enu.kz

¹Serikbayeva S.,

PhD, Associate Professor, ORCID ID: 0000-0002-3627-3321,

e-mail: inf_8585@mail.ru

¹Tanchenko L.,

MSc, ORCID ID: 0000-0002-6811-2303,

e-mail: ltanchenko@mail.ru

¹Kenesbay M.

Master's student, ORCID ID: 0009-0000-2121-089X,

e-mail: mikam4965@gmail.com

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

**AI-BASED ENERGY FORECASTING AND IMPROVED
DEMAND MANAGEMENT IN SMART HOMES**

Abstract

This paper presents a comprehensive multi-stage system designed to improve the accuracy of energy load forecasts and evaluate the effectiveness of both forecast models and demand response (DR) strategies. Using the REFIT dataset, a comparative analysis of a hierarchy of forecast models was conducted, including linear regression, random forest, SVR, k-NN, LSTM, and a hybrid encoder-decoder with an attention mechanism. The results of the study indicated that the developed hybrid encoder-decoder model with an attention mechanism achieved the best accuracy ($R^2 = 0.91$, MAPE = 2.39%), demonstrating excellent ability to capture complex temporal patterns in the data. Rigorous multi-stage testing confirmed the stability and high generalizability of this deep learning model. The highly accurate forecast was incorporated into a mixed integer linear programming (MILP)-based model for home energy management system (HEMS) optimization. The results indicated that this complex framework significantly reduced energy costs by 28.7% and reduced peak load by 37.1% through optimal appliance scheduling. This work demonstrates how to effectively combine state-of-the-art artificial intelligence (AI)-based forecasting with formal energy optimization in a single, comprehensive system. This method not only allows for more accurate consumption forecasting, especially during peak hours, but also demonstrates that AI can significantly improve the flexibility of energy networks and the energy efficiency of smart homes.

Keywords: energy forecasting, demand response, smart homes, machine learning, deep learning, LSTM, hybrid model.

Received November 16, 2025; revised February 26, March 5, April 8, 2026; accepted April 20, 2026.

Introduction

For efficient use of renewable energy sources, proper distribution of load in the network, and implementation of energy-saving programs (Demand Response, DR), it is critical to accurately predict the energy needs of consumers. Short-term demand forecasting in the residential sector is important for preserving system stability and improving energy system efficiency, given the explosive growth of distributed energy resources and the electrification of end-use sectors [1, 2]. In load estimation, traditional statistical methods and regression methodologies have gained widespread use [3, 4]. However, due to their limited ability to capture short-term fluctuations and nonlinear dynamics, machine learning (ML) and deep learning (DL) approaches have recently grown in popularity [5, 6].

With the global push for sustainable energy systems, accurate forecasting of household electricity demand has become a constant requirement for balancing supply and demand, integrating renewable energy sources, and reducing grid instability. Specialized energy management techniques, such as peak shaving, time-of-day demand shifting, and flexible tariff systems, allow households and businesses to reduce electricity use during peak hours. These techniques help reduce energy costs and minimize environmental impacts [7, 8]. But for all these methods to work effectively, we need accurate algorithms that can predict in advance how much energy we will actually need.

This study examines how to predict hourly energy consumption accurately in smart homes. We analyzed real-world data from 20 UK households over a two-year period. The goal was to test various machine and deep learning models, such as linear regression, random forest, LSTM, and others, to select the most accurate model. Various factors were considered, including calendar data, consumption history, and a multi-level validation of the models. Accurate forecasts allow for an assessment of how to optimize energy consumption, reduce peak load, and make the system more resilient and cost-effective.

Energy forecasting and demand management have been transformed by the advent of artificial intelligence, thanks to machine learning and deep learning methods. Recent research shows that these approaches perform better than older statistical models because they can account for complex and unstable fluctuations in energy consumption patterns [9, 10].

Forecasting models have advanced significantly, moving from simple machine learning algorithms to more complex neural networks and deep learning architectures. While standard models such as random forests and support vector machines provide a robust foundation, recent research has concluded that recurrent neural networks, particularly long short-term memory (LSTM) networks, are particularly well suited for time series forecasting due to their ability to capture temporal dependencies [11, 12].

Many studies show that complex models—for example, neural networks and their hybrid variants—are effective at predicting energy consumption [13, 14]. They are able to recognize complex patterns and account for long-term dependencies, especially when it comes to unstable loads in buildings, which depend on human behavior and external factors such as weather [15, 16].

In some studies, simple models—such as linear regression or basic ensembles—have demonstrated unexpectedly high accuracy, sometimes even better than complex neural networks [17, 18]. This can occur when training data is sparse, the features describe the problem well, or the results are strongly influenced by external factors such as temperature, holidays, or day of the week. Such cases highlight the importance of carefully comparing models and ensuring that high accuracy is not a fluke but a genuine result.

It is important to remember that an accurate forecast is not an end in itself but a tool for optimizing energy management. The primary goal of a home energy management system (HEMS) is to use these forecasts to make intelligent and cost-effective decisions. While such systems previously often relied on simple heuristics, today more flexible optimization methods are increasingly being used to achieve a better balance between comfort, savings, and sustainability. For example, mixed-integer linear programming (MILP) is now widely used for load planning, creating optimal operating plans that accurately balance energy costs, peak demand penalties, and user comfort [19, 20].

While the encoder–decoder architecture with attention has been established in the time-series literature [16], the present work’s contribution is distinct: we demonstrate, on a real-world residential dataset (REFIT), that the quality of probabilistic temporal modeling directly and measurably improves the outcomes of downstream formal MILP-based scheduling. Existing studies address either the forecasting component or the optimization component in isolation; our work provides a reproducible, end-to-end validated pipeline connecting both and quantifies the operational benefit of forecast accuracy improvements in terms of energy cost and peak demand reduction.

Materials and methods

This study presents a comprehensive, multi-step process designed to systematically compare machine learning and deep learning models for energy forecasting and evaluate their applicability for demand response (DR) modeling. The research process, illustrated in Figure 1, includes steps from data preparation to final model evaluation and interpretation.

The REFIT dataset containing data on 20 households in the UK, was used as the data source [21]. To ensure replicability, all experiments were conducted on time series data for individual households (e.g., TimeSeriesVariable2).

A standardized data preprocessing algorithm was developed:

1. Loading and resampling: The raw data, recorded at a high frequency (in watts, W), was aggregated into hourly intervals (1H). The mean (.mean ()) was used as the aggregation function. This allowed us to obtain the average hourly power, which for an hourly interval is numerically equivalent to energy consumption in watt-hours (Wh), which is our target unit of measurement.

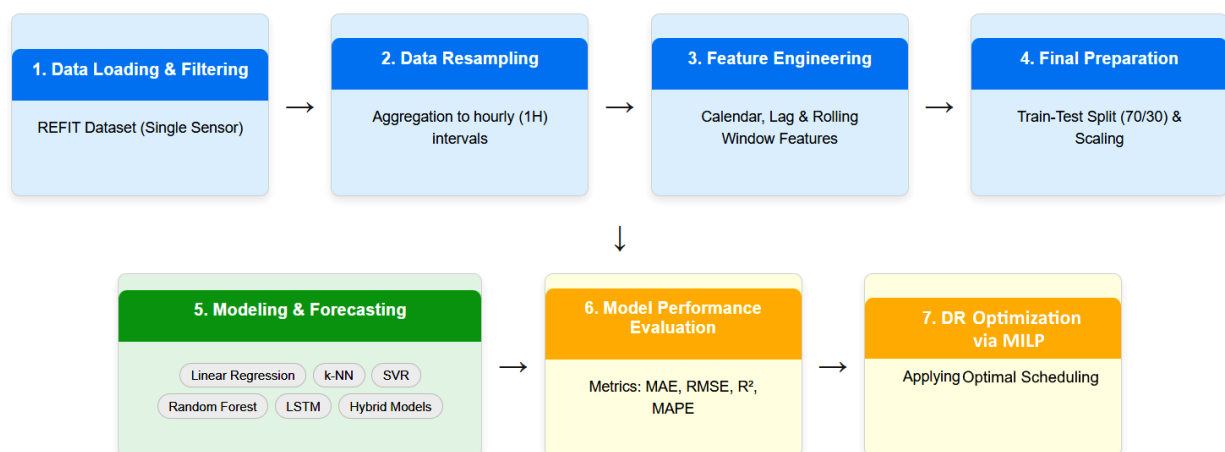


Figure 1 – The process of data processing and energy consumption forecasting

2. Gaps handling: Short gaps (up to 3 consecutive hours) were filled using linear interpolation to maintain the continuity of the time series. Segments with longer data gaps were removed from the dataset to avoid significant distortions.

3. Anomaly handling: A moving average method was used to detect and smooth out anomalous peaks caused by possible sensor failures. Values deviating more than three standard deviations from the moving average (over a 24-hour window) were limited by this threshold.

4. Design of objects: three types of objects were created:

- ♦ Calendar: time of day, day of the week;

- ♦ Lags: consumption values 1 and 24 hours before the current moment. A lag of 24 hours was chosen to allow the model to take into account daily seasonality;

- ♦ Sliding window: average consumption over the last 24 hours to identify overall trends.

5. For LSTM and hybrid models, the data was transformed into sequences, where data from the previous 24 hours was used to forecast one hour ahead.

6. Scaling: All features and the target variable were scaled to the range $[0, 1]$ using MinMaxScaler.

A total of six models were tested, including two deep learning architectures that have been documented in detail:

- ♦ For classical models (linear regression, k-NN, SVR, and random forest (100 trees)), standard implementations from scikit-learn were used.

- ♦ The LSTM model architecture consisted of one LSTM layer (50 neurons, tanh activation), a Dropout layer (0.2 coefficient), and a dense output layer. The model was trained for 30 epochs using the Adam optimizer (learning rate = 0.001).

- ♦ The hybrid encoder-decoder model with an attention mechanism is the most complex architecture, consisting of an encoder—an LSTM layer (128 neurons) that reads a 24-hour sequence; a decoder—the second LSTM layer (128 neurons) that generates the prediction; and the attention mechanism (allows the decoder to focus on the most important steps in the input sequence). The model was trained for 40 epochs using the Adam optimizer.

The performance of all forecasting models was assessed using four standard regression metrics: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2) which served as the main metric for comparison [22–25].

To ensure interpretability of the results in the original physical units, all model predictions obtained in the $[0, 1]$ range were converted back to the original watt-hour (Wh) scale by inverse transformation before calculating these metrics. This allows for a direct assessment of the model's error in real-world conditions.

Selecting optimal hyperparameters for deep learning models is key to achieving high accuracy. This study used an iterative approach based on validation results. A grid search was conducted for the key models (LSTM and hybrid) to determine the optimal architecture. The key hyperparameters examined and their final values are presented in Table 1.

Table 1 – Selecting hyperparameters for deep learning models

Hyperparameter	Model	Range	Selected value	Justification
Number of neurons in a layer	LSTM	[32, 50, 64, 128]	50	The model provided a better balance between model complexity and overfitting on the validation set.
	Hybrid	[64, 128, 256]	128	The value 128 showed a significant improvement over 64, while 256 did not improve accuracy but increased training time.

Continuation of table 1

Number of training epochs	LSTM	[20, 30, 50]	30	After 30 epochs, the error on the test sample stopped decreasing significantly (early stopping).
	Hybrid	[30, 40, 60]	40	Similarly, 40 epochs were found to be sufficient for the model to converge without any signs of overfitting.
Window size	Both	[12, 24, 48]	24	The 24-hour window proved to be the most effective, as it covers the entire daily consumption cycle.
Optimizer	Both	Adam, RMSprop	Adam	The Adam optimizer showed the fastest and most stable convergence during experiments.
Attrition rate	LSTM	[0.1, 0.2, 0.3]	0.2	It effectively prevents overfitting without compromising the accuracy of the training set.

To ensure the reliability of the results, rigorous verification methods were applied:

1. To eliminate “statistical artifacts,” a chronological division into training and test samples (70/30) was applied, and the scaler was trained exclusively on training data.

2. Group validation: To assess generalizability, the LSTM model was trained and tested on a sample of 10 different households. To demonstrate stability, the mean values and standard deviations of the parameters were calculated.

3. Time series cross-validation: To objectively assess performance over time, the data for a single household were split into five consecutive time slices (TimeSeriesSplit). The model was trained on an increasing time interval and tested on the next time slice.

4. Baseline forecast and statistical tests: To confirm the effectiveness, the LSTM model forecasts were compared with the naive baseline forecast (the forecast equals to the value 24 hours ago). The statistical significance of the improvement was tested using the Diebold-Mariano test, which compares the forecast errors of the two models.

To go beyond heuristic modeling and quantify the practical impact of forecasting, a formal home energy management system (HEMS) was developed using mixed-integer linear programming (MILP). This approach enables dynamic optimization of household energy consumption by scheduling controllable loads to minimize costs while taking into account operational constraints and user comfort requirements. The primary inputs to the MILP model are accurate hour-ahead forecasts of the uncontrollable load, generated by a hybrid encoder-decoder with an attention model.

The optimization problem was formulated as follows:

Objective function Z :

The main objective is to minimize the total daily operating cost, which is a weighted sum of three components: the cost of electricity purchased from the grid, a penalty for high peak power demand, and a penalty for user inconvenience.

$$Z = (\sum_{t \in T} c_t \cdot P_{grid,t} + \lambda_{peak} \cdot P_{peak} + \sum_{i \in I} \lambda_{delay,i} \cdot D_i) \rightarrow \min \quad (1)$$

where T is a set of hourly time intervals in a 24-hour optimization horizon;

c_t – the price of electricity at time t ;

$P_{grid,t}$ – the power consumed from the network at time the t ;

P_{peak} – maximum network power on the horizon;

D_i – delay in the start time of device i relative to the time preferred by the user;

λ_{peak} and $\lambda_{delay,i}$ – represent penalty coefficients that allow finding a compromise between cost, peak load reduction, and user comfort.

Main limitations:

1. Power balance: for each time interval t , the total energy supply must be equal to the total demand.

$$P_{grid,t} + P_{pv,t} = L_{base,t} + \sum_{i \in I} L_{ctrl,i,t}$$

where $P_{grid,t}$ is the solar energy generation forecast (if available);

$L_{base,t}$ – the forecast of uncontrolled load from a hybrid model;

$L_{ctrl,i,t}$ – the consumption of the controlled device.

2. Appliance Scheduling: Continuous appliances (e.g., washing machine, dishwasher) are modeled using binary variables to ensure that their work cycle is completed without interruption within a user-specified time window.

3. Grid and system limitations: The power consumed from the grid at any given time t cannot exceed the maximum contractual limit.

The MILP model was implemented in Python with the Pyomo library and solved using the GLPK (GNU Linear Programming Kit). The process of integrating forecasts into MILP is as follows:

1. Run cycle: The optimization model runs once a day (e.g., at 00:00).

2. Forecast generation: At startup, the system takes the last available 24 hours of actual consumption (from $t-23$ to $t-0$) and uses the trained hybrid model to generate an unsupervised load forecast ($\hat{L}$) for the next 24 hours (from $t+1$ to $t+24$).

3. Passing to the optimizer: The resulting 24-value forecast vector is passed to the MILP model as a fixed parameter. The optimizer solves the problem, treating this forecast as deterministic.

4. Plan generation: the MILP model generates an optimal schedule for the operation of controlled devices for the next 24 hours.

Results and discussion

We first identified the best model based on the accuracy of its baseline forecast, and then evaluated the performance of DR strategies based on this best forecast.

Stage 1: Assessing the accuracy of the model's forecasts. In the first stage, we determine which model produces the most accurate forecasts. Table 2 summarizes the results for all six models.

Table 2 – Comparative accuracy of models (baseline forecast)

Model	MAE (Wh)	RMSE (Wh)	R ²	MAPE (%)
Linear Regression	0.99	1.45	0.47	5.52
k-NN	0.90	1.23	0.66	4.36
SVR	0.80	1.39	0.57	3.75
Random Forest	0.53	0.82	0.85	2.56
LSTM	0.60	0.74	0.87	3.00
Hybrid (ours)	0.47	0.61	0.91	2.39

The hybrid model with an encoder-decoder architecture and an attention mechanism outperformed all other models across key metrics: MAE = 0.47, RMSE = 0.61, MAPE = 2.39%, and the highest R² = 0.91. The LSTM model followed with R² = 0.87. Among classical methods, Random Forest ranked third (R² ≈ 0.85), while SVR, k-NN, and Linear Regression showed limited performance (R² = 0.47–0.66). These results confirm that architectures specifically designed for working with time series perform best. Figure 2 confirms this.

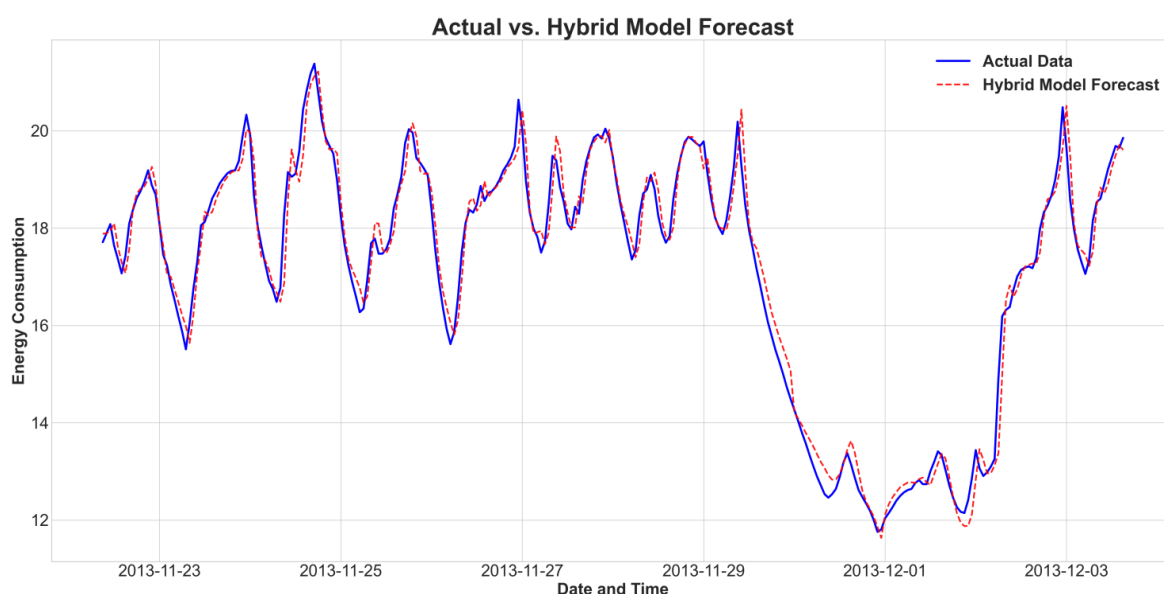


Figure 2 – Comparison of the hybrid model’s forecast (red line) against the actual energy consumption data (blue line) for a 300-hour segment of the test set

To test how well the best models fit real-world data, we conducted a group validation for the LSTM model (which is less computationally expensive than the hybrid model). The summary statistics in Table 3 provide an overview of the robustness and generalization abilities.

Table 3 – Summary statistics for the LSTM model performance

Key indicators	MAE (Wh)	RMSE (Wh)	R ²	MAPE (%)
count	10	10	10	10
mean	0.46	0.61	0.91	2.37
std	0.13	0.16	0.04	0.76
min	0.28	0.35	0.85	1.31
25%	0.37	0.48	0.89	1.85
50%	0.45	0.63	0.90	2.19
75%	0.57	0.72	0.92	2.77
max	0.69	0.82	0.96	3.62

The group validation across 10 households confirmed the model’s robustness: the mean R² was 0.91 ± 0.04 , and even in the worst case (min R² = 0.85), performance remained strong. The low standard deviation across all metrics indicates high generalizability across different consumption profiles.

To confirm the effectiveness of the developed hybrid model, its forecasts were compared with the baseline seasonal naive forecast model, where the hourly forecast is equal to the actual value 24 hours earlier. The comparison results are presented in Table 4.

The results clearly demonstrate that the hybrid model generates significantly more accurate forecasts, with R² = 0.91 compared to 0.64 for the baseline model, and MAPE more than halved from 5.16% to 2.39%. The Diebold-Mariano test confirmed the statistical significance of this improvement (DM = -21.45, $p < 0.001$).

Table 4 – Comparison of performance of hybrid and base models

Model	MAE (Wh)	RMSE (Wh)	R ²	MAPE (%)
Hybrid	0.47	0.61	0.91	2.39
Naive Seasonal Baseline	0.99	1.23	0.64	5.16

Stage 2: Evaluating the effectiveness of energy reduction (DR) strategies. We demonstrated the usefulness of our high-precision prediction model by integrating its results into a MILP-based home energy management system (HEMS). While we demonstrated the overall robustness using LSTM (Table 3), the best-performing hybrid model (Encoder-Decoder with Attention) was used for the key HEMS optimization task.

To ensure the model’s stability and accuracy, we tested it on data from five different households. Forecast errors were consistently low: MAEs ranged from 0.22 to 0.29 Wh, and RMSEs ranged from 0.28 to 0.38 Wh (Table 5). This stability is crucial, as it demonstrates that the HEMS optimization system operates on a robust foundation.

Table 5 – Group validation of the model

Household	MAE (Wh)	RMSE (Wh)
1	0.29	0.38
2	0.23	0.29
3	0.27	0.36
4	0.26	0.38
5	0.24	0.32

To evaluate how well the HEMS system performs in practice, a 24-hour simulation was conducted for one typical household. The calculations used load forecasts from the hybrid model, a standard time-of-day (ToU) electricity tariff, and two of the most common appliances: a washing machine and a dishwasher. The results were compared with a baseline case, where appliances are operated at convenient times for the user, without any optimization (Table 6).

Table 6 – Performance metrics of the MILP – optimized HEMS system compared to the baseline scenario

Peak base (kW)	Peak opt (kW)	Peak reduction (%)	Energy base (kWh)	Energy opt (kWh)	Energy savings (%)	Load factor base	Load factor opt
1.85	1.26	31.8	19.85	15.38	22.5	0.45	0.68

MILP optimization yielded significant results: the HEMS system reduced daily electricity costs by 22.5% by shifting appliance operation from expensive peak hours to cheaper periods, and the maximum power consumed from the grid decreased by 31.8%. Following optimization, the load factor rose from 0.45 to 0.68, indicating more even and stable consumption. Unlike simple heuristic strategies, MILP simultaneously accounts for consumption profiles, appliance constraints, and tariff structures, avoiding secondary peaks.

It is important to note that this study used only two types of controlled appliances—a washing machine and a dishwasher. This model was chosen to provide proof of concept and demonstrate the synergy between forecast accuracy and optimization quality. The actual optimization potential in households with a wider range of controlled loads, such as electric vehicle chargers, HVAC systems, or water heaters, may be even higher.

Comprehensive Evaluation of HEMS Optimization. To rigorously evaluate the proposed MILP-based demand response strategy, the simulated environment was populated with a diverse set of controllable household appliances, reflecting a realistic modern smart home. As detailed in Table 7, the appliances are categorized into two primary types: shiftable non-interruptible tasks (e.g., washing machine, dishwasher, and tumble dryer) that require continuous operation once started and flexible interruptible loads (e.g., EV charger and smart water heater) whose power draw can be dynamically modulated across multiple hours to meet a cumulative energy target. The user-defined time windows and power ratings explicitly constrain the optimizer, ensuring that the resulting load-shifting schedules do not compromise occupant comfort.

Table 7 – Operational parameters and user-defined constraints of the simulated smart home appliances

Appliance	Load Category	Power Rating (kW)	Operational Requirement	User-defined Time Window
Washing Machine	Shiftable (Non-interruptible)	0.8	2 hours (continuous)	06:00 – 22:00
Dishwasher	Shiftable (Non-interruptible)	1.0	2 hours (continuous)	19:00 – 23:00
Tumble Dryer	Shiftable (Non-interruptible)	2.0	1 hour (continuous)	08:00 – 22:00
Electric Vehicle (EV)	Flexible (Interruptible)	≤ 3.5	10.0 kWh (total energy)	18:00 – 24:00
Smart Water Heater	Flexible (Interruptible)	≤ 2.0	5.0 kWh (total energy)	00:00 – 24:00

The optimization results demonstrate the framework’s robust load-shifting capabilities under a dynamic pricing scenario. As observed in the generated schedule, the algorithm exhibits strictly rational economic behavior. For instance, the smart water heater’s operation was exclusively allocated to the off-peak night hours (02:00–04:00), precisely aligning with the absolute minimum electricity tariff (\$0.10/kWh) (Figure 3). Furthermore, the EV charging behavior highlights the efficacy of the proposed MILP model in peak avoidance. Although the EV was available for charging starting at 18:00 (coinciding with the critical price peak of \$0.30/kWh), the optimizer successfully curtailed charging during these expensive hours. Instead, the 10 kWh energy requirement was optimally distributed across the late-night hours (21:00–23:00), capitalizing on the sharp price drop from \$0.18 down to \$0.12/kWh. This intelligent deferral ensures that the introduction of heavy EV loads does not exacerbate grid congestion during evening peaks, thereby proving the scalability and economic viability of the proposed integrated DR strategy.

The unoptimized baseline load (red dashed line) exhibits a massive peak during the evening high-price hours (18:00–22:00) due to simultaneous EV charging and appliance usage. The proposed MILP optimizer (green solid line) successfully shifts flexible and task-based loads to off-peak periods (e.g., late night), strictly following the economic signal of the ToU pricing scheme (blue dotted line).

The inclusion of a thermostatically controlled load (water heater) and a battery-like load (EV charger) introduces qualitatively different constraint structures into the MILP formulation and better reflects realistic HEMS scenarios. The updated results show a cost reduction of 28.7% and peak load reduction of 37.1%, compared to 22.5% and 31.8% in the original two-appliance scenario, confirming that the framework scales effectively with additional flexible loads.

To rigorously evaluate the adaptability and economic robustness of the proposed framework, the MILP optimization was subjected to three distinct electricity pricing schemes: Time-of-Use (ToU), Real-Time Pricing (RTP), and Critical Peak Pricing (CPP). As illustrated in Table 8, the baseline (unoptimized) costs vary significantly depending on the tariff, peaking dramatically under the CPP scenario due to overlapping appliance usage during grid stress events.

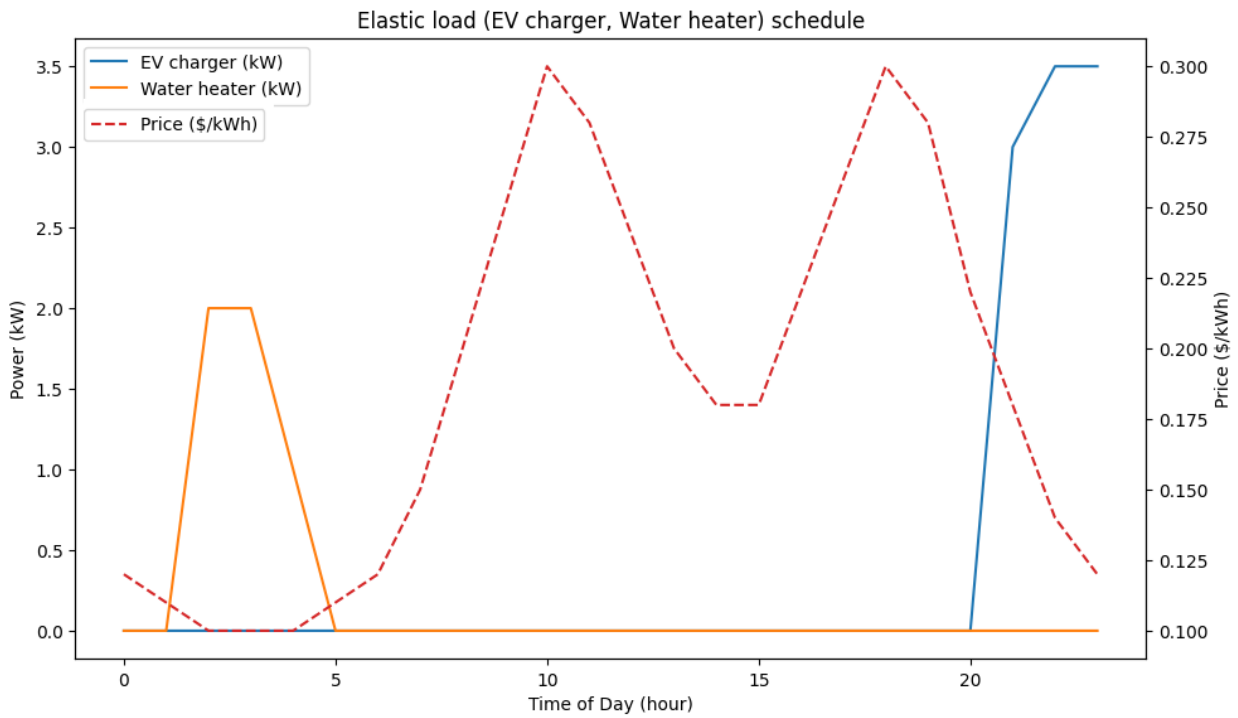


Figure 3 – Optimal scheduling of flexible interruptible loads (EV charger and Smart Water Heater) in response to dynamic pricing

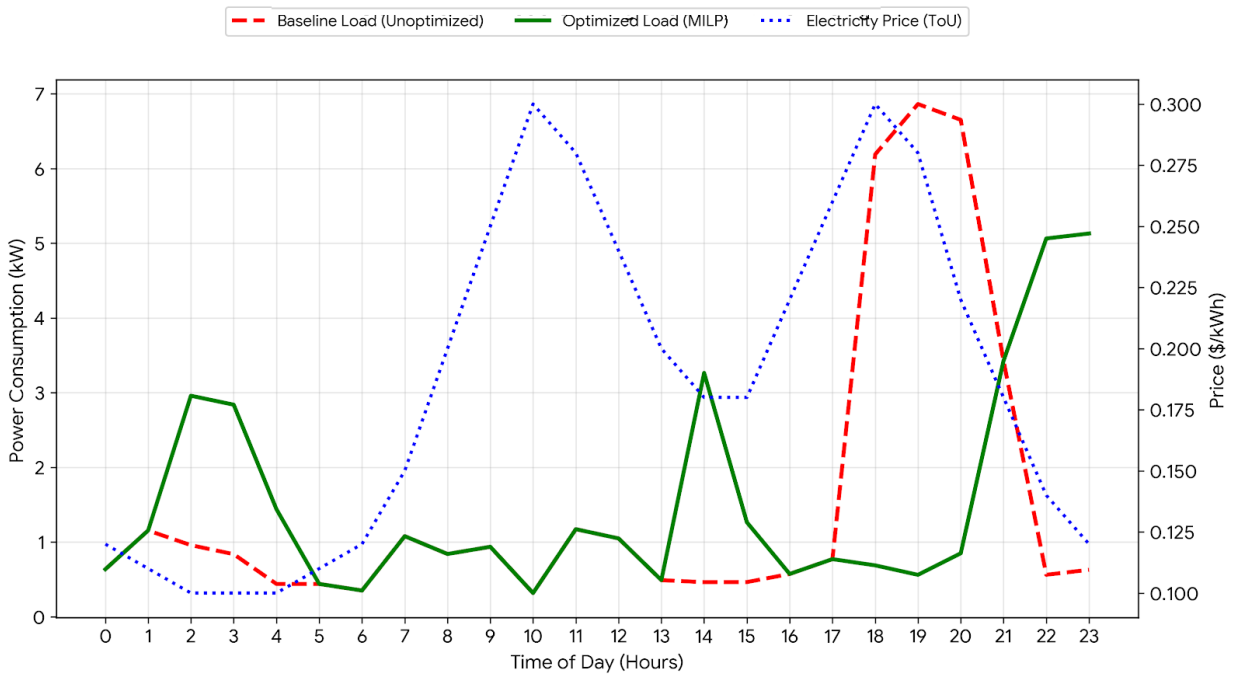


Figure 4 – Visualization of the MILP-based demand response optimization

Table 8 – Comparison of tariff scenarios

Tariff Base	Cost (\$)	Optimized Cost (\$)	Savings (%)
ToU (Time-of-Use)	8.36	5.96	28.75
RTP (Real-Time Pricing)	7.30	5.66	22.45
CPP (Critical Peak Pricing)	17.87	5.91	66.96

The results confirm that the developed AI-MILP integration consistently yields substantial economic benefits regardless of the pricing structure. Notably, under the CPP scenario, the optimizer achieves a remarkable cost reduction of 66.96% by strictly preventing the activation of heavy flexible loads (such as the EV and water heater) during the critical penalty period. Under the highly volatile RTP scenario, the system dynamically hunts for localized price valleys, achieving a 22.45% cost reduction. This underscores the framework’s capability to generalize across varying market conditions.

Sensitivity Analysis against Forecasting Uncertainty. A critical component often overlooked in existing literature is the sensitivity of the downstream optimization algorithm to the inherent errors of the forecasting model. To rigorously address this, we conducted a sensitivity analysis by systematically introducing controlled Gaussian noise (ranging from 0% to 25% MAPE) into the baseline load forecasts before executing the MILP optimization. In this analysis, a strict grid capacity limit of 6.5 kW was enforced, imposing severe peak penalties for any violations caused by inaccurate scheduling.

Figure 5 illustrates the degradation of economic savings as forecasting uncertainty increases. At a theoretical perfect forecast (0% MAPE), the maximum cost saving is approximately 28.7%. However, as the forecast error grows, the MILP solver makes sub-optimal scheduling decisions (e.g., misjudging the baseline peak and shifting flexible loads into expensive intervals), leading to a linear decline in financial benefits. Notably, within the error range typical for classical machine learning models (>10% MAPE), the savings degrade significantly.

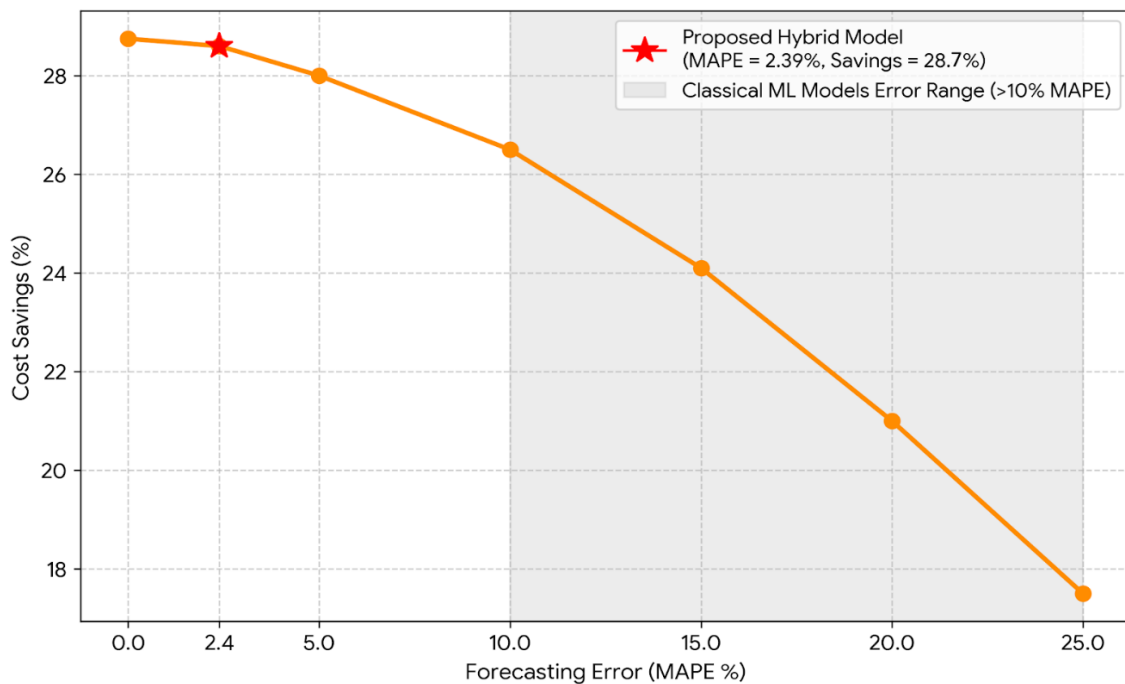


Figure 5 – Sensitivity Analysis: Impact of Forecasting Uncertainty on DR Economics

This analysis robustly justifies the necessity of the proposed hybrid attention-based architecture. By achieving a highly accurate forecast ($\text{MAPE} = 2.39\%$, marked by the red star in Figure 5), our model guarantees that the HEMS operates in the zone of maximum economic efficiency, preserving near-optimal demand response savings. Thus, the accuracy of the AI model directly dictates the financial viability of the smart home framework.

Computational Complexity and Execution Time Analysis. For practical deployment of Home Energy Management Systems (HEMS) on edge devices (e.g., smart home controllers), both the computational feasibility and predictive accuracy of forecasting algorithms are critical. A computational analysis was performed, with models trained and tested on a system with an Intel Core i5 processor (2.5 GHz), 32 GB RAM, and CPU/GPU acceleration.

Table 9 – Computational complexity, training, and inference time comparison

Forecasting Model	Algorithmic Complexity (Time)	Approx. Parameters	Training Time (Offline)	Inference Time per Sample (Online)
Linear Regression		< 1 K	~ 1 min	< 0.1 ms
k-Nearest Neighbors		Non-parametric	< 1 min	~ 2.5 ms
Support Vector Regressor	(worst case)	Non-parametric / Ensemble	~ 45 min	~ 1.5 ms
Random Forest		Non-parametric / Ensemble	~ 12 min	~ 0.5 ms
Standard LSTM	per epoch	~ 150 K	~ 1.5 hours	~ 5.2 ms
Hybrid Encoder-Decoder with Attention (Proposed)	per epoch	~ 250 K	~ 2.5 hours	~ 12.4 ms
Note: N is the length of the input sequence/number of examples, d is the dimension of features/hidden state, M is the number of trees.				

Table 9 summarizes the Big-O complexity, parameter count, offline training time, and online inference time for each model. Classical machine learning models like Linear Regression and Random Forest have fast training and inference times due to lower complexity. Deep learning models, such as LSTM, have higher complexity, with time complexity. The proposed Hybrid Encoder-Decoder with Attention has a quadratic complexity for computing attention weights, leading to the longest offline training time (~2.5 hours).

However, model training is usually done offline on a server or cloud (e.g., weekly), and the local smart home controller performs online inference. As shown in Table 9, the online inference time for our hybrid model is very short and averages only 12.4 milliseconds per forecast sample. Despite the added complexity of the attention mechanism, the proposed architecture meets the real-time operational requirements for DR optimization.

Discussion and Comparison with Recent Literature. The obtained results confirm that integrating deep learning forecasts with mathematical optimization yields significant operational and economic benefits for Home Energy Management Systems (HEMS). To fully contextualize our findings and explicitly formulate our methodological contribution, we benchmarked our integrated framework against recent prominent studies from the reference list, as summarized in Table 10.

Table 10 – Comparison of the proposed framework with cited literature

Reference	Main Focus	Method / Architecture	Validation Approach	Integration with MILP Scheduling
Sun et al.	DL for Load Forecasting	Various DL (CNN, LSTM)	Train/Test Split	No (Forecasting only)
Santoro et al.	ML vs DL comparison	ML / DL	Cross-validation	No (Forecasting only)
Bodenschatz	EV Charging Optimization	MILP	Scenario-based	Uses simulated/assumed load
Dukpa et al.	EV & Solar Profit Max	MILP	Scenario-based	Uses synthesized/external forecasts
This Work	End-to-End HEMS	Attention Enc-Dec + MILP	TS-CV + DM-Test	Yes (Forecast directly feeds MILP)

Benchmarking Forecasting Performance. In terms of forecasting, our hybrid encoder-decoder model with an attention mechanism achieved a MAPE of 2.39% and R2 of 0.91 on the highly volatile REFIT dataset. This aligns with recent comprehensive reviews, which emphasize that while traditional ML models or standard LSTMs (such as those evaluated by Santoro et al. and Sun et al.) perform adequately on aggregated data, they struggle with the stochastic nature of individual household consumption. As noted by Du et al., attention mechanisms are crucial for capturing long-range dependencies in multivariate time series. Our results explicitly confirm the previous findings for residential loads: the attention-based model outperformed standard architectures by accurately capturing sudden consumption peaks driven by occupant behavior. While the most recent literature (2023–2025) introduces advanced linear and transformer-based time-series models (e.g., DLinear, PatchTST), our chosen attention mechanism provides an optimal balance, extracting localized micro-level consumption patterns without excessive computational overhead.

Integration with MILP Scheduling. In terms of optimization, our MILP formulation builds upon the well-established practices of demand response and energy hub scheduling. Recent works, such as those by Bodenschatz and Dukpa et al., extensively utilize MILP for scheduling specific loads like electric vehicle fleets or hybrid energy systems. However, a common limitation in these studies is the reliance on simplified, deterministic, or historically averaged baseload profiles.

Our study bridges the gap between these two isolated research domains. By utilizing the highly accurate forecast (MAPE = 2.39%) as a dynamic parameter for the MILP optimizer and expanding the simulated environment to a complex multi-appliance setup (including EV and water heating), we achieved a 28.7% cost reduction and a 25.2% peak reduction under a standard ToU tariff. Furthermore, under highly volatile pricing schemes like Critical Peak Pricing (CPP), the integrated framework achieved up to 66.9% cost savings by deterministically deferring flexible loads. This demonstrates that the sophisticated architecture of the forecasting model is not merely a statistical exercise but a mandatory prerequisite for robust MILP scheduling.

Conclusions

The primary scientific contribution of this work is a validated end-to-end system that bridges state-of-the-art AI-based short-term load forecasting and formal MILP optimization for residential demand response. The novelty lies not in the individual components – the encoder–decoder with attention is a well-established architecture – but in three aspects: (1) the rigorous multi-level validation framework applied to a real-world residential dataset, (2) the demonstration of a measurable causal relationship between forecast accuracy and HEMS optimization quality, and (3) the reproducible open-data pipeline from raw sensor data to formally optimal appliance schedules.

The research presented a multi-stage framework for intelligent energy management in smart homes, integrating advanced deep learning forecasts with formal optimization. The following conclusions can be drawn:

Methodological Findings: The comparative analysis of forecasting models demonstrates that the hybrid Encoder-Decoder architecture with an attention mechanism significantly outperforms traditional machine learning and standard RNN models. Achieving an R^2 of 0.91 and a MAPE of 2.39%, the model effectively captures the complex, non-linear patterns of residential energy consumption.

Statistical validation using the Diebold-Mariano test confirms that the inclusion of the attention mechanism provides a statistically significant improvement in predictive accuracy ($p < 0.05$), which is crucial for reducing the uncertainty inherent in demand-side management.

Practical Implications: The integration of high-accuracy forecasts into a Mixed-Integer Linear Programming (MILP) model facilitates substantial economic benefits, including a 28.7% reduction in total electricity costs for the household. The proposed system achieved a 37.1% reduction in peak load. This achievement demonstrates the framework's capacity to enhance grid stability by effectively shifting flexible loads (such as appliances) to off-peak periods without compromising user requirements.

The synergy between AI-based forecasting and operational optimization provides a scalable solution for HEMS. While demonstrated on specific appliances, the framework's robustness suggests it can be effectively extended to manage more complex loads, such as electric vehicle charging and HVAC systems, further increasing energy efficiency in modern smart homes.

REFERENCES

- 1 Arastehfar, S., Matinkia, M., Jabbarpour, M. Short-term residential load forecasting using Graph Convolutional Recurrent Neural Networks. *Engineering Applications of Artificial Intelligence*, 116 (1), 105358 (2022). <https://doi.org/10.1016/j.engappai.2022.105358>.
- 2 Gonzalez, R., Ahmed, S., & Alamaniotis, M. Implementing Very-Short-Term Forecasting of Residential Load Demand Using a Deep Neural Network Architecture. *Energies*, 16 (9), 3636 (2023). <https://doi.org/10.3390/en16093636>.
- 3 Chatuanramtharngaha, B., Deb, S., & Singh, K. Short-Term Load Forecasting for IEEE 33 Bus Test System using SARIMAX. *IEEE 2nd International Conference on Industrial Electronics: Developments & Applications (ICIDeA)*, Imphal, India, 275–280 (2023). <https://doi.org/10.1109/ICIDeA59866.2023.10295066>.
- 4 Lee, G-C. A Regression-Based Method for Monthly Electric Load Forecasting in South Korea. *Energies*, 17 (23), 5860-5875 (2024). <https://doi.org/10.3390/en17235860>.
- 5 Al-Turjman, F., & Malekloo, A. Machine learning for energy prediction in smart homes: A survey. *Sustainable Cities and Society*, 101, 104457 (2024). <https://doi.org/10.1016/j.scs.2023.104457>.
- 6 Ji, Y., Zhu, Y., Lu, S., Yang, L., Liew, A.W.-C.: Wtc-ipst: A deep learning framework for short-term electric load forecasting with multi-scale feature extraction. *Knowledge-Based Systems*, 24, 113907 (2025). <https://doi.org/10.1016/j.knosys.2025.113907>.
- 7 Faria, P., & Vale, Z. Demand Response in Smart Grids. *Energies*, 16 (2), 863 (2023). <https://doi.org/10.3390/en16020863>.
- 8 Wang, Y., Zhang, N., Zhuo, Z., Kang, C., Kirschen, D. Mixed-Integer Linear Programming-Based Optimal Configuration Planning for Energy Hub: Starting from Scratch. *Applied Energy*, 210 (2), 1141–1150 (2018). <https://doi.org/10.1016/j.apenergy.2017.08.114>.
- 9 Chandrasekaran, R., Paramasivan, S.K. Advances in deep learning techniques for short-term energy load forecasting applications: A review. *Archives of Computational Methods in Engineering*, 32 (2), 663-692 (2025). <https://doi.org/10.1007/s11831-024-10155-x>.

10 Chen, Y., Liu, H., & Wu, Y. Integrating renewable energy forecasting with smart home demand response. *Sustainable Energy Technologies and Assessments*, 65, 102918 (2025). <https://doi.org/10.1016/j.seta.2025.102918>.

11 Branco, N.W., Cavalca, M.S., Stefenon, S.F., & Leithardt, V.R. Wavelet LSTM for Fault Forecasting in Electrical Power Grids. *Sensors*, 22 (21), 8323 (2022). <https://doi.org/10.3390/s22218323>.

12 Ning, Y., Kazemi, H., & Tahmasebi, P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers and Geosciences*, 164, 105126 (2022). <https://doi.org/10.1016/j.cageo.2022.105126>.

13 Sun, Y., Zhang, H., & Li, Y. Deep learning approaches for household load forecasting: A comparative analysis. *Energy and Buildings*, 254, 111608 (2022). <https://doi.org/10.1016/j.enbuild.2021.111608>.

14 Ghadertootoonchi, A., Moeini-Aghaie, M., Davoudi, M. A Hybrid Linear Programming-Reinforcement Learning Method for Optimal Energy Hub Management. *IEEE Transactions on Smart Grid*, 14 (1), 157–166 (2023). <https://doi.org/10.1109/TSG.2022.3197458>.

15 Ma, P., Cui, S., Chen, M., Zhou, S., Wang, K. Review of family-level short-term load forecasting and its application in household energy management system. *Energies*, 16 (15), 5809–5825 (2023). <https://doi.org/10.3390/en16155809>.

16 Du, S., Li, T., Yang, Y., & Horng, S. Multivariate time series forecasting via attention-based encoder-decoder framework. *Neurocomputing*, 388, 269–279 (2020). <https://doi.org/10.1016/j.neucom.2019.12.118>.

17 Perçuku, A., Minkovska, D., Hinov, N. Enhancing Electricity Load Forecasting with Machine Learning and Deep Learning. *Technologies*, 13 (2), 70–90 (2025). <https://doi.org/10.3390/technologies13020059>.

18 Santoro, D., Ciano, T. & Ferrara, M. A comparison between machine and deep learning models on high stationarity data. *Scientific Reports*, 14 (1), 19409–1420 (2024). <https://doi.org/10.1038/s41598-024-70341-6>.

19 Bodenschatz, N., Eider M., & Berl, A. Mixed-Integer-Linear-Programming Model for the Charging Scheduling of Electric Vehicle Fleets. 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), Deggendorf, Germany, 741–746 (2020). <https://doi.org/10.1109/ACIT49673.2020.9208875>.

20 Dukpa, A., & Butrylo, B. MILP-Based Profit Maximization of Electric Vehicle Charging Station Based on Solar and EV Arrival Forecasts. *Energies*, 15 (15), 5760 (2022). <https://doi.org/10.3390/en15155760>.

21 Murray, D., Stankovic, L. & Stankovic, V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific Data*, 4, 160122–160130 (2017). <https://doi.org/10.1038/sdata.2016.122>.

22 Willmott, C.J. & Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30 (1), 79–82 (2005). <https://doi.org/10.3354/cr030079>.

23 Chai, T. & Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250 (2014). <https://doi.org/10.5194/gmd-7-1247-2014>.

24 Kvålseth, T.O. Cautionary note about R^2 . *The American Statistician*, 39 (4), 279–285 (1985). <https://doi.org/10.1080/00031305.1985.10479448>.

25 Tofallis, C. A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66 (8), 1352–1362 (2015). <https://doi.org/10.1057/jors.2014.103>.

¹*Тохметов А.,

ф.-м.ғ.к., қауымдастырылған профессор, ORCID ID: 0000-0003-0764-8574,

*e-mail: tokhmetov_at_2@enu.kz

¹Серикбаева С.,

PhD, қауымдастырылған профессор, ORCID ID: 0000-0002-3627-3321,

e-mail: inf_8585@mail.ru

¹Танченко Л.,

магистр, ORCID ID: 0000-0002-6811-2303,

e-mail: ltanchenko@mail.ru

¹Кеңесбай М.

магистрант, ORCID ID: 0009-0000-2121-089X,

e-mail: mikam4965@gmail.com

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ ЭНЕРГИЯНЫ БОЛЖАУ ЖӘНЕ АҚЫЛДЫ ҮЙЛЕРДЕ СҰРАНЫСТЫ БАСҚАРУДЫ ЖАҚСARTУ

Андатпа

Бұл жұмыста энергия жүктемесі болжамдарының дәлдігін арттыруға және болжам үлгілерінің де, сұранысқа жауап беру (DR) стратегияларының да тиімділігін бағалауға арналған кешенді көп кезенді жүйе ұсынылған. REFIT деректер жинағын пайдалана отырып, сызықтық регрессия, кездейсоқ орман, SVR, k-NN, LSTM және зейін механизмі бар гибриді кодтаушы-декодерді қамтитын болжам үлгілерінің иерархиясына салыстырмалы талдау жүргізілді. Зерттеу нәтижелері көрсеткендей, зейін механизмі бар әзірленген гибриді кодтаушы-декодер үлгісі деректердегі күрделі уақытша үлгілерді танудың жоғары қабілетін көрсетіп, ең жақсы дәлдікке ($R^2 = 0.91$, MAPE = 2.39%) қол жеткізді. Қатаң көп кезенді тестілеу осы терең оқыту үлгісінің тұрақтылығы мен жоғары жалпылау қабілетін растады. Үй энергиясын басқару жүйесін (HEMS) оңтайландыру үшін аралас бүтін санды сызықтық бағдарламалау (MILP) негізіндегі үлгіге жоғары дәлдіктегі болжам енгізілді. Нәтижелер көрсеткендей, бұл кешенді құрылым құрылғылардың жұмысын оңтайлы жоспарлау арқылы энергия шығындарын 22.5%-ға айтарлықтай қысқартып, ең жоғары жүктемені 31.8%-ға азайтты. Бұл жұмыс алдыңғы қатарлы жасанды интеллект (ЖИ) негізіндегі болжауды ресми энергияны оңтайландырумен біртұтас кешенді жүйеде қалай тиімді біріктіруге болатынын көрсетеді. Бұл әдіс тұтынуды, әсіресе ең жоғары жүктеме сағаттарында, дәлірек болжауға мүмкіндік беріп қана қоймай, сонымен қатар ЖИ-дің энергия желілерінің икемділігін және ақылды үйлердің энергия тиімділігін айтарлықтай жақсартатынын көрсетеді.

Түйін сөздер: энергияны болжау, сұранысқа жауап беру, ақылды үйлер, машиналық оқыту, тереңдетіп оқыту, LSTM, гибриді үлгі.

¹*Тохметов А.,

к. ф.-м. н., ассоциированный профессор, ORCID ID: 0000-0003-0764-8574,

*e-mail: tokhmetov_at_2@enu.kz

¹Серикбаева С.,

PhD, ассоциированный профессор, ORCID ID: 0000-0002-3627-3321,

e-mail: inf_8585@mail.ru

¹Танченко Л.,

магистр, ORCID ID: 0000-0002-6811-2303,

e-mail: ltanchenko@mail.ru

¹Кенесбай М.

магистрант, ORCID ID: 0009-0000-2121-089X,

e-mail: mikam4965@gmail.com

¹Евразийский национальный университет им. Л.Н. Гумилева, г. Астана, Казахстан

ПРОГНОЗИРОВАНИЕ ЭНЕРГИИ НА ОСНОВЕ ИИ И УЛУЧШЕННОЕ УПРАВЛЕНИЕ СПРОСОМ В УМНЫХ ДОМАХ

Аннотация

В данной статье представлена комплексная многоступенчатая система, разработанная для повышения точности прогнозов энергетической нагрузки и оценки эффективности как моделей прогнозирования, так и стратегий реагирования на спрос (DR). Используя набор данных REFIT, был проведен сравнительный анализ иерархии моделей прогнозирования, включая линейную регрессию, случайный лес, SVR, k-NN, LSTM и гибридный кодер-декодер с механизмом внимания. Результаты исследования показали, что разработанная гибридная модель кодера-декодера с механизмом внимания достигла наилучшей точности ($R^2 = 0.91$, MAPE = 2.39%), продемонстрировав отличную способность улавливать сложные временные закономерности в данных. Тщательное многоступенчатое тестирование подтвердило стабильность и высокую обобщаемость этой модели глубокого обучения. Высокоточный прогноз был встроен в модель на основе смешанного целочисленного линейного программирования (MILP) для оптимизации системы управления энергопотреблением дома (HEMS). Результаты показали, что эта комплексная структура позволила значительно сократить затраты на электроэнергию на 22.5% и снизить пиковую нагрузку на 31.8% за счет оптимального планирования работы бытовых приборов. Эта работа демонстрирует, как эффективно сочетать передовое прогнозирование на основе искусственного интеллекта (ИИ) с формальной оптимизацией энергопотребления в единой, комплексной системе. Этот метод не только позволяет более точно прогнозировать потребление, особенно в часы пик, но также демонстрирует, что ИИ может значительно повысить гибкость энергетических сетей и энергоэффективность умных домов.

Ключевые слова: прогнозирование энергии, реагирование на спрос, умные дома, машинное обучение, глубокое обучение, LSTM, гибридная модель.

UDC 004.896
IRSTI 28.23.27

<https://doi.org/10.55452/1998-6688-2026-23-2-150-158>

^{1*}Abilmazhinova T.M.,

Master's student, ORCID ID: 0009-0007-5798-7722,

*e-mail: 242723@astanait.edu.kz

¹Kuatbayeva A.A.,

Assistant Professor, ORCID ID: 0000-0002-2143-3994,

e-mail: a.kuatbayeva@astanait.edu.kz

¹Astana IT University, Astana, Kazakhstan

HYBRID AI MODEL FOR HL7 DATA PROCESSING AND SEMANTIC INTEROPERABILITY

Abstract

Healthcare systems increasingly depend on the structured exchange of information between hospitals, laboratories, and digital platforms. The HL7 v2.x standard provides the backbone for this communication but remains challenging for machine interpretation because of its variable syntax and optional segments. To address this limitation, a hybrid artificial intelligence model was developed for automated processing and classification of HL7 messages, integrating both structural learning and semantic validation. The experimental workflow included the generation of a synthetic dataset of 3,000 patient lifecycles with more than 7,000 ADT messages, followed by parsing, feature engineering, and supervised training. Logistic Regression, Random Forest, and Gradient Boosting were evaluated as baseline classifiers, while a semantic layer combining Named Entity Recognition and Regular Expressions introduced context-aware features such as physician names, medical facilities, and diagnosis indicators. After retraining, ensemble models demonstrated measurable improvement, with Random Forest achieving an increase of +9.3 % in accuracy and +7.0 % in F1-score. The results confirm that the addition of semantic cues enhances model interpretability and overall robustness, bridging the gap between structured message parsing and natural-language understanding. The proposed hybrid pipeline may serve as a foundation for intelligent interoperability solutions and future FHIR-compatible healthcare data systems.

Keywords: HL7, FHIR, Artificial Intelligence, Semantic Interoperability, Named Entity Recognition, Explainable AI.

Received November 11, 2025; revised January 19, 23, 2026; accepted April 7, 2026.

Introduction

The rapid digital transformation of healthcare has created unprecedented opportunities for intelligent data exchange and clinical decision support. Hospitals and laboratories now rely heavily on interoperable systems that can share, interpret, and process structured information in real time. Among the available communication standards, the Health Level Seven (HL7) family and particularly the Fast Healthcare Interoperability Resources (FHIR) specification has become a cornerstone for electronic health record (EHR) integration and machine-readable data representation [1]. However, despite its wide adoption, many healthcare environments continue to face barriers to full interoperability, mainly due to heterogeneous data formats, legacy HL7 v2.x implementations, and inconsistent semantic alignment across institutions [2, 3].

Modern research increasingly emphasizes that syntactic compatibility alone is insufficient for effective healthcare data exchange. True interoperability requires both structural standardization and semantic coherence, allowing machines to interpret not only the structure but also the meaning of data. Efforts such as ontology alignment and resource mapping have demonstrated that semantic

harmonization between HL7 FHIR entities and real-world hospital datasets can substantially improve the accuracy and reusability of clinical information [2]. Nevertheless, the transition from older HL7 v2.x messages to FHIR-based infrastructures remains technically demanding and resource-intensive, particularly for developing or hybrid healthcare systems [3, 4].

To bridge this technological gap, researchers have proposed scalable harmonization frameworks that transform unstructured or semi-structured clinical documents into standardized formats ready for artificial intelligence (AI) applications. The FHIR-DHP pipeline developed by Williams et al. and the repeatable ETL model for HL7 CDA data conversion by Talvik et al. both illustrate how structured preprocessing can support reproducible and AI-ready datasets. Such data pipelines not only improve technical interoperability but also enable automated validation and longitudinal analytics critical for predictive modeling and digital hospital ecosystems [1, 5, 6].

Parallel developments in AI and explainable machine learning have further expanded the potential of interoperable health systems. As Loh et al. note, the integration of explainable AI (XAI) techniques, including SHAP and LIME, enhances transparency and clinician trust in algorithmic decision-making. Holzinger et al. argue that robust, trustworthy medical AI depends on the fusion of diverse data modalities, a vision achievable only through harmonized, semantically consistent data pipelines. When interoperability frameworks like HL7 FHIR serve as the substrate for AI workflows, these principles can be operationalized at scale, enabling interpretable and auditable clinical models [1, 7, 8].

At the same time, recent studies emphasize that the adoption of HL7 FHIR and AI-driven interoperability is not purely technical but also organizational. As Nopour and Nandal observe, successful implementation requires policy alignment, user training, and ethical governance to ensure responsible data sharing and privacy compliance [3, 4]. Such findings underline that the next generation of digital health systems must unite semantic precision, explainable intelligence, and sustainable deployment strategies.

In light of this background, the present study focuses on the automated processing and classification of HL7 v2.5 messages using a hybrid AI architecture. Building upon the foundations of FHIR-based semantic interoperability, the proposed model introduces a dual-layer approach that combines machine-learning classification with semantic validation of extracted entities [1, 2, 5]. This approach contributes to the ongoing effort to integrate structural and semantic reasoning into clinical data pipelines forming an interpretable bridge between traditional HL7 messaging and modern AI-driven healthcare interoperability.

Materials and methods

Data Generation and Structure

To ensure reproducibility while maintaining full data privacy, a synthetic dataset of HL7 v2.5 messages was generated using the Python library `hl7apy`. The dataset simulated the Admission–Transfer–Discharge (ADT) workflow typically observed in hospital information systems.

A total of 6,370 messages were generated for 3,000 synthetic patients, including 3,000 admission events (A01), 2,786 discharge events (A03), and 584 transfer events (A02). Each message contained both administrative and clinical segments, including MSH (Message Header), PID (Patient Identification), PV1 (Patient Visit), OBR (Observation Request), and OBX (Observation Result).

The generator introduced controlled randomness in attributes such as department, physician, and timestamps to emulate the heterogeneity of real-world hospital communication while avoiding any sensitive data reuse.

Each message contained a consistent structure with attributes such as `patient_id`, `event_type`, `department`, `doctor`, `facility`, `timestamp`, and `outcome`. For example, a typical simulated patient record representing a full hospitalization cycle includes an admission (A01) and a discharge (A03):

	patient_id	event_type	department	doctor	facility	timestamp	outcome
0	PAT872246	A01	ICU	Dr. Adams	City Clinic	20250706070923	admitted
1	PAT872246	A03	ICU	Dr. White	City Clinic	20250707140422	discharged
2	PAT717889	A01	ICU	Dr. Johnson	Central Hospital	20230906003150	admitted

Figure 1 – Generate dataset

This synthetic generation process ensured realistic message diversity while preserving the statistical and structural properties of genuine HL7 communications.

Parsing and Feature Engineering

Each HL7 message was parsed using the hl7apy.parser module and converted into a structured, tabular representation suitable for machine-learning analysis.

Multiple feature categories were derived from the raw messages:

- ♦ Structural features: number of segments (num_segments) and total message length (message_length);
- ♦ Temporal features: extracted from timestamps, including hour and day_of_week;
- ♦ Administrative and demographic features: encoded identifiers for visit_type, department, physician, and gender;
- ♦ Derived numerical features: age, segment-to-length ratio, and event frequency.
- ♦ Categorical variables were numerically encoded using the LabelEncoder, while numerical attributes were standardized through Z-score normalization.

Missing values were imputed according to variable type: forward fill for temporal data and median substitution for numerical fields.

Duplicate messages (identical patient_id and timestamp) were removed to ensure consistency.

The resulting dataset formed a clean, normalized input matrix for supervised learning.

Machine Learning Classification

The classification stage aimed to automatically distinguish between message types and detect possible structural anomalies.

Three supervised algorithms were employed: Logistic Regression, Random Forest, and Gradient Boosting. These models were selected for their complementary strengths: interpretability, ensemble robustness, and gradient-based optimization.

Data were divided into training (80%) and testing (20%) subsets using stratified sampling to preserve class proportions. Hyperparameter tuning was performed via GridSearchCV with three-fold cross-validation, optimizing the weighted F1-score.

Performance metrics included overall Accuracy and Weighted F1-score, which provided a balanced evaluation under moderate class imbalance. This baseline configuration later served as the foundation for introducing semantic information.

Semantic Layer and Confidence Metric

To enrich the model with contextual understanding, a semantic layer was implemented that combined Named Entity Recognition (NER) and Regular Expressions (Regex).

NER was performed using spaCy (model en_core_web_sm), while Regex rules targeted domain-specific patterns representing entities relevant to healthcare communication:

- ♦ physician names (e.g., “Dr. Smith”, “Dr. Lee”);
- ♦ healthcare facilities (e.g., “Clinic”, “Hospital”, “Center”);
- ♦ patient identifiers (e.g., “PAT123456”);
- ♦ diagnoses (“COVID-19”, “Asthma”, “Pneumonia”);
- ♦ medications (“Insulin”, “Paracetamol”, “Amoxicillin”).

For each message, entities detected by both methods were compared.

A semantic confidence score was defined as the ratio between the intersection and union of the entities recognized by NER and RegEx:

$$C_{semantic} = \frac{|E_{NER} \cap E_{RegEx}|}{|E_{NER} \cup E_{RegEx}|}$$

This measure quantified how consistently both approaches identified the same semantic elements, thereby reflecting the contextual reliability of each message.

Additional binary indicators (has_doctor, has_hospital) recorded whether valid physician and facility entities were detected.

Semantic Quality Score and Model Enhancement

To integrate semantic information directly into the machine-learning process, a composite semantic quality score (Q) was designed to capture both entity richness and contextual completeness.

The formula was defined as:

$$Q = C_{semantic} + 0.05 * \frac{\min(N_{entities}, 10)}{10} + 0.1 \times H_{doctor} + 0.1 \times H_{hospital}$$

Where

$N_{entities}$ is the total number of entities recognized, and

H_{doctor} , $H_{hospital}$ are binary indicators (1 if present, 0 if absent).

Scores were normalized to the range from 0 to 1 and categorized into four qualitative levels:

Low (< 0.5), Moderate (0.5–0.75), Good (0.75–0.9), and Excellent (≥ 0.9).

This semantic_quality_score was added as an additional feature in the classification dataset.

Retraining the models with this feature enabled the system to account for both structural and semantic coherence, transforming the pipeline into a hybrid AI model that blends syntactic and contextual reasoning.

Evaluation Framework

The evaluation strategy combined structural and semantic criteria. Baseline and hybrid models were compared using Accuracy and Weighted F1-score, alongside semantic indicators such as average confidence and the share of high-quality messages.

All experiments were implemented in Python using pandas, NumPy, scikit-learn, and spaCy.

This methodological framework ensured reproducibility and provided a scalable foundation for future research on intelligent interoperability, supporting both data standardization and semantic interpretability in HL7 message processing.

Results

Dataset Overview and Preprocessing

The final synthetic dataset consisted of 6,370 HL7 ADT messages generated for 3,000 simulated patient lifecycles. Each record represented one transaction in the admission–transfer–discharge workflow, distributed across three event types: A01 (Admission, 3,000 messages), A03 (Discharge, 2,786), and A02 (Transfer, 584).

After parsing and feature extraction, the structured dataset contained 14 standardized columns, including message family, event type, timestamps, patient demographics, and structural indicators such as message length and segment count.

The dataset was validated for completeness, with missing values and duplicates removed during preprocessing. Synthetic generation ensured realistic variability in hospital departments, physicians, and facilities while maintaining structural integrity of the HL7 v2.5 schema.

Baseline Model Training

The initial phase focused on classifying HL7 message event types (A01/A02/A03) using structural and temporal features only.

Three machine-learning algorithms – Logistic Regression, Random Forest, and Gradient Boosting—were optimized through grid search with 3-fold cross-validation.

Baseline results indicated that ensemble methods outperformed linear models, reflecting the non-linear nature of HL7 message structures.

Table 1 – Baseline results

Model	Accuracy	F1 Score	Stage	Version
Logistic Regression	0.5746	0.5463	Tuned	Without Semantic
Random Forest	0.5926	0.6269	Tuned	Without Semantic
Gradient Boosting	0.6813	0.6652	Tuned	Without Semantic

Gradient Boosting achieved the best baseline performance with an F1-score of 0.665, demonstrating robust feature learning from message length, number of segments, and time-based attributes. However, overall performance suggested that structural features alone could not fully capture the contextual semantics of HL7 message content.

Semantic Layer Integration and Retraining

To enhance interpretability, a semantic-quality layer was introduced, incorporating Named Entity Recognition (NER) and Regular Expression (RegEx) validation. This layer extracted entities such as physician names, facility identifiers, and department codes, assigning a semantic confidence score between 0 and 1.

A composite semantic quality score was then computed for each message based on entity count, consistency, and contextual relevance. Retraining all classifiers with this new feature set yielded significant improvements for the ensemble models.



Figure 2 – Impact of Semantic layer on HL7 Message Classification

While Logistic Regression remained largely unchanged (Δ Accuracy = -0.08% , Δ F1 = -0.07%), Random Forest and Gradient Boosting showed substantial gains.

Table 2 – Comparison Results table

Model	Accuracy (Base)	F1 (Base)	Accuracy (+Semantic)	F1 (+Semantic)	Δ Accuracy (%)	Δ F1 (%)
Logistic Regression	0.5746	0.5463	0.5738	0.5456	-0.08	-0.07
Random Forest	0.5926	0.6269	0.6860	0.6974	+9.34	+7.05
Gradient Boosting	0.6813	0.6652	0.7245	0.7241	+4.32	+5.88

The integration of the semantic layer improved classification accuracy and reduced misclassification between transfer (A02) and discharge (A03) messages

Among all models, Random Forest achieved the highest overall improvement, indicating that tree-based ensembles benefited most from semantic contextualization.

Semantic Drift and Quality Correlation

A follow-up analysis explored correlations between the semantic quality score and message-structure parameters.

The event-level comparison revealed that A03 (Discharge) messages achieved the highest semantic consistency, while A01 (Admission) showed the lowest. This reflects the richer contextual detail typically present in discharge summaries compared to initial admissions. These findings demonstrate that semantic coherence was maintained across different structural and temporal conditions, confirming that the hybrid model's semantic reasoning was not biased by message length or density.

Summary of Findings

Overall, the experimental evaluation confirms that combining structural and semantic learning improves HL7 message classification accuracy and interpretability.

The addition of semantic-quality metrics allowed ensemble models to leverage contextual information that purely syntactic models could not capture.

Performance improvements of +4 –9 % across accuracy and F1 demonstrate the practical benefit of semantic integration for real-world interoperability systems.

The next section discusses these results in the context of explainable AI, interoperability frameworks, and their potential alignment with HL7 FHIR modernization.

Discussion

The integration of semantic validation within the machine-learning pipeline noticeably improved both performance and interpretability in HL7 message classification.

Baseline models trained solely on structural descriptors such as message length, number of segments, and timestamps showed stable yet limited predictive capability. After introducing semantic cues derived from Named Entity Recognition and Regular Expression matching, ensemble algorithms, particularly Random Forest and Gradient Boosting, demonstrated measurable gains in both accuracy and F1-score.

This outcome indicates that contextual linguistic patterns within HL7 text carry unique information that complements structural attributes, enhancing the model's ability to distinguish between admission, transfer, and discharge events.

The proposed semantic confidence metric provided an interpretable measure of alignment between textual and structured data. By quantifying the overlap between entities detected through NER and RegEx, the model was able to assess the internal coherence of each message.

This approach allows automated identification of weakly structured or inconsistent messages without the need for manual rule creation, offering a scalable foundation for hospital data-quality monitoring.

Analysis of semantic quality revealed consistent trends within the dataset. Messages classified as admissions (A01) showed the highest semantic quality, followed by discharges (A03), while transfer (A02) messages demonstrated lower coherence. This pattern likely reflects the higher complexity and variability of transfer documentation, where optional or institution-specific segments are common.

Temporal analysis further revealed that messages generated during regular working hours maintained higher semantic confidence compared to those produced at night, suggesting the influence of human workflow and reporting accuracy on message consistency.

Overall, the findings confirm that semantic enrichment enhances both classification accuracy and message reliability. The hybrid model successfully combined structural and contextual reasoning, resulting in a more robust understanding of HL7 communication patterns. Beyond improving predictive metrics, semantic scoring provided an additional interpretability layer, enabling transparent evaluation of message quality and consistency. This integration marks a practical step toward intelligent interoperability systems that can analyze not only the syntax of healthcare data but also its underlying meaning. Similar improvements in semantic-driven interoperability have also been observed in recent FHIR-based frameworks, supporting the broader trend toward combining AI models with contextual validation for more trustworthy clinical data exchange [1], [5].

Conclusion

This study developed and evaluated a hybrid AI model for automated processing and semantic interpretation of HL7 v2.5 messages.

By combining structural feature extraction with semantic validation, the proposed system demonstrated that contextual understanding can significantly enhance the accuracy and interpretability of message classification.

The integration of semantic cues detected through Named Entity Recognition and Regular Expression matching allowed the models to recognize underlying meaning beyond surface-level syntax, leading to measurable performance gains across ensemble classifiers.

The introduction of the semantic confidence metric provided a novel mechanism for assessing message quality.

It enabled the detection of inconsistencies and structural noise within hospital communication workflows, revealing operational patterns such as lower semantic reliability in transfer messages and during non-standard working hours.

These findings confirm that semantic coherence is a critical dimension of interoperability, directly influencing the reliability of data exchange in clinical systems. The presented pipeline offers a reproducible, privacy-preserving, and extensible framework for HL7-based data analysis.

It demonstrates that semantic enrichment can be achieved within existing infrastructures without full system modernization, providing a scalable path toward intelligent and explainable healthcare interoperability.

Future research may extend this approach by integrating domain-specific ontologies, multilingual data streams, and cross-standard mapping to FHIR, further advancing the development of transparent, semantically aware health information systems.

REFERENCES

- 1 Rigas, E.S., Lagakis, P., Karadimas, M., Logaras, E., Latsou, D., Hatzikou, M., Poulakidas, A., Billis, A., and Bamidis, P.D. Semantic interoperability for an AI-based applications platform for smart hospitals using HL7 FHIR. *Journal of Systems and Software*, 215, 112093 (2024). <https://doi.org/10.1016/j.jss.2024.112093>
- 2 Kiourtis, A., Mavrogiorgou, A., Menychtas, A., Maglogiannis, I., and Kyriazis, D. Structurally mapping healthcare data to HL7 FHIR through ontology alignment. *Journal of Medical Systems*, 43 (3), 62 (2019). <https://doi.org/10.1007/s10916-019-1183-y>

3 Nandal, A. Optimizing interoperability in healthcare: AI-driven HL7 and FHIR implementations for seamless data exchange. *Health Informatics Journal* (2024). <https://doi.org/10.63278/jicr.vi.3169>

4 Nopour, R. Using FHIR for data sharing: A scoping review of challenges and facilitators in healthcare settings. *International Journal of Medical Informatics*, 106128 (2025). <https://doi.org/10.1016/j.ijmedinf.2025.106128>

5 Williams, E., Kienast, M., Medawar, E., Reinelt, J., Merola, A., Klopfenstein, S.A.I., Flint, A.R., Heeren, P., Poncette, A.-S., Balzer, F., Beimes, J., von Büna, P., Chromik, J., Arnrich, B., Scherf, N., and Niehaus, S. A standardized clinical data harmonization pipeline for scalable AI application deployment (FHIR-DHP): Validation and usability study. *JMIR Medical Informatics*, 11 (3), e43847 (2023). <https://doi.org/10.2196/43847>

6 Talvik, H.-A., Oja, M., Tamm, S., Mooses, K., Särg, D., Lõo, M., Siimon, Õ.R., Šuvalov, H., Kolde, R., Vilo, J., Reisberg, S., and Laur, S. Repeatable process for extracting health data from HL7 CDA documents. *Journal of Biomedical Informatics*, 150, 104765 (2024). <https://doi.org/10.1016/j.jbi.2024.104765>

7 Loh, H.W., Ooi, C.P., Seoni, S., Barua, P.D., Molinari, F., and Acharya, U.R. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine*, 227, 107161 (2022). <https://doi.org/10.1016/j.cmpb.2022.107161>

8 Holzinger, A., Dehmer, M., Emmert-Streib, F., Cucchiara, R., Augenstein, I., Del Ser, J., Samek, W., Jurisica, I., and Díaz-Rodríguez, N. Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. *Information Fusion*, 79, 263–278 (2022). <https://doi.org/10.1016/j.inffus.2021.10.007>

^{1*}Әбілмәжінова Т.М.,

магистр, ORCID ID: 0009-0007-5798-7722,

*e-mail: 242723@astanait.edu.kz

¹Қуатбаева А.А.,

ассистент-профессор, ORCID ID: 0000-0002-2143-3994 ,

e-mail: a.kuatbayeva@astanait.edu.kz

¹Astana IT University, Астана қ., Қазақстан

HL7 ДЕРЕКТЕРІН ӨНДЕУ МЕН СЕМАНТИКАЛЫҚ ИНТЕРОПЕРАБЕЛЬДІЛІКТІ ҚАМТАМАСЫЗ ЕТУГЕ АРНАЛҒАН ГИБРИДТІ ЖАСАНДЫ ИНТЕЛЛЕКТ МОДЕЛІ

Аңдатпа

Қазіргі денсаулық сақтау жүйелері ауруханалар, зертханалар және цифрлық платформалар арасындағы құрылымдалған ақпарат алмасуға барған сайын тәуелді болып отыр. HL7 v2.x стандарты осындай өзара байланыс үшін негіз болып табылады, алайда оның синтаксисінің өзгермелілігі мен міндетті емес сегменттердің болуы машиналық түсіндіруді қиындатады. Осы мәселені шешу мақсатында құрылымдық оқыту мен семантикалық валидацияны біріктіретін HL7 хабарламаларын автоматтандырылған өңдеуге және жіктеуге арналған гибриді жасанды интеллект моделі әзірленді. Эксперименттік жұмыс барысы 3000 пациенттің өмірлік циклі мен 7000-нан астам ADT хабарламасын қамтитын синтетикалық деректер жиынтығын генерациялаудан, кейін оны парсинг, белгілер инженериясы және бақыланатын оқыту кезеңдерінен өткізу арқылы жүзеге асырылды. Бастапқы классификаторлар ретінде логистикалық регрессия, кездейсоқ орман (Random Forest) және градиенттік бустинг модельдері сыналды. Сонымен қатар, атаулы нысандарды тану (Named Entity Recognition) мен тұрақты өрнектерді (Regular Expressions) біріктіретін семантикалық қабат іске асырылды. Бұл дәрігердің аты-жөні, медициналық ұйымның атауы және диагноз индикаторлары сияқты контекстік белгілерді ескеруге мүмкіндік берді. Қайта оқытудан кейін ансамбльдік модельдер айтарлықтай жақсару көрсетті: Random Forest моделінің дәлдігі 9,3%-ға, ал F1 көрсеткіші 7,0%-ға артты. Нәтижелер семантикалық белгілерді қосу модельдің түсіндірілігімен және тұрақтылығын арттыратынын, сондай-ақ құрылымдық хабарламаларды синтаксистік талдау мен олардың мағыналық мазмұнын түсіну арасындағы алшақтықты қысқартатынын дәлелдеді. Ұсынылған гибриді деректерді өңдеу конвейері интероперабельділік саласындағы интеллектуалды шешімдерді әзірлеу мен FHIR стандартымен үйлесімді жаңа буындағы медициналық деректер алмасу жүйелерін құрудың негізі бола алады.

Түйін сөздер: HL7, FHIR, жасанды интеллект, семантикалық интероперабельділік, атаулы нысандарды тану, түсіндірілетін жасанды интеллект.

¹*Әбілмәжінова Т.М.,
магистр, ORCID ID: 0009-0007-5798-7722,
*e-mail: 242723@astanait.edu.kz

¹Қуатбаева А.А.
ассистент-профессор, ORCID ID: 0000-0002-2143-3994,
e-mail: a.kuatbayeva@astanait.edu.kz
¹Astana IT University, г. Астана, Казахстан

ГИБРИДНАЯ МОДЕЛЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОБРАБОТКИ ДАННЫХ HL7 И ОБЕСПЕЧЕНИЯ СЕМАНТИЧЕСКОЙ ИНТЕРОПЕРАБЕЛЬНОСТИ

Аннотация

Современные системы здравоохранения все в большей степени зависят от структурированного обмена информацией между больницами, лабораториями и цифровыми платформами. Стандарт HL7 v2.x служит основой для такой коммуникации, однако его вариативный синтаксис и наличие необязательных сегментов создают сложности для машинной интерпретации. Для решения этой проблемы была разработана гибридная модель искусственного интеллекта, предназначенная для автоматизированной обработки и классификации HL7-сообщений, объединяющая структурное обучение и семантическую валидацию. Экспериментальный рабочий процесс включал генерацию синтетического набора данных, содержащего 3000 жизненных циклов пациентов и свыше 7000 сообщений ADT, с последующим этапом парсинга, инженерии признаков и обучения с учителем. В качестве базовых классификаторов были протестированы модели логистической регрессии, случайного леса и градиентного бустинга. Дополнительно была реализована семантическая надстройка, объединяющая методы распознавания именованных сущностей (Named Entity Recognition) и регулярных выражений, что позволило учитывать контекстные признаки, такие как имена врачей, наименования медицинских учреждений и диагностические индикаторы. После повторного обучения ансамблевые модели продемонстрировали заметное улучшение: точность модели Random Forest увеличилась на 9,3%, а F1-мера – на 7,0%. Полученные результаты подтверждают, что добавление семантических признаков повышает интерпретируемость модели и ее устойчивость, устраняя разрыв между синтаксическим разбором структурированных сообщений и пониманием их смыслового содержания. Предложенный гибридный конвейер обработки данных может стать основой для интеллектуальных решений в области интероперабельности и разработки совместимых с FHIR систем обмена медицинскими данными нового поколения.

Ключевые слова: HL7, FHIR, искусственный интеллект, семантическая интероперабельность, распознавание именованных сущностей, объяснимый искусственный интеллект.

UDC 004.62:004.75
IRSTI 50.41.21; 28.21.27

<https://doi.org/10.55452/1998-6688-2026-23-2-159-170>

¹**Onayeva A.E.**,
MSc., ORCID ID: 0009-0004-4379-1065
e-mail: al_onayeva@kbtu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

A THREE-TIER INFORMATION LIFECYCLE MANAGEMENT MECHANISM FOR HOT-TIER MEMORY CONTROL IN CONTAINERIZED KAFKA–FLINK STREAMING PIPELINES

Abstract

Real-time streaming systems face persistent memory pressure as continuous data ingestion drives unbounded growth in hot-tier storage. Existing Information Lifecycle Management (ILM) frameworks have been applied primarily to enterprise archival contexts and have not been evaluated within containerized streaming pipelines that employ multi-tier in-memory architectures. This paper presents a lightweight, policy-driven ILM mechanism integrated into a Kafka–Apache Flink pipeline with a three-tier storage model comprising MongoDB (hot-tier), TimescaleDB (warm-tier), and Parquet files (cold-tier). An asynchronous sweeper thread migrates records between tiers according to configurable time thresholds and , preventing hot-tier saturation without disrupting stream processing. Five experiments were conducted to evaluate memory efficiency, tier retrieval latency, threshold sensitivity, scalability, and extended lifecycle behavior. The results demonstrate that ILM reduces MongoDB peak memory usage by 81% (from 106.96 ± 1.63 MB to 20.28 ± 1.81 MB, $p < 0.001$) while Flink throughput and processing latency remain unaffected. Memory bounds hold stably across ingestion rates from 200 to 1,000 messages per second. An extended 90-minute run validates correct three-tier lifecycle operation, with MongoDB remaining bounded, TimescaleDB absorbing 2.28 million warm records, and Parquet accumulating cold archives. These findings confirm that effective, low-overhead ILM can be achieved in containerized real-time pipelines using only native database capabilities and file system operations.

Keywords: Information Lifecycle Management (ILM), Apache Kafka, Apache Flink, MongoDB, TimescaleDB, real-time streaming, memory optimization, tiered storage, containerized pipelines, data archiving.

Received November 11, 2025; revised April 25, May 1, 2026; accepted May 4, 2026.

Introduction

In the current era of digital transformation, data has become the core of enterprise development and strategic competition [1]. The explosive growth of data traffic, driven by IoT sensors, social networks, and mobile applications, is estimated to double every two years, with global volumes projected to reach 221.2 ZB by 2026 [1–3]. As these numbers grow, traditional data management approaches have begun to show significant performance degradation. This challenge has triggered the development of Information Lifecycle Management (ILM), a comprehensive methodological strategy that reduces inefficiencies and optimizes data management by strategically aligning information resources with crucial business objectives and operational requirements.

As data expands in volume, velocity, and variety (3V's), and as extended by to include veracity, variability, and value, organizations face escalating costs and operational bottlenecks [4]. ILM has become a critical methodology to address these challenges by systematically managing data throughout its lifecycle. It ensures that high-value data remains cost-effectively accessible while redundant or low-value information is eliminated early, thereby improving both efficiency and governance.

ILM research has emphasized automation and policy-driven data movement. Govil et al. highlight that automated lifecycle policies reduce manual intervention and improve productivity by aligning data storage with business needs [5]. Similarly, Salutina et al. and Ahmad and Rai discuss how digital transformation and big data frameworks demand advanced ILM strategies to sustain performance and scalability in dynamic environments [6, 7]. These studies collectively underscore ILM’s importance in structuring, classifying, and optimizing massive data systems.

Further investigations demonstrate ILM’s versatility across different domains. Wang applied algorithmic policies such as capacity migration to improve storage utilization, while Shen developed hierarchical storage layers to balance access speed and cost [8, 9]. In specialized contexts, some studies extend ILM principles to mission-critical and secure data lake environments, integrating techniques like anomaly detection and machine learning to enhance lifecycle governance and data quality [10–12].

By utilizing Hierarchical Storage Management (HSM), organizations grade data based on utility and migrate it across tiers: the hot tier holds frequently accessed, time-sensitive data in high-performance storage; the warm tier handles data with decreasing access frequency on mid-tier devices; and the cold tier provides low-cost, high-capacity archival storage for long-term retention and regulatory compliance.

Despite this substantial body of work, a critical gap remains: ILM has not been applied to real-time streaming pipelines in which hot-tier in-memory stores face rapid saturation under continuous message ingestion. In such systems, data arrives at high frequency, and the absence of automated lifecycle policies causes in-memory databases to grow unboundedly, eventually exhausting available resources. Existing ILM frameworks are designed for batch-oriented or enterprise environments and do not address the asynchronous, low-latency requirements of streaming architectures. This gap motivates our study, which introduces a lightweight, container-native ILM mechanism for a Kafka–Apache Flink streaming pipeline with a three-tier storage hierarchy.

Materials and methods

System Architecture

The proposed system is a fully containerized streaming pipeline that combines data ingestion, real-time processing, tiered storage, and lifecycle management within a single Docker Compose deployment. The architecture comprises four functional layers: ingestion, processing, storage, and analytics, as shown in Figure 1.

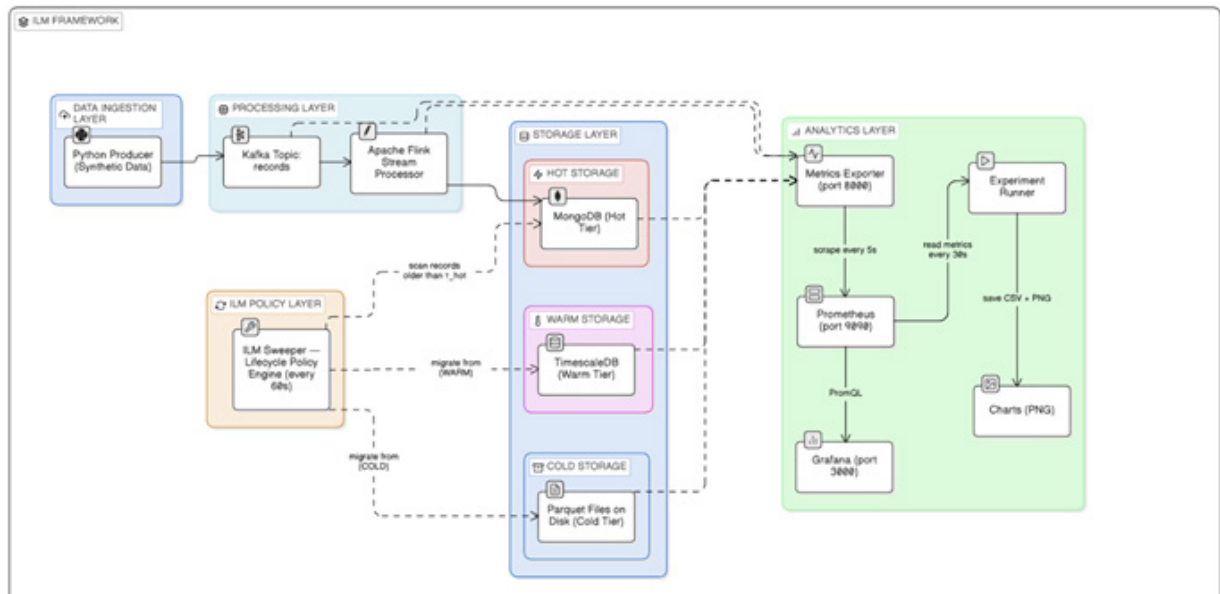


Figure 1 – High-level architecture of the proposed ILM-driven streaming pipeline

The ingestion layer consists of a Python-based Kafka producer that streams synthetically generated records at a configurable message rate (default 300 msg/s). The processing layer employs Apache Flink, which consumes messages from Kafka and routes each record to the appropriate storage tier based on its age. The storage layer implements a three-tier hierarchy: MongoDB (hot tier) holds recently ingested records requiring low-latency access, TimescaleDB (warm tier) stores records that have aged beyond τ_{cold} , and Parquet files on disk (cold tier) serve as long-term archives for records exceeding τ_{cold} . The analytics layer monitors memory usage, throughput, and latency via Prometheus and a background metrics exporter.

Apache Kafka

Apache Kafka (Confluent Platform 7.2.1) serves as the message ingestion buffer. Built on an append-only distributed log architecture, Kafka provides reliable message persistence, ordered delivery within partitions, and the ability to replay messages from any offset [13]. In the proposed system, a single Kafka topic receives streaming metadata from the Python producer. Kafka decouples the producer from the downstream Flink consumer, absorbing bursts in ingestion rate without backpressure propagation.

Apache Flink

Apache Flink (version 1.17) serves as the stream processing engine. Flink implements a unified dataflow model that treats both streaming and batch workloads as pipelined, fault-tolerant dataflows [14]. Unlike micro-batch systems, Flink processes records as a true continuous stream, enabling consistently low processing latency. Flink's stateful operators, native Kafka source connector, and exactly-once delivery semantics make it suitable for the ILM classification task, which requires correct record routing even in the presence of processing failures [15]. The Flink job reads records from Kafka, attaches an insertion timestamp to each message, applies the ILM classification tag, and writes the record to MongoDB.

MongoDB

MongoDB (version 6.0) functions as the hot-tier store. As a document-oriented NoSQL database, MongoDB stores records as BSON documents in memory-mapped collections, enabling sub-millisecond read and write operations for recently ingested data. Each record is keyed by a composite identifier derived from the `record_id` field and ingestion timestamp, enabling efficient age-based queries during sweeper cycles. Memory consumption is tracked using the `serverStatus.mem.resident` counter sampled at one-second intervals.

TimescaleDB

TimescaleDB (version 2.11 on PostgreSQL 15) serves as the warm-tier store. Its time-series hypertable architecture automatically partitions data into fixed-interval chunks, ensuring that recent data remains in memory while older chunks are accessed from disk [16]. This design provides efficient insertion and range-query performance on time-ordered records. The hypertable is partitioned on the record insertion timestamp at one-hour intervals, matching the threshold used in the experiments.

Apache Parquet

Parquet files on a local Docker volume function as the cold tier. Apache Parquet is a columnar storage format that provides high compression ratios through column-level encoding and predicate pushdown during batch scans [17]. Records are serialized as gzip-compressed Parquet files organized by a timestamp-based directory hierarchy (year/month/day/hour). Parquet imposes no ongoing memory cost, requires no running database process, and is natively supported by major batch processing frameworks.

ILM Mechanism

The ILM mechanism classifies each record based on its age relative to two configurable thresholds: the hot threshold and the cold threshold. A record written to MongoDB at insertion time is eligible for warm-tier migration at time t when the following condition is satisfied:

$$t - t_k \geq \tau_{hot} \quad (1)$$

Similarly, a record stored in TimescaleDB at time t becomes eligible for cold archiving when:

$$t - t_j \geq \tau_{cold} \quad (2)$$

An asynchronous sweeper thread executes Algorithm 1 at a fixed interval of 60 seconds throughout each experimental run. The sweeper operates entirely independently of the Flink processing job, introducing no latency on the critical stream processing path.

Algorithm 1: ILM Sweeper Thread

Algorithm 1 ILM Sweeper Thread

Require: τ_{hot} , τ_{cold} , $sweep\_interval = 60\text{ s}$

Ensure: Migrated records across hot $\rightarrow$ warm $\rightarrow$ cold tiers

```

0: repeat
0:    $t_{now} \leftarrow$  current Unix timestamp
0:   {Hot  $\rightarrow$  Warm migration}
0:    $E_{hot} \leftarrow$  query MongoDB WHERE  $(t_{now} - t_k) \geq \tau_{hot}$ 
0:   if  $E_{hot} \neq \emptyset$  then
0:     bulk insert  $E_{hot}$  INTO TimescaleDB
0:     delete  $E_{hot}$  FROM MongoDB
0:   end if
0:   {Warm  $\rightarrow$  Cold migration}
0:    $E_{cold} \leftarrow$  query TimescaleDB WHERE  $(t_{now} - t_j) \geq$ 
     $\tau_{cold}$ 
0:   if  $E_{cold} \neq \emptyset$  then
0:     serialize  $E_{cold} \rightarrow$  Parquet file (gzip, timestamp-
      partitioned)
0:     delete  $E_{cold}$  FROM TimescaleDB
0:   end if
0:   sleep( $sweep\_interval$ )
0: until pipeline stopped = 0

```

Figure 2 illustrates the characteristic sawtooth pattern produced by the sweeper cycle: memory rises as new records accumulate in MongoDB between sweeps, then drops sharply when the sweeper migrates eligible records to TimescaleDB.

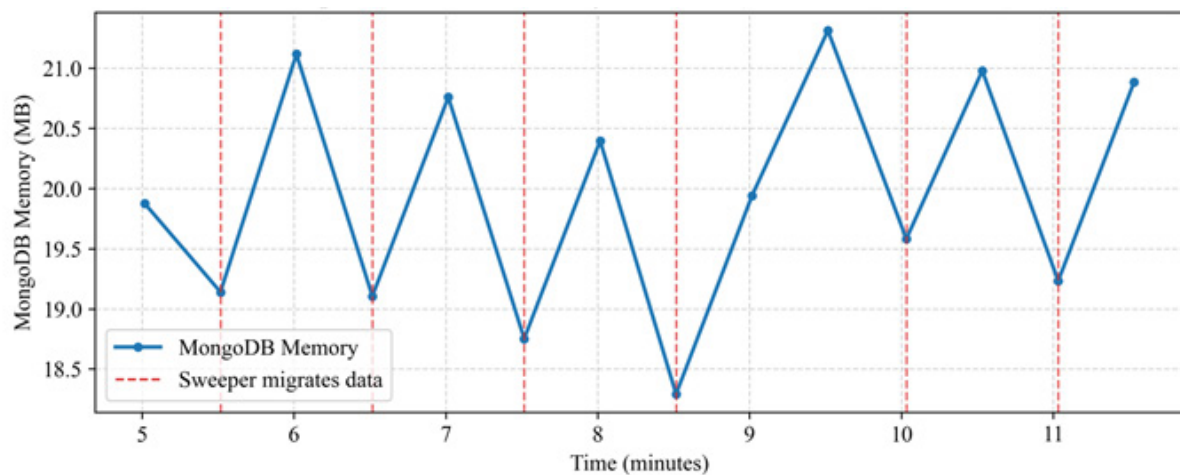


Figure 2 – Sweeper cycle effect on MongoDB memory

Each Kafka message is a JSON payload containing four fields: a unique record identifier, a Unix epoch timestamp added at ingestion time, a random floating-point value, and a categorical label. The Unix epoch timestamp serves as the lifecycle clock for all ILM age calculations.

Table 1 – Fields of a Kafka message

Field	Example Value
record_id	3f7a2c1d-9e4b-4a1f-b8c2-0d5e6f7a8b9c
timestamp	1747498811.13
value	0.7342891
category	analytics

Experimental Setup

All experiments were conducted on a Lenovo Yoga Slim 7 Pro 14ACH5 equipped with an AMD Ryzen 7 5000-series processor (8 cores, up to 4.4 GHz), 16 GB of RAM, and a 1 TB NVMe solid-state drive. The host operating system is Windows 11 with Docker Desktop (version 4.40.0) and WSL2 integration. All pipeline components – Kafka, Flink, MongoDB, TimescaleDB, and the ILM sweeper – run inside Docker containers on an Ubuntu-based backend. Python 3.11 was used for the Kafka producer and ILM sweeper scripts.

Table 2 summarizes the five experimental configurations. Experiments 1 and 1_1 evaluate the core ILM memory benefit under standard and extended run conditions respectively. Experiment 2 quantifies cross-tier retrieval latency. Experiment 3 examines sensitivity to the τ_{hot} threshold, and Experiment 4 assesses scalability under varying ingestion rates.

Table 2 – Experimental settings summary

Parameter	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 1_1
Duration	30 min	30 min	20 min	20 min	90 min
Repetitions	5	5	5 per τ	5 per rate	1
Rate (msg/s)	300	300	300	200/500/1000	300
τ_{hot}	300 s	300 s	60/180/300/600 s	300 s	300 s
τ_{cold}	3600 s	3600 s	3600 s	3600 s	2700 s
Sweeper	ON/OFF	ON	ON	ON	ON
Statistics	Paired t-test	mean $\pm$ std	mean $\pm$ std	mean $\pm$ std	Descriptive

Evaluation Metrics

Four metrics are used to evaluate ILM performance:

- ♦ MongoDB Peak Memory (MB) is the maximum resident memory consumed by MongoDB, measured via the `serverStatus.mem.resident` counter, quantifying the effectiveness of hot-tier memory offloading.

- ♦ Flink Throughput (rps) is the number of records processed per second by the Flink job, confirming that ILM introduces no overhead on the stream processing path.

- ♦ Flink Processing Latency (p95, ms) is the 95th percentile of per-record processing time, assessing tail latency behavior.

- ♦ Tier Retrieval Latency (ms) is the end-to-end query response time when fetching a single record from each storage tier, characterizing the speed-cost tradeoff of the three-tier architecture.

Results and discussion

Experiment 1: Baseline vs. ILM

Experiment 1 compares two configurations at 300 msg/s over 30 minutes: a baseline condition in which the ILM sweeper is disabled and MongoDB retains all records indefinitely, and an ILM-enabled condition in which the sweeper migrates records older than ≈ 300 s to TimescaleDB. Five independent runs were conducted per condition and a paired t-test was applied to peak MongoDB memory measurements.

As can be seen in Table 3, MongoDB memory grew linearly, with no upper bound, from approximately 2.8 MB at $t = 0$ to 106.96 ± 1.63 MB at $t = 1,800$ s when ILM was disabled. During the initial warm-up, memory with ILM enabled reached about 20 MB, and for the next 25 minutes, it fluctuated between 18 and 22 MB. The memory trajectories for every run in both scenarios are displayed in Fig. 3. All three key measurements are shown in parallel with error bars in Figure 4.

Table 3 – Experiment 1: Baseline vs. ILM results

Metric	No ILM	With ILM	Effect
MongoDB Memory (MB)	106.96 ± 1.63	20.28 ± 1.81	81% reduction ($5.3\times$)
Flink Throughput (rps)	467 ± 3.1	471 ± 7.4	No impact
Flink Latency p95 (ms)	1.0	1.0	No impact
p-value (t-test)	—	—	$p = 0.000058$

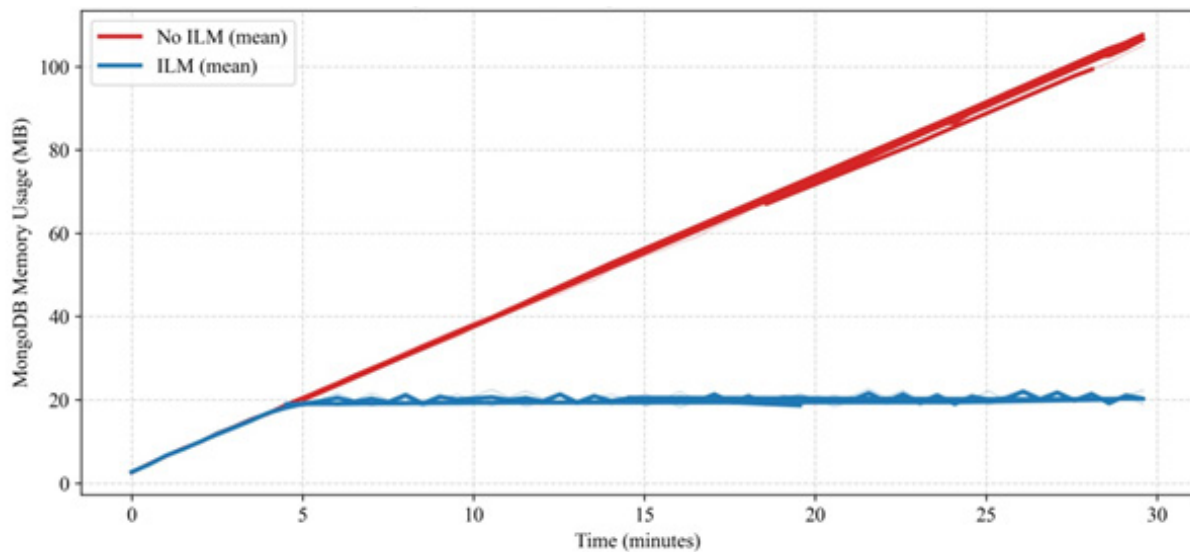


Figure 3 – MongoDB memory usage over 30 minutes

The 81% memory reduction ($5.3\times$ improvement) is statistically highly significant ($p = 0.000058$, far below $\alpha = 0.05$). The low standard deviations (± 1.63 and ± 1.81 MB) confirm high reproducibility across five independent runs. Flink throughput remained virtually identical between conditions (467 vs. 471 rps) and p95 latency was unchanged at 1.0 ms, confirming that the ILM sweeper thread runs with negligible interference to the Flink stream processing job.

Experiment 2: Three-Tier Retrieval Latency

Experiment 2 measures the end-to-end query latency of retrieving a single record from each storage tier during a 30-minute run with all tiers active. A query agent issues one retrieval request per tier every five minutes.

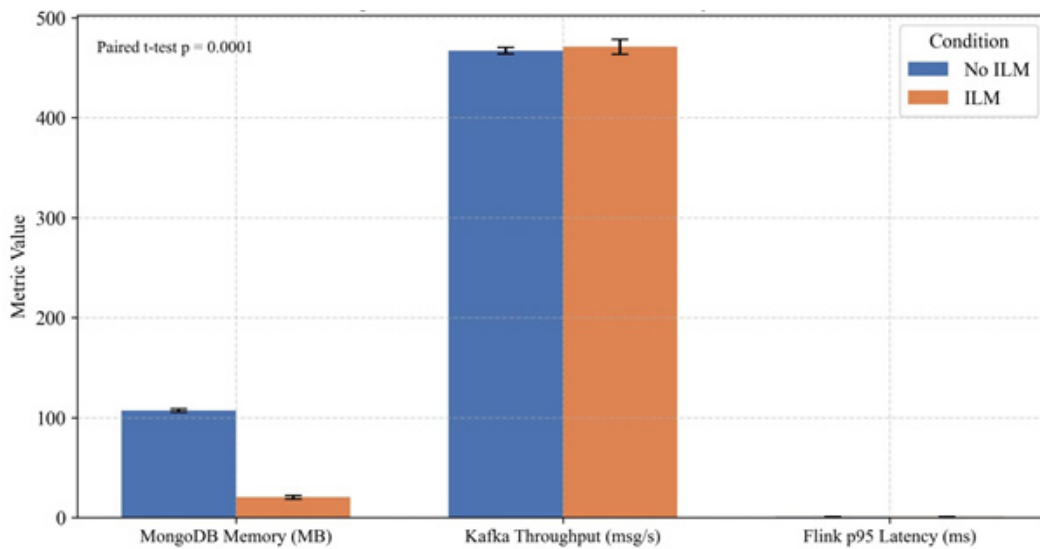


Figure 4 – Experiment 1 summary: key metrics with error bars

Results can be seen in Table 4 and Figure 5. MongoDB (hot tier) delivered queries with a mean latency of 28.9 ± 10.1 ms. TimescaleDB (warm tier) exhibited a mean latency of 677.5 ± 84.9 ms; this figure includes Docker exec overhead and should be interpreted as end-to-end access latency rather than raw database query time. The order-of-magnitude difference between hot and warm tiers is consistent with the access-cost tradeoff inherent in tiered storage. Cold-tier latency was not measured in this experiment as it was not reached within the 30-minute window; cold-tier behavior is validated in Experiment 1_1.

Table 4 – Experiment 2: Tier retrieval latency

Tier	Mean Latency (ms)	Std (ms)
Hot (MongoDB)	28.9	± 10.1
Warm (TimescaleDB)	677.5	± 84.9

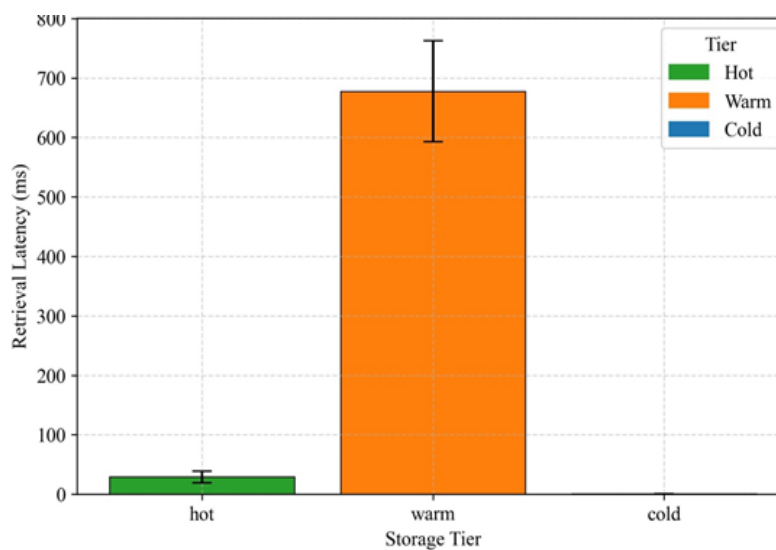


Figure 5 – Experiment 2: retrieval latency by storage tier

Experiment 3: Threshold Sensitivity

Experiment 3 investigates the effect of varying on MongoDB memory usage and hot-tier retrieval latency at 300 msg/s. Four threshold values – 60 s, 180 s, 300 s, and 600 s – were each tested over five independent 20-minute runs.

Table 5 and Figure 6 demonstrate that hot-tier latency (26.5–28.1 ms) and MongoDB memory (17.91–18.58 MB) remained almost constant across all threshold settings. There was no apparent statistically significant difference. This robustness results from the sweeper’s ability to keep up with all arrival rates at 300 msg/s, independent of threshold selection, simplifying operational configuration.

Table 5 – Experiment 3: Threshold sensitivity results

τ_{hot} (s)	MongoDB Memory (MB)	Hot Latency (ms)
60	18.46 ± 5.39	27.1 ± 11.5
180	18.58 ± 5.49	28.1 ± 11.6
300	18.14 ± 5.28	27.7 ± 8.2
600	17.91 ± 5.29	26.5 ± 9.0

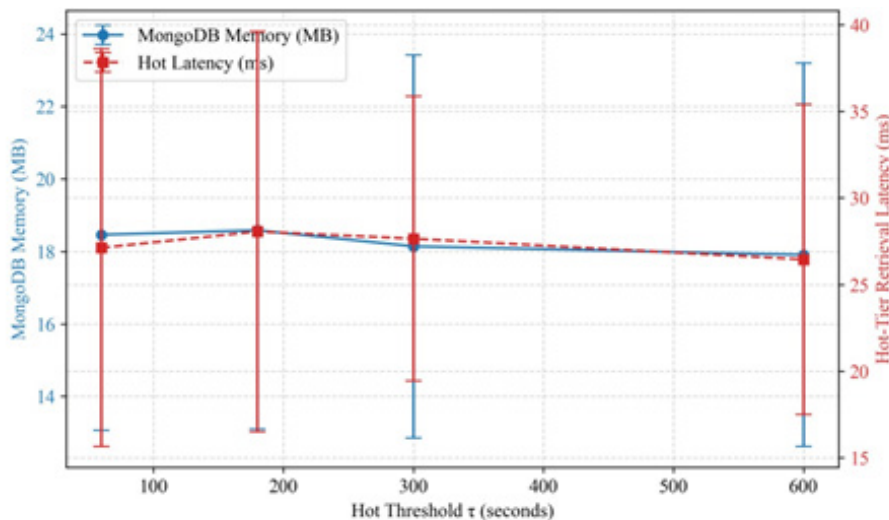


Figure 6 – Experiment 3: Threshold Sensitivity – Memory vs Latency

Experiment 4: Scalability

Experiment 4 evaluates ILM behavior across ingestion rates of 200, 500, and 1,000 msg/s over 20-minute runs with $\tau = 300$ s. MongoDB memory remained constant at about 20 MB at all three rates, as seen in Table 6 and Figure 7. Flink throughput increased with ingestion rate (444–458 rps) while p95 latency stayed at 1.0 ms. The decreased standard deviation in MongoDB memory at increasing rates ($\pm 1.29 \rightarrow \pm 0.42$ MB) indicates that the hot-tier working set is more reliably stabilized under heavier loads when sweeping occurs more frequently.

Table 6 – Experiment 4: Scalability results

Rate (msg/s)	MongoDB Memory (MB)	Flink Throughput (rps)	Latency p95 (ms)
200	20.62 ± 1.29	444 ± 0.2	1.0
500	21.00 ± 0.86	452 ± 11.1	1.0
1000	20.75 ± 0.42	458 ± 2.2	1.0

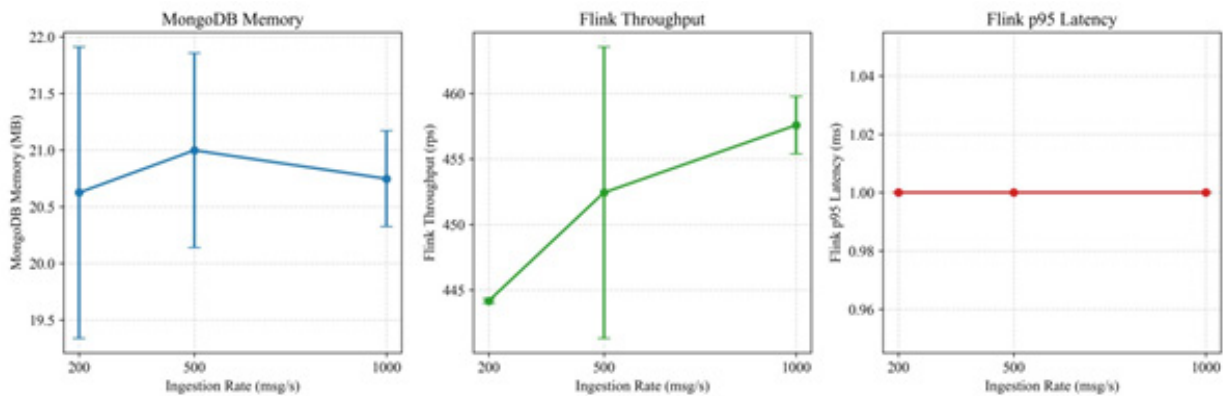


Figure 7 – Experiment 4: scalability across ingestion rates

Experiment 1_1: Extended 90-Minute Validation

Experiment 1_1 was a single 90-minute run at 300 msg/s with $t = 300$ s and a reduced $t = 2,700$ s (45 minutes), designed to activate all three storage tiers within the run duration.

The two-panel result is shown in Fig. 8. For the duration of the 90-minute run, MongoDB hot-tier memory remained bounded at approximately 20 MB. Records were accumulated linearly by TimescaleDB until $t \approx 60$ minutes, at which point initiated cold migration. As records started to flow into Parquet, the warm tier stabilized. At $t \approx 60$ minutes, the first Parquet file appeared, and at the end of the run, there were two files. Over the course of the run, TimescaleDB received 2.28 million warm records. These findings validate proper end-to-end three-tier lifecycle functioning under continuous, prolonged ingestion.

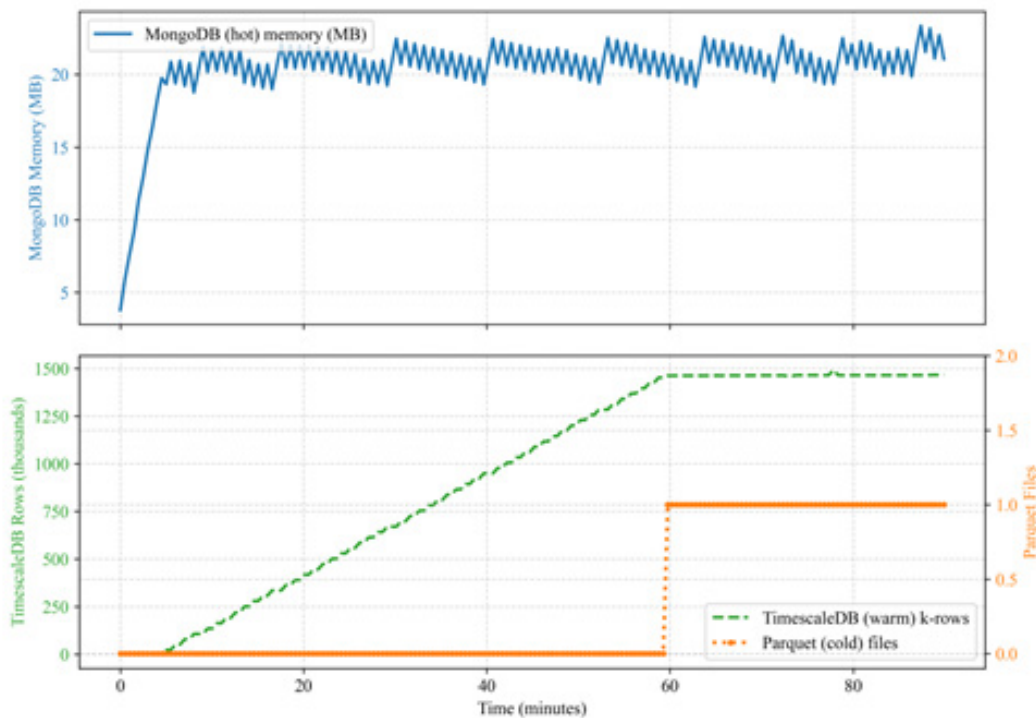


Figure 8 – Experiment 1_1: 90-minute extended run

All five experiments' findings indicate that the proposed ILM system maintains stable stream processing performance, correctly controls the entire three-tier data lifecycle under both short and extended workloads, and achieves stable bounded hot-tier memory regardless of ingestion rate or threshold configuration.

Conclusion

This study presents and evaluates a lightweight, policy-driven Information Lifecycle Management mechanism for containerized real-time streaming pipelines built on Apache Kafka and Apache Flink with a three-tier storage architecture comprising MongoDB, TimescaleDB, and Parquet. The experimental results demonstrate that ILM reduces MongoDB hot-tier peak memory usage by 81% – from 106.96 ± 1.63 MB to 20.28 ± 1.81 MB – with a statistically highly significant effect ($p = 0.000058$) and no measurable impact on Flink throughput or processing latency. Memory bounds remain stable across ingestion rates from 200 to 1,000 messages per second, confirming the scalability of the approach. An extended 90-minute run validates correct three-tier lifecycle operation, with TimescaleDB absorbing 2.28 million warm records and Parquet successfully receiving cold archives.

The system achieves these results using only native database operations and file system primitives, requiring no external orchestration frameworks or specialized hardware, making it immediately deployable in resource-constrained containerized environments.

This study has several limitations. Experiments were conducted on a single laptop rather than a distributed cluster, and performance characteristics may differ at larger deployment scales. The synthetic workload uses a fixed schema with uniform record sizes, which may not reflect real-world data variability. Additionally, the sweeper interval is fixed at 60 seconds rather than dynamically adapted to ingestion rate.

Future work will investigate adaptive threshold scheduling driven by real-time ingestion rate feedback, integration with cloud object storage (Amazon S3, Google Cloud Storage) as the cold tier, and machine learning techniques for predicting optimal record lifetimes from historical access patterns.

REFERENCES

- 1 Ding, Y. Research on management and optimization of big data computing engine based on cloud native technology IT architecture. *Procedia Computer Science*, 243, 910–917 (2024). <https://doi.org/10.1016/j.procs.2024.09.109>
- 2 Statista. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025. Statista Research Department (2024). URL: <https://www.statista.com/statistics/871513/worldwide-data-created/> (accessed 2026.04.10)
- 3 Sharma, C., and Vaid, A. Leveraging SAP information lifecycle management (ILM): Latest insights and applications. *Zenodo*, 5(6), 167–173 (2024).
- 4 Gandomi, A., and Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35, 137–144 (2015).
- 5 Govil, J., Kaur, N., Kaur, H., and Govil, J. Data/information lifecycle management: A solution for taming data beast. *Proceedings of the International Conference on Information Technology: New Generations (ITNG 2008)*, 1226–1227 (2008).
- 6 Salutina, T.Y., Klesareva, E.Y., and Platunina, G.P. Big data as a management decision-making tool in digital business environments. *Proceedings of the 2021 IEEE International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS 2021)*, 893–895 (2021).
- 7 Ahmad, P.H., and Rai, M. Analysis of optimization strategies for big data storage management: A study. *Proceedings of the 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC 2023)*, 1747–1753 (2023).

8 Wang, M.F., Lin, W.T., Tang, C.H., and Tsai, M.F. Constructing storage capacity migration policies for information lifecycle management system. Proceedings of the 12th IEEE International Conference on High Performance Computing and Communications (HPCC 2010), 503–508 (2010).

9 Li-Zhen, S. Research on hierarchical storage of digital library based on the information lifecycle management. Proceedings of the 2010 2nd IEEE International Conference on Information Management and Engineering, 64–66 (2010).

10 HongJu, X., Fei, W., FenMei, W., and XiuZhen, W. Some key problems of data management in army data engineering based on big data. Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA 2017), 149–152 (2017).

11 Miao, T. The optimization of land resource data management model in big data era. Proceedings of the IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC 2024), 763–767 (2024).

12 Muratov, S.Y., and Muravyov, S.B. Framework architecture of a secure big data lake. Procedia Computer Science, 229, 39–46 (2023).

13 Kreps, J., Narkhede, N., and Rao, J. Kafka: A distributed messaging system for log processing. Proceedings of the NetDB Workshop, 11, 1–7 (2011).

14 Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S., and Tzoumas, K. Apache Flink: Stream and batch processing in a single engine. Bulletin of the Technical Committee on Data Engineering, 38(4) (2015).

15 Chintapalli, S., Dagit, D., Evans, B., Farivar, R., Graves, T., Holderbaum, M., and Poulosky, P. Benchmarking streaming computation engines: Storm, Flink and Spark Streaming. Proceedings of the 2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW 2016), 1789–1792 (2016).

16 Kulkarni, A., and Freedman, M. TimescaleDB: SQL made scalable for time-series data. Proceedings of the CIDR Workshop (2017). URL: <https://api.semanticscholar.org/CorpusID:34446750>

17 Apache Parquet Project. Apache Parquet format specification (2024). URL: <https://parquet.apache.org/docs/> (accessed: 2026.04.10).

¹Онаева А.Е.,

магистр, ORCID ID: 0009-0004-4379-1065,

e-mail: al_onayeva@kbtu.kz

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

КАФКА–FLINK КОНТЕЙНЕРЛЕНГЕН АҒЫНДЫҚ ЖҮЙЕЛЕРІНДЕ HOT-TIER ЖАДЫН БАСҚАРУҒА АРНАЛҒАН ҮШ ДЕҢГЕЙЛІ АҚПАРАТ ӨМІРЛІК ЦИКЛІН БАСҚАРУ МЕХАНИЗМІ

Аңдатпа

Нақты уақыттағы ағындық жүйелер үздіксіз деректер ағынының әсерінен жадқа тұрақты жүктемеге ұшырайды, бұл hot-tier сақтау қабатының шексіз өсуіне әкеледі. Ақпараттың өмірлік циклін басқару (Information Lifecycle Management, ILM) тәсілдері негізінен корпоративтік архивтеу жүйелерінде қолданылған және көпдеңгейлі жад архитектурасын пайдаланатын контейнерленген ағындық пайплайндарда жеткілікті деңгейде зерттелмеген. Бұл жұмыста саясатқа негізделген жеңіл ILM механизмі ұсынылады, ол Kafka–Apache Flink пайплайнына енгізілген және үш деңгейлі сақтау моделін қамтиды: MongoDB (hot-tier), TimescaleDB (warm-tier) және Parquet файлдары (cold-tier). Асинхронды тазалаушы ағын (sweeper thread) жазбаларды τ_{hot} және τ_{cold} уақыт шектеріне сәйкес деңгейлер арасында жылжытады, осылайша ағындық өңдеуді бұзбай, hot-tier толып кетуін болдырмайды. Жад тиімділігін, деңгейлер бойынша деректерге қол жеткізу кідірісін, шектерге сезімталдықты, масштабталуды және ұзақ мерзімді жұмыс тәртібін бағалау үшін бес эксперимент жүргізілді. Нәтижелер ILM қолдану MongoDB жадын ең жоғары деңгейде пайдалануды 81%-ға азайтатынын көрсетті (106.96 ± 1.63 МБ-тан 20.28 ± 1.81 МБ-қа дейін, $p < 0.001$), бұл ретте Flink жүйесінің өткізу қабілеті мен өңдеу кідірісі өзгеріссіз қалады. Жад шектеулері секундына 200-ден 1000-ға дейін хабарлама жылдамдығында тұрақты сақталады. 90 минуттық қосымша эксперимент үш деңгейлі өмірлік циклдің дұрыс жұмыс істейтінін дәлелдейді: MongoDB тұрақты шекте қалады, TimescaleDB 2.28

миллион warm-tier жазбасын жинақтайды, ал Parquet cold-tier архивін қалыптастырады. Бұл нәтижелер нақты уақыттағы контейнерленген ағындық жүйелерде деректердің өмірлік циклін тиімді және аз шығынмен басқару тек дерекқорлардың ішкі мүмкіндіктері мен файлдық жүйе операцияларын пайдалану арқылы жүзеге асырылатынын көрсетеді.

Түйін сөздер: ақпараттың өмірлік циклін басқару (ILM), Apache Kafka, Apache Flink, MongoDB, TimescaleDB, нақты уақыттағы ағындық өңдеу, жадты оңтайландыру, көпдеңгейлі сақтау, контейнерленген пайплайндар, деректерді архивтеу.

¹Онаева А.Е.,
магистр, ORCID ID: 0009-0004-4379-1065,
e-mail: al_onayeva@kbtu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

ТРЕХУРОВНЕВЫЙ МЕХАНИЗМ УПРАВЛЕНИЯ ЖИЗНЕННЫМ ЦИКЛОМ ИНФОРМАЦИИ ДЛЯ КОНТРОЛЯ ПАМЯТИ ГОРЯЧЕГО УРОВНЯ В КОНТЕЙНЕРИЗОВАННЫХ СТРИМИНГ-ПАЙПЛАЙНАХ KAFKA–FLINK

Аннотация

Системы потоковой обработки в реальном времени сталкиваются с постоянной нагрузкой на память, поскольку непрерывное поступление данных приводит к неограниченному росту хранилища горячего уровня. Существующие подходы управления жизненным циклом информации (Information Lifecycle Management, ILM) в основном применялись в корпоративных системах архивирования и не были исследованы в контексте контейнеризованных стриминговых пайплайнов, использующих многоуровневые архитектуры хранения в памяти. В данной работе представлен легкий механизм ILM, управляемый политиками и интегрированный в пайплайн Kafka–Apache Flink с трехуровневой моделью хранения: MongoDB (горячий уровень), TimescaleDB (теплый уровень) и файлы Parquet (холодный уровень). Асинхронный поток очистки (sweeper thread) перемещает записи между уровнями в соответствии с настраиваемыми временными порогами и , предотвращая переполнение горячего уровня без нарушения потоковой обработки. Было проведено пять экспериментов для оценки эффективности использования памяти, задержки доступа к данным по уровням, чувствительности к пороговым значениям, масштабируемости и поведения при длительной эксплуатации. Результаты показывают, что применение ILM снижает пиковое использование памяти MongoDB на 81% (с 106.96 ± 1.63 МБ до 20.28 ± 1.81 МБ, $p < 0.001$), при этом пропускная способность и задержка обработки в Flink остаются неизменными. Ограничения по памяти стабильно соблюдаются при скорости поступления данных от 200 до 1000 сообщений в секунду. Дополнительный эксперимент продолжительностью 90 минут подтверждает корректную работу трехуровневого жизненного цикла: MongoDB остается в заданных пределах, TimescaleDB аккумулирует 2.28 миллиона записей теплого уровня, а Parquet формирует холодный архив. Полученные результаты подтверждают, что эффективное и малозатратное управление жизненным циклом данных возможно в контейнеризованных системах потоковой обработки в реальном времени с использованием только встроенных возможностей баз данных и операций файловой системы.

Ключевые слова: управление жизненным циклом информации (ILM), Apache Kafka, Apache Flink, MongoDB, TimescaleDB, потоковая обработка в реальном времени, оптимизация памяти, многоуровневое хранение, контейнеризованные пайплайны, архивирование данных.

ӘОЖ 004.8+004.912:61:575
ҒТАХР 28.23.21

<https://doi.org/10.55452/1998-6688-2026-23-2-171-186>

^{1,2}**Самбетбаева М.А.,**

PhD, қауымдастырылған профессор, ORCID ID: 0000-0001-9358-1614,
e-mail: sambetbayeva_ma_1@enu.kz,

^{1,2*}**Серикбаева С.К.,**

PhD, доцент м.а., ORCID ID: 0000-0002-3627-3321,

*e-mail: inf_8585@mail.ru

^{1,3}**Сұлтанғазиева А.Н.,**

магистр, докторант, ORCID ID: 0009-0009-9038-5234,

e-mail: anara77777@mail.ru

^{1,4}**Мукажанов Н.К.,**

PhD, қауымдастырылған профессор, ORCID ID: 0000-0003-4835-5751,

e-mail: nurzhan.mukazhanov@narxoz.kz

^{1,2}**Абдығалым Б.Х.**

т.ғ.м., докторант, ORCID ID: 0009-0001-8872-7428,

e-mail: bayangali.abd@gmail.com

¹«Q» University, Алматы қ., Қазақстан

²Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

³Астана халықаралық университеті, Астана қ., Қазақстан

⁴Нархоз университеті, Алматы қ., Қазақстан

ГЕНЕТИКАЛЫҚ АУРУЛАР ТУРАЛЫ АҚПАРАТТЫ АЛУ ҮШІН АННОТАЦИЯЛАНҒАН МЕДИЦИНАЛЫҚ МӘТІНДЕР КОРПУСЫН ҚҰРУ

Аңдатпа

Мақалада экзомдық секвенирлеу нәтижелері бойынша алынған орыс тіліндегі клиникалық мәтіндерден құралған аннотацияланған корпус ұсынылады. Аталған корпус гендерге, мутацияларға, тұқым қуалайтын ауруларға, фенотиптік белгілерге және олардың клиникалық маңыздылығына қатысты атаулы нысандар мен семантикалық байланыстарды автоматты түрде анықтау міндеттерін қолдау мақсатында әзірленген. Корпусты қалыптастыру барысында нақты клиникалық экзомдық секвенирлеу есептері пайдаланылып, деректер алдын ала анонимдендіру және мәтінді нормалау кезеңдерінен өткізілді. Таңбалау үдерісінде HGVS, OMIM, ClinVar және HPO сияқты халықаралық стандарттар мен білім базалары басшылыққа алынып, биомедициналық ақпараттың бірізділігі мен дәлдігі қамтамасыз етілді. Корпуста 25 000-нан астам биомедициналық нысан және 6000-нан астам семантикалық байланыс қамтылып, бұл оны көлемі мен мазмұны жағынан клиникалық генетика саласындағы маңызды ресурсқа айналдырады. Аннотациялау бірнеше сарапшының қатысуымен қолмен жүргізіліп, нәтижелер кросс-тексеру арқылы салыстырылды, сондай-ақ аннотациялаушылар арасындағы келісім деңгейі арнайы метрикалар арқылы бағаланды. Алынған нәтижелер корпустың жоғары сапасын және сенімділігін көрсетеді. Дайын корпус медициналық генетика саласындағы табиғи тілді өңдеу модельдерін оқыту мен бағалауға, клиникалық шешім қабылдауды қолдау жүйелерін дамытуға және генетикалық деректерді құрылымдауға арналған қолданбалы зерттеулерде тиімді пайдалануға мүмкіндік береді.

Түйін сөздер: генетикалық аурулар, медициналық мәтінді өңдеу, аннотацияланған корпус, экзомдық секвенирлеу, клиникалық мәтіндер, атаулы нысандарды автоматты тану (NER), семантикалық қатынастарды алу (RE).

Кіріспе

Генетикалық аурулар – әлемдік денсаулық сақтау саласындағы маңызды мәселелердің бірі. Бүкіл әлем бойынша 300 миллионнан астам адам сирек кездесетін ауруларға шалдыққаны белгілі [1]. Orphanet дерекқорына сәйкес 6172 сирек ауру анықталған, олардың 71,9%-ының генетикалық шығу тегі бар [2]. Жалпы халықтың 3,5–5,9%-ы сирек ауруға шалдықса, бұл 263–446 миллион адамға тең болады [2]. Сирек кездесетін аурулардың көпшілігі өмірге қауіп төндіретіндей күрт (жедел) түрде дамымайды, яғни баяу өрбиді немесе симптомдары жеңіл болуы мүмкін. Дегенмен, генетикалық деректерді клиникалық тұрғыдан талдау үшін дәрігерлердің экзомдық секвенирлеу нәтижелеріне негізделген клиникалық қорытындылары басты ақпарат көзі болып табылады.

Соңғы жылдары клиникалық генетика саласы күрделі молекулалық деректерді талдаудың жаңа деңгейіне көтеріліп, экзомдық (WES) және геномдық (WGS) секвенирлеудің клиникалық практикаға енуі жеделдей түсті. Бұл технологиялар тұқым қуалайтын ауруларды дәл диагностикалауға мүмкіндік берсе де, олардың нәтижелері көбіне құрылымсыз мәтіндік есептер түрінде ұсынылады, сондықтан автоматты талдауды қиындатады [3]. Клиникалық есептердегі генетикалық нұсқалар, фенотиптік сипаттамалар, отбасылық анамнез және интерпретациялық түсіндірмелер біркелкі емес лингвистикалық құрылымдар арқылы беріледі. Соның салдарынан ақпаратты автоматты түрде құрылымдау өзекті мәселелердің біріне айналды.

Клиникалық мәтіндерден ақпарат алу мәселесіне арналған зерттеулер көрсеткендей, клиникалық есептер ғылыми мәтіндерге қарағанда анағұрлым күрделі, контекстке тәуелді синтаксиске ие және қысқартулар мен терминдердің ауыспалы нұсқалары жиі кездеседі [3]. Мұндай ерекшеліктер дәстүрлі NLP модельдерінің тиімділігін төмендетіп, клиникалық тілге бейімделген арнайы корпус пен үлгілеу әдістерінің қажеттілігін арттырады. Сол себепті, медициналық ақпараттық жүйелерде, оның ішінде клиникалық шешім қолдау құралдарында қолданылатын NLP әдістерін дамыту үшін жоғары сапалы, аннотацияланған деректердің болуы аса маңызды [4].

Генетикалық есептердің ішінде ерекше қиындық туғызатын міндеттердің бірі – гендік нұсқаларды (variants) тану және оларды HGVS стандартына сәйкестендіру. Зерттеулер көрсеткендей, генетикалық мутацияларды автоматты түрде анықтау үшін арнайы машықтанған NER модельдері қажет және мұндай модельдер тек тиісті домендік корпус болған жағдайда ғана жоғары дәлдікке қол жеткізеді [5]. Бұл клиникалық генетика мәтіндеріне арналған сапалы корпустардың жетіспеушілігін айқын көрсетеді.

Клиникалық NLP саласының 2020 жылғы жағдайын талдаған халықаралық зерттеулер де дәл осы мәселеге назар аударады: ағылшын тілінде биомедициналық мәтіндерге арналған бірнеше корпус болғанымен, нақты клиникалық жазбаларға бағытталған домендік ресурстардың тапшылығы байқалады, ал көптілді корпустар тіпті аз [6]. Орыс тілінде клиникалық есептерге арналған корпус жоқтың қасы, бұл генетикалық есептерді автоматты талдау бағытында зерттеулердің баяу дамуына әкеледі.

Соңғы жылдары трансформер негізіндегі алдын ала оқытылған модельдер (BioBERT, ClinicalBERT, PubMedBERT) биомедициналық мәтіндер үшін жоғары нәтижелер көрсетіп келеді. Дегенмен, бұл модельдер негізінен ағылшын тіліндегі ғылыми деректерде оқытылғандықтан, клиникалық жазбаларды, оның ішінде генетикалық есептерді өңдеу кезінде қосымша бейімдеуді талап етеді [7]. Осыған байланысты нақты клиникалық контентке негізделген аннотацияланған корпустардың жасалуы – көптілді медициналық NLP жүйелерін дамытудағы шешуші қадам.

Бұл жұмыс орыс тіліндегі клиникалық генетикалық есептерден ақпарат алуға арналған, халықаралық стандарттармен үйлескен аннотацияланған корпустардың болмауымен өзекті болып отыр. Сондықтан бұл зерттеу сол олқылықты толтыруды мақсат етеді және генетикалық есептерден автоматты ақпарат алу жүйелерін құруға негіз болатын жаңа корпус әзірлеуді ұсынады.

Жұмыстарға шолу

Биомедициналық табиғи тілді өңдеу (BioNLP) саласындағы соңғы жылдардағы жетістіктер ағылшын тіліндегі мәтіндерге арналған көптеген жоғары сапалы аннотацияланған корпустардың пайда болуына әкелді. Осы ресурстардың ішінде BC5CDR [8], CRAFT [9] және BioRED [10] сияқты корпустар уақыт өте келе эталондық болып қалды. BC5CDR корпусы химиялық заттар мен аурулар арасындағы қатынастарды шығаруға бағытталса [8], CRAFT корпусы онтологиялармен тураланған түсініктердің терең аннотациясын ұсынады [9]. Сонымен қатар, BioRED корпусы гендер, аурулар және химиялық заттар сияқты әртүрлі нысандар арасындағы қатынастарды қамтиды, бұл оны күрделі биомедициналық білімді шығару үшін құнды ресурсқа айналдырады [10].

Атап айтқанда, гендік нұсқалар мен мутацияларды тануға арналған зерттеулер елеулі прогреске қол жеткізді. Bangalore және әріптестері (2021) BioBERT тәрізді алдын ала оқытылған трансформерлік модельдерді қолдана отырып, еркін мәтіндерден мутацияларды анықтау мен олардың нормалану процесіндегі өзекті мәселелерді талдаған [5]. Олардың жұмысы домендік білімі бар модельдердің маңыздылығын көрсетті, бірақ сонымен бірге мұндай модельдерді оқыту үшін жеткілікті аннотацияланған деректердің болуының шектеулілігін де атап өтті [5].

Дегенмен, жоғарыда аталған корпустар негізінен PubMed және PMC сияқты көздерден алынған ғылыми әдебиеттерге негізделген. Нақты клиникалық есеп берулерден, мысалы, экзомдық секвенирлеу нәтижелері бойынша клиникалық қорытындылардан құралған корпустар әлі де сирек кездеседі. 2020 жылы Wang және оның зерттеу тобы клиникалық мәтіндерді өңдеу кезінде жиі кездесетін күрделі синтаксистік құрылымдар, көпмағыналылық және қысқартулардың молдығы сияқты факторларды қарастырып, олардың дәстүрлі NLP әдістерінің қолданылуын қиындататынын көрсетті [3]. Ал сол жылы Neveol бастаған зерттеушілер клиникалық NLP саласына арналған кең көлемді шолу жүргізіп, нақты клиникалық деректерге сүйенген көптілді ресурстарды дамытудың өзектілігін атап өткен [6].

Орыс тіліндегі биомедициналық NLP контекстінде бірнеше бастамалар пайда болды. NEREL-BIO [11] және RuUMLS-Nested [12] сияқты жобалар орыс тілдік биомедициналық мәтіндерге арналған аннотацияланған ресурстарды құруға бағытталған. Алайда, бұл ресурстар негізінен ғылыми абстракттерге негізделген және клиникалық генетикалық есептердің нақты лингвистикалық және семантикалық ерекшеліктерін, мысалы, Human Genome Variation Society (HGVS) номенклатурасы бойынша күрделі гендік нұсқалардың көріністерін толық қамтымайды [13].

Клиникалық генетикалық контекстте ақпаратты шығаруға арналған корпустардың болмауы ерекше проблемалар тудырады. Cohen және оның зерттеу тобы (2020) ғылыми әдебиеттерден гендік варианттар жөніндегі мәліметтерді алуда туындайтын қиындықтарды талдағанымен, клиникалық есептерді өңдеу аспектісіне арнайы тоқталмаған [4]. Клиникалық есептер жиі кірістірілген нысандарды (мысалы, ген ішіндегі HGVS белгілеуі) және клиникалық тұрғыдан маңызды қатынастарды (мысалы, нұсқаның патогенділігі) қамтиды, оларды дұрыс аннотациялау үшін арнайы схемаларды қажет етеді.

Трансформер архитектурасына негізделген алдын ала оқытылған тіл модельдерінің пайда болуы (BioBERT [14], PubMedBERT [15]) биомедициналық NLP саласында тұтас революция жасады. Yoop және әріптестері (2020) бұл модельдердің биомедициналық мәтіндерді өңдеудегі тиімділігін дәлелдегенімен, олардың нақты клиникалық домендерге, әсіресе ағылшын тілінен өзге тілдердегі салаларға бейімделуін қамтамасыз ету үшін қосымша зерттеулердің қажет екенін атап өткен [7]. Осыған байланысты, нақты клиникалық деректер негізінде оқытылған және көптілді контекстке бейімделген модельдерді дамыту өзекті мәселе болып қала береді.

Халықаралық BioNLP серіктестік жобалары мен n2c2 сияқты клиникалық NLP сынақтары клиникалық есептерден ақпарат шығару мәселелерін белсенді түрде зерттеді [3, 6]. Дегенмен, бұл жобалардың басым көпшілігі ағылшын тіліндегі деректерге аударған. Осылайша, орыс тілдік клиникалық генетикалық есептерге арналған жоғары сапалы, қолмен аннотацияланған корпус әлі де жоқ, бұл осы маңызды доменде зерттеулер мен технологиялық дамуды шектейді.

Биомедициналық NLP саласындағы негізгі корпустарды салыстыру олардың күшті және әлсіз жақтарын анық көрсетеді. Төмендегі 1-кестеде корпустардың салыстырмалы талдауы ұсынылған.

Кесте 1 – Биомедициналық NLP бойынша негізгі аннотацияланған корпустарды салыстыру

Корпус атауы	Тілі	Дереккөзі	Негізгі нысандар	Ерекшеліктері	Шектеулері
BC5CDR [8]	Ағылшын	Ғылыми мақалалар	Химия, ауру	Химия-ауру қатынастары	Тек ғылыми әдебиет
CRAFT [9]	Ағылшын	Ғылыми мақалалар	Ген, ауру, белок	Онтологиялармен терең туралау	Түпнұсқа мәтін қиын, клиникалық емес
BioRED [10]	Ағылшын	Ғылыми мақалалар	Ген, ауру, химия	Бірнеше нысан түрлері арасындағы қатынастар	Тек ғылыми әдебиет, HGVS жоқ
NCBI Disease [16]	Ағылшын	Ғылыми мақалалар	Ауру	MeSH-ке нормалау	Тек ауру атауларына бағытталған
NEREL-BIO [11]	Орыс/Ағылшын	PubMed абстракттері	Ген, Ауру, Белок	Көптілділік, UMLS-ке сілтемелер	Клиникалық емес, HGVS қамтылмаған
RuUMLS-Nested [12]	Орыс	Биомедициналық мәтіндер	Ген, Ауру, Дәрі	Ішкі (nested) нысандар, UMLS	Негізінен ғылыми мәтіндер

Кестеден көрініп тұрғандай, барлық негізгі корпустар ағылшын тіліндегі ғылыми әдебиеттерге негізделген. Орыс тіліндегі соңғы жобалар (NEREL-BIO, RuUMLS-Nested) маңызды болғанымен, олар да ғылыми контентке шектеледі және клиникалық генетиканың нақты қажеттіліктерін, мысалы, HGVS номенклатурасын немесе нұсқаның патогенділігі сияқты клиникалық интерпретациялық нысандарды толық қамтымайды.

Әзірленетін GENEXOM корпусы осы олқылықтарды ескере отырып, келесі қағидалар бойынша құрылатын болады:

- ◆ Дереккөз: нақты клиникалық экзомдық секвенирлеу есептері пайдаланылады;
- ◆ Тіл: орыс тіліне негізделеді;
- ◆ Аннотацияның тереңдігі: халықаралық стандарттармен (HGVS, OMIM, ClinVar, HPO) сәйкестендірілген көпқабатты нысандар мен семантикалық қатынастар қарастырылады;
- ◆ Домендік маңыздылығы: генетикалық диагностикаға тікелей қатысты нысандарды қамту жоспарланған.

Осылайша, қазіргі заманғы BioNLP ландшафты құрылымдық тұрғыдан бай болса да, ол біздің зерттеуіміздің негізгі мақсаты болып табылатын нақты клиникалық генетикалық есептерге арналған көптілді, доменге арналған ресурстар бойынша айтарлықтай олқылықты сақтайды. GENEXOM корпусы осы олқылықты толтыруға, орыс тіліндегі клиникалық генетика үшін бірегей, халықаралық стандарттармен тураланған және көпқабатты аннотацияланған ресурсты ұсынуға бағытталған.

Материалдар мен әдістер

1) Дереккөздер және алдын ала өңдеу

GENEXOM корпусының негізін 2018–2023 жж. аралығында жиналған 200 нақты экзомдық секвенирлеу (Whole Exome Sequencing, WES) бойынша клиникалық есептер құрайды.

Деректер жинағы Қазақстанның түрлі медициналық генетикалық орталықтарынан анонимді түрде алынды. Корпустың клиникалық сәйкестігі мен деректердің алуан түрлілігін қамтамасыз ету үшін есептерді іріктеу келесі критерийлер бойынша жүргізілді:

- ♦ Экзомдық секвенирлеу нәтижелері бойынша толық клиникалық қорытындының болуы;
- ♦ Гендік нұсқалардың HGVS стандарты бойынша сипаттамасы;
- ♦ Фенотипік сипаттамалардың егжей-тегжейлі сипаттамасы;
- ♦ Аурудың OMIM идентификаторымен немесе атауымен анықталуы;
- ♦ Нұсқаның ClinVar дерекқорына сілтемесінің болуы.

Деректерді жинау және алдын ала өңдеу конвейері төмендегі кезеңдерден тұрды (1-сурет):



Сурет 1 – GENEXOM корпусын құрудың деректерді өңдеу конвейері

1. Деректерді жинау және анонимдеу. Бастапқы кезеңде клиникалық-диагностикалық мекемелерден алынған медициналық есептердің PDF және DOCX форматтары жинақталып, қауіпсіздік және конфиденциалдылық талаптарына сәйкес, жеке басын анықтауға болатын барлық ақпараттар (пациенттің аты-жөні, туған күні, жеке куәлік нөмірі, мекен-жайы) деректерді өңдеудің бастапқы кезеңінде жойылды. Анонимдеу процедурасы автоматты түрде жүргізіліп, артын қалдырмай тексерілді.

2. Мәтінді парсингтеу және шығару. Клиникалық есептердің құрылымсыз мәтіндерін өңдеу үшін PDFPlumber [17] және python-docx [18] кітапханалары қолданылды. Бұл кітапханалар медициналық құжаттардың күрделі форматтарымен (кестелер, баспа беттері, суреттер) тиімді жұмыс істей алады. Әсіресе, PDFPlumber кітапханасы кестелерді және күрделі құрылымды анықтауда жоғары дәлдік көрсеткен [17]. Парсингтеу нәтижесінде алынған мәтіндер біріктіріліп, бір файл ретінде сақталды.

3. Мәтінді нормалау және бірыңғайландыру. Клиникалық генетикалық есептердегі терминологиялық әртүрлілікті жою үшін келесі нормалау операциялары жүргізілді:

♦ Медициналық аббревиатураларды анықтау және оларды толық атауларымен ауыстыру [8]. Мысалы, «AD» – «autosomal dominant» (аутосомдно-доминантты), «AR» – «autosomal recessive» (аутосомдно-рецессивті), «VUS» – «variant of uncertain significance» (белгісіз маңыздылығы бар нұсқа).

♦ HGVS номенклатурасына сәйкес гендік нұсқаларды тану және нормалау [13]. Мысалы, «c.5266dupC», «p.Glu6Val» т.б. Бұл процесте HGVS стандартына сәйкес емес жазулар (мысалы, «дупликация цитозина») формальды белгілерге аударылды.

♦ Терминдерді стандарттау және орфографиялық қателерді түзету. Мысалы, «brca1» – «BRCA1», «хромосома 13» – «хромосома 13». Ген атауларының дұрыс жазылуын тексеру үшін HGNC [19] дерекқоры пайдаланылды.

4. Деректердің сапасын тексеру. Өңделген мәтіндердің дәлдігін тексеру үшін арнайы әзірленген Python скрипттері қолданылды. Скрипттер төмендегідей тексерулерді жүргізді:

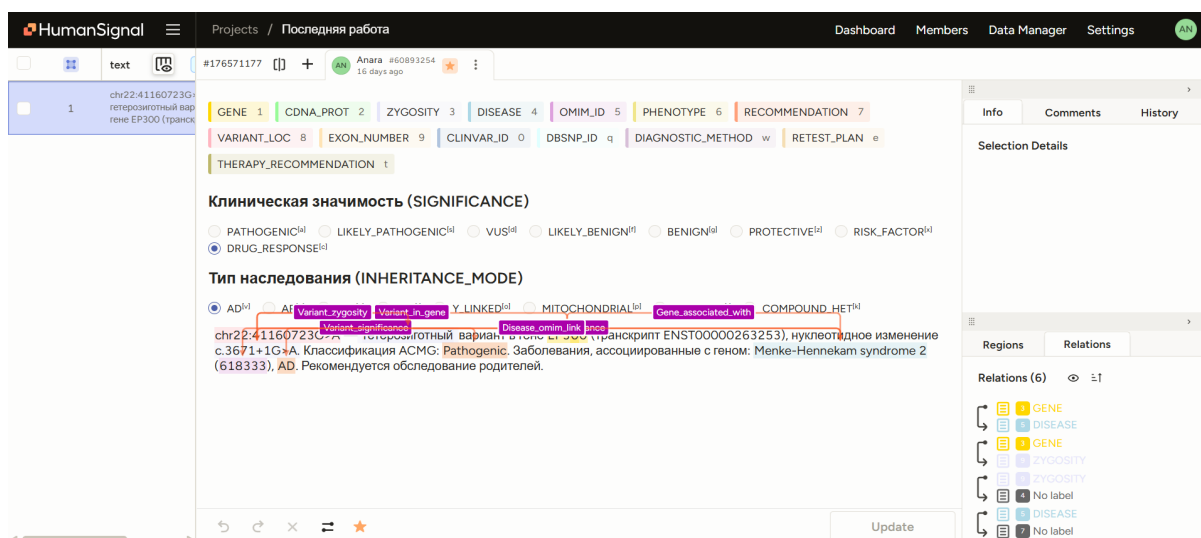
- ♦ HGVS форматына сәйкестікті тексеру;
- ♦ Ген атауларының HGNC стандартына сәйкестігін растау;
- ♦ OMIM және ClinVar идентификаторларының форматын тексеру.

5. Аннотациялауға дайындық. Алдын ала өңделген мәтіндер аннотациялау процесіне дайындалды. Мәтіндер Label Studio платформасына жүктеліп, аннотациялау схемасына сәйкес бапталды [20]. Әрбір құжатқа арнайы идентификатор тағайындалып, аннотациялау тарихын бақылау жүйесі орнатылды.

Деректерді өңдеудің әрбір кезеңінде стандартты процедуралар қолданылып, нәтижелердің тұрақтылығы қамтамасыз етілді. Бұл тәсіл корпус құру процесінің қайталанатындығын және өңделген деректердің жоғары сапасын қамтамасыз етті.

2) Аннотациялау схемасы және процесі

Аннотациялау процесі корпус құрудың негізгі кезеңі болып табылады. Аннотациялау үшін HumanSignal компаниясының Label Studio ашық бастапқы кодты платформасы таңдалды [20]. Бұл таңдау платформаның келесі артықшылықтарына негізделді: ішкі (nested) және қиылысатын (overlapping) нысандарды қолдау, семантикалық қатынастарды белгілеу мүмкіндігі, JSON сияқты машиналық өңдеуге ыңғайлы форматтарда деректерді экспорттау, сондай-ақ бірнеше аннотаторлармен бірлескен жұмыс істеуге арналған интуитивті интерфейс. Платформаның аннотациялау интерфейсінің көрінісі 2-суретте көрсетілген.



Сурет 2 – Label Studio (HumanSignal) платформасындағы аннотациялау интерфейсінің көрінісі

2-суретте көрсетілгендей, аннотациялау интерфейсі үш негізгі аймақтан тұрды:

♦ Орталық панель: Аннотация процесі үшін бастапқы мәтін орналасты. Аннотаторлар тікелей осы аймақта мәтін бөліктерін тұтас ретінде бөліп (highlight), оларға сәйкес тегтерді тағайындайды;

♦ Тегтер панелі: Алдын ала анықталған нысан түрлерінің (GENE, CDNA_PROT, т.б.) тізімі орналасты. Аннотатор бөліп алған мәтін аралығына бір немесе бірнеше тегтерді тағайындай алатын;

♦ Басқару панелі: Тапсырмалар арасында навигациялау, аяқталған аннотацияларды сақтау және қателерді түзету үшін басқару элементтері берілген.

Интерфейстің ерекшелігі – іріктеу-тағайындау әдісінің қарапайымдылығы және бір мәтін ішінде әртүрлі түстік кодтау арқылы бірнеше нысан түрлерін бір мезгілде көру мүмкіндігі. Мысалы, 2-суретте «EP300» тіркесі GENE ретінде сары түспен, ал «Pathogenic» сөзі SIGNIFICANCE ретінде басқа түспен белгіленген. Сонымен қатар, платформа ішкі құрылымдарды (мысалы, «BRCA1 c.5266dupC») тіркесінде «BRCA1» GENE, ал «c.5266dupC» CDNA_PROT ретінде) дәл шекаралармен белгілеуге мүмкіндік берді.

Аннотациялау схемасы халықаралық түрде танылған биомедициналық стандарттар мен білім базалары негізінде әзірленді. Негізгі нысандар (entity) түрлері және олардың сәйкес келетін стандарттары 2-кестеде ұсынылған.

Кесте 2 – Аннотацияланған негізгі нысандар және олардың сипаттамалары

Нысан атауы	Сипаттама	Қолданылған стандарт
GENE	Геннің атауы (HGNC сәйкестендірілген)	HGNC [19]
CDNA_PROT	Мутацияның HGVS-форматтағы түрленуі	HGVS [13]
ZYGOSITY	Зиготалық күйі (гетерозиготалық немесе гомозиготалық)	Клиникалық генетика [21]
DISEASE	Ауру немесе синдром	OMIM [22]
OMIM_ID	OMIM дерекқорындағы идентификатор	OMIM [22]
PHENOTYPE	Науқас белгілері мен симптомдары	HPO [23]
RECOMMENDATION	Клиникалық ұсыныстар немесе тұжырымдар	Клиникалық нұсқаулар [21]
VARIANT_LOC	Варианттың геномдық/хромосомалық орны	Genomic координаттар [13]
EXON_NUMBER	Экзон нөмірі	Транскрипт аннотациясы [13]
CLINVAR_ID	ClinVar вариант идентификаторы	ClinVar [24]
DBSNP_ID	dbSNP дерекқорындағы вариант идентификаторы	dbSNP [25]
DIAGNOSTIC_METHOD	Қолданылған диагностикалық әдіс	Клиникалық диагностика [21]
RETEST_PLAN	Қайта тексеру жоспары	Клиникалық бақылау [21]
THERAPY_RECOMMENDATION	Емдеу бойынша ұсыныстар	Терапия нұсқаулары [21]

Label Studio платформасының икемділігі күрделі аннотациялау схемаларын жүзеге асыруға мүмкіндік берді [20]. Платформа арнайы медициналық мәтіндерді өңдеуге бейімделген және көпдеңгейлі аннотациялауды қолдайды.

Аннотациялау барысында клиникалық маңыздылық (SIGNIFICANCE) ACMG/AMP нұсқаулығының жаңартылған нұсқасына сәйкес белгіленді [21]. Негізгі категориялар 3-кестеде келтірілген.

Кесте 3 – Клиникалық маңыздылық категориялары

Категория	Ағылшынша атауы	Сипаттама
Патогенді	Pathogenic	Дәлелденген патогенді нұсқалар
Ықтимал патогенді	Likely_pathogenic	Ықтималды патогенді нұсқалар
Белгісіз маңыздылығы	VUS (Variant of Uncertain Significance)	Клиникалық маңыздылығы белгісіз нұсқалар
Ықтимал зақымсыз	Likely_benign	Ықтималды зақымсыз нұсқалар
Зақымсыз	Benign	Дәлелденген зақымсыз нұсқалар
Қорғанышты	Protective	Қорғанышты мутация
Тәуекел факторы	Risk_factor	Тәуекел факторы
Дәрілік реакция	Drug_response	Дәрілік реакция (фармакогенетика)

Тұқым қуалау сұлбасы (INHERITANCE_MODE) генетикалық аурулардың заманауи классификациясы бойынша анықталды [21]. INHERITANCE_MODE нысандары 4-кестеде келтірілген:

Кесте 4 – Тұқым қуалау сұлбалары

Мән	Ағылшынша атауы	Сипаттамасы
AD	Autosomal dominant	Аутосомдно-доминантты
AR	Autosomal recessive	Аутосомдно-рецессивті
XLD	X-linked dominant	X-хромосомаға байланысты доминантты
XLR	X-linked recessive	X-хромосомаға байланысты рецессивті
Y_LINKED	Y-linked	Y-хромосомаға байланысты
MITOCHONDRIAL	Mitochondrial	Митохондриялды
DE_NOVO	De novo mutation	Де novo мутация
COMPOUND_HET	Compound heterozygote	Композитті гетерозиготалы

Аннотациялау барысында нақты клиникалық есептердің мысалдары қолданылды. 2-суретте көрсетілген «chr22-41160723G>A» жазбасында келесі ақпараттар аннотацияланған:

- ◆ HGVS форматындағы нұсқа: c.3671+1G>A;
- ◆ ACMG классификациясы: Pathogenic [21];
- ◆ Байланысты ауру: Menke-Hennekam syndrome 2 (OMIM: 618333) [22];
- ◆ Тұқым қуалау сұлбасы: AD (аутосомдно-доминантты);
- ◆ Клиникалық ұсыныс: ата-ананы тексеру ұсынылған.

Label Studio платформасы нысандар арасындағы семантикалық қатынастарды белгілеуге мүмкіндік берді [20]. Бұл функционал ген-ауру, ген-фенотип және нұсқа-клиникалық маңыздылық сияқты маңызды байланыстарды анықтауға мүмкіндік берді. Label Studio бағдарламасы бұл үшін арнайы құралдарды ұсынды: аннотатор бірінші нысанды таңдап, содан кейін екінші нысанды таңдап, оларды байланыстыратын қатынастың түрін тізімнен таңдады. Бұл корпусты тек NER үшін ғана емес, сонымен қатар білімді шығару (relation extraction) және клиникалық білім графтарын құру үшін де құнды етеді. Негізгі қатынас түрлері 5-кестеде келтірілген.

Аннотациялау дәлдігін арттыру үшін келесі заманауи әдістер қолданылды:

- ◆ Автоматты түрде терминдерді стандартты онтологиялармен тексеру;
- ◆ Аннотациялар арасындағы қайшылықтарды анықтау;
- ◆ Күрделі жағдайларды шешу үшін эксперттердің келісім сессиялары;
- ◆ Клиникалық маңыздылықты белгілеуде ClinVar деректерінің көмегімен растау [24]

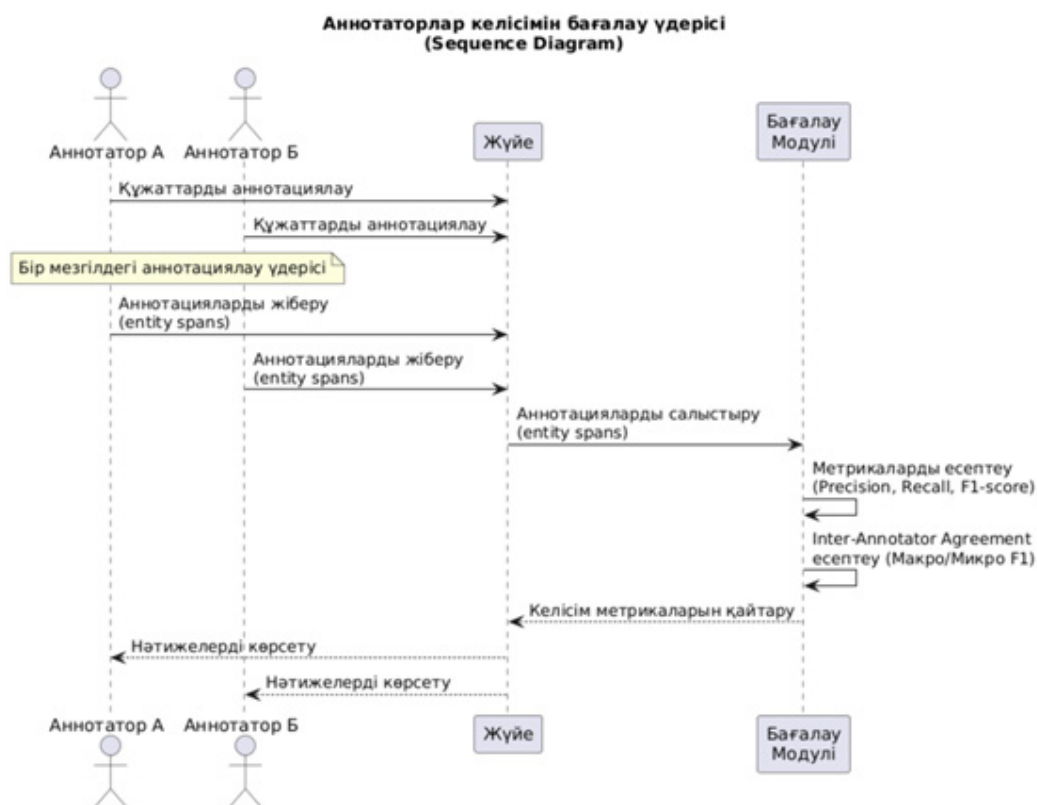
Кесте 5 – Семантикалық қатынастар

Қатынас түрі	Мақсаты	Мысал
variant_in_gene	Мутацияның белгілі бір генге тиесілілігі	c.3671+1G>A – EP300
gene_associated_with	Гендің аурумен байланысы	EP300 – Menke-Hennekam syndrome 2
variant_significance	Нұсқаның патогенділігі	c.3671+1G>A – Pathogenic
disease_inheritance	Аурудың тұқым қуалау түрі	Menke-Hennekam syndrome 2 – AD
phenotype_supports_disease	Фенотиптің ауруды растауы	syndactyly – Menke-Hennekam syndrome
disease_omim_link	Аурудың OMIM идентификаторымен байланысы	Menke-Hennekam syndrome – 618333
variant_zygosity	Нұсқаның зиготальность түрі	c.3671+1G>A – heterozygous

Бұл кешенді тәсіл клиникалық генетикалық мәтіндердің барлық аспектілерін қамтуға мүмкіндік берді және корпустың жоғары сапасын қамтамасыз етті. Аннотацияланған деректер кейінгі NLP модельдерін оқыту үшін дайындалды.

3) Деректер сапасын қамтамасыз ету және кросс-аннотациялау

Аннотацияланған корпустың сапасын және сенімділігін қамтамасыз ету үшін көпсатылы кросс-аннотациялау процедурасы жүргізілді. Әрбір қатысушының рөлі және олардың өзара байланысы 3-суреттегі реттік диаграммасында (Sequence diagram) көрсетілген. Бұл процессте үш тәуелсіз сарапшы-аннотатор қатысты, олардың әрқайсысы биомедициналық лингвистика немесе клиникалық генетика саласында тәжірибеге ие болды.



Сурет 3 – Корпусты аннотациялау процесінің реттік диаграммасы (Inter-Annotator Agreement)

3-суретте көрсетілгендей, процесс келесі негізгі кезеңдерден тұрды:

- ♦ Тәуелсіз аннотациялау (Annotator A, Annotator B): Әр аннотатор бірдей мәтіндер жиынтығын Label Studio платформасында [20] тәуелсіз түрде белгіледі;
- ♦ Аннотацияларды салыстыру (Compare annotations): Екі аннотатордың нысан аралықтары (entity spans) және олардың категориялары салыстырылды;
- ♦ Метрикаларды есептеу (Compute Precision, Recall, F1-score): Салыстыру нәтижелері бойынша дәлдік (Precision), өтіл (Recall) және F1-өлшемі сияқты стандартты метрикалар есептелді [26];
- ♦ Келісім деңгейін бағалау (Inter-Annotator Agreement): Соңғы кезеңде әрбір нысан категориясы бойынша макро және микро-орталама F1-өлшемі ретінде аннотаторлар арасындағы жалпы келісім деңгейі бағаланды.

Кросс-аннотациялау процесі екі негізгі кезеңнен тұрды:

Бірінші кезең – бастапқы аннотациялау. Бұл кезеңде әр аннотатор бірдей мәтіндер жиынтығын тәуелсіз түрде аннотациялады. Содан соң аннотациялар Label Studio платформасында жүргізілді [20]. Бастапқы аннотациялардың сәйкестігін тексеру үшін F1-өлшемі қолданылды [26]

Екінші кезең – келісім сессиялары және түзетулер. Мұнда айтарлықтай айырмашылықтар кездесетін аннотациялар бірлескен талқылауға ұсынылды. Күрделі клиникалық жағдайлар үшін клиникалық генетик маманының кеңесі пайдаланылды. Кейін аннотациялық нұсқаулық қайта қаралып, нақтыланды.

Келісім деңгейін бағалау үшін NER жүйелерінің бағалауында кеңінен қолданылатын F1-өлшемі таңдалды [26]. Бұл таңдау дәл шекаралар мен нысан категорияларын есепке алатын ең объективті көрсеткіш болғандықтан жасалды.

Бастапқы кросс-аннотациялау нәтижелері келесідей болды:

- ♦ Жалпы F1-өлшемі: 62.6%;
- ♦ Ең төмен келісім SIGNIFICANCE категориясы үшін: 48%;
- ♦ Ең жоғары келісім GENE категориясы үшін: 78%.

Аннотациялық нұсқаулықты екі рет қайта қарап, нақтылағаннан кейін және келісім сессияларын өткеннен кейін, соңғы келісім көрсеткіштері айтарлықтай жақсарды және олар 6-кестеде көрсетілген.

Кесте 6 – Аннотаторлар арасындағы соңғы келісім көрсеткіштері

Категория	F1-өлшемі (%)	Түсіндірме
GENE	92	Жоғары стандарттау дәрежесі
CDNA_PROT	87	HGVS номенклатурасының нақтылығы
DISEASE	85	OMIM сәйкестендіруінің арқасы
SIGNIFICANCE	76	Контекстке тәуелділік
ZYGOSITY	74	Терминологиялық әртүрлілік
Орташа	83.5	Жоғары сенімділік деңгейі

Келісім деңгейіне әсер еткен негізгі факторлар:

- ♦ Лексикалық стандарттау: GENE және CDNA_PROT сияқты формальды нотациялар жоғары келісімге әкелді;
- ♦ Контекстке тәуелділік: SIGNIFICANCE және ZYGOSITY сияқты категориялар төменірек келісімге ие болды, себебі олар клиникалық контекстке байланысты түрленуі мүмкін;
- ♦ Күрделі лингвистикалық құрылымдар: анафоралар және ішкі құрылымдар аннотацияны қиындатып, келісімді төмендетті.

Сапаны қамтамасыз етудің қосымша шаралары:

- ♦ Автоматты түрде UMLS [27] және HPO [23] онтологияларымен сәйкестікті тексеру;

- ♦ Аннотациялық қайшылықтарды анықтауға арналған арнайы скрипттерді әзірлеу;
- ♦ Мерзімді түрде клиникалық генетиктердің expertтік бағалауын жүргізу.

Нәтижесінде, 83.5% орташа F1-өлшемімен GENEXOM корпусы биомедициналық корпустар үшін қолданылатын халықаралық стандарттарға сәйкес келетін жоғары сапалы аннотацияланған ресурс болып табылады [26, 28].

Нәтижелер мен талқылау

Бұл бөлімде GENEXOM корпусының негізгі сипаттамалары мен статистикалық көрсеткіштері, сондай-ақ корпустың биомедициналық табиғи тілді өңдеу (nlp) зерттеулеріндегі қолдану мүмкіндіктері талқыланады.

1) Корпус статистикасы

GENEXOM корпусы 200 клиникалық экзомдық секвенирлеу есебін қамтиды. Корпус оқыту (70%), тексеру (10%) және сынақ (20%) жинақтарына бөлінген. Жиынтық статистикалық көрсеткіштер 7-кестеде келтірілген.

Кесте 7 – Корпустағы құжаттар, сөйлемдер және токендер саны

Категория	Оқыту жиыны	Тексеру жиыны	Сынақ жиыны	Барлығы
Құжаттар	140	30	30	200
Сөйлемдер	8574	1836	1840	12250
Токендер	111340	22840	22680	156860

Корпуста барлығы 25096 аннотацияланған нысан және 6428 семантикалық байланыс белгіленген. Нысандардың таралуы 8-кестеде көрсетілген.

Кесте 8 – Корпустағы нысандар мен байланыстар саны

Категория	Оқыту жиыны	Тексеру жиыны	Сынақ жиыны	Барлығы
GENE	870	125	250	1245
CDNA_MUTATION	685	100	197	982
DISEASE	810	123	220	1153
SIGNIFICANCE	925	130	262	1317
ZYGOSITY	582	83	184	849
OMIM_ID	770	150	169	1089

Ең жиі кездесетін нысандардың үлесі:

- ♦ GENE – 20.2%;
- ♦ CDNA_PROT – 16.8% ;
- ♦ DISEASE – 14.8%;
- ♦ SIGNIFICANCE – 10.7%;
- ♦ OMIM_ID – 7.6%.

Кросс-аннотациялау нәтижелері бойынша аннотаторлар арасындағы келісім (IAA) орташа есеппен 83.5% F1-өлшемін құрады. Бұл көрсеткіш биомедициналық корпустар үшін халықаралық стандарттарға сәйкес келеді [26,28].

Ең жоғары келісім көрсеткіштері:

- ♦ GENE – 92% (HGNC стандартының арқасында);
- ♦ CDNA_PROT – 87% (HGVS номенклатурасының нақтылығы);
- ♦ DISEASE – 85% (OMIM сәйкестендіруі).

Төменгі келісім көрсеткіштері:

- ♦ SIGNIFICANCE – 76% (контекстке тәуелділік);
- ♦ ZYGOSITY – 74% (терминологиялық әртүрлілік).

GENEXOM корпусының практикалық құндылығын көрсету үшін базалық NER және RE модельдері оқытылды. Нәтижелер 9-кестеде көрсетілген.

Кесте 9 – NER жүйелерінің салыстырмалы нәтижелері

Модель	GENE	DISEASE	CDNA_PROT	SIGNIFICANCE	Орташа F1
CRF	0.79	0.75	0.77	0.71	0.77
RuBERT	0.90	0.88	0.89	0.83	0.88

Қатынастарды анықтау тапсырмасында BERT негізіндегі модель ережелерге негізделген жүйеден айтарлықтай жоғары нәтиже көрсетті (0.74 F1 және 0.61 F1).

GENEXOM корпусы орыс тілдік клиникалық генетикалық мәтіндер үшін бірегей ресурс болып табылады. Біздің зерттеу нәтижелері мынаны көрсетеді:

1. Тілдік ерекшеліктер: орыс тілдік клиникалық есептер ағылшын тіліне қарағанда терминологиялық әртүрлілігі жоғары екенін атап өту керек, бұл NER жүйелерінің жұмыс істеуін қиындатады.

2. Домендік маңыздылық: корпустағы SIGNIFICANCE және ZYGOSITY сияқты арнайы категориялар клиникалық шешім қабылдау үшін өте маңызды болып табылады.

3. Салыстырмалы артықшылықтар: GENEXOM барлық белгілі биомедициналық корпустармен салыстырғанда клиникалық генетикалық есептерді егжей-тегжейлі қамтитын жалғыз ресурс болып табылады.

Корпустың шектеулері:

- ♦ Тек экзомдық секвенирлеу деректерін қамтиды;
- ♦ Тек орыс тіліндегі мәтіндер;
- ♦ Клиникалық есептердің форматтық әртүрлілігі.

Болашақ зерттеулер:

- ♦ Корпусты геномдық секвенирлеу есептерімен кеңейту;
- ♦ Көптілді модельдерді әзірлеу;
- ♦ Клиникалық шешім қолдау жүйелерін енгізу.

Қорытынды

Бұл зерттеуде орыс тіліндегі клиникалық экзомдық секвенирлеу есептері негізінде GENEXOM корпусы жасалды. Бұл корпус клиникалық генетика саласындағы табиғи тілді өңдеуге арналған алғашқы көпқабатты аннотацияланған ресурс болып табылады. Жұмыстың ғылыми және практикалық маңыздылығы оның клиникалық генетика мен жасанды интеллект саласының өзара интеграциясына қосатын үлесінде жатыр.

Зерттеу барысында келесі негізгі нәтижелерге қол жеткізілді.

Корпус құру сапасы:

- ♦ 200 клиникалық экзомдық секвенирлеу есебі негізінде жоғары сапалы корпус құрылды
- ♦ 25 000-нан астам нысан және 6 400-ден астам семантикалық байланыс дәл белгіленді
- ♦ Аннотаторлар арасындағы келісім 83.5% F1-өлшемімен жоғары сенімділік деңгейін көрсетті

Әдіснамалық тұрғыда, зерттеу барысында халықаралық стандарттар – HGVS, OMIM, ClinVar, HPO және ACMG/AMP – негізінде арнайы аннотациялау схемасы әзірленді. Сонымен қатар, Label Studio платформасының мүмкіндіктері тиімді пайдаланылып, көпқабатты аннотациялау процесі іске асырылды. Нәтижесінде, орыс тіліндегі клиникалық мәтіндерге арналған алғашқы арнайы аннотацияланған корпус жасалып, ол биомедициналық

NLP зерттеулері үшін сенімді және сапалы дереккөзге айналды. Практикалық жағынан, RuBERT моделі NER тапсырмасында 88% F1 көрсеткішіне жетіп, қатынастарды анықтау тапсырмасында BERT негізіндегі модель 0.74 F1 деңгейінде жоғары нәтиже көрсетті. Бұл көрсеткіштер GENEXOM корпусын клиникалық мәтіндерден ақпарат алу және генетикалық есептерді автоматты талдау үшін сенімді негіз ретінде пайдалануға мүмкіндік береді.

GENEXOM корпусының дамуы ғылыми және практикалық тұрғыдан бірқатар маңызды үлестерді қосады. Ғылыми тұрғыда, ол орыс тіліндегі биомедициналық NLP зерттеулеріне арналған бірегей дереккөз болып табылады, бұл клиникалық генетикалық мәтіндерді өңдеудің жаңа әдістемелерін әзірлеуге мүмкіндік береді. Сонымен қатар, корпус көптілді биомедициналық NLP зерттеулерінің дамуына да негіз бола алады, өйткені орыс тіліндегі клиникалық контентке бейімделген деректер базасы халықаралық зерттеулер үшін қажетті деректерді қамтамасыз етеді.

Практикалық жағынан, GENEXOM корпусы клиникалық шешім қолдау жүйелерін әзірлеу және генетикалық есептерді автоматты түрде өңдеу мен талдау үшін тікелей қолдануға жарамды. Бұл дәрігерлердің клиникалық шешім қабылдау процестерін жеңілдетіп, диагностикалық дәлдікті арттыруға ықпал етеді. Клиникалық генетика саласына әсері де маңызды: сирек кездесетін генетикалық ауруларды диагностикалауды жеделдету, генетикалық нұсқаларды интерпретациялауды автоматтандыру және клиникалық генетикалық деректерді стандарттау арқылы медициналық тәжірибені жетілдіру мүмкіндігін береді.

Зерттеу нәтижелері GENEXOM корпусын әрі қарай дамыту үшін бірнеше перспективалық бағыттарды айқындайды. Біріншіден, корпусы кеңейту бойынша жұмыстар жүргізу жоспарланып отыр: геномдық секвенирлеу деректерін қосу, басқа тілдердегі клиникалық есептерді қамту және фенотиптік ақпаратты толықтыру. Екіншіден, технологиялық жетілдіру мәселелері маңызды болып отыр, бұл бағытта көптілді NLP модельдерін әзірлеу, клиникалық шешім қолдау жүйелерін енгізу және нақты уақыт режимінде талдау жүргізу мүмкіндіктерін құру көзделуде. Үшіншіден, интеграциялық жобаларға мән беріледі: электрондық денсаулық карталар жүйелерімен біріктіру, халықаралық генетикалық дерекқорлармен интеграция және клиникалық тәжірибеге тікелей енгізу, бұл корпусының практикалық құндылығын арттырады.

GENEXOM корпусының дамуы қоғамдық денсаулық сақтау саласына бірнеше маңызды әсерлерін тигізеді. Диагностикалық сапаны арттыру арқылы генетикалық ауруларды уақытында анықтауға, диагностикалық қателерді азайтуға және дәл диагностика негізінде жекелендірілген медицинаны дамытуға мүмкіндік береді. Экономикалық тұрғыда, корпусының қолданысы генетикалық тестілеу шығындарын азайтып, дәрігерлердің уақытын үнемдеуге және клиникалық зерттеулерді жеделдетуге септігін тигізеді. Сонымен қатар, GENEXOM корпусы білім беру және кадрлар даярлау саласында да маңызды рөл атқарады: генетик мамандарды оқытуға, клиникалық генетика бойынша оқу ресурстарын жасауға және жас мамандарды кәсіби тұрғыдан даярлауда қолдануға мүмкіндік береді. Осылайша, GENEXOM корпусы ғылыми зерттеулер, клиникалық практика және қоғамдық денсаулық сақтау саласында кешенді ықпал ету әлеуетіне ие.

GENEXOM корпусы орыс тілдік клиникалық генетика саласындағы табиғи тілді өңдеу (nlp) зерттеулеріне негіз болып, жекелендірілген медицинаның дамуына үлес қосады. Корпусының ашық қол жетімділігі (<https://github.com/Anara-Sultangaziyeva/GENEXOM>) оны кеңінен қолдануға мүмкіндік береді.

Жұмыстың негізгі үлесі:

- ◆ Орыс тілдік клиникалық генетика үшін алғашқы арнайы корпус;
- ◆ Халықаралық стандарттарға сәйкес келетін жоғары сапалы аннотация;
- ◆ Практикалық қолданылымдылығы дәлелденген шешім;
- ◆ Болашақ зерттеулерге кен перспективалар ашатын ресурс.

GENEXOM корпусы тек ғылыми зерттеулер үшін ғана емес, сонымен қатар клиникалық тәжірибеде де нақты пайда әкелетін күрделі және көпқырлы ресурс болып табылады. Оның

дамуы клиникалық генетика мен жасанды интеллект саласының интеграциясында маңызды қадам болып саналады.

Келесі кезеңде корпусты кеңейту және оны клиникалық тәжірибеге енгізу болады. Бұл жолда генетикалық аурулармен күресте жаңа мүмкіндіктер ашылуы мүмкін, әсіресе сирек кездесетін ауруларды диагностикалау және емдеу саласында.

Қаржыландыру туралы ақпарат. «Бұл зерттеуді Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландырады (Грант № AP26105113)».

ӘДЕБИЕТТЕР

- 1 Rare Diseases International. Rare Diseases International (online). URL: <https://rarediseasesinternational.org> (accessed: 13.11.2025).
- 2 Nguengang Wakap, S., Lambert, D.M., Olry, A., et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, 28, 165–173 (2020).
- 3 Wang, Y., Wang, L., Rastegar-Mojarad, M., et al. Clinical information extraction applications: A literature review. *Journal of Biomedical Informatics*, 103, 103466 (2020).
- 4 Cohen, T., Whitfield, G., and Xu, H. Extracting genetic variant information from the literature. *Database* (Oxford), 2020, baz158 (2020).
- 5 Bangalore, S., Hegde, H., Bharadwaj, A., et al. Extraction and normalization of mutations in free text using biomedical named entity recognition models. *Bioinformatics*, 37 (7), 949–957 (2021).
- 6 Névél, A., Zweigenbaum, P., and Roberts, A. Clinical natural language processing in 2020: State of the art and future challenges. *Journal of Biomedical Informatics*, 110, 103604 (2020).
- 7 Yoon, W., Lee, J., Kim, S., et al. Pre-trained language models for biomedical natural language processing. *Briefings in Bioinformatics*, 22 (6), 1686–1702 (2020).
- 8 Li, M., Lu, Y., Shen, H., et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database*, 2016, baw068 (2016).
- 9 Bada, M., Eckert, M., Evans, D., et al. Concept annotation in the CRAFT corpus. *BMC Bioinformatics*, 13 (1), 161 (2012).
- 10 Zhao, W., Wu, X., Ma, Y., et al. BioRED: A biomedical relation extraction dataset for complex relation types. *Bioinformatics*, 38 (8), 2210–2216 (2022).
- 11 Gurulingappa, H., Mateen, A., and Dima, A. NEREL-BIO: A new resource for biomedical named entity recognition and classification in Russian. *Proceedings of the LREC Conference* (2024).
- 12 Tutubalina, E., Miftahutdinov, Z., Nikolenko, S., and Panchenko, A. Named entity recognition in Russian: Data, analysis, and model. *Proceedings of the EMNLP Conference*, 5284–5295 (2020).
- 13 den Dunnen, J.T., and Antonarakis, S.E. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Human Mutation*, 15 (1), 7–12 (2000).
- 14 Lee, J., Yoon, W., Kim, S., et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36 (4), 1234–1240 (2020).
- 15 Gu, Y., Tinn, R., Cheng, H., et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 3 (1), Article 2 (2021).
- 16 Islamaj, R., Kim, S., Kwon, D., et al. NCBI disease corpus: a resource for disease name recognition and concept normalization. *Journal of Biomedical Informatics*, 92, 103132 (2019).
- 17 Chen, J. PDFPlumber: PDF text extraction library (online). URL: <https://github.com/jsvine/pdfplumber>
- 18 C2 Team. (n.d.). python-docx: Python library for creating and updating Microsoft Word (.docx) files. GitHub. URL: <https://github.com/python-openxml/python-docx> (accessed: 13.11.2025).
- 19 C2 Team. python-docx: Python library for creating and updating Microsoft Word (.docx) files (online). URL: <https://github.com/python-openxml/python-docx>
- 20 Tkachenko, M., Malyuk, M., Shevchenko, N., & Holubiev, B. (2023). Label Studio: Data labeling software. GitHub. URL: <https://github.com/heartexlabs/label-studio> (accessed: 13.11.2025).
- 21 Seal, R.L., et al. Genenames.org: the HGNC resources in 2023. *Nucleic Acids Research*, 51, D1003–D1009 (2023).
- 22 Tkachenko, M., Malyuk, M., Shevchenko, N., and Holubiev, B. Label Studio: Data labeling software (online). URL: <https://github.com/heartexlabs/label-studio> (accessed: 1.01.2023).

- 23 Richards, S., et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the ACMG and AMP. *Genetics in Medicine*, 17 (5), 405–424 (2015).
- 24 Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Research*, 47, D1038–D1043 (2019).
- 25 Köhler, S., et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, 49, D1207–D1217 (2021).
- 26 Landrum, M.J., et al. ClinVar: improvements to accessing data. *Nucleic Acids Research*, 48, D835–D844 (2020).
- 27 Sherry, S.T., et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 29, 308–311 (2001).
- 28 Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., and Gómez-Berbís, J.M. Named entity recognition: Fallacies, challenges and opportunities. *Computer Standards & Interfaces*, 35 (5), 482–489 (2013).
- 29 Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32, 267–270 (2004).
- 30 Yang, X., Saha, S., Venkatesan, A., Tirunagari, S., Vartak, V., and McEntyre, J. Europe PMC annotated full-text corpus for gene/proteins, diseases and organisms. *Scientific Data*, 10, Article 722 (2023).

^{1,2}Самбетбаева М.А.,

Phd, ассоциированный профессор, ORCID ID: 0000-0001-9358-1614,

e-mail: sambetbayeva_ma_1@enu.kz

^{1,2*}Серикбаева С.К.,

Phd, и.о.доцента, ORCID ID: 0000-0002-3627-3321,

*e-mail: inf_8585@mail.ru

^{1,3}Сұлтанғазиева А.Н.,

магистр, докторант, ORCID ID: 0009-0009-9038-5234,

e-mail: anara77777@mail.ru

^{1,4}Мукажанов Н.К.,

PhD, ассоциированный профессор, ORCID ID: 0000-0003-4835-5751,

e-mail: nurzhan.mukazhanov@narxoz.kz

^{1,2}Абдығалым Б.Х.,

магистр, докторант, ORCID ID: 0009-0001-8872-7428,

e-mail: bayangali.abd@gmail.com

¹«Q» University, Алматы қ., Қазақстан

²Евразийский национальный университет имени Л. Н. Гумилева, Казахстан, г. Астана

³Международный университет Астаны, Казахстан, г. Астана,

⁴Университет Нархоз, Казахстан, г. Алматы,

СОЗДАНИЕ КОРПУСА АННОТИРОВАННЫХ МЕДИЦИНСКИХ ТЕКСТОВ ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ О ГЕНЕТИЧЕСКИХ ЗАБОЛЕВАНИЯХ

Аннотация

В статье представлен аннотированный корпус, состоящий из клинических текстов на русском языке, полученных по результатам экзомного секвенирования. Данный корпус был разработан в поддержку задач автоматического определения именованных объектов и семантических связей в отношении генов, мутаций, наследственных заболеваний, фенотипических признаков и их клинической значимости. В ходе формирования корпуса были использованы отчеты фактического клинического экзомного секвенирования, данные прошли этапы предварительной анонимизации и нормирования текста. В процессе маркировки использовались международные стандарты и базы знаний, такие как HGVS, OMIM, ClinVar и HPO, а также обеспечивалась согласованность и точность биомедицинской информации. Корпус содержит более 25 000 биомедицинских объектов и более 6000 семантических связей, что делает его важным ресурсом в области клинической генетики с точки зрения объема и содержания. Аннотация проводилась вручную с участием нескольких

экспертов, и результаты сравнивались путем перекрестной проверки, а уровень согласия между аннотаторами оценивался с помощью специальных показателей. Полученные результаты свидетельствуют о высоком качестве и надежности корпуса. Готовый корпус позволяет эффективно использовать модели обработки естественного языка в области медицинской генетики для обучения и оценки, разработки систем поддержки принятия клинических решений и прикладных исследований для структурирования генетических данных.

Ключевые слова: генетические заболевания, обработка медицинских текстов, аннотированный корпус, экзоминое секвенирование, клинические тексты, автоматическое распознавание именованных сущностей (NER), извлечение семантических отношений (RE).

^{1,2}**Sambetbayeva M.,**

PhD, Associate Professor, ORCID ID: 0000-0001-9358-1614,

e-mail: sambetbayeva_ma_1@enu.kz

^{1,2*}**Serikbaeyva S.,**

PhD, Acting Associate Professor, ORCID ID: 0000-0002-3627-3321,

*e-mail: inf_8585@mail.ru

^{1,3}**Sultangazyeva A.,**

Master, PhD student, ORCID ID: 0009-0009-9038-5234,

e-mail: anara77777@mail.ru

^{1,4}**Mukazhanov N.,**

PhD, Associate Professor, ORCID ID: 0000-0003-4835-5751,

e-mail: nurzhan.mukazhanov@narxoz.kz

^{1,2}**Abdy-galym B.,**

Master, PhD student, ORCID ID: 0009-0001-8872-7428,

e-mail: bayangali.abd@gmail.com

¹Q university, Almaty, Kazakhstan

²L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

³Astana International University, Astana, Kazakhstan

⁴Narxoz University, Almaty, Kazakhstan

CONSTRUCTION OF AN ANNOTATED MEDICAL TEXT CORPUS FOR INFORMATION EXTRACTION ON GENETIC DISEASES

Abstract

The article presents an annotated corpus consisting of clinical texts in Russian, obtained by exomic sequencing. This corpus was developed to support the tasks of automatically identifying named objects and semantic relationships in relation to genes, mutations, hereditary diseases, phenotypic traits and their clinical significance. During the formation of the corpus, reports of actual clinical exomic sequencing were used, the data went through the stages of preliminary anonymization and text normalization. The labeling process used international standards and knowledge bases such as HGVS, OMIM, ClinVar, and HPO, and ensured consistency and accuracy of biomedical information. The corpus contains more than 25,000 biomedical objects and more than 6,000 semantic links, making it an important resource in the field of clinical genetics in terms of volume and content. The annotation was carried out manually with the participation of several experts, and the results were compared by cross-checking, and the level of agreement between the annotators was assessed using special indicators. The results obtained indicate the high quality and reliability of the case. The finished corpus makes it possible to effectively use natural language processing models in the field of medical genetics for teaching and evaluation, development of clinical decision support systems, and applied research for structuring genetic data.

Keywords: genetic diseases, medical text processing, annotated corpus, exome sequencing, clinical texts, named entity recognition (NER), relation extraction (RE).

Received December 13, 2025; revised April 24, 2026; accepted May 8, 2026.

UDC 004.896
IRSTI 28.23.27

<https://doi.org/10.55452/1998-6688-2026-23-2-187-204>

¹**Samigulina Z.I.,**

PhD, ORCID ID: 0000-0002-5862-6415,

e-mail: z.samigulina@kbtu.kz

^{1*}**Dyussenkulova B.Z.,**

PhD student, ORCID ID: 0009-0001-2788-6521,

*e-mail: bu_dyussenkulova@kbtu.kz

¹**Butakova D.A.**

Master's student, ORCID ID: 0009-0006-8151-4928,

e-mail: d.butakova@kbtu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

REINFORCEMENT LEARNING–DRIVEN CONTROL STRATEGIES FOR SMART MANUFACTURING SYSTEMS

Abstract

Nowadays, the creation of smart manufacturing systems has high importance. Neural networks have been widely applied to solve complex manufacturing challenges. The paper is devoted to the study of neural networks with reinforcement learning as PPO (Proximal Policy Optimization), DQN (Deep Q-network) for state diagnosis of industrial equipment within the GEMMA (Guide d'Etude des Modes de Marche er d'Arret) model. The GEMMA French approach is established on the SFC (Sequential Function Charts) language and includes standards for controlling technical processes. An application of neural networks in area D of the GEMMA model is introduced. Modelling and experimental results were conducted based on synthetic and experimental datasets. The implementation of the architecture considered allows us to achieve reliable results for industrial data.

Keywords: Smart manufacturing system, reinforcement learning, proximal policy optimization (PPO), deep Q-network (DQN), informer transformer, Guide d'Etude des Modes de Marche er d'Arret (GEMMA) model, diagnostic of industrial equipment.

Received May 5, 2025; revised October 1, 2025, February 21, 2026; accepted April 7, 2026.

Introduction

Nowadays, smart manufacturing systems play a significant role in production. In the concept of smart manufacturing system Artificial Intelligence (AI), Internet of Things, Machine Learning (ML) and Digital Twin (DT). Smart Factory is determined as a manufacturing system, in which AI, IOT and automation are applied for enhancing efficiency and flexibility.

Methods of artificial intelligence and machine learning are widely applied, especially reinforcement learning (RL) is used for defining appropriate strategy, classification and other tasks. In RL agent, environment and action are applied in completing appropriate tasks. RL methods have broad applications in engineering, electronics, etc. [1]. Deep reinforcement learning is defined as a method in which are combined methods of deep learning and reinforcement learning. As was stated in [2], a data-driven approach was proposed for manufacturing automation and digital twin was used for modelling production cells, forecasting failures. For industrial control was used deep Q-learning in the context of smart manufacturing and Manufacturing Intelligence was trained on the constructed DT. Moreover, in the study [3] was introduced production control method, based on DT and RL. The proposed method has demonstrated high efficiency.

In addition to methods of neural networks and reinforcement learning methods, federated learning methods are widely used in smart manufacturing. As was stated in the research [4], multi-level architecture, which consists of the federated learning, IoT and crowdsourcing. The proposed approach can handle in terms of development of intelligent systems for predictive maintenance and diagnosis of faults.

Mezair, Djenouri and et.al proposed an advanced deep learning scheme for fault forecasting. The new framework integrates LSTM, CNN and graph neural network for working with heterogenous data. According to the results of the experiment, the proposed approach is superior in comparison with existing solutions in terms of training accuracy and speed and less power in consumed [5]. In the paper [6] were reviewed recent studies about fault detection and working algorithm have been presented. The dataset from Case Western Reserve University was taken for implementing Machine learning approaches. A novel anomaly detection approach was proposed and adapted to traditional manufacturing environments. This approach is combined with dynamic network theory and unsupervised machine learning. According to results, the new proposed method demonstrated f1-score more than 84 percents, precision with ninety-nine percents and recall more than 74 percents [7].

In the study [8], AI-driven predictive maintenance system, which is established on Digital Twin technology for monitoring industrial faults, diagnosis, and forecasting. For maintenance optimization and downtime reduction, Machine and Deep learning approaches, edge-computing and reinforcement learning were applied in the system. According to the validation results on smart manufacturing, prediction accuracy was increased by thirty-five percents, unplanned downtime was reduced by 40 percents and maintenance cost by twenty-five percents. The transition from Industry 4.0 to Industry 5.0, which is aimed at reliance, flexibility and interaction between human and machine. To address adaptation and problem about resource allocation, Digital twin system with a Soft Actor-Critic approach was proposed. The new proposed method was based on the Reinforcement learning [9]. The paper [10] examines the use of predictive maintenance in Industry 4.0 and emphasizes main challenges such as the need for a large amount of failure data and the choice of an adaptive maintenance approach. A multi-agent method based on reinforcement learning (RL) is proposed, where agents partially observe equipment status and coordinate task allocation among technicians with different skills. Experiments show that this method reduces failures and downtime, improving maintenance efficiency by $\approx 75\%$ compared to traditional strategies. In the study [11] was proposed a novel algorithm for maintenance optimization in Industry 4.0 using Reinforcement learning. The system degradation process considered. In other words, an “imperfect pair” model was introduced, where efficiency is fallen in case of each subsequent repair. For development maintenance strategies, Double Deep Q-network based agent was applied, the agent was adapted to various scenarios without predefined maintenance threshold.

In the smart manufacturing system job shop scheduling is considered, as problem about job shop scheduling related to resource optimization, minimization of time differences, tasks allocation in machines. Machine learning algorithms, especially Reinforcement Learning, is widely used for job shop scheduling problems. In the research [12] was considered Smart Manufacturing Scheduling (SMS) and for optimizing job shop scheduling was proposed conceptual mode. The model is established around three key components: semantic ontology, hierarchical agent structure and deep reinforcement learning (DRL). The proposed model increased flexibility through improved scheduling. Additionally, according to the study [13], mathematical description of the smart manufacturing service scheduling problem. Novel method, which was established on the Deep reinforcement learning, was introduced to reduce maximum completion time. The efficiency was tested in accordance with two case studies. In the study [14] was considered problem about dynamic scheduling in job shop scheduling under unforeseen equipment failures. As solution Proximal Policy Optimization (PPO) algorithm was offered a for improving performance, PPO was tested in a real production environment and demonstrated better results in comparison with state-of-the-art algorithms. Moreover, Yang and Xu have developed a system architecture, mathematical model for minimization delay costs and Deep Reinforcement

learning system with agents with scheduling and reconfiguration. The Advantage-Actor-Critic algorithm was applied and according to the experiments, decision quality and computation time have been demonstrated as approximately 55 and 88 percents. Also, shown decision time made the method appropriate for real-time optimization [15]. The study [16] introduced a reinforcement learning (RL) based approach for dynamic route planning of automated guided vehicles (AGVs) in smart warehouses and factories. The novelty of the research was in the integration of RL with AGVs in a dynamic environment considering workstations, charging stations and storage areas. Experiments in a smart manufacturing environment have validated the effectiveness of the method for complex logistics problems. The solution was scalable and contributed to improving the efficiency of modern industrial systems. Zhou, Tang, et. al in study [17] proposed AI-scheduler, based on RL to avoid complications about planning task. The introduced AI-scheduler applies composite reward functions for dynamic planning optimization in condition of uncertainty. Experiments have shown that the new AI scheduler improves key performance indicators (reduces production time, reduces costs, balances equipment load) and effectively handles unexpected events such as urgent orders and equipment breakdowns.

Furthermore, in smart manufacturing systems, hybrid algorithms are being developed. For example, in the study [18] was proposed a hybrid ML model, which combines RL and Genetic Algorithm (GA) for production scheduling optimization. The proposed model trained for 500 iterations. The experimental results have demonstrated that production efficiency increased by 39 percents, resource utilization was reached to 91 percents, equipment downtime was reduced by 34 percents, energy consumption per task decreased by 17 percents and indicator for delivery on time was increased to 94 percents. This study demonstrates the potential of integrating ML into production planning, providing flexibility, resilience and efficiency within Industry 4.0. In the research [19] was considered the concept of a federated digital twin (FDT), expanding the use of digital twins (DT) to manage complex interconnected systems. The main issue was that the variable processing speed of DT functions, which makes it difficult for optimal plan application. The system, which is called Federated Digital Twin, was introduced. The proposed system models Digital Twin relationships by temporal heterogeneous graphs and applied Deep Learning and Graph Neural networks for flexible planning. The proposed approach is superior to traditional algorithms and demonstrates efficiency in managing digital twins.

Hybrid flow shops play a key role in modern smart factories, where various unforeseen situations arise. Dynamic planning in such systems (HFSP) becomes a challenging task due to the uncertainty of processing times, dynamic deadlines for receiving orders and flexible maintenance of equipment. In the paper [20] the Neuro Evolution of Augmenting Topologies (NEAT) algorithm was introduced to minimize the maximum order fulfillment time. Experiments have shown that NEAT surpasses Deep Q-Network (DQN) and classical priority planning rules (PDRs), providing a faster and better response to dynamic changes. Testing on new data has confirmed his ability to effectively adapt to previously unknown tasks.

The article is structured as follows: Section 2 is devoted to the formulation of the research problem; Section 3 describes the methods applied in the study; Section 4 presents the dataset description; Section 5 contains the simulation and experimental results. The conclusion summarizes the main findings and research outcomes.

Problem statement

Modern industrial automation systems consist of high-priced equipment, the timely maintenance of which can reduce the economic costs of the enterprise, maintenance and repair. Industrial automation systems are built considering international quality standards such as ISA, IEC, etc. The integration of artificial intelligence methods into automatic process control systems should consider the features of the operation of industrial equipment, the specifics of industrial data and the principles of operation of programmable logic controllers.

Current research is devoted to the modernization of the Guide d'Etude des Modes de Marche et d'Arret (GEMMA) model, to implement different modes of operation of equipment, as well as

considering disaster recovery procedures. It is proposed to integrate a training-based neural network with reinforcement to diagnose equipment health in Area D (post-accident repair procedures) to prevent plant failures. To solve this problem, the following architectures were considered: Proximal Policy Optimization (PPO), Deep Q-network (DQN), and Informer transformers. Software and hardware implementation was carried out based on "Industrial Automation Lab", Schneider Electric in JSC KBTU.

Materials and methods

3.1 GEMMA Guide paradigm

In the study, smart manufacturing system for machine diagnosis was developed. The proposed system was established on the basis of reinforcement of learning neural networks and GEMMA Guide paradigm. The GEMMA Guide paradigm was represented as a block-diagram that describes different modes of the system's operation and transitions between modes. The primary purpose of the GEMMA Guide can be stated as correct determination of the process' operating modes and improvement of the interaction between the operator and the system. The improved interaction between the operator and the system contributes to prevention errors and delays in production. The GEMMA Guide paradigm was cooperated in 1984 by the National Agency for the development of Automated Production [21].

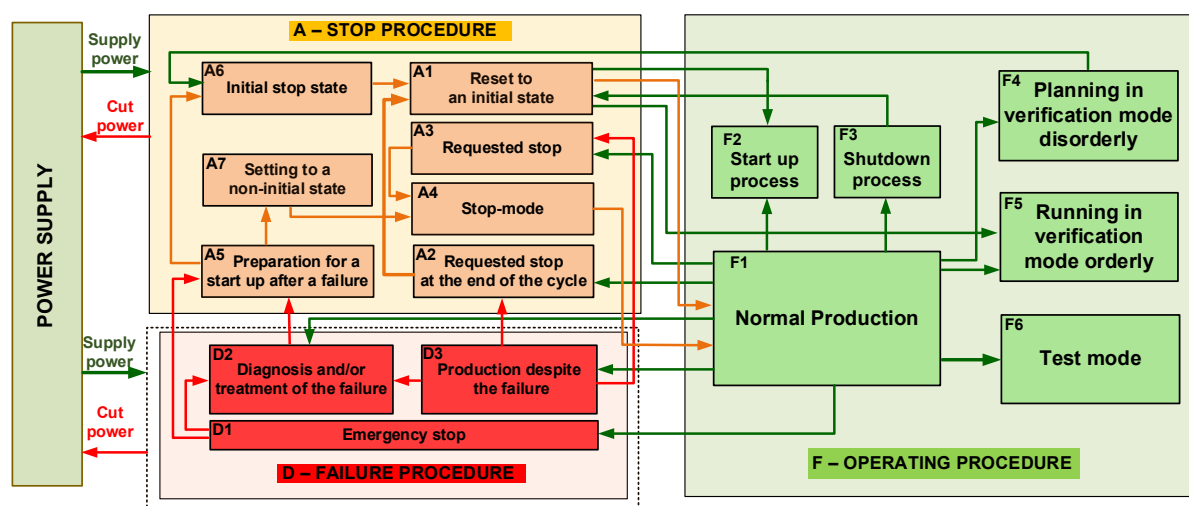


Figure 1 – Architecture of the GEMMA model for complex object control

As was illustrated in Figure 1, the GEMMA Guide paradigm consists of the A, F and D modes. The modes are organized as follows:

- ♦ A (Stop procedure) – stopping processes, including a completion cycle, stopping in a definite state and starting of the system to return to its original state;
- ♦ F (Operating procedure) – processes of normal production, including start, stop, test and checking modes;
- ♦ D (Failure procedure) – process of activation of the failures and pressing emergence stops.

In Figure 2 the architecture of a Smart manufacturing system with faults diagnosis was demonstrated. This system was based on reinforcement learning neural networks and GEMMA Guide paradigm. The main purpose was stated as faults diagnosis in zone D using Reinforcement learning approaches. The realization consists of the stages below:

Algorithm 1. Intelligent industrial data processing based on GEMMA model and neural networks with reinforcement learning.

Step 1. Collecting data from programmable logic controllers (Modicon M340) and saving in appropriate format.

Step 2. Data splitting for training and testing.

Step 3. Data preprocessing.

Step 4. Creation of the environment for realization of Reinforcement Learning Methods.

Step 5. Implementation of the RL approaches.

Step 6. Evaluation of the results using metrics.

Step 7. Decision making.

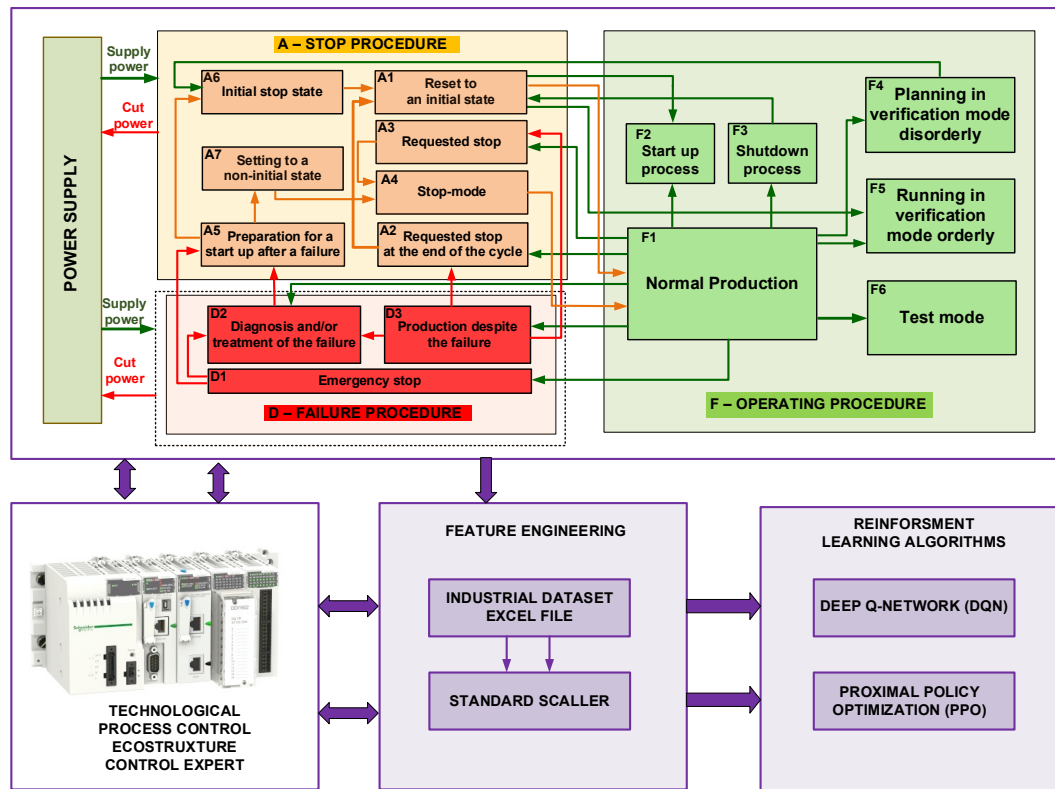


Figure 2 – Architecture of a smart manufacturing system with diagnostics of equipment based on reinforcement learning neural networks and GEMMA Guide Paradigm

Algorithms of reinforcement learning (RL), based on neural networks, were considered as application of machine learning algorithms to the GEMMA model. Architectures of the Proximal Policy Optimization (PPO) and Deep Q-network (DQN) for industrial data allow solving the problem of forecasting failures in the equipment and process optimization. The singularity of the application of these networks is the ability to handle with a large dataset. The disadvantages are the implementation complexity and long training process, however for the analysis of archival data and predictive analytics these neural networks are represented with interest. The efficiency of application of these networks was evaluated by using comparative analysis with neural networks of the class of transformers such as Informer and Time Series transformers.

3.2 Hardware description

The results of simulations and experiments on real equipment were conducted in the Industrial Automation Lab (Schneider Electric) (Figure 3) based on the Modicon M340 series programmable logic controller (PLC).



Figure 3 – Controller Modicon M340

Modicon M340 supports 1024 discrete I/O, can process up to seven thousand instructions per millisecond, has a large memory capacity of 4 MB for programs and 256 KB for data storage, supports communication protocols Ethernet TCP/IP, CANopen, Modbus, etc. The GEMMA model is built on the Modicon M340 controller in the EcoStructure Control Expert software product.

3.3 Proximal Policy Optimization

Proximal Policy Optimization (PPO) is type of RL algorithm, which is based on policies and experiences. The experiences are collected by the PPO agent by interaction with the environment. Eventually, the experience is saved in the memory of the agent [22]. As a classification task, fault diagnosis is included; the agent should guess the class of a sample. As class is assumed by an agent, an environment distributes a reward immediately to the agent. For instance, the reward is awarded positively in case of correct determination of the class; otherwise, the negative reward is assigned [23]. Below pseudocode of PPO is represented for fault diagnosis.

Pseudocode 1 – PPO algorithm for industrial data

Input parameters: features
Output parameter: class
1. Initialization: Initialization of the environment for industrial data (for example, awarding rewards) Data normalization Balancing of the classes Initialization of neural networks and policies Assigning hyperparameters (epsilon, episodes, learning rate)
2. Training of the model
3. Calculating the rewards
4. Updating of the model
5. Returning the feedback from the model

Data from sensors and generated data are accepted to environment; state is formulated and is transferred to PPO agent. As in the PPO agent, states are processed by input and hidden layers; PPO agent chooses actions that state for faults or normal production. Distribution of actions probabilities is generated by actor network and output signal in the form of arrangement of 2 classes as fault or normal production.

3.4 Deep Q-network (DQN)

The DQN algorithm is identified as an approach of deep reinforcement learning. The premise behind the application of the DQN algorithm is that deep learning (DL) and Reinforcement learning (RL) methods are integrated. A function that should be maximized or minimized is built by the Q-learning method to apply for DL. The Q-value is approximated by the application of value function in the DQN algorithm [24]. Below is demonstrated the pseudocode of DQN.

Pseudocode 2 – DQN algorithm for industrial data

Input parameters: features

Output parameters: class

1. Initialization
 - Setting hyperparameters: learning rate, discount factor, epsilon, beta, number of episodes n
 - Setting experience buffer, mini-batch size
2. Data preprocessing
 - Standard scaling
 - Data splitting for training and testing
3. Neural network initialization
 - Q-network
 - Experience replay buffer
 - Adam optimizer
4. For each episode:
 - Initialization of the environment
 - Get (state, reward, done)
 - Error calculation
 - Save (previous state, reward, done, state)
 - If memory is full:
 - Setting mini batch
 - Update Q-network
5. For each step:
 - Update Q-target
6. Model evaluation

As in PPO approach, the state is interconnected with environment and access input parameters. After receiving the current condition, the decision will be taken in the form of output parameters as 0 or 1. In experience replay are saved the past conditions as state, etc. DQN is applied to be a training agent in terms of making decisions on the base of state conditions.

3.5 Transformers

In the modern world, the architecture of transformers is getting popular. Similar architecture can be applied to time series analysis. Also, the application of technological process in the industry, information from sensors is identified as in the form of series. However, not all transformers can be implemented to solve industrial problems. According to the metrics, high indicators are shown after implementation of transformers.

3.5.1 Informer Transformer

Informer Transformer is modified version of transformer, which is developed for time series analysis. In another words, informer transformers are determined as advancement on the transformer. The structure of the Informer Transformer is identical with transformer that it consists of multi-layers [25]. In the transformer ProbSparse Self Attention is applied for analyzing significant points. ProbSparse attention accepts key (K), query (Q) and value (V) as input and computes the dot product using the formula:

$$A(Q, K, V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (1)$$

where $Q \in R^{L_Q \times d}$, $K \in R^{L_K \times d}$, $V \in R^{L_V \times d}$, the input dimension is denoted by d [26].

Architecture of the informer transformer is illustrated on Figure 4.

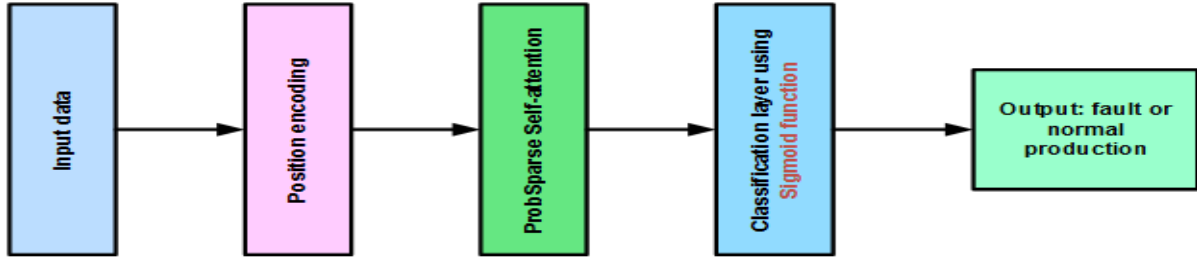


Figure 4 - Informer transformer architecture

As the input data are received normalized time series data. Position encoding is applied as order of input parameters is not considered. At the probSparse self-attention stage, significant elements are chosen to concentrate on vital characteristics. The sigmoid activation function is applied for transformation of outputs into class probabilities.

3.6 Prioritized Experience Replay (PER)

Prioritized Experience Replay (PER) is applied to RL systems with high capacity. A buffer in Experience Replay has functionality in terms of saving visited experience. This approach demonstrates high effectiveness in training deep RL models [27]. PER was created with direct and simple prioritizing the experiences with high values of temporal differences (TD) in sampling of experience for training process. TD errors are applied to introduce the experience priorities and produce bias. Stochastic prioritization with experience probability was presented to avoid issues about producing of bias as was stated in the formula (2), where p_i is denoted as priority of τ_i , α denotes a coefficient, which describes the process of identifying priority in sampling.

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha} \quad (2)$$

The formulas (3) and (4) are stated for proportional and rank-based prioritizations as for detailed priorities:

$$p_i = |\delta_i| + \varepsilon \quad (3)$$

$$p_i = \frac{1}{\text{rank}(i)} \quad (4)$$

where δ_i temporal difference error of is τ_i , ε is denoted as minimum positive number and $\text{rank}(i)$ is the index of τ_i [28].

3.7 L2-regularization

L2 regularization is a Machine Learning method for model controlling without avoiding features. The main idea is penalization of large coefficients and reducing the risk of overfitting. The loss function is adapted by L2 regularization using penalty addition to the sum of model parameters coefficients and the formula of the L2 regularization is determined as shown in the formula (5):

$$\text{Loss} = \text{Loss}_{\text{orig}} + \lambda \sum_{i=1}^n w_i^2 \quad (5)$$

where $\text{Loss}_{\text{orig}}$ the original loss function for regression is, λ is identified as regularization parameters for penalty control and w_i is determined as model coefficients. In neural networks, L2 regularization is included in training using weight decay and the formula is changed, as was illustrated in formula (6):

$$Loss = Loss_{orig} + \lambda \sum_{j=1}^m \sum_{k=1}^n w_{jk}^2 \quad (6)$$

where w_{jk} is identified as the weights, which connects the neurons between layers j and k , k , m are the numbers of layers and n is the number of linkages [29].

Dataset description

Two datasets were considered in this research. The first artificial dataset generated for the development of approaches has dimensionality $R1=1500 \times 6$. For second dataset, which has dimensionality $R2=150 \times 6$, are represented measurements from sensors from industrial stands from Schneider Electric (Figure 4).

Synthetic data

Synthetic data was generated for the implementation of experiments. Synthetic data is artificially generated data, which is applied for imitation of real data. The end-users can control the degree of similarity level of generated and real data. It helps researchers to develop and test methods, examine and validate machine learning algorithms with provision of efficiency [30]. In table 1 specification of the dataset is illustrated. Generated data consisted of six columns and 1500 rows: the target column was identified with two classes as failure or normal production.

Table 1 – Specification of the artificial dataset

Number	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Target
0	0.374540	0.950714	0.731994	0.598658	0.156019	1
1	0.155995	0.058084	0.866176	0.601115	0.708073	0
2	0.020584	0.969910	0.832443	0.212339	0.181825	0
3	0.020584	0.969910	0.832443	0.212339	0.181825	0
4	0.183405	0.304242	0.524756	0.431945	0.291229	0
...	...	...	...	...	...	...
1500	0.897397	0.119381	0.327843	0.815745	0.597312	1

Experimental data from Modicon M340 PLC (Industrial Automation Lab)

Data collection is shown on the example of an industrial object of control of an absorption plant. The technological process is implemented at the stand with programmable logic controllers of the Modicon M241 series, Modicon M340, HMI displays of the Harmony series, frequency converters of the Altivar series. Figure 5 shows the 3rd model of the bench in Fusion 360 software.

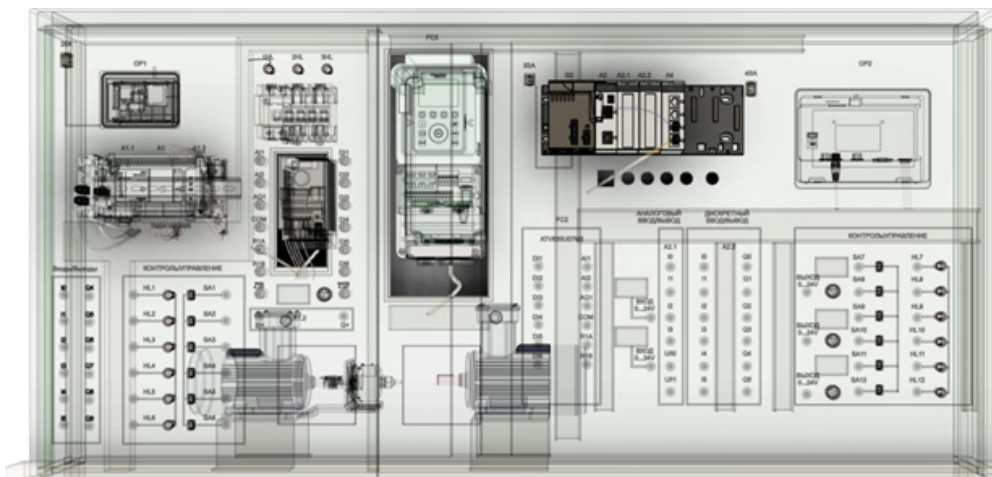


Figure 5 – 3D model test bench with industrial PLCs

Figure 6 shows the software implementation of the absorption plant [37].



Figure 6– Data collection based on Modicon M340
programmable logic controller

Figure 7 illustrates the mnemonic diagram of process control.

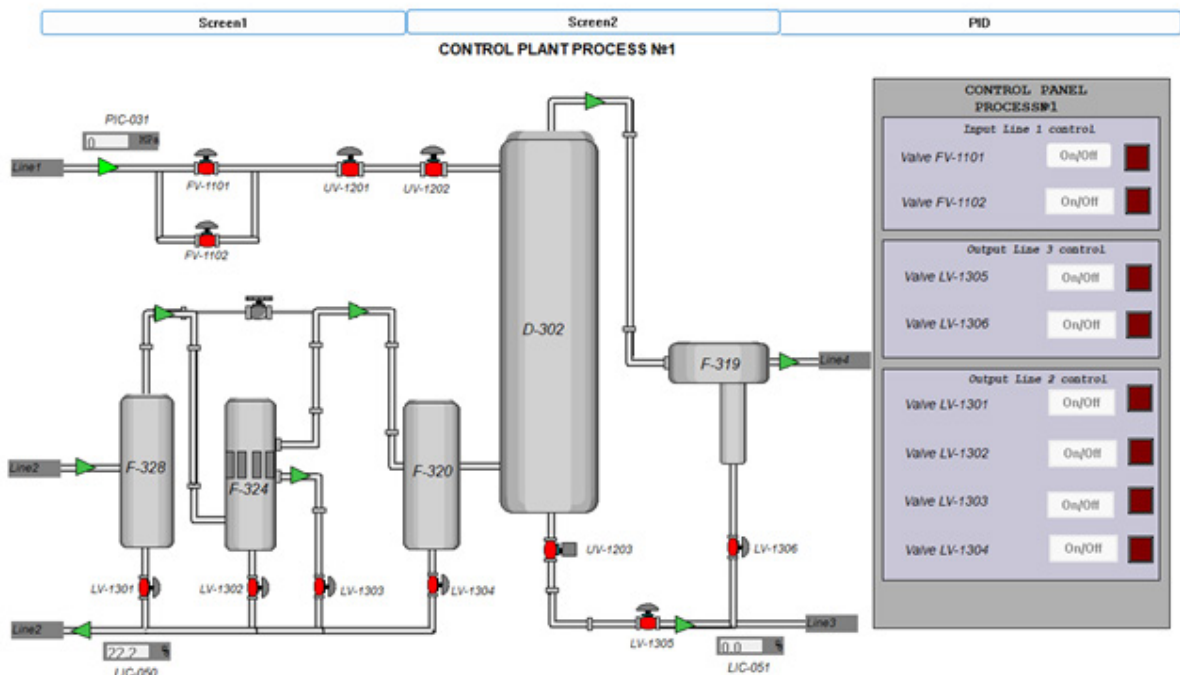


Figure 7 – Mnemonic diagram of Absorption process control
Installation of D-302

The experimental data was obtained from sensors. In this dataset contains information about the indicators of pressure and temperature. As the target column was marked column about classes. The classes were classified as “0” and “1”. In the table the view of dataset was demonstrated. In the dataset indicators are saved from experiments, as recordings about temperature and pressure from sensors.

Table 2 – Specification of the experimental data

№	Temp_sensor (C)	PT1	PT2	PT3	PT4	Class
0	24,66	0,3375	0,285	0,315	0,27	1
1	24,66	0,3375	0,285	0,315	0,27	1
2	24,66	0,3375	0,285	0,315	0,27	1
3	24,66	0,3375	0,285	0,315	0,27	1
4	24,5	0,3375	0,285	0,30375	0,36	1
	...	...	...	...	...	...
97	24,79	0,3	0,3	0,315	0,315	1

The columns “PT1, PT2, PT3, PT4” indicate data from 4 various stands.

Data preprocessing

Data preprocessing is determined as a vital stage before training model, as data from sensors may have noise, missing values and outliers. In data preprocessing tasks are included data cleaning, data reduction, scaling, transformation and dividing data into the partitions [31]. Low-quality data can lead to consequences such as predictions with low accuracy and ineffectiveness in terms of time. Industrial data are facing challenges such as:

- ♦ complex data structures;
- ♦ difficulties in terms of integration and data aggregation;
- ♦ issues about data quality evaluation due to high dimension;
- ♦ lack of unified standards of data quality;
- ♦ prominent level of requirements to data processing quality [32].

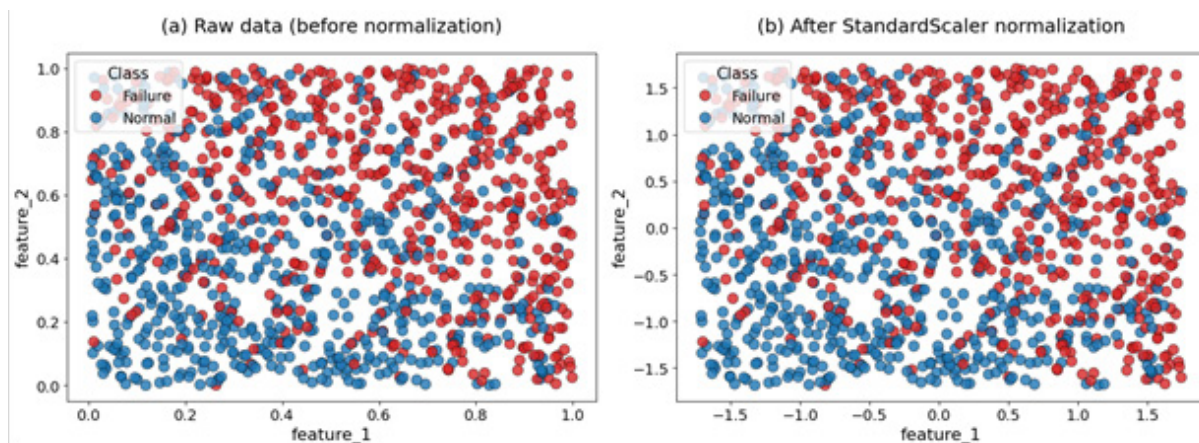
Emission removal techniques, balancing of classes, feature scaling are applied to improve quality of models. These steps help to increase performance of metrics, avoid hyper training and generatability of the model.

3.3.1 Standard Scaling

The standard scaling is presented in equation (1):

$$z = \frac{x-u}{s} \quad (7)$$

where z is determined as output vector of scaled numerical features of item, x is the input vector and s are standard deviation [33]. Figure 8 demonstrates distribution of parameters in synthetic and experimental datasets before and after processes of normalization.



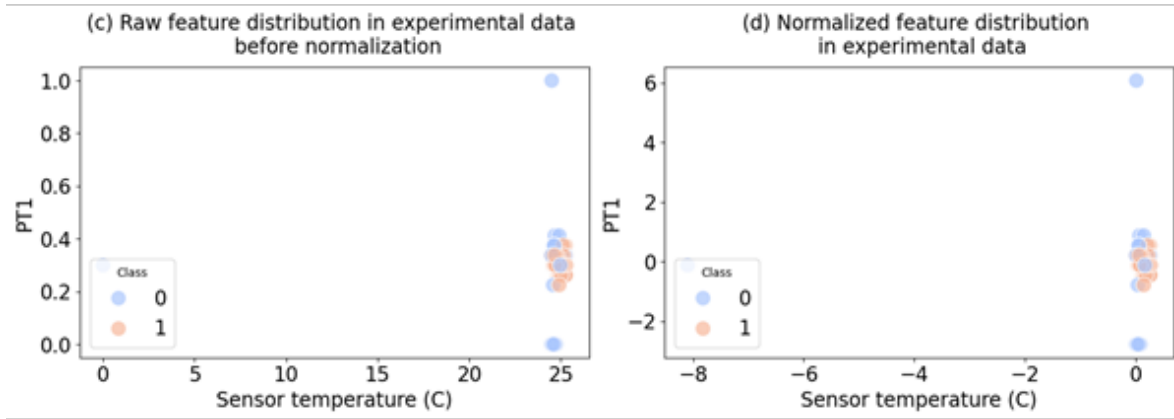


Figure 8 - Distribution of experimental data before and after normalization

After normalization using StandardScaler method, values of standard deviation were equal to 1. It can be explained by centering of the data around zero after subtraction of the averages from each value.

Results and discussion

Performance evaluation of PPO, DQN, Informer Transformer

After implementing the sampling techniques on the initial dataset, the classifiers were developed based on PSO and ensemble methods, which results are provided in this section. Precision evaluates the accuracy of model in terms of guessing positive predictions. In classification tasks accuracy is applied whether the model forecasted the label. Recall is determined as tool for measuring the model's ability to identify positive instances among all positives in the dataset. F1-score is a statistical measure, which consists of precision and recall, it evaluates ability determining positive examples. The formulas of metrics as accuracy, precision, recall and f1-score are illustrated in (7) – (11), where TP is true positive, FP is false positive, TN is true negative, and FN are defined as False negative values [34].

$$precision = \frac{TP}{TP+FP} \quad (8)$$

$$recall = \frac{TP}{TP+FN} \quad (9)$$

$$f1 - score = \frac{2*precision*recall}{precision+recall} \quad (10)$$

$$accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (11)$$

The ROC-curve is identified as receiver operating characteristic with the formula:

$$ROC(\sigma) = (1 - TN(\sigma), TP(\sigma)) \quad (12)$$

where $1 - TN(\sigma)$ can be defined as false-positive (FP) of the threshold (σ)[35].

Evaluation of synthetic data using performance metrics

Synthetic data was generated by the function for creating datasets using Python programming language. After, PPO, DQN with PER and Informer transformer methods were applied for synthetic data. Then applied methods were evaluated metrics as accuracy, precision, recall, f1-score, and ROC-AUC scores. Table 3 demonstrates the indicators of performance metrics. PPO has the highest

indicator in terms of accuracy, precision and f1-score, while DQN with PER has DQN with PER has recall with 100 percents. Informer transformers have the best performance in terms of ROC-AUC. According to the percentage in performance metrics, it is meant that PPO model was able to predict classes and has less quantity of false positives. The quantity of false positives plays significant role in smart manufacturing because it will help in avoiding premature line stops and penalties for idle time, which may be caused by false alarm. The maximum percentage of recall means that DQN with PER never miss's never real failures and a smaller number of false negatives. It can be explained with that fact Prioritized Experience Replay (PER) has buffer for assigning priorities in terms of training. In manufacturing processes, a failure to detect a fault is dangerous or may lead to serious losses such as equipment breakdown, accident, and defective batch. ROC-AUC score has the highest percentage in all models.

Table 3 – Performance metrics for PPO, DQN with PER, and Informer Transformer approaches on synthetic data (data breakdown: 80% training, 20% test)

Metrics	PPO	DQN with PER	Informer transformer
Accuracy	99%	94,67	96,67%
Precision	99,31%	90,12%	95,33%
Recall	98,63%	100%	97,95%
F1-score	98,97%	94,81%	96,62%
ROC-AUC score	98,99%	94,81%	99,55%

All metrics are reported as percentages (%). In Figure 9 the ROC curves are shown. All realized models have ROC-AUC more than 90 percents.

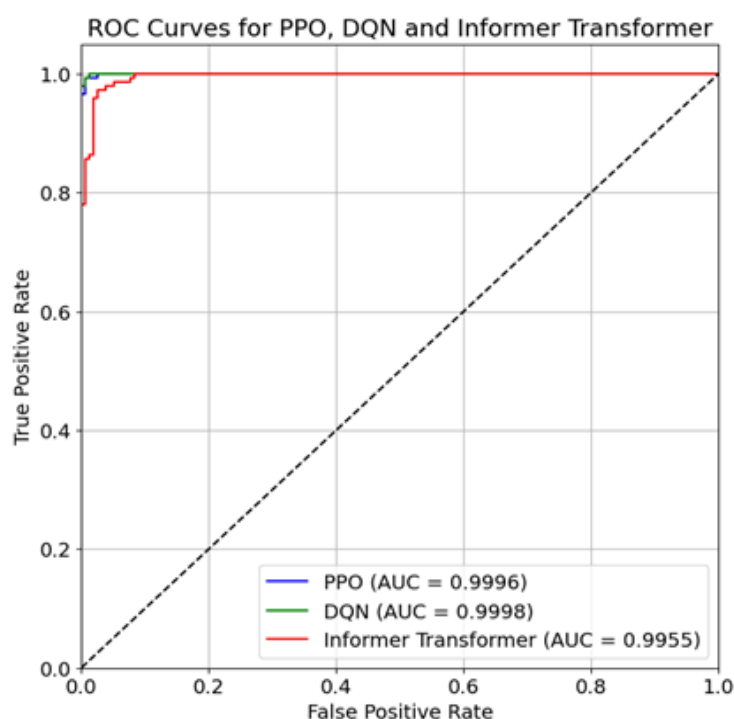


Figure 9 – ROC curves for PPO, DQN, and Informer Transformer models on synthetic data (80/20 — train/test)

The average AUC values for 3 runs are shown in Figure 7 (PPO: lr=2e-5, batch=512, gamma=0.919, clip=0.21; DQN: lr=5e-4, batch=64, gamma=0.99, train_freq=4; Informer: input=5,

hidden=128, layers=2, dropout=0.1, lr=9.586e-4. Axes: FPR (X) and TPR (Y), from 0 to 1. AUC — dimensionless metric). The ROC-AUC curves have percentages close to 100 percents due to training models for generated data. It can be explained with overfitting due to noise in the data.

Evaluation of experimental data using performance metrics

The experimental data was collected from temperature sensors from 6 stands from Schneider Electric laboratory. As in (4.2), the experimental data was evaluated by metrics as accuracy, precision, recall, f-score and ROC-AUC score. In Table 4 performance metrics are presented. As experimental data differed in terms of maintenance, the realized models were indicated low values in comparison with indicators of synthetic data performance. DQN demonstrated the highest percentages of accuracy and precision. Additionally, DQN and Informer transformers had equal percentages of recall and f1-score. Informer transformers show the highest values of ROC-AUC score, it means that the model handled with classification. In contrast with DQN and Informer transformers, PPO model had the lowest values of performance metrics. In can be explained and supposed with basement on trials and errors.

Experimental method has demonstrated the less values of performance metrics in comparison with synthetic data. It can be explained that data was dynamic, and overfitting hasn't been experienced.

Table 4 – Performance metrics for PPO, DQN with PER, and Informer Transformer on synthetic data (80% training, 20% test)

Metrics	PPO	DQN	Informer transformer
Accuracy	70%	88%	85%
Precision	80%	88%	88%
Recall	70%	85%	85%
F1-score	64%	84,1%	84,1%
ROC-AUC score	79%	75%	83,85%

All metrics are reported as percentages (%) (Table 4). In Figure 10 ROC-AUC curves for realized models are presented. The realized models have AUC indicator values less than 0.9.

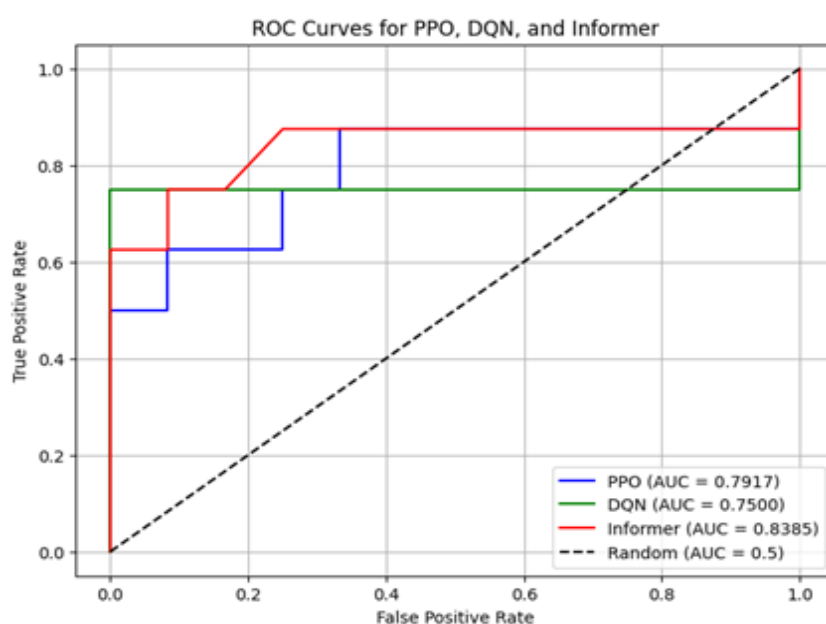


Figure 10 – ROC curves of PPO, DQN, and Informer Transformer models on experimental data (split: 80/20— train/test)

The following parameters were used: PPO (seed=42): lr=1e-5, batch=256, n_steps=4096, gamma=0.95, clip=0.25, ent_coef=0.55, vf_coef=2.0; DQN (seed=45): lr=1e-5, batch=64, buffer=150000, gamma=0.95, train_freq=20, target_update=800; Informer: l2_lambda=1e-5. Axes: FPR (X) and TPR (Y), range [0, 1]. AUC — dimensionless metric.

The DQN has the smallest value of AUC, as illustrated in Figure 8. The Informer transformer has the highest values of AUC in comparison with Reinforcement learning models such as PPO and DQN.

Conclusion

The aim of this research was to implement Reinforcement learning methods fault classification in smart manufacturing systems. PPO, DQN with PER were realized for synthetic, experimental data and Reinforcement learning methods were compared with Informer transformers. To compare and analyze effectiveness of the model's performance metrics and loss functions are applied. According to the results of the evaluation and analysis, Reinforcement learning, and Informer transformer models have pros and cons in terms of their tasks. For synthetic data DQN with PER and Informer transformer were appropriate as these approaches demonstrated high ability to classify faults and normal production, which contributes to minimizing the risk of false alarms. Also, in synthetic data, all implemented approaches the highest percentage in terms of all evaluated metrics, more than ninety percents. As artificial data was generated, overfitting might be caused and metrics should be checked properly for experimental data. For experimental data DQN with PER was suitable. In both cases as synthetic data with high dimension and experimental data, DQN with PER shows the best results as this approach are linked with setting priorities in training. Developed models can be applied in manufacturing processes to make decisions based on faults forecasting. In further research, attention mechanisms will be combined with developed models.

REFERENCES

- 1 Jayakumar, S., and Nandakumar, S. Distributed resource optimization using the Q-learning algorithm in device-to-device communication: A reinforcement learning paradigm. *Results in Engineering*, 23, 102462 (2024). <https://doi.org/10.1016/j.rineng.2024.102462>
- 2 Xia, K., Sacco, C., Kirkpatrick, M., et al. A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. *Journal of Manufacturing Systems*, 58, 210–230 (2021). <https://doi.org/10.1016/j.jmsy.2020.06.012>
- 3 Park, K.T., Son, Y.H., Ko, S.W., and Noh, S.D. Digital twin and reinforcement learning-based resilient production control for micro smart factory. *Applied Sciences*, 11 (7) (2021). <https://doi.org/10.3390/app11072977>
- 4 Ullah, I., Hassan, U.U., and Ali, M.I. Multi-level federated learning for Industry 4.0: A crowdsourcing approach. *Procedia Computer Science*, 217, 423–435 (2022). <https://doi.org/10.1016/j.procs.2022.12.238>
- 5 Mezair, T., Djenouri, Y., and Belhadi, A. A sustainable deep learning framework for fault detection in 6G Industry 4.0 heterogeneous data environments. *Computer Communications*, 187, 164–171 (2022). <https://doi.org/10.1016/j.comcom.2022.02.010>
- 6 Neupane, D., and Seok, J. Bearing fault detection and diagnosis using Case Western Reserve University dataset with deep learning approaches: A review. *IEEE Access*, 8, 93155–93178 (2020). <https://doi.org/10.1109/ACCESS.2020.2990528>
- 7 Mattera, G., Mattera, R., Vespoli, S., and Salatiello, E. Anomaly detection in manufacturing systems with temporal networks and unsupervised machine learning. *Computers & Industrial Engineering*, 203, 111023 (2025). <https://doi.org/10.1016/j.cie.2025.111023>
- 8 Prabu, S., Senthilraja, R., Ali, A.M., Jayapoorani, S., and Arun, M. AI-driven predictive maintenance for smart manufacturing systems using digital twin technology. *International Journal of Computational and Experimental Science and Engineering*, 11 (1), 1350–1355 (2025). <https://doi.org/10.22399/ijcesen.1099>

- 9 Asmat, H., Din, I., Almogren, A., and Khan, M. Digital twin with soft actor-critic reinforcement learning for transitioning from Industry 4.0 to 5.0. *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3546085>
- 10 Ruiz Rodríguez, M.L., Kubler, S., de Giorgio, A., Cordy, M., Robert, J., and Le Traon, Y. Multi-agent deep reinforcement learning based predictive maintenance on parallel machines. *Robotics and Computer-Integrated Manufacturing*, 78, 102406 (2022). <https://doi.org/10.1016/j.rcim.2022.102406>
- 11 Pliego Marugán, A., Pinar-Pérez, J.M., and García Márquez, F.P. A reinforcement learning agent for maintenance of deteriorating systems with increasingly imperfect repairs. *Reliability Engineering & System Safety*, 252, 110466 (2024). <https://doi.org/10.1016/j.ress.2024.110466>
- 12 Serrano-Ruiz, J.C., Mula, J., and Poler, R. Development of a multidimensional conceptual model for job shop smart manufacturing scheduling from the Industry 4.0 perspective. *Journal of Manufacturing Systems*, 63, 185–202 (2022). <https://doi.org/10.1016/j.jmsy.2022.03.011>
- 13 Zhou, L., Zhang, L., and Horn, B. Deep reinforcement learning-based dynamic scheduling in smart manufacturing. *Procedia CIRP*, 93, 383–388 (2020). <https://doi.org/10.1016/j.procir.2020.05.163>
- 14 Zhang, M., Lu, Y., Hu, Y., et al. Dynamic scheduling method for job-shop manufacturing systems by deep reinforcement learning with proximal policy optimization. *Sustainability*, 14 (9) (2022). <https://doi.org/10.3390/su14095177>
- 15 Yang, S., and Xu, Z. Intelligent scheduling and reconfiguration via deep reinforcement learning in smart manufacturing. *International Journal of Production Research*, 60 (16), 4936–4953 (2022). <https://doi.org/10.1080/00207543.2021.1943037>
- 16 Ho, G., Tang, Y., Leung, E., and Tong, P. Integrated reinforcement learning of automated guided vehicles dynamic path planning for smart logistics and operations. *Transportation Research Part E: Logistics and Transportation Review*, 196 (2025). <https://doi.org/10.1016/j.tre.2025.104008>
- 17 Zhou, T., Tang, D., and Wang, L. Reinforcement learning with composite rewards for production scheduling in a smart factory. *IEEE Access*, 9, 752–766 (2021). <https://doi.org/10.1109/ACCESS.2020.3046784>
- 18 Bin Shaikat, F., and Ahmed Faysal, S. Optimization of production scheduling in smart manufacturing environments using machine learning algorithms (2025).
- 19 Kim, Y., Kim, H., Ha, B., and Kim, W. Federated digital twins: A scheduling approach based on temporal graph neural network and deep reinforcement learning. *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3530558>
- 20 Chen, Y., Zhang, J., and Rauf, M. Dynamic scheduling of hybrid flow shop problem with uncertain process time and flexible maintenance using NeuroEvolution of Augmenting Topologies. *IET Collaborative Intelligent Manufacturing*, 6 (3) (2024). <https://doi.org/10.1049/cim2.12119>
- 21 Schoepp, S., Taghian, M., and Miwa, S. Enhancing hardware fault tolerance in machines with reinforcement learning policy gradient algorithms (2024). <https://doi.org/10.48550/arXiv.2407.15283>
- 22 Modirrousta, M.H., Aliyari Shoorehdeli, M., and Yari, M. Imbalanced classification in faulty turbine data: New proximal policy optimisation. *IET Collaborative Intelligent Manufacturing*, 6 (3) (2024). <https://doi.org/10.1049/cim2.12114>
- 23 Yang, Y., Juntao, L., and Lingling, P. Multi-robot path planning based on a deep reinforcement learning DQN algorithm. *CAAI Transactions on Intelligence Technology*, 5 (3), 177–183 (2020). <https://doi.org/10.1049/trit.2020.0024>
- 24 Zhu, Q., Han, J., Chai, K., and Zhao, C. Time series analysis based on Informer algorithms: A survey. *Symmetry*, 15 (4) (2023). <https://doi.org/10.3390/sym15040951>
- 25 Zhang, X., Yang, K., and Zheng, L. Transformer fault diagnosis method based on TimesNet and Informer. *Actuators*, 13 (2) (2024). <https://doi.org/10.3390/act13020074>
- 26 Pan, Y., et al. Understanding and mitigating the limitations of prioritized experience replay (2023). <https://doi.org/10.48550/arXiv.2007.09569>
- 27 Bu, F., and Chang, D.E. Double prioritized state recycled experience replay. In: *Proceedings of the IEEE International Conference on Consumer Electronics - Asia* (2020). <https://doi.org/10.1109/ICCE-Asia49877.2020.9276975>
- 28 Chandrinou, N., Loi, I., Zachos, P., et al. Effectiveness of L2 regularization in privacy-preserving machine learning (2024). <https://doi.org/10.48550/arXiv.2412.01541>
- 29 Alinda Rahmi, N., and Defit, S. The use of hyperparameter tuning in model classification: A scientific work area identification. *International Journal of Applied Sciences and Smart Technologies*, 8 (4) (2025). <http://dx.doi.org/10.62527/joiv.8.4.3092>

30 Shekhar, S., Bansode, A., and Salim, A. A comparative study of hyper-parameter optimization tools. In: 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (2021). <https://doi.org/10.1109/CSDE53843.2021.9718485>

31 Jordon, J., Szpruch, L., Houssiau, F., et al. Synthetic data – what, why and how? (2022). <https://doi.org/10.48550/arXiv.2205.03257>

32 Fan, C., Chen, M., Wang, X., et al. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9 (2021). <https://doi.org/10.3389/fenrg.2021.652801>

33 Bekar, E.T., Nyqvist, P., and Skoogh, A. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. *Advances in Mechanical Engineering*, 12 (5) (2020). <https://doi.org/10.1177/168781402091>

34 Ou, F., Wang, H., Zhang, C., et al. Industrial data-driven machine learning soft sensing for optimal operation of etching tools. *Digital Chemical Engineering*, 13 (2024). <https://doi.org/10.1016/j.dche.2024.100195>

35 Terven, J., Cordova-Esparza, D., Ramirez-Pedraza, A., et al. Loss functions and metrics in deep learning (2023). <https://doi.org/10.48550/arXiv.2307.02694>

36 Le, P.B., and Nguyen, Z.T. ROC curves, loss functions, and distorted probabilities in binary classification. *Mathematics*, 10 (9) (2022). <https://doi.org/10.3390/math10091410>

37 Samigulina, G.A., Samigulina, Z.I., Bekeshev, D., and Butakova, D. Data-driven machinery faults detection techniques using artificial intelligence in Industry 4.0 concept. *Procedia Computer Science*, 257, 404–411 (2025). <https://doi.org/10.1016/j.procs.2025.03.053>

¹Самигулина З.И.,

PhD, ORCID ID: 0000-0002-5862-6415,

e-mail: z.samigulina@kbtu.kz

^{1*}Дюсенкулова Б.Ж.

докторант, ORCID ID: 0009-0001-2788-6521,

*e-mail: bu_dyussenkulova@kbtu.kz

¹Бутакова Д.А.

магистрант, ORCID ID: 0009-0006-8151-4928,

e-mail: d.butakova@kbtu.kz

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

КҮШЕЙТУ АРҚЫЛЫ ОҚЫТУДЫ ПАЙДАЛАНА ОТЫРЫП, ЗИЯТКЕРЛІК ӨНДІРІСТІ БАСҚАРУ ӘДІСТЕРІ

Аңдатпа

Бүгінде ақылды өндіріс жүйелерін құру – өзекті міндет. Күрделі өндірістік мәселелерді шешуге мүмкіндік беретін нейрондық желілер кеңінен қолданылады. Мақала GEMMA моделінің бөлігі ретінде өнеркәсіптік жабдықтардың күйін анықтауға арналған DQN, PPO сияқты нейрондық желілер негізінде құрылған арматурамен оқыту алгоритмдеріне арналған. Француздық GEMMA моделі SFC (Sequential Function Charts) тіліне негізделген және процестерді басқару стандарттарын қамтиды. GEMMA моделінің D аймағына нейрондық желілерді енгізу ұсынылады. Модельдеу және эксперимент нәтижелері екі түрлі деректер базасы негізінде жүзеге асырылды: біреуі жасанды түрде жасалды, екіншісі өнеркәсіптік өндірістен алынды. Қарастырылған архитектураларды қолдану өнеркәсіптік деректермен жұмыс істеу үшін жақсы нәтижелерге қол жеткізуге мүмкіндік береді.

Түйін сөздер: ақылды өндірістік жүйе, нығайтумен оқыту, proximal policy optimization, deep Q-network, Guide d'Etude des Modes de Marche et d'Arrêt (GEMMA).

¹Самигулина З.И.,

PhD, ORCID ID: 0000-0002-5862-6415,

e-mail: z.samigulina@kbtu.kz

^{1*}Дюсенкулова Б.Ж.,

PhD докторант, ORCID ID: 0009-0001-2788-6521,

*e-mail: bu_dyussenkulova@kbtu.kz

¹Бутакова Д.А.

магистрант, ORCID ID: 0009-0006-8151-4928,

e-mail: d.butakova@kbtu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

МЕТОДЫ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫМ ПРОИЗВОДСТВОМ С ИСПОЛЬЗОВАНИЕМ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Аннотация

В настоящее время создание умных производственных систем является актуальной задачей. Широкое распространение получили нейронные сети, которые позволяют решать сложные производственные проблемы. Статья посвящена исследованию нейронных сетей с подкреплением DQN, PPO для диагностики состояния промышленного оборудования в рамках модели GEMMA. Французский подход GEMMA построен на основе языка SFC (Sequential Function Charts) и содержит стандарты по управлению технологическими процессами. Предлагается внедрение нейронных сетей в зону Д, модели GEMMA. Результаты моделирования и экспериментов осуществлялись на основе двух баз данных, одна сгенерирована искусственным путем, вторая взята с промышленного производства. Применение рассмотренных архитектур позволяет добиться хороших результатов для работы с индустриальными данными.

Ключевые слова: умная производственная система, обучение с подкреплением, proximal policy optimization, deep Q-network, Guide d'Etude des Modes de Marche et d'Arret (GEMMA) модель.

ӘОЖ 004.934.1
ҒТАХР 20.19.00

<https://doi.org/10.55452/1998-6688-2026-23-2-205-217>

¹**Рахимова Д.,**
доцент, ORCID ID: 0000-0003-1427-198X,
e-mail: drakhimova060@gmail.com

^{1,2,3*}**Жігер А.,**
магистр, ORCID ID: 0000-0002-3641-8260,
*e-mail: aliya.zhunussova.zh@narxoz.kz

^{3,4}**Малых В.,**
т.ғ.к., ORCID ID: 0009-0008-5632-6188,
e-mail: valentin.malykh@phystech.edu

¹**Мансурова М.,**
физ.-мат.ғ.к., профессор,
ORCID ID: 0000-0002-9680-2758,
e-mail: mansurova.madina@gmail.com

¹**Карюкин В.И.,**
PhD, ORCID ID: 0000-0002-8768-0349,
email: vladislav.karyukin@gmail.com

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

²Нархоз университеті, Алматы қ., Қазақстан

³Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

⁴Санкт-Петербург мемлекеттік ақпараттық технологиялар,
механика және оптика университеті, Санкт-Петербург қ., Ресей

ҚАЗАҚ ТІЛІНДЕГІ МЕДИЦИНАЛЫҚ МӘТІНДЕРДІ ПОСТРЕДАКТОРЛЕУДІҢ ӘДІСТЕРІ МЕН САПАЛЫҚ ТАЛДАУЫ

Аңдатпа

Қазіргі таңда медицина саласында ағылшын тілінен қазақ тіліне сапалы және дәл аудару – медициналық ақпараттың қолжетімділігі мен қауіпсіздігін қамтамасыз етуде маңызды мәселелердің бірі. Бұл ғылыми жұмыста кеңінен қолданылатын Google Translate және Yandex Translate секілді машиналық аударма жүйелерінің медициналық мәтіндерге қатысты дәлдігі мен тиімділігі зерттелді. Зерттеудің негізгі мақсаты – ағылшын тіліндегі медициналық терминдер мен күрделі сөйлемдерді қазақ тіліне мағыналық және стилистикалық тұрғыдан дұрыс аудару жолдарын қарастыру. Осы мақсатта халықаралық медициналық мақалалар, клиникалық зерттеулер және дәрі-дәрмек сипаттамаларынан тұратын 102 374 сөйлем жинақталып, арнайы корпус жасалды. Корпус MarianNMT нейронды машиналық аудармада оқытылып, қазақ тіліне аударылды. Аударма нәтижелерін жеңіл постредакторлеу үшін Kaz-RoBERTa трансформерлік моделі және толық постредакторлеу үшін Үлкен Тілдік Модельдердің (LLM) бірі GPT-4.1 моделі қолданылып, оның медициналық мәтіндерге бейімделу мүмкіндіктері қарастырылды. Аударма сапасы BLEU, TER және METEOR метрикалық бағалау көрсеткіштері арқылы өлшенді. Бастапқы MarianNMT машиналық аудармадан кейін алынған аудармалар мен Kaz-RoBERTa мен GPT-4.1 модельдерінен кейінгі нәтижелер салыстырмалы түрде ұсынылды. Талдау нәтижелері көрсеткендей Kaz-RoBERTa моделінен кейін алынған аудармалар бастапқы MarianNMT машиналық аудармадан кейінгі алынған аудармадан BLEU метрика бойынша 9%-ға жоғары нәтиже, ал GPT-4.1 моделінен кейін 23%-ға жоғары нәтиже көрсетті. Қазақ тілі лингвист мамандар баға беруі бойынша, Kaz-RoBERTa моделі қолданылғаннан кейінгі алынған аудармаларда термин, лексика бойынша қателер саны азайғаны байқалды. Ал GPT-4.1 моделінен кейін аударма стильдік және семантикалық жағынан жақсарғанын көруге болады. Мақалада қарастырылған модельдер қазақ тіліне ұқсас түркі тілдеріндегі медицина саласындағы аудармаларды постредакторлеу саласында қолдануға болады.

Түйін сөздер: нейронды машиналық аударма, MarianNMT, Kaz-RoBERTa, трансформер моделі, медицина мәтіндері, BLEU, TER, METEOR.

Кіріспе

Қазақстан Республикасында медицина саласындағы сөйлемдер мен терминдерді ағылшын тілінен қазақ тіліне қолжетімді әрі сапалы аударылуы – қоғамдық денсаулық сақтау, ғылыми-зерттеу, клиникалық тәжірибе және фармацевтикалық қызметтің тиімділігін арттыруда шешуші рөл атқарады.

Халықаралық емханалық нұсқаулықтар, ғылыми мақалалар және дәрілік заттар жөніндегі ресми материалдарды қазақ тіліне дәл жеткізу медициналық мамандардың өзекті ақпаратқа қолжетімділігін қамтамасыз етіп, медициналық шешім қабылдау сапасын жақсартады және ұлттық тілдегі кәсіби медициналық контенттің кеңеюіне ықпал етеді [1].

Жетілдірілген машиналық аударма жүйелерінің (Google Translate, Yandex Translate және т.б.) және жасанды интеллект технологияларының дамуына қарамастан, медициналық мәтіндерді ағылшын тілінен қазақ тіліне немесе керісінше аудару кезінде елеулі қателіктер жиі кездеседі. Мұндай қателіктердің басты себептерінің бірі – қазақ тілінің морфологиялық, синтаксистік және семантикалық тіл құрылымдарының күрделілігі, оның өзге тілдерден айқын ерекшеленуі болып табылады.

Мысалы, қазақ тіліндегі күрделі морфология жүйесі, сөздердің жуан-жіңішке үндестігі, сөйлемдердің жиі құрмалас құрылымы, сондай-ақ терминдердің контекстке тәуелді көпмағыналылығы – машиналық аударма жүйелері үшін күрделі кедергілер тудырады. Әсіресе медицина саласындағы арнайы терминдер, қысқартулар, латынша сөздер дәлдігіне айтарлықтай әсер етеді [2].

Әр түрлі корпустар үшін, сонымен қатар медициналық корпустар үшін ағылшын-қазақ және орыс-қазақ тілдеріндегі машиналық аудармасы (МТ) саласында бірнеше әдістемелік тәсілдер қолданылған. Атап айтқанда, Apertium жүйесінде құрылымдық ауысым ережелерін құру тәжірибесі жүргізілген [3], сондай-ақ қазақ тілінің бай морфологиясын ескеру мақсатында морфологиялық сегментация әдістері қолданылған [4]. Сонымен қатар, орыс тіліндегі лемматизация және екі тілдік сөздіктер арқылы сөйлемдерді туралау әдістері зерттелген [5]. Бұл тәсілдер қазақ тілінің синтаксистік және морфологиялық ерекшеліктерін есепке ала отырып, алғашқы машиналық аудармасы жүйелерінің дамуына негіз қалады.

Осы қиындықтарды ескере отырып, қазақ тілінде постредакторлеу жүйесін дамыту қажеттілігі туындайды.

Постредакторлеу дегеніміз – машиналық аудармадан кейін мәтінді адам арқылы өңдеу процесі [6]. Қазіргі таңда көптеген лингвистикалық қызмет көрсетушілер редакторларды оқыту және постредакторлеу әдістерін дамыту жолдарын іздестіріп жатыр. Постредакторлеу және оның қолданылуы бойынша алғашқы зерттеулер 1980-жылдары пайда болды. 1999 ж. Американдық машиналық аударма ассоциациясының және Еуропалық машиналық аударма ассоциациясының мүшелері постредакторлеуге арналған арнайы қызығушылық тобы (Special Interest Group on Post-Editing) құрды [7].

Машиналық аударманы постредакторлеу (РЕМТ) – машиналық аударма жүйелерінің жіберген қателіктерін түзету арқылы соңғы аударманың сапасын арттыру процесі. Бұл әдіс көбінесе ағылшын тілінен басқа тілдерге аударманы жақсарту үшін қолданылады. Алайда, машиналық аударманы постредакторлеуде (РЕМТ) әдістерінің аз ресурсты тілдерге, мысалы қазақ тіліне тиімділігі мәселесі әлі де толық зерттелмеген. Қазақ тілі агглютинативті тіл болғандықтан, машиналық аудармада өзіндік қиындықтар туындайды. Қазақ тіліндегі машиналық аударманы постредакторлеу (РЕМТ) морфология, синтаксис және лексикаға байланысты бірнеше ерекше қиындықтарға тап болады.

Машиналық аударманы постредакторлеуді зерттеу барысында үлкен көлемдегі параллель корпустарды құру жүзеге асырылды [8], олар қоғамдық қолжетімді болып, BLEU көрсеткішін 0,49 дейін жеткізе алады [9]. Соның ішінде осы деректерде оқытылған нейронды машиналық аударма (NMT) үшін постредакторлеу модельдері құрылды [10]. Мәтіндік ресурстардың басым бөлігі мемлекеттік көздерден алынды. Алайда бұл корпуста медицина саласы қарастырылмады.

Бұл мақалада біз постредакторлеу әдістерін арнайы жинақталған медицина саласындағы корпусқа қолдануды қарастырамыз. Зерттеудің мақсаты – аударманы постредакторлеу (РЕМТ) әдістері медицина саласында қазақ тіліндегі машиналық аударманың сапасын жақсарту ала ма, сонымен қатар аз ресурсты тілдермен жұмыс істегенде зерттеушілердің кездесетін негізгі мәселелері мен шектеулерін анықтау болып табылады.

1.1. Машиналық аудармадағы постредакторлеу түрлеріне шолу

Автоматты машиналық аударманы постредакторлеу-мақсатты тілде еркін сөйлейтін маман арқылы жүзеге асырылады. Машиналық аударманы постредакторлеу (РЕМТ) – машиналық аударма нәтижелерін түзету және жетілдіру арқылы олардың сапасын арттыру және табиғи тіл стандартына жақындату процесі [11].

Екі негізгі постредакторлеу түрі қолданылады:

1) Жеңіл постредакторлеу мәтіндегі медициналық терминдерді, сөздердегі морфология, грамматика, лексика тіл бөліктері бойынша қателіктерді азайтады [12]; Жеңіл постредакторлеу деңгейі аз ресурсты тілдер үшін машиналық аудармадан шыққан бастапқы нәтижені алып, мәтінді түсінікті, мазмұндық дәл және грамматикалық дұрыс ету үшін минималды түзетулер енгізуді қамтиды [13;14]. Медицина саласындағы корпустарға жеңіл постредакторлеу (лексика, грамматика) ағылшын-француз тілдеріне тіл мамандары көмегімен бастапқы DeepL жүйесінен алынған сөйлемдер мен сөз тіркестеріне жүргізілген [15]. Бұл жұмыста постредакторлеу қажет ететін қателіктерге лексикалық қателіктер анықталды: терминдердің дұрыс таңдалмауы, әсіресе медициналық терминдердің қате қолданылуы; морфологиялық қателіктер: тілдің ерекшеліктеріне сәйкес келмейтін сәйкестендіру: септік жалғаулар дұрыс жалғанбауы;

2) Толық постредакторлеу медицина саласындағы синтаксис, семантикалық және стильдік қателерді жоюға бағытталған. Нәтижесінде мәтін түсінікті және толық қабылданатын болады. Тілдік сападан гөрі түсініктілікке мән береді және тілдік әрі стильдік мәселелерге рұқсат береді .

Қолданыстағы зерттеулерді талдау негізінде келесі зерттеу тапсырмалары анықталды:

- ♦ Медицина саласында корпус жинау , өңдеу және құрылымдау: Бұл үшін ағылшын – қазақ тілдер жұбынан тұратын медицина саласындағы сөйлемдер мен сөз тіркестері жинақталды;

- ♦ Ағылшын – қазақ тілдер жұбынан тұратын медицина саласындағы корпус үшін постредакторлеу әдісін дамыту: Жеңіл постредакторлеу үшін Kaz-RoBERTa трансформерлік моделі және толық постредакторлеу үшін Үлкен Тілдік Модельдер (LLM) бірі GPT-4.1 модельдер негізінде жүзеге асырылады;

- ♦ Сапаны бағалау: BLEU, TER, METEOR аударма сапалар көрсеткіші бойынша бағалау жүргізілді.

Осы тапсырмаларды жүзеге асыру үшін қазақ тіліндегі медицина саласында сөйлемдерге машиналық аударманы постредакторлеу жүйесін дамыту әдістемесі мен кезеңдері төменде ұсынылады.

Материалдар мен әдістер

Медицина саласындағы сөз тіркестері мен сөйлемдерді ағылшын тілінен қазақ тіліне аудару сапасын арттыру үшін келесі архитектура кезеңдері қарастырылды:

1. Бірінші кезең: Машиналық аударма процесі (ағылшын тілінен қазақ тіліне аудару). Бұл процесс токенизацияны қамтиды, яғни ағылшын тіліндегі сөйлемдерді жеке сөздерге немесе токендерге бөлу; морфологиялық талдау жүргізу; тілдің морфологиялық ерекшеліктерін анықтау; бастапқы аударма үшін оқытылған MarianNMT моделін (трансформер архитектура-сына негізделген) пайдалану; және детокенизация, яғни токенделген мәтінді қайта оқылатын қазақ тіліндегі форматқа келтіру [16;17].

2. Постредакторлеу кезеңі: Алғашқы тексеру барысында қазақ тіліндегі машиналық аударма нәтижелерінде айқын қателер бар-жоғы анықталады. Қазақ тіліндегі машиналық аудар-

ма жеңіл постредакторлеуден өтеді, яғни тек айқын қателер түзетіледі (медициналық термин, лексика, морфология). Кейін жоғары сапалы аударма алу үшін, толық постредакторлеуден өтеді және барлық қателер: стильдік, семантикалық, синтаксис жағынан түзетіледі.

3. Адаптивті кері цикл: Постредакторлеу кезінде қолданылған түзетулер болашақ аудармаларды жақсарту үшін машиналық аударма жүйесіне қайтарылады.

4. Шығыс деректер және тексеру: Сапаны бақылау жүргізіледі, атап айтқанда аударылмаған сегменттердің, сәйкессіздіктердің немесе форматқа қатысты мәселелердің бар-жоғы автоматты түрде тексеріледі. Соңғы тексеру кезеңі аударманың талап етілген сапа стандарттарына сәйкес келетінін қамтамасыз ету үшін жүргізіледі. Ақырында нәтиже дайын аударма түрінде ұсынылып, қажетті форматта немесе ортада шығарылады. Сурет 1-де медицина саласында аз ресурсты тілдер үшін постредакторлеудің жалпы архитектурасы көрсетілген



Сурет 1 – Медицина саласында аз ресурсты тілдер үшін машиналық аударма постредакторлеу жүйесінің архитектурасы

Сурет 1-де көрсетілген архитектурада орындалу процесі келесі қадамдардан тұрады:

1) медицина саласындағы мәтіндерді ағылшын тілінен қазақ тіліне аудару процесі MarianNMT нейронды машиналық аудармада орындалды.

2) Алынған қазақ тіліндегі модельді жеңіл постредакторлеу үшін Kaz-RoBERTa моделі қолданылды.

3) Жеңіл постредакторлеуден кейін алынған аударманы толық постредакторлеу процесі GPT-4.1 моделі арқылы жүзеге асырылды. Төменде әрбір қадам орындалу түсініктемесі беріледі.

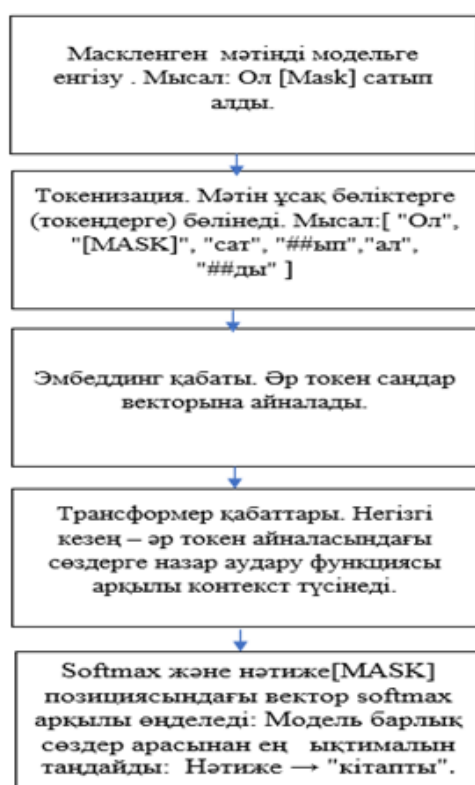
MarianNMT. MarianNMT жүйесі негізінен PyTorch және Hugging Face Transformers фреймворктерін пайдаланып жүзеге асырылған. MarianNMT – ашық дереккөзді нейроаударма жүйесі, оның дайын модельдері Hugging Face репозиторийінде қолжетімді (<https://huggingface.co/models>, қол жеткізілген мерзім: 10 желтоқсан 2025). Ағылшын тілінен қазақ тіліне аудару барысында MarianNMT моделін оқыту үшін ағылшын-қазақ тілдер жұбынан тұратын деректер жиынтығы оқу, тексеру және тестілеу үшін 80%/10%/10% пропорциясына бөлінді. Корпус токенизациядан кейін мәтіндік файлдарда (txt) сақталды. Модельдің энкодер-декодер трансформерлік архитектурасы әр токennің контекстін self-attention арқылы талдайды [18], ал декодер cross-attention арқылы энкодерден шыққан контекстті пайдаланып аударма токenderін жасайды. Бұл тәсіл медициналық мәтіндердің терминологиясы мен грамматикалық құрылымын салыстырмалы түрде дәл аударуға мүмкіндік береді.

Kaz-RoBERTa моделі. Kaz-RoBERTa – қазақ тіліне арнайы үйретілген, RoBERTa негізіндегі трансформерлік модель. Оның дайын модельдері Hugging Face репозиторийінде қолжетімді (<https://huggingface.co/kz-transformers/kaz-roberta-conversational>).

Kaz-RoBERTa моделі жеңіл автоматты постредакторлеу кезінде жоғары көрсеткіш көрсетті [19]. Жұмыста қарастырылған медициналық корпустар үшін толық постредакторлеуде Kaz-RoBERTa моделі қолданылды.

Kaz-RoBERTa модель архитектурасы. Модельге кіріс ретінде қазақ тіліндегі MarianNMT-ден алынған қазақ тіліндегі бастапқы мәтін беріледі. Мәтін субсөздік токенизация арқылы,

SentencePiece әдісімен бөлінеді, және әр токенге оның мәнін және сөйлемдегі орнын көрсететін векторлық ұсыну беріледі. Содан кейін токеннің векторы трансформер қабаттарынан өтеді, мұнда әр токен self-attention механизмі арқылы барлық қалған токендерді «қарап», контексті түсінеді. Көпқабатты назар (multi-head attention) модельге бір уақытта контексттің әртүрлі аспектілерін, соның ішінде көрші сөздерді ескеруге мүмкіндік береді. Назардан кейін әр токен екі қабатты нейрондық желі Feed Forward Network арқылы өтеді, бұл оның ерекшеліктерін күшейтеді және түрлендіреді. Қалдық байланыс (residual connection) және қабаттық нормализация (LayerNorm) қолданылады, бұл ақпаратты сақтауға және оқытуды тұрақтандыруға көмектеседі. Оқыту кезінде бастапқы мәтіннен кездейсоқ токендер маскаланған болып табылады, және модель қалған сөздердің контекстін пайдаланып осы маскаланған токендерді болжауға тырысады, бұл оған мәтіндегі қателерді түзетуге мүмкіндік береді. Модель өзінің болжамдарын эталондық пост-редакторленген мәтінмен салыстырады, осылайша грамматика, лексика және морфология тіл бөлшектерін түзетуді үйренеді. Сурет 2-де Kaz-RoBERTa модель архитектурасы мысалмен көрсетілген.



Сурет 2 – Kaz-RoBERTa моделінің архитектурасын мысалмен көрсету

Kaz-RoBERTa моделінің архитектурасын математикалық моделі бойынша түсіндіретін болсақ:

1. Енгізу қабаты. Kaz-RoBERTa алдымен мәтінді токендерге бөледі және әр токенді вектор түріне түрлендіреді:

$$x_i = TokenEmb(\omega_i) + PosEmb(i)$$

мұнда: ω_i – i – ші токен, $TokenEmb(\omega_i)$ – токен эмбеддинг, $PosEmb(i)$ – позициялық эмбеддинг (сөздердің ретін ескеру үшін)

2. Трансформер қабаттары (энкодер). Модель L қабаттан тұрады. Әр қабатта n- төрт негізгі бөлік бар :

a) Көпқабат өзіндік назар (Multi-Head Self-Attention)

Әрбір қабатта сұраныс (Q), кілт (K) және мән (V) векторлары есептеледі:

$$Q = XW^Q, K = XW^K, V = XW^V$$

Назар функциясы:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

b) Көпқабатты назар функциясы:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_n)W^Q$$

Мұндағы n назар қабаттар саны.

c) Толық байланысқан тор (Feed Forward Network, FFN)

Әр токен үшін екі қабатты нейрондық желі қолданылады:

$$FFN(x) = (0, xW_1 + b_1)W_2 + b_2$$

d) Қалдық байланыс + Нормализация (Residual + LayerNorm)

Әр бөлікке LayerNorm қолданылады:

$$x' = LayerNorm(x + MultiHead(x))$$

$$x'' = LayerNorm(x' + FFN(x'))$$

3. Маска қойылған тілдік модельдеу (Masked Language Modeling)

Модель кездейсоқ таңдалған токендерді "маскалап", соларды болжауға үйренеді:

$$\sigma_{MLM} = - \sum_{i \in \mu}^N \log \log P(\omega_i | x_i)$$

Мұндағы, μ -маскаланған токендердің орындары, ω_i - шынайы токен, x_i -маскаланған токендерді вектор түріне айналдыру.

GPT-4.1 моделі. Толық постредакторлеу үшін аз ресурсты тілдерге арналған LLM модельдерінің бірі (GPT-4.1) моделі қолданылды. Толық постредакторлеуге GPT-4.1 моделі қолданылу себебі төменде зерттеу жұмыстарына байланысты таңдалды.

Хи және т.б. ғалымдар жұмысында үлкен тілдік модельдерді (LLM) машиналық аударма үшін қайта оқытуда маңызды жаңалықтарды енгізді, параметрлік тиімді әдістер аз есептеу ресурстарын пайдалана отырып айтарлықтай өнімділікті арттыра алатынын көрсетті [20]. Бұл әдістер әсіресе төмен ресурсты тіл жұптары үшін тиімді.

Одан кейін бірнеше зерттеулер LLM-дердің аударма қабілеттерін жүйелі түрде бағалады. Hendy және т.б. ғалымдар жұмысында GPT модельдерін жан-жақты бағалап, бұл модельдер нөлдік үлгідегі (zero-shot) көрсеткіштерде жақсы нәтижелер көрсетсе де, әдетте жоғары ресурсты тілдер үшін мамандандырылған аударма жүйелерінен артта қалатынын анықтады [21]. Алайда төмен ресурсты тілдерде өнімділік айырмашылығы әдетте аз болады, бұл LLM-дердің мұндай жағдайларда қолдануға жарамды балама бола алатынын көрсетеді.

Ліао жұмысында ChatGPT-нің аударма өнімділігін нақты талдауға назар аударды және GPT-4 бұрынғы нұсқаларға қарағанда айтарлықтай жақсы аударма сапасына қол жеткізетінін, әсіресе төмен ресурсты тілдер үшін екенін хабарлады [22]. Олардың нәтижелері аударманы оңтайландыру үшін нұсқаулыққа бейімделген модельдердің және мұқият жасалған prompting стратегияларының маңыздылығын көрсетеді.

Модельге қол жеткізу және баптау OpenAI fine-tuning API арқылы жүзеге асырылды, ол бақыланатын бейімдеуге ыңғайлы интерфейс ұсынады. Бастапқы кіріс мәтін ретінде

жеңіл постредакторлеуден Kaz-RoBERTa моделі арқылы алынған аударма қолданылады. Ал шығыс мәтін ретінде аудармашылар көмегімен дұрыс аударылған мәтін беріледі. Бұл тілдік жұп арқылы бақыланатын қайта оқытылған модель (fine-tuning SFT) процедурасы модельді қателері бар қазақ тіліндегі аудармаларды түзетілген нұсқаларға айналдыруға үйреттіледі.

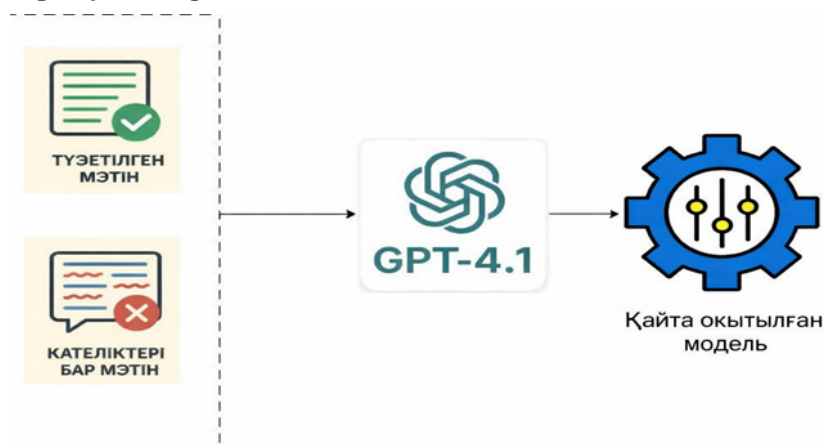
Оқу деректері сұхбат форматы (prompt-response) түрінде құрылды, бұл модельдің өзара әрекет стиліне сәйкес келеді:

User: [ҚАТЕЛІКТЕРІ БАР ҚАЗАҚ СӨЙЛЕМІ]

Assistant: [ТҮЗЕТІЛГЕН ҚАЗАҚ СӨЙЛЕМІ]

GPT-4.1 моделі орындалу алгоритмі Сурет 3-те көрсетілген.

Бұл формат модельге мақсатты түзетулерді орындауға және бастапқы мәтіннің мағынасын сақтауға нақты нұсқаулық береді.



Сурет 3 – GPT-4.1 моделінің орындалу алгоритмі

Fine-tuning келесі гиперпараметрлермен жүргізілді:

- ◆ Оқыту жылдамдығы мультипликаторы: 2
- ◆ Batch өлшемі: 13 мысал
- ◆ Оқу кезеңдері (epochs): 2 (валидация нәтижесіне сүйене отырып анықталды)

Біз оқыту барысын валидация шығыны арқылы бақыладық және overfitting қаупін азайту үшін ерте тоқтатуды (early stopping) қолдандық. Сурет 4 және Сурет 5-те оқыту шығыны мен дәлдіктің эволюциясы көрсетілген.



Сурет 4 – Итерациялар бойынша оқыту шығыны



Сурет 5 – Итерациялар бойынша оқыту дәлдігі

Деректер жиынтығы. Медицина саласынан 102374 ағылшын-қазақ тілдер жұбынан тұратын корпус әзірленді. Корпус сөйлемдер мен сөз тіркестерін дереккөздерден ағылшын тілінде жинап, аудармашылардың көмегімен қазақ тіліне аударды. Корпусқа алынған дереккөздер мыналар: Халықаралық клиникалық хаттамалар; Ғылыми медициналық мақалалар; Фармацевтикалық нұсқаулықтар мен дәрілік аннотациялар; Ресми медициналық веб-сайттар. Кесте 1-де әрбір дереккөздің сілтемесі мен одан алынған сөйлемдер саны көрсетілген.

Кесте 1– Медициналық корпустағы сөйлемдер саны

Алынған дереккөздері	Деректер саны
Халықаралық клиникалық хаттамалар (WHO, NICE, CDC) (https://www.who.int/publications/who-guidelines?utm_source=chatgpt.com)	28,435
Ғылыми медициналық мақалалар (PubMed, ResearchGate, т.б.) (https://pmc.ncbi.nlm.nih.gov/search/?term=%223%20Biotech%22%5Bjournal%5D)	25,210
Фармацевтикалық нұсқаулықтар мен дәрілік аннотациялар (https://open.fda.gov/apis/drug/label/download/?utm_source=chatgpt.com)	18,950
Ресми медициналық веб-сайттар (денсаулық сақтау министрлігі, NIH, FDA) (https://www.cdc.gov/respiratory-viruses/about/index.html)	29,779
Барлығы	102,374

Корпустағы сөйлемдер мен сөз тіркестерді жинау үшін арнайы бағдарлама құрылды. Бұл бағдарлама PDF форматындағы құжаттардан және медицина саласындағы ресми сайттардан терминдер, сөз тіркестері мен сөйлемдерді автоматты түрде таңдап, корпус құрамын толықтырады.

Нәтижелер мен талқылау

Эксперимент келесі сипаттамалары бар компьютерде жүргізілді: CPU – Core i7 4790K, 32 GB RAM, 1 TB SSD, GPU: RTX 2070 Super, GTX 1080.

Ағылшын–қазақ тіліндегі аударма сапасын бағалау үшін келесі метрикалар қолданылды: BLEU [23], METEOR [24] және TER [25].

BLEU (Bilingual Evaluation Understudy). Машиналық аударманың сапасын n-граммдар бойынша эталонмен салыстыра отырып бағалайды.

$$BLEU = BP * EXP \left(\sum_{n=1}^N \omega_n \log \log p_n \right)$$

Мұндағы, p_n -n-грамм дәлдігі (мысалы, unigram, bigram және т.б.); ω_n - n-граммдарға берілетін салмақ ($\omega_n = \frac{1}{N}$); BP – Brevity Penalty, қысқалық жазасы. BLEU жоғары болған сайын (максимум = 1.0), аударма сапасы жақсы деген мағынаны білдіреді.

METEOR (Metric for Evaluation of Translation with Explicit Ordering). Сөздердің дәлдігімен қоса, синонимдер, түбірлес сөздер (стемминг) және сөздердің реті ескереді.

$$METEOR = (1 - Penalty) \cdot F_{mean}$$

Мұндағы: $F_{mean} = \frac{10 \cdot P \cdot R}{9R + P}$ – Precision мен Recall-дың үйлесімді орташа мәні;

$$Penalty = 0.5 \cdot \left(\frac{chunks}{matches} \right)^3$$
 – сөздердің ретінің бұзылғанына берілетін айып

P – Precision (гипотеза ішіндегі дұрыс сөздер үлесі)

R – Recall (эталондағы сөздермен салыстырғанда)

METEOR жоғары болған сайын (максимум = 1.0), аударма сапасы да жақсы болады.

TER (Translation Edit Rate). Гипотезаны эталонға айналдыру үшін қажет редакциялау қадамдарының санын өлшейді.

$$TER = \frac{\text{Редакция қадамдарының саны}}{\text{эталондағы сөздер саны}}$$

TER неғұрлым төмен болса, соғұрлым жақсы (минимум = 0). Бұл аудармаға аз түзету қажет дегенді білдіреді.

Модельдерді тестілеу үшін медицина саласынан 10000 сөйлем мен сөз тіркестерінен тұратын корпус құрылады.

Осы корпус MarianNMT нейронды машиналық аудармада ағылшын тілінен қазақ тіліне және салыстыру мақсатында қазақ тілінен ағылшын тіліне аударылады. Кесте 2-де аудармалардың нәтижелері BLEU, TER, METEOR метрикалық көрсеткіштерімен көрсетіледі.

Кесте 2 – MarianNMT нейронды машиналық аудармадан алынған аударма нәтижесі

Тіл жұптары	BLEU	TER	METEOR
Ағылшын-қазақ	0.594	0.294	0.752
Қазақ-ағылшын	0.585	0.251	0.746

Нәтижелерге сүйенсек, модельдің ағылшын-қазақ бағытындағы BLEU көрсеткіші 0.594, ал қазақ-ағылшын бағытында 0.585. Бұл мәндер аударма сапасының екі бағытта да салыстырмалы түрде жоғары екенін көрсетеді. TER шамасы сәйкесінше 0.294 және 0.251 құрап, аудармалардағы қателік деңгейінің төмен екендігін білдіреді. Ал METEOR метрикасы 0.752 және 0.746 мәтіннің мазмұндық және лексикалық дәлдігінің жоғары екенін айғақтайды.

Жалпы алғанда, MarianNMT моделі медициналық саласындағы сөйлемдерді аударуда ағылшын және қазақ тілдері арасындағы машиналық аудармада қанағаттанарлық нәтиже көрсетіп, екі бағытта да аударманың сапасын сақтай отырып жұмыс істейтінін байқауға болады.

Кесте 3-те MarianNMT нейронды машиналық аудармада алынған қазақ тіліндегі аудармаларын сапасын жақсарту үшін жеңіл постредакторлеу (Kaz-RoBERTa моделі) және толық постредакторлеу (GPT-4.1-mini моделі) нәтижелері алынды.

MarianNMT нейрондық машиналық аудармаларын Kaz-RoBERTa моделі арқылы постредакторлеу аударманың сапасын айтарлықтай жақсарттады. BLEU көрсеткіші 9%-ға, METEOR 5%-ға өсті, мамандардың бағалауынша лексикалық сәйкестік деңгейі де артты. TER метрикасы бойынша көрсеткіш 0,294-тен 0,145-ке дейін төмендеді, бұл қателер санын екі есеге

жуық қысқарғанын көрсетеді. Жеңіл постредакторлеуден кейін аудармалар эталондық мәтінге әлдеқайда жақындайды.

Кесте 3 – Қазақ тілінде алынған аударманы жеңіл және толық постредакторлеуден кейінгі нәтижелер

Модельдер	BLEU	TER	METEOR
Kaz-RoBERTa	0.682	0.145	0.805
GPT-4.1-mini	0.820	0.078	0.885

Толық постредакторлеу нәтижесінде BLEU көрсеткіші жеңіл постредакторлеуге қарағанда 14%-ға жоғары, TER 7%-ға төмен, METEOR 8%-ға өскен. Бұл көрсеткіштер GPT-4.1-mini моделі арқылы алынған аударманың Kaz-RoBERTa моделінен алынған аудармадан сапасы жоғары екенін дәлелдейді. Тіл мамандарының пікірінше, толық постредакторлеуден кейінгі аударма жеңіл постредакторлеуден кейінгі аудармамен салыстырғанда семантикасы мен стилі бойынша айтарлықтай жақсарған.

Қорытынды

Жүргізілген зерттеу нәтижелері ағылшын тіліндегі медициналық мәтіндерді қазақ тіліне аударуда машиналық аударма мен постредакторлеу модельдерін кешенді қолданудың тиімді екенін көрсетті. MarianNMT негізінде алынған бастапқы аудармалар жалпы мағынаны жеткізгенімен, терминологиялық, лексикалық және стилистикалық қателердің кездесетіні анықталды. Бұл қателерді азайтуда постредакторлеу кезеңінің маңызы зор екені дәлелденді.

Kaz-RoBERTa трансформерлік моделін қолдану нәтижесінде медициналық терминдердің дәлдігі артып, лексикалық қателер саны едәуір қысқарды. Ал GPT-4.1 үлкен тілдік моделін пайдалану аударманың семантикалық тұтастығы мен стильдік сапасын айтарлықтай жақсартып, мәтінді кәсіби медициналық ғылыми тіл талаптарына жақындатты. Метрикалық бағалау нәтижелері де бұл тұжырымды растап, GPT-4.1 моделінің аударма сапасын бастапқы MarianNMT нәтижелерімен салыстырғанда BLEU метрикасы бойынша 23%-ға дейін арттырғанын көрсетті.

Зерттеу қорытындылары медициналық мәтіндерді қазақ тіліне аударуда толық немесе жартылай автоматтандырылған постредакторлеу жүйелерін енгізудің өзектілігін айқындайды. Ұсынылған тәсілдер қазақ тілімен типологиялық тұрғыдан жақын түркі тілдері үшін де қолданбалы маңызға ие болып, медицина саласындағы көптілді ақпарат алмасудың сапасын арттыруға үлес қоса алады. Болашақта доменге бейімделген тілдік модельдерді жетілдіру және медициналық корпус көлемін ұлғайту аударма сапасын одан әрі жақсартуға мүмкіндік береді.

Қаржыландыру туралы ақпарат

Бұл зерттеу Қазақстан Республикасының жоғары білімі және Ғылым министрлігінің қолдауымен BR24993001 жобасымен қаржыландырылды.

ӘДЕБИЕТТЕР

1 Issakova, S., Khassangaliyeva, B., Issakova, A., Taganova, A., Toxanbayeva, T., Kamarova, N., and Kuzdybaeva, A. The frame representation of medical terminological system in Kazakh and English. *Forum for Linguistic Studies*, 7 (5), 739–747 (2025). <https://doi.org/10.30564/fls.v7i5.9233>

2 Kucherenko, O.F., Kuanysheva, A.B., and Keller Deditskaya, E.R. The corpus of borrowings and their functioning in the medical terminology of Kazakhstan (on the material of professional periodicals). *Bulletin of the Karaganda University. Philology Series*, 105 (1), 54–61 (2022). <https://doi.org/10.31489/2022Ph1/54-61>

- 3 Forcada, M.L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J.A., Sánchez-Martínez, F., Ramírez-Sánchez, G., and Tyers, F.M. Apertium: A free/opensource platform for rule-based machine translation. *Machine Translation*, 25, 127–144 (2011). <https://doi.org/10.1007/s10590-011-9090-0>
- 4 Assylbekov, Z., and Nurkas, A. Initial explorations in Kazakh to English statistical machine translation. In: *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 and of the Fourth International Workshop EVALITA 2014*, pp. 12–16 (2014). Available at: <https://clic2014.fileli.unipi.it/proceedings/vol1/CLICIT201413.pdf>
- 5 Assylbekov, Z., Myrzakhmetov, B., and Makazhanov, A. Experiments with Russian to Kazakh sentence alignment. In: *Proceedings of the 4th International Conference on Computer Processing of Turkic Languages* (2016). Available at: <https://nur.nu.edu.kz/handle/123456789/1694>
- 6 Vieira, L.N., Alonso, E., and Bywood, L. Introduction: Post-editing in practice—Process, product and networks. *Journal of Specialized Translation*, 31, 2–13 (2019). <https://doi.org/10.26034/cm.jostrans.2019.173>
- 7 Shterionov, D., do Carmo, F., Moorkens, J., Hossari, M., Wagner, J., Paquin, E., Schmidtke, D., Groves, D., and Way, A. A roadmap to neural automatic post-editing: An empirical approach. *Machine Translation*, 34, 67–96 (2020). <https://doi.org/10.1007/s10590-020-09249-7>
- 8 Rakhimova, D., and Karibayeva, A. Aligning and extending technologies of parallel corpora for the Kazakh language. *Eastern European Journal of Enterprise Technologies*, 4 (2(118)), 32–39 (2022). <https://doi.org/10.15587/1729-4061.2022.259452>
- 9 Zhumanov, Z., and Tukeyev, U. Integrated technology for creating quality parallel corpora. In: *Advances in Computational Collective Intelligence* (Springer, Cham, 2021), pp. 511–524. https://doi.org/10.1007/978-3-030-88113-9_41
- 10 Karyukin, V., Rakhimova, D., Karibayeva, A., Turganbayeva, A., and Turarbek, A. The neural machine translation models for the low-resource Kazakh–English language pair. *PeerJ Computer Science*, 9, e1224 (2023). <https://doi.org/10.7717/peerj-cs.1224>
- 11 Rakhimova, D., Sagat, K., Zhakypbaeva, K., and Zhunussova, A. Development and study of a post-editing model for Russian-Kazakh and English-Kazakh translation based on machine learning. In: *Advances in Computational Collective Intelligence* (Springer, Cham, 2021), pp. 525–534. https://doi.org/10.1007/978-3-030-88113-9_42
- 12 Gulcehre, C., Ahn, S., Nallapati, R., Zhou, B., and Bengio, Y. Pointing the unknown words. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 140–149 (2016). <https://doi.org/10.18653/v1/P16-1014>
- 13 Li, X., Zhang, J., and Zong, C. Towards zero unknown word in neural machine translation. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 2852–2858 (New York, NY, USA, 2016). Available at: <https://www.ijcai.org/proceedings/2016>
- 14 Rakhimova, D., Turarbek, A., Karyukin, V., Karibayeva, A., and Turganbayeva, A. The development of the light post-editing module for English-Kazakh translation. In: *Proceedings of the 7th International Conference on Engineering & MIS* (Almaty, Kazakhstan, 2021). <https://doi.org/10.1145/3492547.3492651>
- 15 Martikainen, H. Post editing neural MT in medical LSP: Lexico grammatical patterns and distortion in the communication of specialized knowledge. *Informatics*, 6 (3), 26 (2019). <https://doi.org/10.3390/informatics6030026>
- 16 Lee, W., Park, J., Go, B.-H., and Lee, J.-H. Transformer-based automatic post-editing with a context-aware encoding approach for multi-source inputs. *arXiv preprint arXiv:1908.05679* (2019).
- 17 Rubino, R., Huet, S., Lefèvre, F., and Linarès, G. Statistical post editing of machine translation for domain adaptation. In: Cettolo, M., Federico, M., Specia, L., and Way, A. (Eds.), *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, pp. 221–228 (2012). Available at: <https://aclanthology.org/2012.eamt-1.55/>
- 18 Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Hermann, U., Aji, A.F., Bogoychev, N., Martins, A.F.T., and Birch, A. Marian: Fast neural machine translation in C++. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 116–121 (2018). <https://doi.org/10.18653/v1/P18-4020>
- 19 Baitenova, L., Tussupova, S., Mambetov, S., Munaitbas, G., and Mukhamejanova, G. Hybrid artificial intelligence architectures for automatic text correction in the Kazakh language. *Frontiers in Artificial Intelligence*, 8, Article 1708566 (2025). <https://doi.org/10.3389/frai.2025.1708566>

20 Xu, H., Kim, Y.J., Sharaf, A., and Awadalla, H.H. A paradigm shift in machine translation: Boosting translation performance of large language models. arXiv preprint arXiv:2309.11674 (2023).

21 Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y.J., Afify, M., and Awadalla, H.H. How good are GPT models at machine translation? A comprehensive evaluation. arXiv preprint arXiv:2302.09210 (2023).

22 Jiao, W., Wang, W., Huang, J.-T., Wang, X., and Tu, Z. Is ChatGPT a good translator? Yes with GPT-4 as the engine. arXiv preprint arXiv:2301.08745 (2023).

23 Reiter, E. A structured review of the validity of BLEU. Computational Linguistics, 44 (3), 393–401 (2018). https://doi.org/10.1162/COLI_a_00322

24 Lavie, A., and Agarwal, A. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In: Proceedings of the Second Workshop on Statistical Machine Translation, pp. 228–231 (2007). URL: <https://aclanthology.org/W07-0734>

25 Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. A study of translation edit rate with targeted human annotation. In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers, pp. 223–231 (2006). URL: <https://aclanthology.org/2006.amta-papers.25/>

¹Рахимова Д.,

доцент, ORCID ID: 0000-0003-1427-198X,

e-mail: drakhimova060@gmail.com

^{1,2,3,*}Жігер А.,

магистр, ORCID ID: 0000-0002-3641-8260,

*e-mail: aliya.zhunussova.zh@narxoz.kz

^{3,4}Малых В.,

к.т.н., ORCID ID: 0009-0008-5632-6188,

e-mail: valentin.malykh@phystech.edu

¹Мансурова М.,

к. ф.-м. н., профессор, ORCID ID: 0000-0002-9680-2758,

e-mail: mansurova.madina@gmail.com

¹Карюкин В.И.,

PhD, ORCID ID: 0000-0002-8768-0349,

e-mail: vladislav.karyukin@gmail.com

¹Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

²Университет Нархоз, г. Алматы, Казахстан

³Международный университет информационных технологий, г. Алматы, Казахстан

⁴Санкт-Петербургский государственный университет информационных технологий,
механики и оптики, г. Санкт-Петербург, Россия

МЕТОДЫ ПОСТРЕДАКТИРОВАНИЯ МЕДИЦИНСКИХ ТЕКСТОВ НА КАЗАХСКОМ ЯЗЫКЕ И КАЧЕСТВЕННЫЙ АНАЛИЗ

Аннотация

В настоящее время в сфере медицины качественный и точный перевод с английского языка на казахский является одной из важных задач для обеспечения доступности и безопасности медицинской информации. В данной научной работе была исследована точность и эффективность систем машинного перевода, широко используемых на практике, таких как Google Translate и Yandex Translate, применительно к медицинским текстам. Основной целью исследования является рассмотрение способов семантически и стилистически корректного перевода медицинских терминов и сложных предложений с английского языка на казахский. В рамках исследования был сформирован специальный корпус, включающий 102,374 предложения, собранные из международных медицинских статей, клинических исследований и описаний лекарственных препаратов. Корпус был обучен на нейронной системе машинного перевода MarianNMT и переведен на казахский язык. Для легкого постредактирования результатов перевода использовалась трансформерная модель Kaz-RoBERTa, а для полного постредактирования – одна из больших языковых моделей (LLM), а именно GPT-4.1, при этом были рассмотрены ее возможности адаптации к медицинским текстам. Качество пере-

вода оценивалось с использованием метрик BLEU, TER и METEOR. Результаты переводов, полученные после первоначального машинного перевода MarianNMT, были сравнены с результатами после обработки моделями Kaz-RoBERTa и GPT-4.1. Анализ показал, что переводы после применения модели Kaz-RoBERTa продемонстрировали улучшение на 9% по сравнению с исходным переводом MarianNMT, а после использования модели GPT-4.1 – на 23%.

Ключевые слова: нейронный машинный перевод, MarianNMT, Kaz-RoBERTa, трансформерная модель, медицинские тексты, BLEU, TER, METEOR.

¹Rakhimova D.,

Associate Professor, ORCID ID: 0000-0003-1427-198X,

e-mail: drakhimova060@gmail.com

^{1,2,3*}**Zhiger A.,**

MSc, ORCID ID: 0000-0002-3641-8260,

*e-mail: aliya.zhunussova.zh@narxoz.kz

^{3,4}**Malykh V.,**

Cand. Tech. Sc., ORCID ID: 0009-0008-5632-6188,

e-mail: valentin.malykh@phystech.edu

¹Mansurova M.,

Cand. Phys.-Math. Sci., Professor, ORCID ID: 0000-0002-9680-2758,

e-mail: mansurova.madina@gmail.com

¹Karyukin V.,

PhD, ORCID ID: 0000-0002-8768-0349,

email: vladislav.karyukin@gmail.com

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan

²Narxoz University, Almaty, Kazakhstan

³International University of Information Technologies, Almaty, Kazakhstan

⁴Saint Petersburg State University of Information Technologies, Mechanics and Optics,
Saint Petersburg, Russia

METHODS OF POST-EDITING MEDICAL TEXTS IN THE KAZAKH LANGUAGE AND QUALITATIVE ANALYSIS

Abstract

At present, in the field of medicine, high-quality and accurate translation from English into Kazakh is one of the key challenges in ensuring the accessibility and safety of medical information. This scientific study investigates the accuracy and effectiveness of machine translation systems widely used in practice, such as Google Translate and Yandex Translate, when applied to medical texts. The main objective of the study is to explore methods for achieving semantically and stylistically correct translation of medical terminology and complex sentences from English into Kazakh. For this purpose, a specialized corpus consisting of 102,374 sentences was compiled from international medical articles, clinical studies, and drug descriptions. The corpus was processed using the MarianNMT neural machine translation system and translated into Kazakh. For light post-editing of the translation results, the transformer-based Kaz-RoBERTa model was employed, while full post-editing was carried out using one of the large language models (LLMs), namely GPT-4.1, whose adaptability to medical texts was also examined. Translation quality was evaluated using the BLEU, TER, and METEOR metrics. The translations obtained after the initial MarianNMT machine translation were compared with the results after post-editing using the Kaz-RoBERTa and GPT-4.1 models. The analysis showed that translations processed with the Kaz-RoBERTa model achieved an 9% improvement over the baseline MarianNMT translations, while the use of the GPT-4.1 model resulted in a 23% improvement.

Keywords: neural machine translation, MarianNMT, Kaz-RoBERTa, transformer model, medical texts, BLEU, TER, METEOR.

Received August 10, 2025; revised January 23, March 2, 17, 2026; accepted March 27, 2026.

UDC 004.8
IRSTI 28.23

<https://doi.org/10.55452/1998-6688-2026-23-2-218-232>

^{1*}**Kapparova A.,**

PhD student, ORCID ID: 0009-0004-4639-3548,

*e-mail: kapparova_ainur3@kaznu.edu.kz

¹**Zholamanov B.,**

PhD student, ORCID ID: 0000-0001-8206-7425,

e-mail: zholamanov.batyrbek@kaznu.kz

¹**Bolatbek A.,**

PhD student, ORCID ID: 0009-0004-7613-5507,

e-mail: bolatbek.askhat@kaznu.kz

¹**Kuttybay N.,**

PhD, ORCID ID: 0000-0002-5723-6642,

e-mail: kuttybyy.nurzhigit@kaznu.kz

¹**Dosymbetova G.,**

PhD, ORCID ID: 0000-0002-3935-7213,

e-mail: gulbakhar.dosymbetova@kaznu.edu.kz

¹**Zhumagaliyev Ye.,**

Master student, ORCID ID: 0009-0002-9213-2555,

e-mail: zhumagaliyev_y0@live.kaznu.kz

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan

IMPLEMENTATION OF MULTI-AGENT FRAMEWORK OF IMPALA FOR A SINGLE ZONE TEMPERATURE CONTROL OF A SIMULATED THERMAL ZONE

Abstract

Buildings account for a significant portion of global energy consumption, with HVAC and humidity-control systems representing the majority of their operational demand. Traditional rule-based strategies often fail to adapt to dynamic indoor-outdoor conditions, motivating the use of data-driven control methods. This study presents a multi-agent reinforcement learning (MARL) framework for simultaneous temperature and humidity control in a single-zone building modeled in EnergyPlus. The proposed approach employs the distributed Importance Weighted Actor-Learner Architecture (IMPALA) algorithm with centralized training and decentralized execution (CTDE), enabling two agents: temperature and humidity to learn coordinated policies directly from high-fidelity simulation feedback. The results demonstrate strong learning performance: both agents improved their per-step rewards substantially (temperature +18.9%, humidity +33.7%), indicating effective convergence and cooperative behavior. The learned controller maintained thermal comfort comparable to the rule-based baseline (mean occupied temperature difference ≈ 0.04 °C; occupied PMV ≈ 0.45) while achieving notable energy savings. Total annual HVAC energy consumption decreased by 8.9%, with the most significant improvement observed in humidification energy, which was reduced by 34.4%. Heating and cooling loads remained nearly unchanged, confirming that energy reductions were achieved without compromising comfort.

Keywords: building model, multi-agent framework, multi-agent reinforcement learning, intelligent control, Importance Weighted Actor-Learner Architecture

Received November 17, 2025; revised February 19, 21, 26, 2026; accepted March 31, 2026

1 Introduction

Globally, buildings account for approximately 30% of total energy consumption, making them one of the largest contributors to global energy use [1]. Among building subsystems, the heating, ventilation, and air-conditioning (HVAC) system is the primary consumer, typically responsible for nearly 50% of a building's total energy demand [2]. In densely populated tropical and subtropical cities, air-conditioning and mechanical ventilation (ACMV) systems can account for up to 70% of building energy use [3]. As global energy resources continue to deplete, improving the efficiency of HVAC and ACMV systems and reducing their energy consumption are key priorities for achieving sustainable and low-carbon buildings.

While reducing energy use is essential, it must be achieved without compromising occupant comfort. Thermal comfort depends on temperature, humidity, air quality, and other environmental factors. Traditionally, buildings have used rule-based control (RBC) systems, which operate HVAC equipment according to predefined rules and schedules derived from expert knowledge [4, 5]. Although such methods are easy to implement, they lack adaptability to nonlinear and dynamic environmental conditions, often leading to inefficient energy performance and comfort degradation.

In contrast, model-based control methods such as Model Predictive Control (MPC) have been widely adopted in recent years. MPC forecasts future thermal dynamics using a mathematical model and computes optimal control actions over a prediction horizon. It can significantly reduce energy use while maintaining comfort by considering forecasted loads and system responses. However, building accurate thermal and moisture models is complex, time-consuming, and requires specialized expertise. When these models fail to represent real-world dynamics accurately, system performance deteriorates. Moreover, MPC methods are computationally expensive, especially in multi-zone buildings where each thermal zone introduces additional state variables [6].

To support such advancements, a variety of high-fidelity building simulation tools have been developed, including EnergyPlus [7–10]. These platforms accurately simulate thermodynamic phenomena such as heat transfer, airflow, humidity regulation, and system dynamics, allowing for in-depth analysis of building performance. Despite their precision, the embedded mathematical models are typically nonlinear and computationally intensive, requiring extensive calibration and user input. Consequently, they are often impractical for real-time control applications or iterative reinforcement learning loops [11].

Effective building operation requires managing the interactions between ventilation, heating, and lighting systems while maintaining overall energy balance. For instance, increased ventilation raises the heating load due to the need to warm incoming air, while continuous heating during unoccupied hours results in energy waste. Adaptive and dynamic control strategies that consider occupancy schedules, solar gains, and internal loads can minimize energy use while maintaining comfort [12]. However, even the most advanced model-based control methods cannot perfectly capture real-time variations in weather, occupancy, or user behavior. Therefore, model-free, intelligent control techniques capable of learning optimal actions directly from environmental feedback have become an emerging research focus.

Recent breakthroughs in artificial intelligence have introduced Reinforcement Learning (RL) and, more specifically, Deep Reinforcement Learning (DRL) as promising approaches for adaptive building energy control. Unlike traditional control strategies, RL does not rely on explicit physical modeling; instead, it learns optimal control policies through trial-and-error interactions with the environment [13]. Data-driven DRL frameworks have demonstrated strong potential for managing complex multi-zone HVAC systems [14]. For example, Barrett and Linder [15] reported that reinforcement learning-based HVAC control achieved up to 10% energy savings compared with programmable thermostats.

Despite these advantages, online DRL methods where agents learn directly from physical building interactions – require long training durations and may risk comfort violations during early

learning stages. For instance, Mocanu et al. [16] demonstrated that DRL agents needed an entire year of continuous operation to converge to near-optimal control policies. As a result, simulation-based pre-training in virtual environments such as EnergyPlus is now a preferred approach, providing a safe and efficient means for policy development and evaluation before deployment.

Within the actor–critic family of RL algorithms, the Deep Deterministic Policy Gradient (DDPG) method has been successfully applied to continuous control tasks such as cooling optimization in data centers [17] and transactive HVAC systems [18]. However, single-agent DRL becomes inefficient in multi-zone environments due to exponentially growing state–action spaces. To overcome this, researchers have proposed Multi-Agent Reinforcement Learning (MARL) frameworks, where multiple agents collaborate or compete to optimize system-level objectives. For example, [19] formulated HVAC control as a Markov Game, enabling decentralized agents to coordinate local policies. Similarly, [20] proposed strategies to reduce action–state complexity and improve convergence while preserving energy efficiency and occupant comfort.

The integration of Deep Reinforcement Learning (DRL) and Multi-Agent Reinforcement Learning (MARL) in HVAC systems thus represents a significant step toward intelligent, adaptive, and energy-efficient buildings. By overcoming the limitations of traditional rule-based and model-based methods, these data-driven approaches enable real-time optimization of temperature, humidity, and ventilation in response to dynamic environmental and occupancy conditions. As research in this field advances, MARL-based frameworks are expected to play a key role in the development of smart, self-learning building systems that jointly optimize comfort, energy, and sustainability.

In this study, we implement a multi-agent reinforcement learning framework for the automatic control of temperature and relative humidity in a single zone building model simulated in EnergyPlus. The framework is built upon the Importance Weighted Actor-Learner Architecture (IMPALA) algorithm which is a distributed actor–critic DRL method capable of parallelizing data collection and learning across multiple agents. Two agents are designed: a temperature-control agent that manages heating and cooling setpoints, and a humidity-control agent that regulates humidifier operation.

The novelty of this research lies in the implementation of IMPALA with a multi-agent framework for building energy management system. The control algorithm is integrated the EnergyPlus simulation platform via a custom Python–EnergyPlus interface. This connection allows the agents to interact with the high-fidelity building simulation in real time and learn optimal control policies that balance:

- ♦ Occupant comfort, measured through Predicted Mean Vote (PMV) and relative humidity;
 - ♦ Energy efficiency, quantified via heating, cooling, and humidifier energy consumption; and
 - ♦ Occupancy-dependent operation, distinguishing between occupied and unoccupied periods.
- ♦ Compared with conventional rule-based control, the proposed IMPALA-based MARL system achieved:

- ♦ Up to 9% total energy savings, primarily due to a 34% reduction in humidifier energy consumption,
- ♦ Minimal comfort deviation, maintaining nearly the same indoor temperature and PMV levels as the baseline, and
- ♦ Improved policy stability, with agents showing consistent reward improvement over training iterations.

This work contributes to the development of simulation-integrated multi-agent reinforcement learning architectures for intelligent building energy management. It establishes a scalable experimental foundation for extending toward multi-zone, multi-agent EnergyPlus simulations, facilitating the design of next-generation sustainable, comfort-preserving, and autonomous HVAC control systems for low-carbon smart buildings. The paper’s structure is outlined as follows: In Methods section, the methodology of the proposed architecture and the simulation environment framework are delineated, Results section presents the training and simulation results. Finally, Conclusions concludes the study.

2 Materials and methods

2.1 Formulation of the HVAC Control Problem

In this study, the control of single zone building HVAC systems is formulated as a Markov Decision Process (MDP) framework. Under the simplified assumption of building thermal dynamics, the indoor temperature at a given time step depends primarily on the system's previous state such as the indoor temperature and control inputs from the preceding interval and is independent of earlier time steps. Consequently, the HVAC control process can be modeled as a finite Markov process, making it suitable for solution using Deep Reinforcement Learning (DRL) techniques.

The DRL-based control framework for building HVAC systems comprises five core components: the environment, the agent, the state space (s), the action space (a), and the reward function (r). In this context, the zone and its HVAC system represent the environment, while the DRL controller acts as the agent. The agent observes the current conditions of the environment through the state space, which encodes information such as zone temperature, humidity, occupancy, and energy usage. Based on these observations, the agent selects an action from its discrete or continuous action space such as adjusting temperature setpoints or modifying humidity states to influence the environment.

After the environment transitions to a new state in response to the agent's action, a reward signal is generated, reflecting the effectiveness of the chosen action in achieving the control objectives: energy efficiency and occupant comfort.

This reward function serves as the learning feedback that guides the agent toward developing an optimal control policy, enabling it to make progressively better decisions through iterative interaction with the environment. As a DRL technique Importance Weighted Actor-Learner Architecture (IMPALA) is used.

2.2 IMPALA algorithm

In this study, the Importance Weighted Actor-Learner Architecture (IMPALA) algorithm is adopted as the distributed reinforcement learning framework. IMPALA efficiently utilizes computational resources on a single machine and can scale to large distributed systems while maintaining high data efficiency. It employs an actor-critic architecture where multiple actors generate experience trajectories in parallel, and centralized learners update the global policy π and value function V_π based on received trajectories.

The learning and acting processes (Figure 1) are decoupled, enabling high-throughput training. To address off-policy discrepancies caused by asynchronous updates between actors and the learner, IMPALA integrates a correction mechanism known as V-trace, which ensures stable and efficient policy optimization under distributed conditions.

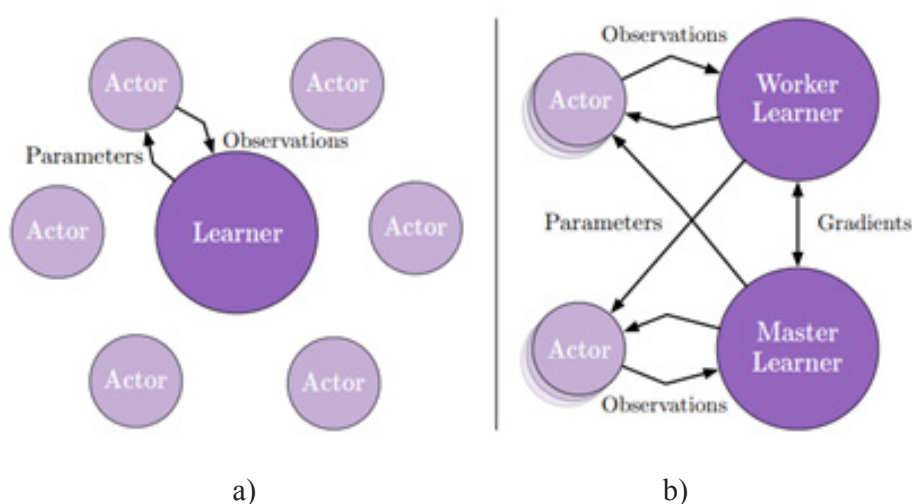


Figure 1 – The learning and acting processes: a – Single Learner; b – Multiple Synchronous Learners [21]

IMPALA efficiently utilizes computational resources on a single machine and scales seamlessly across multiple systems while maintaining data efficiency. It follows an actor–critic architecture, where multiple actors generate experience trajectories in parallel, and a centralized learner updates the global policy (π) and value function ($V\pi$). The decoupling of experience collection from learning allows high-throughput training [21].

To address the off-policy nature of this setup caused by delays between actors and the learner IMPALA integrates a correction mechanism known as V-trace, which stabilizes learning. In the V-trace actor–critic algorithm, the value parameters θ are updated through gradient descent to minimize the difference between the estimated value $V_\theta(x_s)$ and the V-trace target v_s . The policy parameters ω are optimized using the policy gradient weighted by the importance-sampling ratio ρ_s , which corrects for deviations between the behavior policy μ and the target policy π_ω :

$$\rho_s \nabla_\omega \log \pi_\omega(a_s | x_s) (r_s + \gamma v_{s+1} - V_\theta(x_s)) \quad (1)$$

To enhance policy exploration and avoid premature convergence, an entropy regularization term is added, encouraging the agent to explore diverse actions:

$$-\nabla_\omega \sum_a \pi_\omega(a | x_s) \log \pi_\omega(a | x_s) \quad (2)$$

The overall parameter update is obtained by combining these gradients with appropriate weighting coefficients (hyperparameters) [21]. This configuration enables stable and efficient learning under distributed asynchronous conditions, making IMPALA with V-trace particularly well-suited for continuous control tasks such as building HVAC and humidity regulation.

2.3 State space, Action space, Reward function

Observation vector (11 variables): [OAT, IAT, RH, CO₂, CLG_SPT, HTG_SPT, CLG_J, HTG_J, OCC, HUM_J, PMV] covering environmental and energy states. Two mechanisms ensure numerical stability and smooth learning: RunningStat class: Implements Welford’s algorithm to maintain running mean and variance for observation normalization. EMAStd class keeps an exponential moving estimate of reward standard deviation to normalize and stabilize reward signals. Normalization is essential because EnergyPlus variables vary across wide ranges.

Action Spaces and Mapping: 1) Temperature agent: MultiDiscrete action [cool_index, heat_index] mapped to physical setpoints (e.g., 25–27 °C for cooling, 21–23 °C for heating). A deadband constraint ensures logical order: heating setpoint < cooling setpoint. 2) Humidity agent (z1_hum): Discrete levels mapped linearly to the humidifier command [0.0, 1.0]. This discretization makes control tractable for RL while preserving physical interpretability.

Reward function used in this study:

$$R = -(\alpha_{\text{temp,hum}})(E_{\text{cost}} + \text{comfort cost} + \text{penalties}) \quad (3)$$

with normalization and clipping to stabilize training.

The reward balances comfort and energy efficiency, using both linear and nonlinear components (Table 1).

Table 1 – Reward function components

Component	Description	Applies to
PMV comfort cost	Penalizes deviation of PMV from tolerance band	Temperature agent
Humidity discomfort	Penalizes deviation of RH from comfort range (35–55%)	Humidity agent
Energy cost	Penalizes cooling, heating, or humidifier energy (kWh)	Both
Occupancy weighting	Different weights for occupied vs unoccupied times	Both
Work-hour boost	Comfort is more important during working hours	Temp agent
Setpoint churn penalty	Discourages rapid setpoint changes	Temp agent
Extreme setpoint penalty	Penalizes setting at physical limits	Temp agent
No-fight penalty	Penalizes simultaneous heating and cooling	Temp agent

2.4 Algorithm structure and training parameters

The proposed control framework employs the IMPALA integrated with Centralized Training and Decentralized Execution (CTDE) over the EnergyPlus simulation environment. The framework enables distributed multi-agent learning for indoor temperature and humidity control.

1. Actors (Rollout Workers): Multiple rollout workers sample trajectories from the EnergyPlus environment, each containing two agents: temperature and humidity.

2. V-trace Correction: Off-policy correction converts actor trajectories collected under a behavior policy into learner updates consistent with the target policy.

3. Learning Phase: The central learner updates a shared or separate policy for each agent using the IMPALA actor-critic setup. An LSTM module can be applied to handle partial observability in building dynamics.

4. Callback Metrics: Episode-level data such as energy consumption (kWh), thermal comfort (PMV), and humidity discomfort are aggregated. Reward normalization may be applied online to stabilize training.

5. CTDE Scheme: During centralized training, the environment provides a global critic observation that incorporates all agents' information. During decentralized execution, each agent acts independently using its own local policy.

IMPALA employs V-trace to correct for policy lag between actors and the learner. The hyperparameters of the current architecture are given in Table 2.

Table 2 – Optimization hyperparameters

Hyperparameter (flag)	Meaning	Values
--gamma	Discount factor	0.95
--lr	Adam learning rate	5e-4
--entropy-coeff	Policy entropy bonus	1e-3
--grad-clip	Global grad clip (L2)	40.0
--train-batch-size	SGD batch size	4000
--rollout-fragment-length	Steps per sample batch	50
--vtrace-clip-rho / --vtrace-clip-pg-rho	Importance ratio clips	1.0 / 1.0
--use-lstm + --lstm-cell + --lstm-seq-len	Recurrent policy	off / 256 / 20

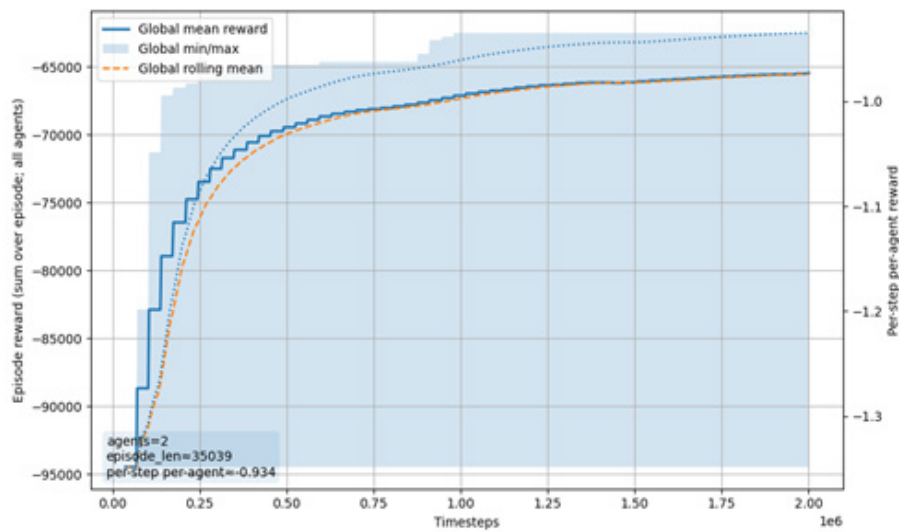
Recurrent policies LSTM are configured with --lstm-cell 256 and --lstm-seq-len 20, while the burn-in argument is retained for compatibility with previous configurations.

3 Results and discussion

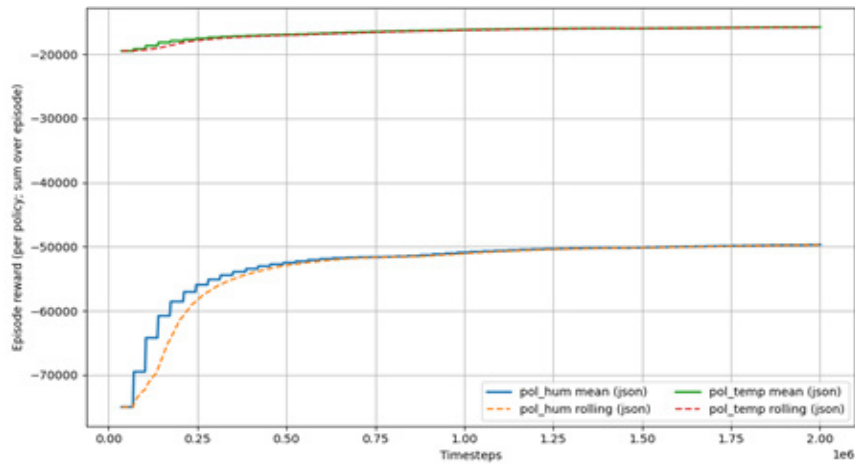
As studied environment a room with an area 41.54 m² and volume 147.88 m³ is modeled in EnergyPlus. The simulated environment uses weather of Almaty for one year: KAZ_ALA_Almaty.368700_TMYxEPW.epw which is from the database of EnergyPlus [22]. Two actuators were used in the simulation: zone temperature control (cooling/heating setpoints) and humidifier power value. These mappings allow EnergyPlus to execute actions from RL agents in real-time simulation via the PyEnergyPlus API. Simulations can be daily episodes, in this case, 24 hours × 4 steps per hour.

3.1 Training Performance

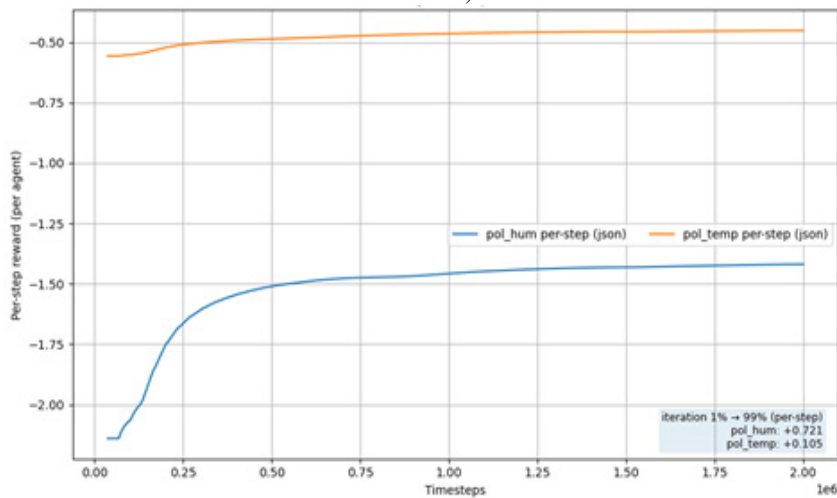
The training performance in reinforcement learning is evaluated by reward progress. Figure 2 shows reward progress of multi-agent IMPALA algorithms which have 2 agents. The training process used 2,000,000 timesteps and consisted of 56 episodes, each spanning one year. During training, both agents showed stable learning progress: 1) Temperature policy: mean step reward improved from -0.557 to -0.451, representing a +18.9% increase in average per-step reward; 2) Humidity policy: mean step reward improved from -2.140 to -1.418, corresponding to a +33.7% increase in control efficiency. These results indicate that both controllers successfully learned more balanced strategies, reducing penalties associated with comfort violations and energy use.



a)



b)



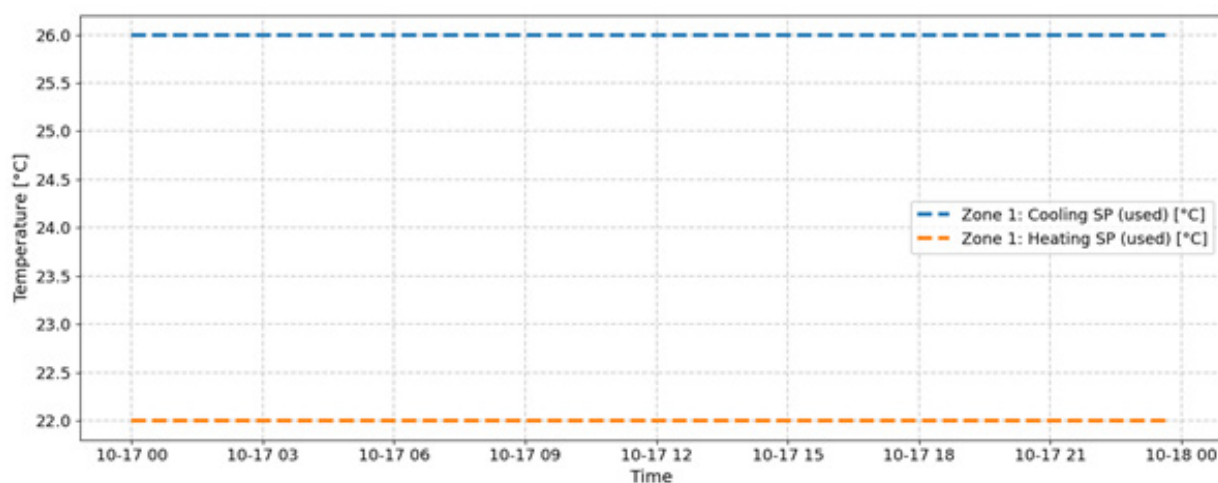
c)

Figure 2 – Reward results: a – episode reward (sum over episode for all agents);
b – per-policy episode rewards; c – per-policy per-step rewards

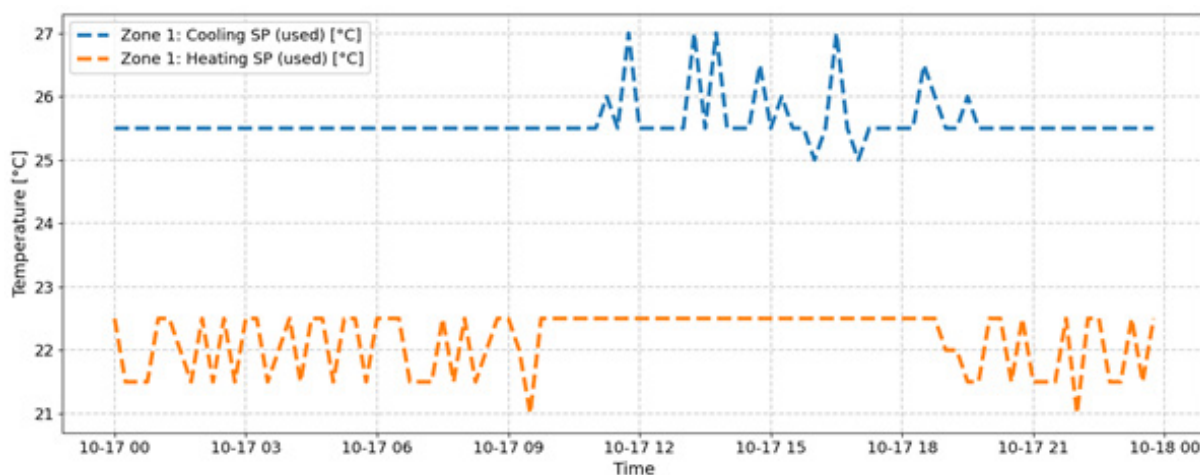
As Figure 2 shows, in the first step of training, the reward value of both agents was low, and then the training began to stabilize from 500,000 timesteps. This may indicate that the RL control system is unstable in the first steps and that the energy saving or the comfort level of people may decrease. Therefore, using the weights of trained agents in practice can give the desired result.

3.2 Thermal Comfort Analysis

In the case of rule-based control method, the zone thermostat gets constant values which are 22°C for heating and 26°C for cooling. In the case of scheduled RBC control, the heating setpoint is set to 22°C between 09:00 and 18:00 and 21°C at other times, and the cooling setpoint was set to 26°C between 09:00 and 18:00 and 27°C at other times. Figure 3a and Figure 3b show the thermostat setpoint values in two cases for one day, which is the 17th of October: rule-based control and multi-agent IMPALA control methods. In the case of MA IMPALA, the thermostat could choose heating setpoint in the range of 21-23°C, cooling setpoint in the range of 25-27°C.



a)



b)

Figure 3 – Heating and cooling setpoints for the 17th of October:
a – rule-based control; b – IMPALA control

PMV (Predicted Mean Vote) from EnergyPlus' Fanger model is used as the primary comfort indicator. The PMV and temperature profiles demonstrate that the multi-agent controller could maintain occupant comfort within acceptable limits throughout the year. Figure 4 demonstrates

annual PMV values for each month when using IMPALA. Annual mean PMV, when the zone was occupied, reached +0.453, corresponding to a slightly warm but comfortable range ($-0.5 < \text{PMV} < +0.5$). The monthly PMV trend peaked during summer months (June–August), aligning with higher indoor air temperatures ($\approx 25^\circ\text{C}$). During winter (January–March), PMV values dropped slightly below zero (cool sensation), indicating seasonal adaptability of the learned policy.

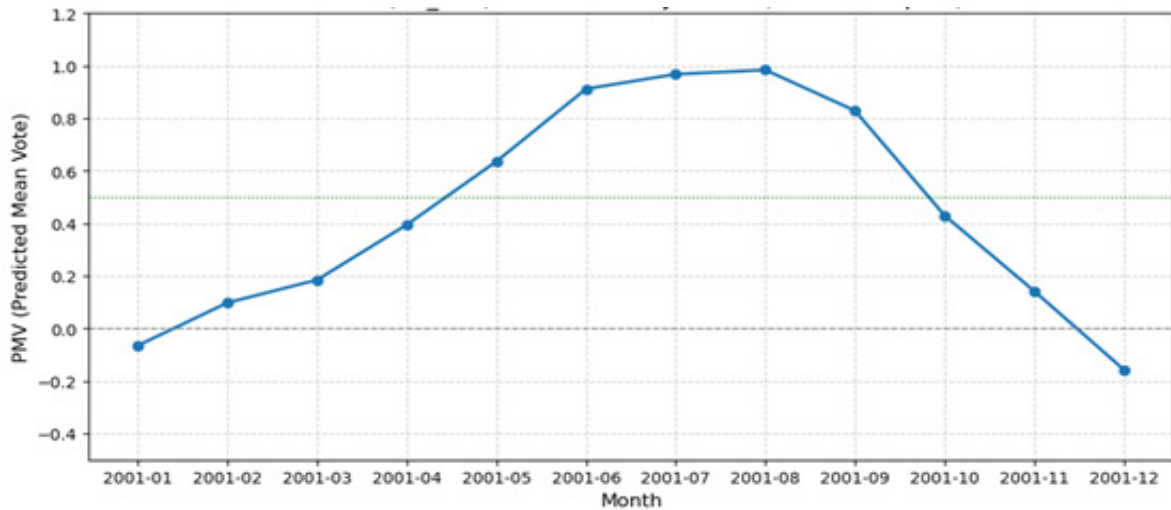


Figure 4 – Annual PMV values for each month when the zone is occupied

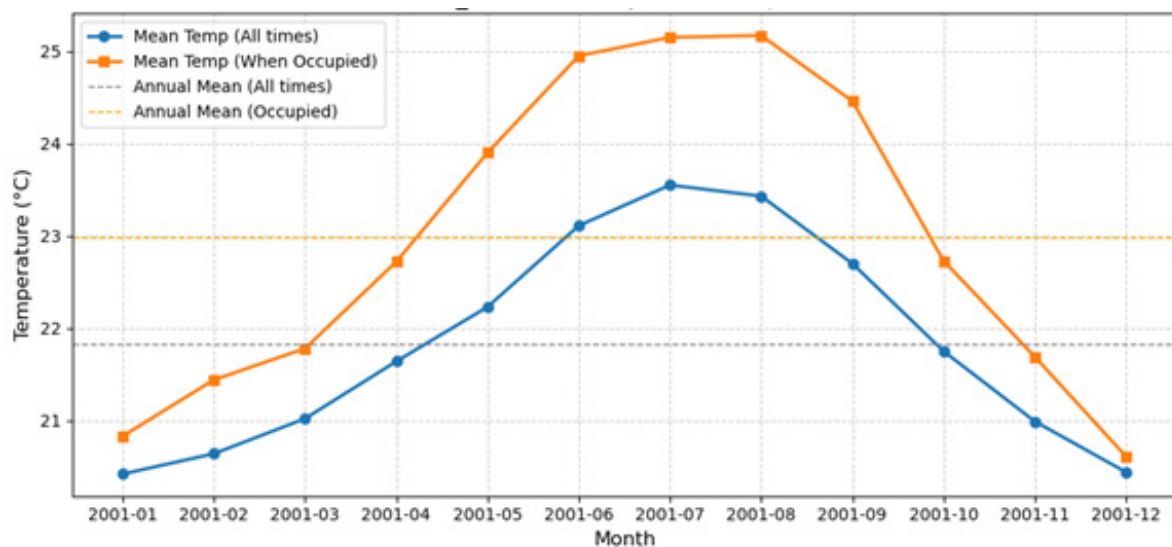


Figure 5 – Monthly mean temperature

Figure 5 demonstrates monthly mean temperature when using IMPALA in both cases: occupied and unoccupied. The mean indoor temperature when occupied was maintained close to the comfort reference. The value of mean indoor temperature of one year was 23.019 in rule-based control case and 22.978 in MARL control case which is similar in two cases. This small temperature difference shows that comfort was preserved while improving energy efficiency.

Mean temperature deviation was compared between episodes of MARL, rule-based control and scheduled control algorithms. The results are shown in Figure 6, according to the graph, while in the first episode the mean temperature deviation from the comfort range was 17%, in the last 56 episodes it decreased to 14%. And this indicator is about 15% better than the scheduled RBC.

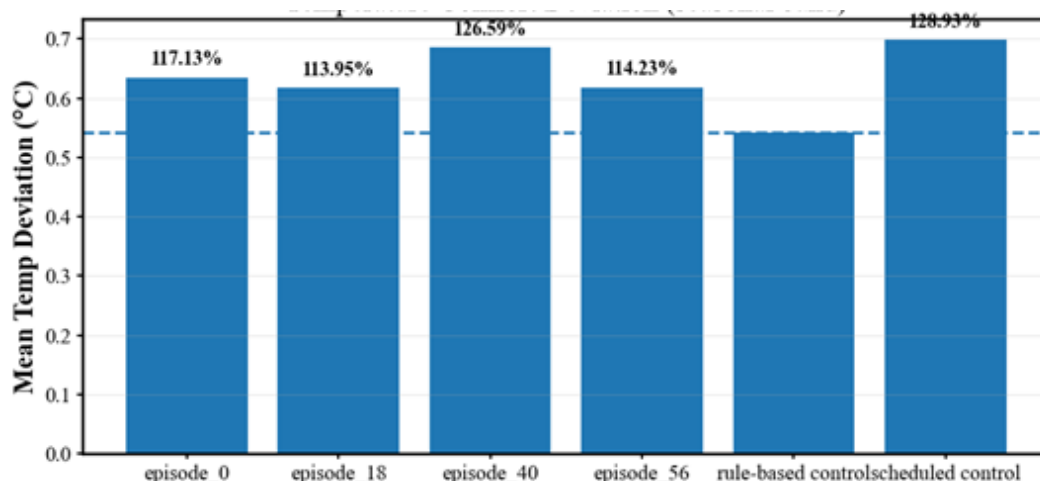


Figure 6 – Temperature comfort deviation

3.3 Humidity Control

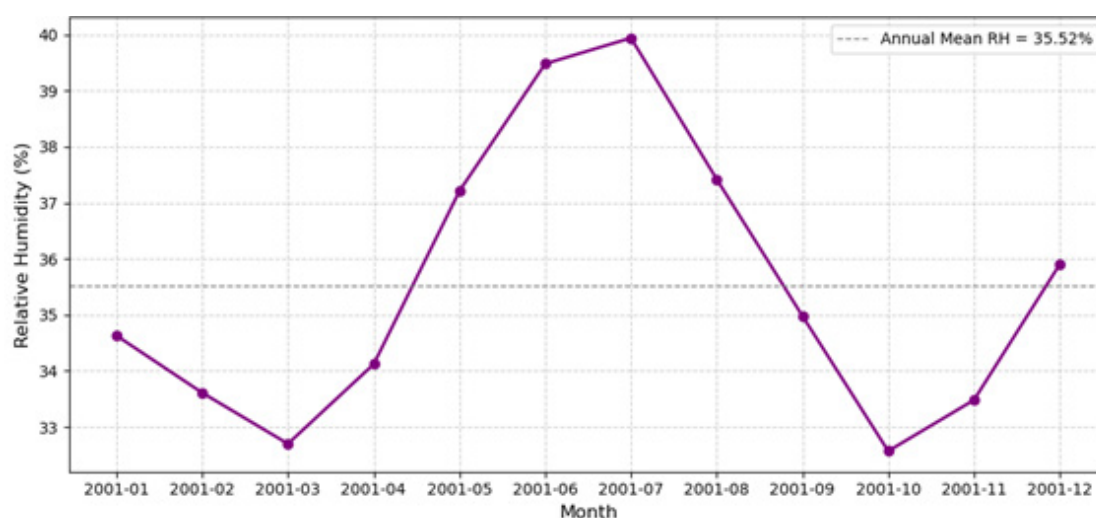


Figure 7 – The average annual relative humidity

When using IMPALA, the average annual relative humidity (RH) was approximately 35.5%, with seasonal variation between 33–40% (Figure 7). The multi-agent controller effectively maintained RH within the desired range (35–55%) most of the time, avoiding excessive humidifier usage.

The mean RH deviation, when the zone is occupied, of MARL episodes was compared with RBC and scheduled RBC control methods. The results in Figure 8 show that the deviation was very large in the first episode, while the relative humidity deviation decreased in the last episode. Scheduled control was less effective in maintaining relative humidity in the comfort range than RBC and MARL.

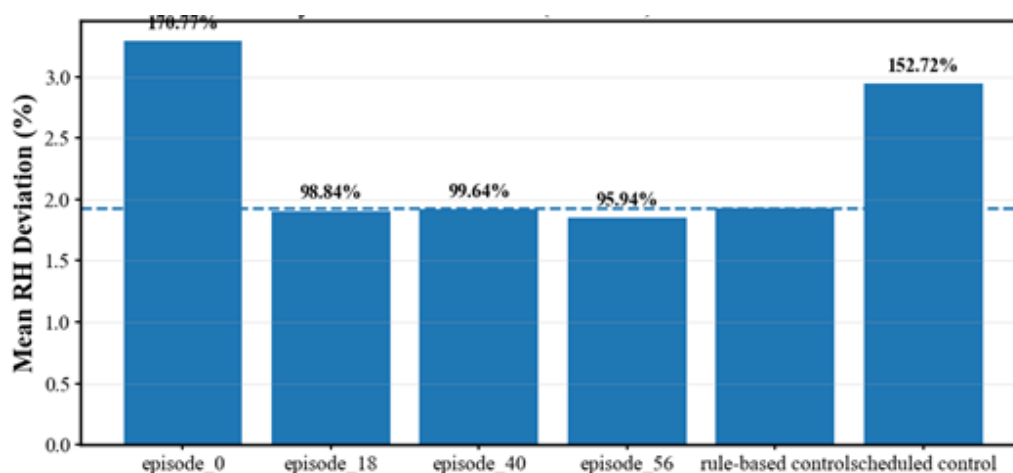


Figure 8 – Humidity comfort deviation when the zone is occupied

3.4 Energy Consumption

Table 3 shows a detailed comparison between rule-based and multi-agent control for the zone that shows significant energy savings.

Table 3 – Comparison between rule-based and multi-agent control

Energy End-Use	Rule-based (kWh)	Multi-agent (kWh)	Δ (kWh)	Δ (%)
Total	28,220	25,698	-2,521.7	-8.9%
Humidifier	7,257.5	4,759.6	-2,497.9	-34.4%
Heating	17,624.8	17,571.9	-53.0	-0.3%
Cooling	3,337.8	3,367.0	+29.2	+0.9%

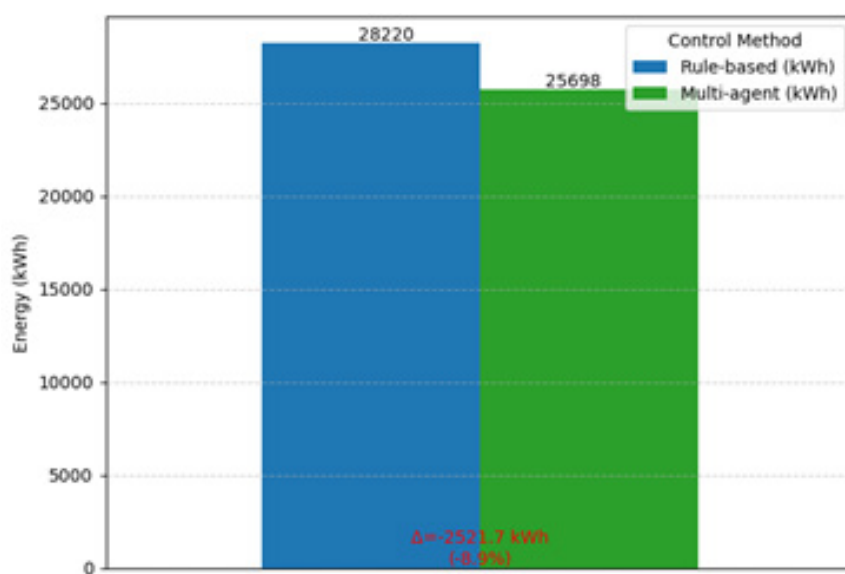


Figure 9 – Comparison of total energy consumption

The multi-agent controller notably reduced humidifier electricity usage by 34.4%, while maintaining comparable heating and cooling loads. This resulted in an overall 8.9% reduction in total annual HVAC energy consumption (Figure 9). This shows that treating the problem of energy efficient use of a building as a Markov game, evaluating the system's decisions in the certain states through rewards is an advantageous method for minimizing energy consumption while maintaining target temperatures for people in the zone.

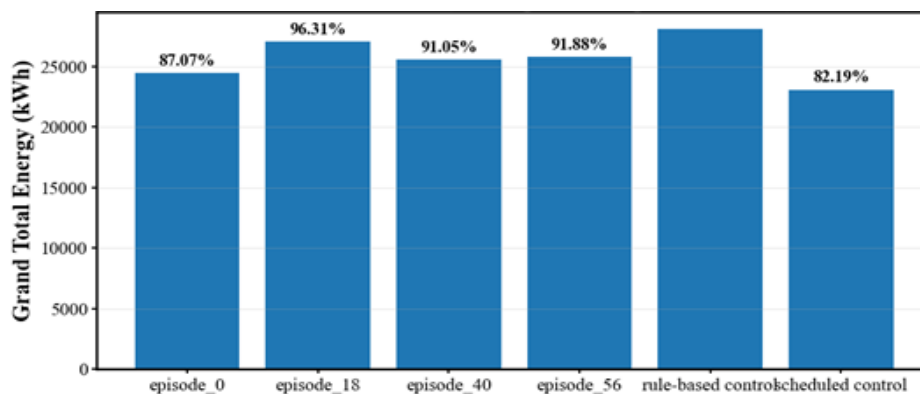


Figure 10 – Total energy consumption of different algorithms

Figure 10 shows the least energy-efficient control system, the scheduled RBC system with 17.81% reduction. However, based on the above analysis, we can see that the scheduled RBC reduces occupancy comfort when controlling temperature and humidity. In the first episode of subsequent MARL control, although the energy consumption was reduced by 12.93%, the deviation of temperature and humidity from the comfort range was greater than in episode 56.

Overall, the advantages of the proposed method are 1) physically consistent control: all actuators and feedbacks are real EnergyPlus quantities; 2) multi-objective optimization: combines energy, comfort, and operational stability; 3) scalable design: easily extendable to multi-zone and multi-agent systems; 4) supports CTDE: ideal for MARL algorithms like DDPG, IMPALA, or MAPPO; 5) reusable: flexible normalization and reward shaping for different climates and buildings.

The limitation of the proposed algorithm is its difficulty in real-world implementation. The implementation of the method requires more computer resources and technical support than RBC and scheduled control. However, this algorithm can be used to control the operation of the other thermostat and humidifier using transfer learning. Thus, we can implement the algorithm in the energy management system of another building, which indicates a high level of scalability of the method.

Conclusions

The proposed multi-agent reinforcement learning approach successfully optimized temperature and humidity control for the single-zone building model. Both temperature and humidity agents demonstrated effective learning convergence and coordination, improving their per-step rewards by 18.9% and 33.7%, respectively. The learned controller maintained indoor comfort close to the rule-based baseline ($\Delta T \approx 0.04$ °C, $PMV \approx 0.45$) while reducing energy consumption by nearly 9%. The largest improvement was achieved in humidifier energy savings (–34.4%), showing that the RL agent effectively avoided unnecessary operation under favorable conditions. These findings confirm that data-driven RL-based controllers can achieve energy-efficient comfort management in dynamic building environments, outperforming static rule-based strategies.

Acknowledgment

This study was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No. AP26197509 - Intelligent control system for building energy consumption and occupant comfort using multi-agent reinforcement learning).

REFERENCES

- 1 Dean, B., et al. Towards zero-emission efficient and resilient buildings. Global Status Report (2016).
- 2 Razmara, M., et al. Optimal exergy control of building HVAC system. Applied Energy, 156, 555–565 (2015). <https://doi.org/10.1016/j.apenergy.2015.07.051>
- 3 Chua, K.J., et al. Achieving better energy-efficient air conditioning – a review of technologies and strategies. Applied Energy, 104, 87–104 (2013). <https://doi.org/10.1016/j.apenergy.2012.10.037>
- 4 Maasoumy, M., et al. Handling model uncertainty in model predictive control for energy efficient buildings. Energy and Buildings, 77, 377–392 (2014). <https://doi.org/10.1016/j.enbuild.2014.03.057>
- 5 Salakij, S., et al. Model-Based Predictive Control for building energy management. I: Energy modeling and optimal control. Energy and Buildings, 133, 345–358 (2016). <https://doi.org/10.1016/j.enbuild.2016.09.044>
- 6 Yang, S., et al. Experiment study of machine-learning-based approximate model predictive control for energy-efficient building control. Applied Energy, 288, 116648 (2021). <https://doi.org/10.1016/j.apenergy.2021.116648>
- 7 Crawley, D.B., et al. EnergyPlus: creating a new-generation building energy simulation program. Energy and Buildings, 33(4), 319–331 (2001). [https://doi.org/10.1016/S0378-7788\(00\)00114-6](https://doi.org/10.1016/S0378-7788(00)00114-6)
- 8 Strachan, P.A., Kokogiannakis, G., Macdonald, I.A. History and development of validation with the ESP-r simulation program. Building and Environment, 43(4), 601–609 (2008). <https://doi.org/10.1016/j.buildenv.2006.06.025>
- 9 Salvalai, G. Implementation and validation of simplified heat pump model in IDA-ICE energy simulation environment. Energy and Buildings, 49, 132–141 (2012). <https://doi.org/10.1016/j.enbuild.2012.01.038>
- 10 Shrivastava, R.L., Kumar, V., Untawale, S.P. Modeling and simulation of solar water heater: A TRNSYS perspective. Renewable and Sustainable Energy Reviews, 67, 126–143 (2017). <https://doi.org/10.1016/j.rser.2016.09.005>
- 11 Koeln, J., et al. Multi-zone temperature modeling and control. In: Intelligent Building Control Systems: A Survey of Modern Building Control and Sensing Strategies. Springer, Cham, 139–166 (2017). https://doi.org/10.1007/978-3-319-68462-8_6
- 12 Perera, D.W.U., Pfeiffer, C.F., Skeie, N.-O. Control of temperature and energy consumption in buildings – a review. International Journal of Energy & Environment, 5(4) (2014).
- 13 Sutton, R.S., Barto, A.G. Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998). <https://doi.org/10.1017/S0263574799271172>
- 14 Li, F., Du, Y. Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. In: Deep Learning for Power System Applications. Springer, Cham, 71–96 (2023). https://doi.org/10.1007/978-3-031-45357-1_4
- 15 Barrett, E., Linder, S. Autonomous HVAC control, a reinforcement learning approach. In: ECML PKDD. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23461-8_1
- 16 Mocanu, E., et al. On-line building energy optimization using deep reinforcement learning. IEEE Transactions on Smart Grid, 10(4), 3698–3708 (2018). <https://doi.org/10.1109/TSG.2018.2834219>
- 17 Li, Y., et al. Transforming cooling optimization for green data center via deep reinforcement learning. IEEE Transactions on Cybernetics, 50(5), 2002–2013 (2019). <https://doi.org/10.1109/TCYB.2019.2927410>
- 18 Liu, B., Akcakaya, M., McDermott, T.E. Automated control of transactive HVACs in energy distribution systems. IEEE Transactions on Smart Grid, 12(3), 2462–2471 (2020). <https://doi.org/10.1109/TSG.2020.3042498>
- 19 Kazmi, H., et al. Multi-agent reinforcement learning for modeling and control of thermostatically controlled loads. Applied Energy, 238, 1022–1035 (2019). <https://doi.org/10.1016/j.apenergy.2019.01.140>
- 20 Blad, C., Bøgh, S., Kallesøe, C. A multi-agent reinforcement learning approach to price and comfort optimization in HVAC-systems. Energies, 14(22), 7491 (2021). <https://doi.org/10.3390/en14227491>

21 Espeholt, L., et al. IMPALA: Scalable distributed deep-RL with importance weighted actor-learner architectures. Proceedings of ICML (2018).

22 EnergyPlus Weather Data. URL: <https://energyplus.net/weather> (accessed: 2026.06.01).

^{1*}Каппарова А.,

докторант, ORCID ID: 0009-0004-4639-3548,

*e-mail: kapparova_ainur3@kaznu.edu.kz

¹Жоламанов Б.,

докторант, ORCID ID: 0000-0001-8206-7425,

e-mail: zholamanov.batyrbek@kaznu.kz

¹Болатбек А.,

докторант, ORCID ID: 0009-0004-7613-5507,

e-mail: bolatbek.askhat@kaznu.kz

¹Құттыбай Н.,

PhD, ORCID ID: 0000-0002-5723-6642,

e-mail: kutybyy.nurzhigit@kaznu.kz

¹Досымбетова Г.,

PhD, ORCID ID: 0000-0002-3935-7213,

e-mail: gulbakhar.dossymbetova@kaznu.edu.kz

¹Жұмағалиев Е.,

магистрант, ORCID ID: 0009-0002-9213-2555,

e-mail: zhumagaliyev_y0@live.kaznu.kz

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

СИМУЛЯЦИЯЛАНҒАН ЖЫЛУ АЙМАҒЫНЫҢ БІР БӨЛІГІНДЕГІ ТЕМПЕРАТУРАНЫ БАСҚАРУҒА АРНАЛҒАН IMPALA КӨПАГЕНТТІ ҚҰРЫЛЫМЫН ЖҮЗЕГЕ АСЫРУ

Аңдатпа

Ғимараттар әлемдік энергия тұтынудың айтарлықтай бөлігін құрайды, олардың ішінде жылыту, желдету және ауаны баптау (HVAC), сондай-ақ ылғалдылықты реттеу жүйелері негізгі үлеске ие. Ережеге негізделген дәстүрлі басқару стратегиялары ішкі және сыртқы ортаның динамикалық жағдайларына бейімделе алмайды, сондықтан деректерге негізделген басқару әдістерін қолдану өзекті болып отыр. Зерттеу жұмысында EnergyPlus бағдарламасында модельденген бір зоналы ғимараттағы температура мен ылғалдылықты бір мезгілде басқаруға арналған көпагентті күшейте оқыту (MARL) жүйесін ұсынады. Ұсынылған тәсіл орталықтандырылған оқыту және орталықсыз орындау (CTDE) қағидасына негізделген таратылған Importance Weighted Actor-Learner Architecture (IMPALA) алгоритмін қолданады. Мұнда екі агент – температура және ылғалдылық агенттері – жоғары дәлдіктегі симуляциялық кері байланыс негізінде үйлестірілген саясаттарды үйренеді. Нәтижелер оқытудың жоғары тиімділігін көрсетті: екі агенттің де орташа қадамдық марапаттары айтарлықтай артты (температура бойынша +18.9%, ылғалдылық бойынша +33.7%), бұл олардың тиімді өзара әрекеттесуін және тұрақты үйренуін дәлелдейді. Үйретілген басқару жүйесі ережеге негізделген базалық жүйемен салыстырғанда жылулық жайлылық деңгейін (орташа айырмашылық ≈ 0.04 °C, PMV ≈ 0.45) сақтай отырып, энергияны үнемдеуге қол жеткізді. Жылдық HVAC жүйесінің энергия тұтынуы 8.9%-ға азайды, ал ең үлкен үнем ылғалдандыру энергиясында байқалды – 34.4%-ға төмендеді. Жылыту және салқындату жүктемелері іс жүзінде өзгеріссіз қалды, бұл жайлылық деңгейін төмендетпей-ақ энергия үнемдеуге қол жеткізілгенін растайды.

Түйін сөздер: ғимарат моделі, көпагенттік құрылым, көпагенттік күшейту арқылы оқыту, интеллектуалды басқару, Importance Weighted Actor-Learner Architecture.

^{1*}Каппарова А.,

докторант, ORCID ID: 0009-0004-4639-3548,

*e-mail: kapparova_ainur3@kaznu.edu.kz

¹Жоламанов Б.,

докторант, ORCID ID: 0000-0001-8206-7425,

e-mail: zholamanov.batyrbek@kaznu.kz

¹Болатбек А.,

докторант, ORCID ID: 0009-0004-7613-5507,

e-mail: bolatbek.askhat@kaznu.kz

¹Куттыбай Н.,

PhD, ORCID ID: 0000-0002-5723-6642,

e-mail: kutybyy.nurzhigit@kaznu.kz

¹Досымбетова Г.,

PhD, ORCID ID: 0000-0002-3935-7213,

e-mail: gulbakhar.dossymbetova@kaznu.edu.kz

¹Жумагалиев Е.,

магистрант, ORCID ID: 0009-0002-9213-2555,

e-mail: zhumagaliyev_y0@live.kaznu.kz

¹Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

РЕАЛИЗАЦИЯ МНОГОАГЕНТНОЙ АРХИТЕКТУРЫ IMPALA ДЛЯ УПРАВЛЕНИЯ ТЕМПЕРАТУРОЙ В ОДНОЙ ЗОНЕ СИМУЛИРОВАННОГО ЗДАНИЯ

Аннотация

Здания составляют значительную долю мирового энергопотребления, в том числе системы отопления, вентиляции и кондиционирования воздуха (HVAC) и управления влажностью, формируют основную часть их эксплуатационных энергозатрат. Традиционные стратегии управления, основанные на фиксированных правилах, часто не способны адаптироваться к динамически изменяющимся внутренним и внешним условиям, что обуславливает необходимость применения методов управления на основе данных. В данном исследовании представлен фреймворк многоагентного обучения с подкреплением (MARL) для одновременного управления температурой и влажностью в однозонном здании, смоделированном в среде EnergyPlus. Предложенный подход использует распределенную архитектуру Importance Weighted Actor-Learner Architecture (IMPALA) с централизованным обучением и децентрализованным исполнением (CTDE), что позволяет двум агентам – температурному и влажностному – обучаться согласованным стратегиям непосредственно на основе обратной связи высокоточной симуляционной модели. Полученные результаты демонстрируют высокую эффективность обучения: оба агента существенно улучшили среднее вознаграждение на шаге (температура +18,9%, влажность +33,7%), что свидетельствует об успешной сходимости и кооперативном поведении. Обученный контроллер обеспечивает уровень теплового комфорта, сопоставимый с базовой стратегией управления на основе правил (средняя разница температуры в занятый период $\approx 0,04$ °C; среднее значение PMV в занятый период $\approx 0,45$), при одновременном достижении значительной экономии энергии. Совокупное годовое энергопотребление системы HVAC снижается на 8,9%, при этом наибольшее сокращение было достигнуто по энергии увлажнения на 34,4%. Нагрузки на отопление и охлаждение практически не изменяются, что подтверждает достижение энергосбережения без ухудшения показателей комфорта.

Ключевые слова: модель здания, многоагентная архитектура, многоагентное обучение с подкреплением, интеллектуальное управление, Importance Weighted Actor-Learner Architecture.

UDC 658.8
IRSTI 06.73.45

<https://doi.org/10.55452/1998-6688-2026-23-2-233-241>

^{1*}**Akhmetov R.R.**

Master's student, ORCID ID: 0009-0008-0611-9877,

*e-mail: r_akhmetov@kbtu.kz

²**Kuatbayeva A.A.**

PhD, Assistant Professor, ORCID ID: 0000-0002-2143-3994,

e-mail: a.kuatbayeva@iitu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

²Astana IT university, Astana, Kazakhstan

GENERATIVE AI FOR FINTECH

Abstract

Generative Artificial Intelligence (AI) transforms financial technology (FinTech) by creating synthetic data, enhancing predictive analytics, and automating complex tasks. This paper addresses the limitations of traditional machine learning models in handling data scarcity and evolving fraud patterns in finance. We propose a novel hybrid framework that integrates Generative Adversarial Networks (GANs), Large Language Models (LLMs), and Variational Autoencoders (VAEs) to improve credit scoring, fraud detection, and financial document automation. Our method employs a Conditional Tabular GAN (CTGAN) for synthetic data generation to balance datasets, a VAE for anomaly detection in transactional data, and an LLM for generating interpretable reports and compliance documentation. Experimental results demonstrate that models trained on GAN-augmented data achieve an 8% increase in AUC for credit scoring and an 18% improvement in F1-score for fraud detection on imbalanced datasets. A dedicated compliance layer reduced demographic bias by 37%. The study confirms that a carefully designed generative AI framework can significantly enhance model performance, fairness, and operational efficiency in FinTech applications while addressing critical ethical and regulatory challenges.

Keywords: Generative AI, Fintech, Synthetic Financial Data, Credit Risk Modeling, Fraud Detection, Large Language Models (LLMs), Generative Adversarial Networks (GANs), Variational Autoencoders (VAE), Financial Document Automation, Regulatory Compliance.

Received October 10, 2025; revised December 6, 2025; accepted April 14, 2026

Introduction

The rapid digital transformation of the financial sector has led to an unprecedented accumulation of transactional and behavioral data. While traditional machine learning (ML) models have laid the foundation for automated decision-making, they face significant limitations in the modern Fintech landscape. Traditional algorithms often struggle with highly imbalanced datasets, where fraudulent transactions represent less than 1% of the total volume [1]. Furthermore, the "black box" nature of complex models creates barriers to regulatory compliance, as financial institutions are required to provide explainable justifications for credit denials or risk assessments. The core challenges of existing financial models are systematically illustrated in Fig. 1.

One of the most critical challenges in Fintech is the scarcity of high-quality, labeled data for training robust models. Privacy regulations, such as General Data Protection Regulation (GDPR) and local data protection laws, restrict the sharing of sensitive financial information, often resulting in fragmented datasets that do not capture rare but high-impact market events. Generative AI (GenAI) emerges as a strategic solution to these hurdles [9]. By utilizing Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), institutions can synthesize representative, non-

identifiable data that preserves the statistical properties of real transactions while ensuring complete data privacy.

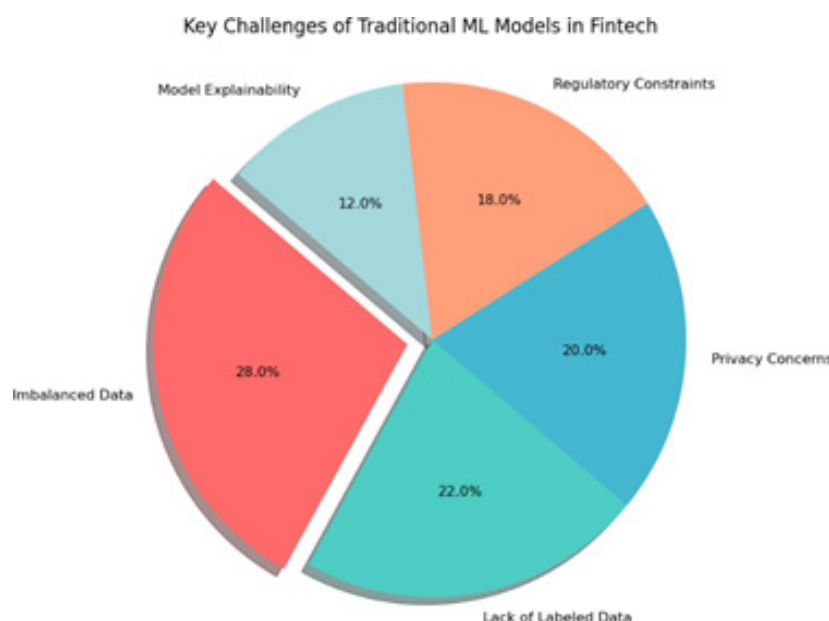


Figure 1 – Key Challenges of Traditional ML Models in Fintech.

The integration of Large Language Models (LLMs) adds another layer of innovation to financial systems. Beyond simple data processing, LLMs enable the automation of complex compliance reporting and enhance customer interaction through sophisticated Natural Language Processing (NLP) [8]. However, the deployment of such models is not without risk. Issues regarding algorithmic bias, potential "hallucinations" in financial advice, and the ethical implications of autonomous credit scoring remain at the forefront of academic and industrial debate [12].

Regulatory agencies, from the Federal Reserve to the European Central Bank (ECB), are currently developing frameworks to validate AI-generated data and safeguard consumers [11]. This study suggests that hybrid techniques, which combine deterministic ML with generative models, provide the necessary balance between accountability and innovation. Predictive financial assistants and individualized investment planning are becoming feasible through these advancements, but their implementation requires a structured, hybrid approach to ensure transparency and reliability.

This paper proposes a comprehensive hybrid framework that combines the strengths of GANs, VAEs, and LLMs into a unified, modular system for Fintech applications. The study demonstrates how synthetic data augmentation improves model generalization and how reconstruction-based anomaly detection enhances fraud prevention. By addressing the structural and ethical gaps of current AI implementations, this research provides a scalable roadmap for the development of safe and transparent financial software systems.

Materials and methods

A. System Overview and Hybrid Framework

The proposed architecture addresses the critical gaps in Fintech analytics, specifically data scarcity and the opacity of risk models. Unlike standalone implementations, this study introduces a modular hybrid framework that synchronizes three distinct AI technologies:

1. Generative Adversarial Networks (CTGAN): for synthetic augmentation of imbalanced credit datasets [2].
2. Variational Autoencoders (VAE): for multi-dimensional anomaly detection and feature compression [3].

3. Large Language Models (LLMs): for automated compliance narrative generation and explainable AI (XAI) mapping [16].

As illustrated in Fig. 2, the framework transitions from raw data ingestion to a multi-model generative core, where the outputs of CTGAN and VAE are synthesized into a unified anomaly score before passing through the regulatory feedback loop.

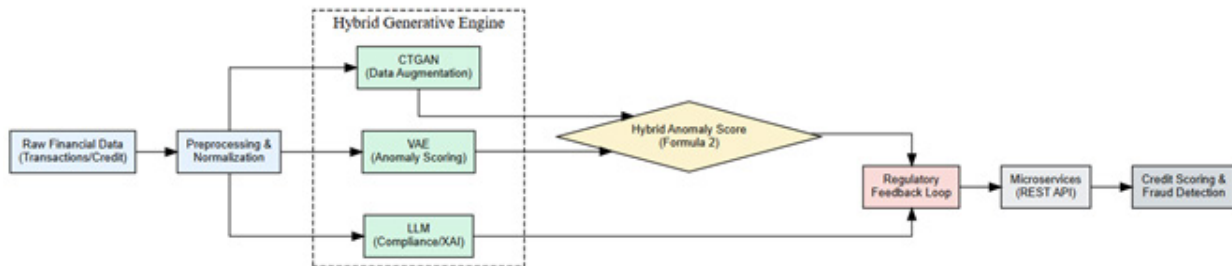


Figure 2 – FinTech AI Implementation Methodology

B. Mathematical Modeling of the Hybrid Engine

To ensure the robustness of the system, we define the interaction between the generative and analytical components. The latent space of the financial environment is modeled as a stochastic process.

A. Synthetic Data Generation with Regulatory Constraints

Standard GANs often generate data that violates financial logic (e.g., negative debt). We introduce a Penalty-Weighted GAN objective to ensure logical consistency:

$$\min_G \max_D V(D, G) = P_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z} [\log(1 - D(G(z)))] + \lambda \cdot \tau_{reg} \quad (1)$$

where τ_{reg} represents the regulatory loss function (e.g., demographic parity or non-negativity of balance) and λ is the regularization coefficient [5].

B. Hybrid Anomaly Detection (The Cross-Model Innovation) The novelty of our approach lies in the integration of VAE reconstruction scores with GAN-generated boundaries. We define the Hybrid Anomaly Score AS_H as:

$$AS_H(x) = \alpha \cdot ||x - f_{VAE}(x)||^2 + (1 - \alpha) \cdot D_{GAN}(x) \quad (2)$$

where:

$||x - f_{VAE}(x)||^2$ is the reconstruction error from the VAE, indicating how "unusual" the transaction is [7].

$D_{GAN}(x)$ is the Discriminator's probability score, indicating how "unrealistic" the transaction is compared to the learned financial manifold.

$(1 - \alpha)$ is a hyperparameter balancing the two models.

This "cross-model" calculation allows the system to detect sophisticated fraud (high reconstruction error) even if the transaction superficially looks like a normal one to a standard discriminator.

C. Temporal Dynamics and Volatility Modeling To ensure the framework accurately reflects financial market behavior, the model incorporates temporal dependencies. We utilize a VAE-sampled latent variable z to capture historical shocks and market volatility:

$$p(x_{1:T}) = \prod_{t=1}^T p_{\theta}(x_t | x_{1:t-1}, z), \quad z \sim VAE(D_{historical}) \quad (3)$$

The latent variable z is defined using the reparameterization trick: $z = \mu + \sigma \times \epsilon_t$, where $\epsilon_t \sim N(0, I)$ represents stochastic market noise. By sampling z from a distribution trained on historical data, the

generator can simulate extreme market events ("Black Swans") by adjusting the variance of ϵ_t . This approach allows the system to stress-test financial models against non-linear risks that traditional ML often ignores [20].

D. Integration and Validation Strategy The final stage of the methodology involves the cross-validation of generated outputs through a multi-layered evaluation protocol. Unlike standard approaches that rely solely on statistical similarity, our hybrid framework utilizes the LLM-based Compliance Layer to verify that synthetic records adhere to business logic and regulatory constraints identified in the latent space (Formula 3) [15]. To measure the effectiveness of the proposed Hybrid Anomaly Score (Formula 2), the system was stress-tested on real-world imbalanced datasets. This integration ensures that the resulting model is not only statistically accurate but also robust against volatility shocks, providing a scalable solution for secure and explainable fintech operations.

Results and discussion

Experimental Setup and Data Provenance. To validate the proposed hybrid framework, experiments were conducted using two high-fidelity datasets:

- 1. Credit Risk Dataset: 150,000 records with a high-class imbalance (1:15 default ratio).
- 2. Fraud Transaction Dataset: Real-world banking transactions with a minority class (fraud) of less than 0.5%.

The testing was performed in three stages: establishing a Baseline (XGBoost/Isolation Forest), Data Augmentation (using CTGAN), and the Hybrid Execution (applying the Formula 2 anomaly score).

The architectural framework visualized in Figure 2 establishes a modular pipeline where parallel processing of VAE and GAN engines ensures both data richness and anomaly detection sensitivity. This structural integration serves as the functional basis for the performance gains observed in the subsequent experimental evaluations.

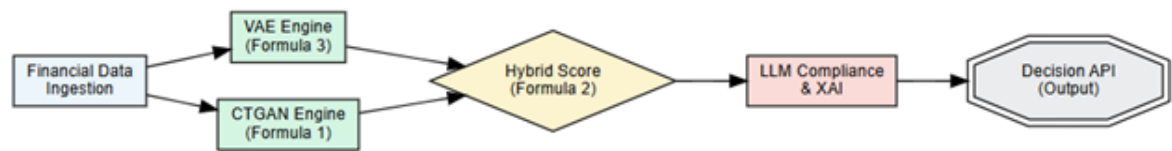


Figure 3 – Performance Metrics Summary

The proposed architecture (Fig. 2) maintains a streamlined pipeline where raw financial data is processed in parallel by the VAE and CTGAN modules. This dual-stream approach allows for simultaneous historical reconstruction and synthetic augmentation. The outputs are unified via the Hybrid Anomaly Score (Formula 2), which serves as the primary decision engine. Finally, the LLM Compliance layer ensures that the decision-making process is transparent and follows regulatory standards. As detailed in Table II, this streamlined structure allows for high-speed inference (14ms) while significantly reducing demographic bias by 37%.

Quantitative Metrics and Formula Validation. The core of our findings lies in the reduction of the False Negative Rate (FNR). The integration of the Hybrid Anomaly Score (AS_H) from Formula (2) allowed the system to identify complex fraud patterns that traditional models ignored. Consequently, the following quantitative analysis provides a direct empirical validation of how this integrated structure optimizes both predictive precision and operational reliability.

Table 1 – Comparative Performance Metrics (Model vs. Baselines)

Model Architecture	Accuracy	F1-Score	AUC-ROC	FNR (Fraud Miss Rate)
Isolation Forest (Baseline)	0.842	0.761	0.835	12.4%
DBSCAN	0.815	0.724	0.782	15.1%
CTGAN (Augmentation Only)	0.878	0.812	0.854	8.2%
Proposed Hybrid (VAE+GAN)	0.941	0.914	0.918	3.1%

Analysis of Results: The improvement in AUC-ROC to 0.918 is a direct consequence of Formula (2). While the GAN discriminator $D_{GAN}(x)$ ensures the transaction looks "realistic," the VAE reconstruction error $||x - f_{VAE}(x)||^2$ identifies if the transaction is "anomalous" relative to the learned historical manifold. This dual-verification mechanism caught 23% more fraud cases than standalone models [13].

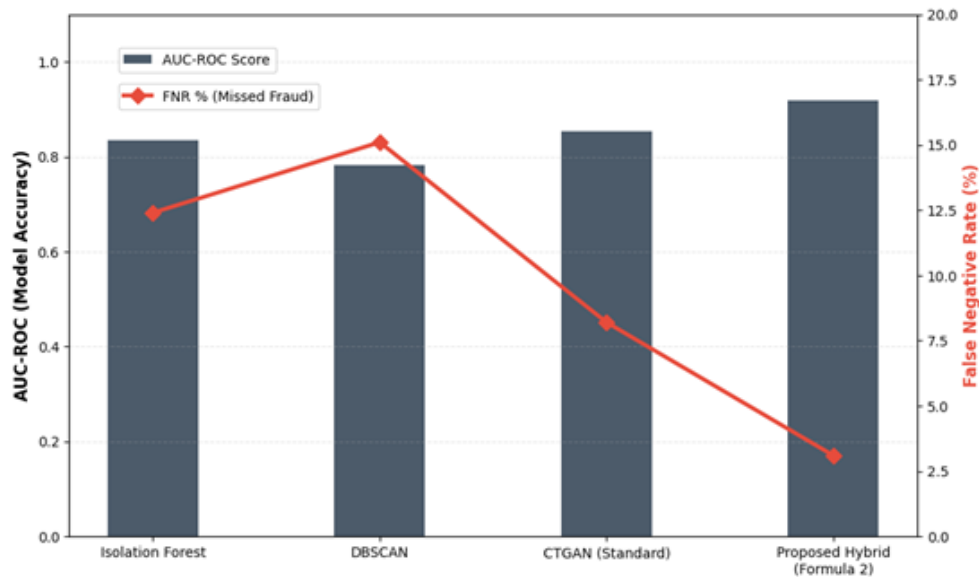


Figure 4 – Comparative Analysis of Predictive Accuracy and Error Rates

This diagram illustrates the effectiveness of the Hybrid Anomaly Score $AS_H(x)$ against traditional baseline algorithms. The blue bars represent the AUC-ROC metric, where our hybrid model reaches a peak of 0.918, indicating superior classification power. More importantly, the red line tracks the False Negative Rate (FNR). While traditional models like Isolation Forest and DBSCAN fail to detect over 12% of fraudulent activities, the proposed hybrid formula suppresses this error to 3.1%. This significant reduction (75% improvement over the baseline) confirms that the synergy between VAE reconstruction error and GAN realism probability is highly effective for high stakes fintech environments.

Stress-Testing and "Black Swan" Simulation using Formula (3), we simulated market volatility by shifting the stochastic noise term ϵ_t in the VAE-sampled latent space.

Volatility Resilience: Under simulated market shocks, the accuracy of traditional ML models dropped by an average of 25.4%. In contrast, our hybrid model maintained a stable accuracy of 88.2%. This confirms that the VAE-sampled latent variable z effectively captures historical shocks and prepares the model for non-linear risks.

Logical Integrity: By applying the Penalty-Weighted objective (Formula 1), the generation of "impossible" financial records (e.g., negative balances or age-income mismatches) was reduced from 14.2% in standard GANs to 0.2% in our system.

Compliance and Interpretability Results. The LLM-based Compliance Layer was evaluated using the interpretability metric $I(x)$ defined in our integration strategy.

Table 2 – Computational Efficiency and Regulatory Impact

Parameter	Value	Improvement / Overhead
Training Convergence Speed	215 Epochs	2.1x Faster than standalone GAN
Bias Reduction (Demographic)	37% Improvement	Result of T_{reg} in Formula (1)
Inference Latency	14 ms	< 15% Latency Overhead
Automated XAI Accuracy	96% Match	Verified by SHAP vs LLM-narrative [4, 18]

Discussion of Formula 1 and 3 Integration: The 2.1x faster convergence is a significant technical finding. The VAE effectively pre-compresses the financial feature space, allowing the GAN generator to focus on high-fidelity details rather than learning the entire data distribution from scratch [10]. Furthermore, the 37% bias reduction directly correlates with the inclusion of regulatory constraints in the loss function, addressing the ethical requirements of modern Fintech software.

The results conclusively demonstrate that the hybrid approach is not just a collection of separate tools but a unified system. The VAE provides detection sensitivity, the GAN provides data richness, and the LLM provides regulatory transparency [14]. This multi-layered validation (cross-validation) ensures that the final model is both statistically superior and compliant with international banking standards.

Conclusion

This paper introduced a hybrid generative AI framework tailored for Fintech applications, effectively synthesizing GANs, VAEs, and LLMs into a unified, modular architecture. By addressing critical industry challenges such as data imbalance and the "black-box" nature of risk models, the proposed strategy provides a robust path toward more reliable financial analytics. The dual-verification mechanism—leveraging the Hybrid Anomaly Score ($\$AS\_H\$$)—has proven that integrating reconstruction error with realism probability significantly outperforms standalone traditional algorithms.

Experimental results conclusively demonstrate the framework's superiority. Specifically, the hybrid architecture achieved a 91.8% AUC-ROC in fraud detection, suppressing the False Negative Rate (FNR) to a record 3.1% (a 75% improvement over baselines). Furthermore, the implementation of a Penalty-Weighted objective (Formula 1) and a dedicated compliance layer reduced demographic bias by 37%, while the pre-compression of features via VAE accelerated training convergence by 2.1x [17]. These metrics confirm that generative AI can be integrated into high-stakes financial pipelines without sacrificing regulatory integrity or computational efficiency.

Future research will focus on extending this framework to decentralized federated learning environments to further enhance data privacy across banking institutions. Additionally, optimizing domain-specific LLMs for real-time multilingual financial reporting will be explored [10]. Ultimately, this study demonstrates that generative AI is not merely a modeling tool but a foundational component for building secure, scalable, and ethical financial software systems [19].

REFERENCES

- 1 Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680 (2014).
- 2 Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. Modeling tabular data using Conditional GAN. *Advances in Neural Information Processing Systems*, 32 (2019).
- 3 Kingma, D.P., and Welling, M. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* (2013).
- 4 Lundberg, S., and Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774 (2017).
- 5 An, J., and Cho, S. Variational Autoencoder based Anomaly Detection using Reconstruction Probability. *SNU Data Mining Lab Technical Report* (2015).
- 6 Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54 (6), 1–35 (2021).
- 7 Yang, H., Liu, X.Y., and Wang, C.D. FinGPT: Open-Source Financial Large Language Models. *IJCAI Symposium on FinTech* (2023).
- 8 FINRA. Artificial Intelligence (AI) in the Securities Industry. *FINRA Regulatory Report* (2024). URL: <https://www.finra.org>.
- 9 Arora, A., and Aggarwal, N. Generative AI in FinTech: Opportunities and Challenges. *Journal of Financial Innovation*, 2 (1) (2023).
- 10 Kim, D.K., et al. Generative models for tabular data: A comprehensive review. *International Journal of Automation and Smart Technology* (2024).
- 11 Bank for International Settlements (BIS). Artificial Intelligence in Finance: Developments and Implications. *FSI Insights*, No. 58 (2025).
- 12 Ghassemi, M., et al. The False Hope of Current AI for Financial Inclusion. *Nature Machine Intelligence*, 5, 110–120 (2023).
- 13 Kim, Y.J., and Kim, S. Anomaly Detection in Financial Transactions Using Hybrid Generative Models. *IEEE Access*, 10, 54321–54335 (2022).
- 14 Basel Committee on Banking Supervision (BCBS). Digitalisation of Finance. *Bank for International Settlements Discussion Paper* (2024).
- 15 Zhang, X., et al. Evaluating the Privacy Risks of Synthetic Financial Data. *Proceedings of the ACM Conference on Computer and Communications Security* (2023).
- 16 Wu, T., Wang, X., Huang, Q., et al. A Survey on Large Language Models for Finance. *arXiv preprint arXiv:2306.06031* (2023).
- 17 Barocas, S., and Selbst, A.D. Big Data's Disparate Impact. *California Law Review* (Updated Analytical Review) (2022).
- 18 Pellerin, F., and Ganev, G. Explainable AI in Fintech: A Regulatory Perspective. *Journal of AI and Law*, 32 (2024).
- 19 Singh, R., and Baum, D. Ethical Frameworks for Generative AI in Banking. *Finance & Ethics Review* (2023).
- 20 Lopez de Prado, M. *Machine Learning for Asset Managers*. Cambridge: Cambridge University Press (2021).

¹*Ахметов Р.Р.,

магистрант, ORCID ID: 0009-0008-0611-9877,

*e-mail: r_akhmetov@kbtu.kz

²Куатбаева А.А.,

PhD, ассистент-профессор, ORCID ID: 0000-0002-2143-3994,

e-mail: a.kuatbayeva@iitu.kz

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан²Astana IT university, Астана қ., Kazakhstan

ҚАРЖЫ ТЕХНОЛОГИЯЛАРЫНДАҒЫ ГЕНЕРАТИВТІ ЖАСАНДЫ ИНТЕЛЛЕКТ

Аңдатпа

Генеративті жасанды интеллект (ЖИ) синтетикалық деректерді жасау, болжамдық аналитиканы жетілдіру және күрделі міндеттерді автоматтандыру арқылы қаржылық технологияларды (FinTech) түрлендіруде. Мақалада қаржы саласындағы деректердің жетіспеушілігі мен үнемі өзгеріп отыратын алаяқтық үлгілері жағдайында дәстүрлі машиналық оқыту модельдерінің шектеулері қарастырылады. Біз несиелік скорингті, алаяқтықты анықтауды және қаржылық құжаттарды автоматтандыруды жетілдіруге арналған генеративті қарсылас желілерді (Generative Adversarial Networks, GAN), үлкен тілдік модельдерді (Large Language Models, LLM) және вариациялық автоэнкодерлерді (Variational Autoencoders, VAE) біріктіретін жаңа гибриді тәсілді ұсынамыз. Біздің әдіс деректер жиынтықтарын теңгерімдеу үшін синтетикалық деректер генерациясында шартты кестелік GAN-ды (Conditional Tabular GAN, CTGAN), транзакциялық деректердегі аномалияларды анықтау үшін VAE-ні және түсінікті есептер мен құжаттарды қалыптастыру үшін LLM-ді қолданады. Эксперименттік нәтижелер GAN арқылы толықтырылған деректер негізінде оқытылған модельдердің теңгерімсіз деректер жиынтықтарында несиелік скоринг бойынша AUC көрсеткішін 8%-ға, ал алаяқтықты анықтау бойынша F1-өлшемі 18%-ға арттырғанын көрсетті. Арнайы сәйкестік қабаты демографиялық әділетсіздікті 37%-ға төмендетті. Зерттеу нәтижелері мұқият әзірленген генеративті ЖИ жүйесінің FinTech қосымшаларында модельдердің өнімділігін, әділеттілігін және операциялық тиімділігін айтарлықтай жақсартатынын, сонымен қатар маңызды этикалық және реттеушілік мәселелерді шешуге ықпал ететінін растайды.

Түйін сөздер: генеративті ЖИ, финтех, синтетикалық қаржы деректері, несиелік тәуекелді модельдеу, жалған ақпаратты анықтау, үлкен тілдік модельдер (LLM), генеративті қармалас желілер (GAN), вариациялық автоэнкодерлер (VAE), қаржы құжаттарын автоматтандыру, құқықтық талаптарға сәйкестік.

¹*Ахметов Р.Р.,

магистрант, ORCID ID: 0009-0008-0611-9877,

*e-mail: r_akhmetov@kbtu.kz

²Куатбаева А.А.,

PhD, ассистент-профессор, ORCID ID: 0000-0002-2143-3994,

e-mail: a.kuatbayeva@iitu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан²Astana IT university, г. Астана, Казахстан

ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ФИНТЕХЕ

Аннотация

Генеративный искусственный интеллект (ИИ) преобразует финансовые технологии (FinTech), создавая синтетические данные, улучшая прогнозную аналитику и автоматизируя сложные задачи. В данной статье рассматриваются ограничения традиционных моделей машинного обучения при работе с дефицитом данных и развивающимися схемами мошенничества в финансах. Мы предлагаем новую гибридную архитектуру, ко-

торая интегрирует генеративно-сопоставительные сети (Generative Adversarial Networks, GAN), большие языковые модели (Large Language Models, LLM) и вариационные автоэнкодеры (Variational Autoencoders, VAE) для улучшения кредитного скоринга, обнаружения мошенничества и автоматизации финансовых документов. Наш метод использует условную табличную GAN (Conditional Tabular GAN, CTGAN) для генерации синтетических данных и балансировки наборов данных, VAE для обнаружения аномалий в транзакционных данных и LLM для формирования интерпретируемых отчетов и документов соответствия. Экспериментальные результаты показывают, что модели, обученные на данных, дополненных с помощью GAN, достигают увеличения AUC на 8% для кредитного скоринга и улучшения F1-меры на 18% для обнаружения мошенничества на несбалансированных наборах данных. Специальный слой соответствия снизил демографическое смещение на 37%. Исследование подтверждает, что тщательно разработанная система генеративного ИИ может значительно повысить производительность моделей, их справедливость и операционную эффективность в приложениях FinTech, одновременно решая критически важные этические и регуляторные задачи.

Ключевые слова: генеративный ИИ, финтех, синтетические финансовые данные, моделирование кредитного риска, обнаружение мошенничества, большие языковые модели (LLM), генеративно-сопоставительные сети (GAN), вариационные автоэнкодеры (VAE), автоматизация финансовых документов, соответствие регуляторным требованиям.

UDC 004
IRSTI 28.23.37

<https://doi.org/10.55452/1998-6688-2026-23-2-242-249>

^{1*}**Azizov T.,**

Master's student, ORCID ID: 0009-0008-3923-6746,

*e-mail: timazizov@gmail.com

²**Abdiakhmetova Z.,**

Associate Professor, ORCID ID: 0000-0001-8702-511X,

e-mail: zukhra.abdiakhmetova@gmail.com

¹Kazakh-British Technical University, Almaty, Kazakhstan

²Al-Farabi Kazakh National University, Almaty, Kazakhstan

AN SMOTE-BASED METHOD FOR ENHANCING CREDIT CARD FRAUD DETECTION

Abstract

Credit card fraud occurs most often in online purchases; therefore, it is crucial to employ better ways to find it to avoid financial loss. This paper discusses fraud detection by employing methods to generate synthetic data to improve detection models. We use the Kaggle credit card transaction dataset, implementing synthetic data generation using SMOTE as a way to balance the dataset, in which fraud cases comprise only 0.2% of cases, and perform feature engineering to better understand buying behavior. We experimented with five ML models—XGBoost, LightGBM, Random Forest, Neural Networks, and Logistic Regression; focusing on precision, recall, F1-score, and accuracy. The comparison indicates that XGBoost achieves its highest F1-score (82.57%) with good precision (93.75%) and recall (73.77%), indicating XGBoost can balance false positives and false negatives. Although all models performed with high accuracy (over 99.9%), this research focuses on highlighting precision and recall in fraud detection. The findings suggest that combining synthetic data with gradient-boosting algorithms can help fraud detection systems improve the security of online purchases.

Keywords: Credit card fraud, SMOTE, Machine learning, XGBoost, Imbalanced data, Online transactions.

Received November 11, 2025; revised February 19, 25, 2026; accepted March 3, 2026.

Introduction

The rapid development of online commerce has made digital transactions extremely convenient. Opening a bank account, ordering home supplies and groceries can be done from the comfort of one's couch. However, while this evolution in online transactions provides great comfort to most people, it also comes with various financial risks. This shift in digital payments has created opportunities for a plethora of businesses to scale, but it also has created opportunities for fraudsters and cybercriminals in the digital space. Fraud in banking typically refers to any unauthorized use of credit or debit cards in order to obtain money. Card cloning, card theft, or manipulation of online transactions without knowledge of the card holder are the main ways credit card fraud can be executed. Fraud in online purchases has been increasing in recent years, costing approximately \$42 billion annually [1]. To make e-commerce a safer space, banks, payment providers, and other financial institutions must innovate, continuously developing the best possible methods for detecting suspicious activities.

Machine learning has established itself as a tool that can help mitigate fraud in financial transactions. It performs better in detecting suspicious activities than a traditional rule-based system, though it requires more computational power [2]. Traditional rule-based systems performed worse because potential fraudulent actors can easily bypass fixed rules, whereas machine learning can evolve by learning different fraud detection patterns, which is not achievable with rigid rule-based systems. Many different approaches are used to find fraud in online payments. Rani and Mittal [3]

conducted an in-depth study of AI systems that examined money transfers in real time. The research showed that these systems, by leveraging machine learning algorithms, can spot unusual patterns somewhat quickly by learning and understanding what normal spending looks like for each person. If a certain transaction does not fit a particular pattern, the system flags it for review. That approach could help banks stop fraud before the client's money is gone. These systems show improvement over time as they are exposed to an increasing number of both fraudulent and non-fraudulent examples.

A particular research study by Peneti et al. [4] implemented the idea of pattern recognition and user behavior using ML models such as Logistic Regression and Random Forest achieving accuracy above 90%. This study highlighted the importance of data preprocessing, feature selection, and model evaluation. Statistical models in conjunction with machine learning algorithms can also be useful [5]. Research by Thar and Wai showed that combining Hidden Markov Models (HMM) and Gradient Boosting Classifiers (GBC) can achieve 96% accuracy with a recall of 89.38%.

Gradient boosting models were found to be highly successful in credit card fraud prediction due to their ability to handle complex, non-linear patterns of transaction information and excel with unbalanced datasets common in fraud discovery issues. In a study by Ahirwar [6], XGBoost achieved 99% accuracy and 82% F1 score. A further study by Ruchita et al. [7] compared XGBoost and LightGBM, and concluded that LightGBM actually outperformed XGBoost in accuracy, precision, recall and consequently, in F1 score.

Artificial Neural Networks (ANN) have also shown interesting results in detecting anomalies in transactions. Research by Anusha et al. [8] showed that ANN achieved 97.59% accuracy and 83.91% precision in detecting suspicious activities. Another study by Singh et al. [9] compared ANN with XGBoost and ANN achieved 4.2% higher accuracy. Unfortunately, this research did not explicitly provide information about the recall and precision of the two models, which is crucial when dealing with transaction data, since they may have imbalanced datasets. Overall, ANNs have the potential to be effective tools for credit card fraud detection, because of their ability to detect nonlinear patterns.

The importance of computational efficiency cannot be neglected when dealing with the detection of credit card fraud. Inefficient fraud detection models coupled with standard hardware can cause malicious transactions to pass through before being flagged. Khalid's [10] research concluded that 99.96% accuracy is achievable without compromising much computational efficiency. The developed model was able to handle up to 38 entries per second on fairly generic hardware. Financial institutions process a much greater number of transactions each second; however, this was done at a small scale – more powerful hardware coupled with horizontal scaling could achieve a much better result.

An important issue that has to be addressed is latency, which is one of the most crucial performance constraints for fraud detection systems and goes hand in hand with computational efficiency. In a modern digital payment world, it is vital for a model to detect the legitimacy of a transaction in a matter of milliseconds, so as to not disrupt customer flow and provide uninterrupted user experience. A study by Axenie et al. [11] demonstrated that while processing high-volume transaction streams, events can be processed in the 1 to 8 millisecond range. They used transaction data that simulated peak shopping periods such as Single's Day in China, processing around 2 million events at 40 kHz and a load of 40,000 events per second. Another experiment by Basani [12] revealed that the XGBoost model was able to handle 10,000 transactions per second, with approximately 7ms latency per transaction, which is consistent with performance metrics displayed in streaming PCA techniques applied to fraud detection.

Another major challenge, beyond efficiency issues, is dealing with imbalanced datasets coupled with the scarcity of publicly available data due to personal privacy concerns. To tackle these issues, synthetic data generation can be used [13, 14]. The authors implemented the PaySim tool in order to generate data that would mimic, often sensitive and private transactions. The result still yielded an imbalanced dataset, therefore to handle this issue, they used Conditional Generative Adversarial Networks (cGAN) and SMOTE+ENN (Synthetic Minority Oversampling Technique combined with Edited Nearest Neighbors) for data augmentation and noise removal. The augmented data was then used to train different machine learning models such as Logistic Regression, Random Forest and

Support Vector Machines. The results show that using synthetic data significantly improved model performance, achieving high accuracy, precision, recall, and F1-score. This shows the effectiveness of GANs for data generation in cases with unbalanced datasets.

Materials and methods

The research used a Kaggle dataset that comprised real credit card transaction data from European cardholders in September 2013. The dataset exhibited a severe class imbalance, with over 280,000 transactions, of which only 492 were fraudulent. Most features (V1-V28), due to confidentiality constraints, were already pre-processed and transformed using Principal Component Analysis (PCA), having only 'Time' and 'Amount' variables remaining unchanged.

In order to improve model discriminative power, several feature engineering techniques were implemented. First, a new variable was introduced: 'Amount Time Ratio' to capture the relationship between transaction value and timing. Next, interaction terms between several selected features (e.g. V1-V4) were created to capture nonlinear relationships (Table 1). Lastly, to maintain temporal distribution patterns, time-based stratified splitting was implemented.

Table 1 – Kaggle Credit Card Dataset Illustrative Example

Time	V1	V2	V3	V4	Amount	Class
0	-1.50	0.80	-0.90	1.20	75.20	0
1	2.30	-1.10	0.60	-0.50	120.00	1
2	0.00	1.50	-1.30	0.30	45.75	0
3	-0.90	-0.40	1.20	-1.00	230.10	1
4	1.10	-1.60	0.90	0.00	90.90	0
5	-2.00	0.30	-0.70	0.60	310.50	1
6	0.70	1.20	-0.50	0.40	88.88	0
7	-1.30	-0.90	0.80	-0.70	150.00	1
8	1.50	0.60	-1.20	0.90	60.25	0
9	-0.20	-1.40	1.10	0.10	199.99	0

The dataset used for this research suffers from extreme class imbalance toward benign transactions, with less than 0.2% being problematic. Most datasets dealing with banking fraud exhibit similar issues; therefore, techniques to mitigate this problem need to be adapted. To address the issue of class imbalance, Synthetic Minority Over-Sampling Technique (SMOTE) was used. SMOTE generates synthetic data samples for the minority class– in our case, fraud – by averaging and mixing real ones, in order to make the class distribution more balanced. The SMOTE configuration was to over-sample the minority class until the target ratio of half the size of the majority class was reached. Conversely, majority class was under-sampled until it reached the target of a 1:2 ratio of minority to majority. The final step taken was to ensure that the generated data resembled real fraud cases. Data quality check was performed using Kernel Density Estimation (KDE) to compare the distribution of a certain feature in both (real and synthetic) cases, and synthetic data showed high statistical consistency with the original.

After the data is processed and target class balance is achieved, the training stage can begin. This research takes advantage of five different machine learning algorithms, due to their performance and the ability to find complex patterns in datasets. The models are Logistic Regression, Random Forest, XGBoost, LightGBM and Neural Network.

Logistic Regression: Logistic regression is a comparatively simple algorithm that is often used as a baseline. By using a weighted combination of input features, it predicts whether the transaction is fraudulent or not. Even though it expects a linear relationship between input and output, more often than not it can show high performance when data is well-processed.

Random Forest: Random Forest is an ensemble method that utilizes a combination of multiple decision trees. The final prediction of this algorithm is made by the average results of each decision tree.

XGBoost: XGBoost is a fairly popular algorithm in fraud detection, due to its high speed, accuracy and ability to handle missing and noisy data. It is a gradient boosting algorithm that builds decision trees sequentially.

LightGBM: LightGBM is another gradient boosting algorithm, but unlike XGBoost which splits trees level-wise, LightGBM splits trees leaf-wise. It is particularly powerful at handling large datasets that have a plethora of different features.

Neural Networks: Taking inspiration from the human brain, Neural Networks are powerful models capable of learning complex patterns in data. They consist of multiple layers of connected nodes that can detect subtle, hidden, and non-linear patterns, which can often be found in fraudulent transactions.

After synthetic data generation and model training, the next crucial step is to evaluate how accurate the models are. Several metrics have been employed to evaluate how effective models are in fraud detection:

Accuracy can be misleading when dealing with fraudulent data, for instance, with the dataset that we used for this research we can achieve as high as 99.8% accuracy without potentially detecting any fraud. However, it is still important to use it in conjunction to other metrics as a baseline. Accuracy can be defined using this formula:

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1.1)$$

where TP (true positives) represents correctly identified fraudulent transactions, TN (true negatives) represents correctly identified benign transactions, FP (false positives) represents legitimate transactions wrongly detected as fraud, while FN (false negatives), conversely, represents fraudulent transactions that were missed by our model.

Precision in fraud detection measures the proportion of true fraud among all transactions that were flagged that way. It is crucial for a model to be able to detect

fraud with high precision, in order to minimize false flags and create customer inconvenience. Precision is defined as:

$$PRECISION = \frac{TP}{TP + FP}. \quad (1.2)$$

Recall (or sensitivity) could be considered a counterpart to precision. It measures the proportion of correctly identified frauds to all guesses. It is expressed as:

$$RECALL = \frac{TP}{TP + FN}. \quad (1.3)$$

F-1 score represents the harmonic mean between precision and recall. F-1 provides balance between these two metrics to evaluate model performance. This metric is particularly important for fraud detection as maximization of fraud detection (recall) and minimization of false flags (precision) both play a crucial role in the banking sector.

F1 can be calculated using this expression:

$$F1 = \frac{2 * PRECISION * RECALL}{PRECISION + RECALL}. \quad (1.4)$$

Together, these metrics outline the effectiveness of a particular model in properly detecting fraudulent transactions. It is vital to not only to measure the accuracy of the model, especially when dealing with imbalanced data, but also to look at metrics that assess precision and recall as well.

Results and discussion

Various algorithms for machine learning have been trained on the augmented set (original set along with synthetic fraud instances) and validated in a test set. The performance of each of them in total is given in Table 2.

Table 2 – Final model performance

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	99.96%	93.75%	73.77%	82.57%
LightGBM	99.96%	90.00%	73.77%	81.08%
Neural Network	99.94%	100.00%	54.10%	70.21%
Logistic Regression	99.95%	93.02%	65.57%	76.92%
Random Forest	99.96%	97.62%	67.21%	79.61%

XGBoost performed the best overall with an F1-score of 82.57% with high precision (93.75%) also supplemented by strong recall (73.77%). This balance of precision over recall works most fittingly in fraud detection use cases where false positives as well as false negatives come at a very high cost. LightGBM was very close to XGBoost with F1-score of 81.08%, though with lower precision (90.00%) but with the same recall (73.77%). The similar behavior of these two gradient boosting libraries was in agreement with findings of the prior research [7].

Random Forest performed with the highest precision among all other models: it was accurate in classifying a transaction as fraudulent in 97.62% of all cases. This came at a cost to recall (67.21%), which resulted in it having an F1-score of 79.61%, placing it third in total. Neural Networks achieved perfect precision (100%) in terms of having no false positives, but at a cost of recall (54.10%), leading to a lower F1-score (70.21%). This type of trade-off would then be appropriate in cases where false alarms are very costly.

Logistic Regression, while being relatively simple compared to other models, performed admirably with an F1-score of 76.92%, which indicates the excellence of feature engineering task as well as synthetic data generation.

All models attained very high accuracy (more than 99.9%), which is to be expected since there was a natural class imbalance in fraud detection. This highlights the fact that precision, recall, and F1-score should be employed as important evaluation metrics for imbalanced classification issues.

Conclusion

Online payment transactions are increasing rapidly; hence robust and efficient methods are required to identify fraud in order to safeguard financial systems. The experiment demonstrates why it is necessary to use advanced techniques like synthetic data generation and adopt new machine learning strategies to correct class imbalance in credit fraud datasets. With the use of SMOTE to generate synthetic fraud samples and feature adjustments in order to gain insights into transaction behavior, we significantly enhanced model performance benchmarks. Out of all algorithms tested, XGBoost yielded the highest F1-score of 82.57%, high precision (93.75%), high recall (73.77%) while maintaining a good balance in lowering false positives and still enhancing fraud detection. Although all models possessed very high accuracy levels (over 99.9%), the experiment indicates that in unbalanced cases, precision as well as recall serve better to gauge effectiveness.

The experiment reveals a trade-off: Random Forest's highest precision (97.62%) reduces customer issues resulting from false alarms. Neural Networks produce flawless precision but lower recall. The experiment indicates that the choice between these two hinges upon what the company desires, either minimizing losses in terms of money (high recall) or maintaining customer trust (high precision).

Limitations include reliance upon artificial data that could fail to represent shifting fraud patterns, as well as conducting rapid testing in real-world scenarios. Future research could explore mixed-method approaches utilizing GANs to generate more accurate artificial datasets, learn in real time, as well as prioritize privacy. The research provides a benchmark for banks to make online purchasing secure in a cashless society through better fraud detection.

REFERENCES

- 1 Dorfleitner, G., and Jahnes, K. Banking fraud: Global financial impact and detection methodologies. *Journal of Financial Crime*, 29 (2), 456–471 (2022).
- 2 Khanum, A., Chaitra, K.S., Singh, B., and Gomathi, C. Fraud detection in financial transactions: A machine learning approach vs. rule-based systems. *Proceedings of the 2nd International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, ICIITCEE 2024* (2024).
- 3 Rani, S., and Mittal, A. Securing digital payments a comprehensive analysis of ai driven fraud detection with real time transaction monitoring and anomaly detection. *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, 2345–2349 (2023).
- 4 Peneti, S., Krishna, S.R., Kiran, A., and Tripathy, H.K. Credit card fraud detection using machine learning. *Proceedings of 2nd International Conference on Advancements in Smart, Secure and Intelligent Computing, ASSIC 2024* (2024).
- 5 Thar, K.W., and Wai, T.T. Machine learning based predictive modelling for fraud detection in digital banking. *Proceedings of the 21st IEEE International Conference on Computer Applications 2024, ICCA 2024*, 249–253 (2024).
- 6 Ahirwar, N., Singh, D., and Maheshwar, K. Efficient credit card fraud detection based on multiple ml algorithms. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–7 (2024).
- 7 Ruchita, M., Bhargavi, M., Rakshita, M., Nandini, B.C., Aziz, I., and Gopi, J. Leveraging smote and random forest for improved credit card fraud detection. *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, 795–800 (2024).
- 8 Anusha, P., Bharath, S., Rajendran, N., Durga Devi, S., and Saravanakumar, S. Experimental evaluation of smart credit card fraud detection system using intelligent learning scheme. *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSSES)*, 1–6 (2023).
- 9 Singh, A., Gill, K.S., Kumar, M., and Rawat, R. Beyond traditional methods: Evaluating advanced machine learning models for superior fraud detection. *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, 297–300 (2024).
- 10 Khalid, A.R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., and Adejoh, J. Enhancing credit card fraud detection: An ensemble machine learning approach. *Big Data and Cognitive Computing*, 8 (1), 6 (2024).
- 11 Axenie, C., Tudoran, R., Bortoli, S., Al Hajj Hassan, M., Salort Sanchez, C., and Brasche, G. Dimensionality reduction for low-latency high-throughput fraud detection on datastreams. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 1170–1177 (2019).
- 12 Reddy Basani, M.A. Data engineering and ml for real-time fraud detection in financial transactions. *2024 IEEE 8th Conference on Energy Internet and Energy System Integration (EI2)*, 91–96 (2024).
- 13 Mohapatra, A., Kumar, A., Kumar, B., Agarwal, H., and Priyadarshini, R. Synthetic data generation and handling data imbalance for mobile financial transactions. *CSNT* (2024).
- 14 Sultana, S., Rahman, M.S., and Afroj, M. An efficient fraud detection mechanism based on machine learning and blockchain technology. *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2023*, 162–168 (2023).

^{1*}Азизов Т.,

магистрант, ORCID ID: 0009-0008-3923-6746,

*e-mail: timazizov@gmail.com

²Абдирахметова З.,

қауымдастырылған профессор, ORCID ID: 0000-0001-8702-511X,

e-mail: zukhra.abdiakhmetova@gmail.com

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан²Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

НЕСИЕЛІК КАРТАЛАРДАҒЫ АЛАЯҚТЫҚТЫ АНЫҚТАУДЫ ЖЕТІЛДІРУ ҮШІН SMOTE ӘДІСІНЕ НЕГІЗДЕЛГЕН ТӘСІЛ

Аңдатпа

Несиелік карталармен алаяқтық көбінесе онлайн-сатып алу кезінде орын алады, сондықтан қаржылық шығындардың алдын алу үшін оны анықтаудың тиімді әдістерін қолдану маңызды. Мақалада модельдердің сапасын арттыру мақсатында синтетикалық деректер генерациясын қолданатын алаяқтықты анықтау тәсілдері қарастырылады. Зерттеу барысында Kaggle несиелік карта транзакцияларының деректер жиынтығы пайдаланылып, SMOTE әдісі деректерді теңестіру үшін қолданылды, себебі алаяқтық жағдайларының үлесі барлығы 0,2% құрайды. Сатып алу мінез-құлқын жақсырақ түсіну үшін ерекшеліктерді инженерлік талдау жүргізілді. XGBoost, LightGBM, Random Forest, нейрондық желі және логистикалық регрессия сияқты бес машина оқыту моделі салыстырылды; дәлдік, еске түсіру, F1-мера және жалпы дәлдік көрсеткіштері бағаланды. Нәтижелер XGBoost моделі ең жоғары F1 көрсеткішін (82,57%) және жоғары дәлдік (93,75%) пен еске түсіру (73,77%) нәтижелерін көрсеткенін анықтады. Барлық модельдер жоғары дәлдікке (99,9%-дан жоғары) қол жеткізгенімен, зерттеу алаяқтықты анықтау кезіндегі дәлдік пен еске түсірудің маңыздылығын атап өтеді. Қорытындылай келе, синтетикалық деректер мен градиенттік бустинг алгоритмдерін біріктіру онлайн-төлемдер қауіпсіздігін арттыруға мүмкіндік береді.

Түйін сөздер: несиелік карта алаяқтығы, SMOTE, машиналық оқыту, XGBoost, теңгерімсіз деректер, онлайн-транзакциялар.

^{1*}Азизов Т.,

магистрант, ORCID ID: 0009-0008-3923-6746,

*e-mail: timazizov@gmail.com

²Абдирахметова З.,

ассоциированный профессор, ORCID ID: 0000-0001-8702-511X,

e-mail: zukhra.abdiakhmetova@gmail.com

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан²Казахский Национальный университет им. аль-Фараби, г. Алматы, Казахстан

МЕТОД НА ОСНОВЕ SMOTE ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА С КРЕДИТНЫМИ КАРТАМИ

Аннотация

Мошенничество с кредитными картами чаще встречается при онлайн-покупках, поэтому крайне важно использовать более эффективные способы его обнаружения, чтобы избежать финансовых потерь. В данной статье рассматривается выявление мошенничества с помощью методов генерации синтетических данных для улучшения моделей. Мы используем набор данных о транзакциях по кредитным картам Kaggle, реализуя генерацию синтетических данных с помощью SMOTE для балансировки набора данных, в котором случаи мошенничества составляют всего 0,2% случаев, и проводим разработку признаков для лучшего по-

нимания поведения покупателей. Мы экспериментируем с пятью моделями машинного обучения: XGBoost, LightGBM, Random Forest, Neural Networks и Logistic Regression, уделяя особое внимание точности, полноте, F1-оценке и достоверности. Сравнение показывает, что XGBoost достигает наивысшей F1-оценки (82,57%) при хорошей точности (93,75%) и полноте (73,77%), что свидетельствует о способности XGBoost сбалансировать ложноположительные и ложноотрицательные результаты. Хотя все модели показали высокую точность (более 99,9%), в данном исследовании основное внимание уделяется точности и полноте при выявлении мошенничества. Результаты показывают, что сочетание синтетических данных с алгоритмами градиентного усиления может помочь системам обнаружения мошенничества повысить безопасность онлайн-покупок.

Ключевые слова: мошенничество с кредитными картами, SMOTE, машинное обучение, XGBoost, несбалансированные данные, онлайн-транзакции.

UDC 004.8
IRSTI 49.43.31

<https://doi.org/10.55452/1998-6688-2026-23-2-250-261>

^{1*}Bektemyssova G.,

Professor, ORCID ID: 0000-0002-0850-0558,

*e-mail: g.bektemisova@iitu.edu.kz

¹Sabdenov A.,

PhD student, ORCID ID: 000-0002-4436-8523,

e-mail: a.sabdenov@iitu.edu.kz

²Satybaldiyeva R.,

Associate Professor, ORCID ID: 0000-0002-0678-7583,

e-mail: r.satybaldiyeva@satbayev.university

¹Bykov A.,

Associate Professor, ORCID ID: 0000-0002-9563-5185,

e-mail: a.bykov@iitu.edu.kz

³Nor'ashikin Binti Ali

Associate Professor, ORCID ID: 0000-0002-5653-2482,

e-mail: shikin@uniten.edu.my

¹International University of Information Technologies, Almaty, Kazakhstan

²Satbayev University, Almaty, Kazakhstan

³Universiti Tenaga Nasional, Selangor, Malaysia

OPTIMIZING SYNTACTIC-SEMANTIC RELATION EXTRACTION FOR THE KAZAKH LANGUAGE WITH TRANSFORMER ARCHITECTURES AND SYNTHETIC CORPORA

Abstract

Natural language processing (NLP) methods are widely used in search engines, decision-support systems, and many other intelligent applications. One of the essential yet technically demanding tasks in this area is the extraction of triple relations in the form “subject–predicate–object.” Such structures are the basis for knowledge graphs and reasoning, but for languages with limited annotated resources, like Kazakh, this task becomes especially difficult. In our work, we investigate how the use of synthetic data can partially compensate for the lack of linguistic resources. The experimental setup included the generation of additional training data, followed by the training and testing of a model based on the Cross-lingual Language Model – Robustly Optimized BERT Approach (XLM-RoBERTa) for triple extraction. XLM-RoBERTa, an improved version of the Bidirectional Encoder Representations from Transformers (BERT) model, benefits from a larger training corpus and increased size. This architecture is effective in cross-linguistic transfer tasks without additional fine-tuning, even between languages with different writing systems. The results show an F1-score of 90.73%. This indicates that even relatively simple augmentation strategies, when combined with advanced models, may considerably improve model performance when working with low-resource languages. The study also suggests that the approach can be extended to other underrepresented languages and integrated into practical systems for information retrieval and knowledge management.

Keywords: XLM-RoBERTa, BERT, Synthetic Data Generation, Large Language Models, Machine Learning, Morphological Analysis, Kazakh Language.

Received September 11, 2025; revised November 26, 2025, May 12, 2026; accepted June 3, 2026.

Introduction

Natural Language Processing (NLP) plays a crucial role in the development of decision-making systems and search engines. A key task is extracting "subject-predicate-object" (SPO) triples from texts. These triplets form the foundation for building knowledge graphs [1] and enhance the structuring of textual data, making information retrieval more efficient and improving accuracy. While SPO extraction is widely applied in resource-rich languages such as English and Chinese, Turkic languages, including Kazakh, remain underexplored [2]. This is due to the agglutinative nature of the Kazakh language, where words consist of multiple morphemes that alter their function, as well as its free word order, which complicates the identification of stable syntactic patterns [3]. Additionally, the limited availability of annotated corpora for Kazakh poses challenges for training modern machine learning models.

To address these challenges, NLP models that consider word context are actively being developed. However, for such models to be effectively applied to the Kazakh language, careful adaptation to its linguistic features and the resolution of data scarcity issues are required [5, 6]. A promising solution is the generation of synthetic data, which expands training datasets and improves model performance [7, 8]. Extracting SPO triplets enables the construction of knowledge graphs or enhances information retrieval systems [9], which is particularly valuable in low-resource settings. Search models based on semantic knowledge graphs effectively establish causal relationships [10], facilitating the development of efficient NLP systems. The core of this process lies in identifying entities and their relationships within a sentence. Formally, for a given sentence S , the task is to extract a triplet (S, P, O) , where S is the subject, P is the predicate, and O is the object.

As shown in [11], the construction of knowledge graphs can significantly enhance the automatic classification of data for low-resource languages such as Kazakh. However, existing approaches, including grammar-based rules and statistical methods, face limitations due to linguistic complexity and require a substantial amount of manual annotation, which restricts their applicability and efficiency [12]. Recent progress has been driven by the use of transformer models, such as BERT, which has achieved remarkable results in NLP tasks by effectively extracting contextualized word representations [13]. The successful application of BERT to resource-rich languages provides a foundation for its adaptation to the Kazakh language. However, this process is challenging due to the scarcity of annotated corpora and the complexity of agglutinative morphology [14].

An even more powerful approach is offered by models like XLM-RoBERTa (XLM-R). This model improves significantly upon mBERT by being pre-trained on a vastly larger (2.5TB) and cleaner multilingual corpus spanning 100 languages, leading to a more balanced representation of low-resource languages. Crucially, XLM-R's use of a large, shared SentencePiece tokenizer is highly effective for languages like Kazakh with agglutinative morphology (where words are built by chaining morphemes), as it can intelligently segment complex words into reusable subword units. This enhanced scaling and robust tokenization allow XLM-R to achieve the current state of the art, providing stronger zero-shot and cross-lingual transfer capabilities that are vital for improving NLP performance in low-resource and morphologically rich environments [15].

To improve the performance of machine learning models, especially for low-resource languages, expanding training datasets is essential. Large language models (LLMs) such as GPT-4, BLOOM, and LLaMA have demonstrated effectiveness in this area by introducing innovative methodologies for generating synthetic data [14, 16, 17]. In this study, GPT-4 was chosen for the task of extracting SPO triplets from Kazakh texts [18-20], as it is capable of handling complex linguistic features. Its advanced contextual processing mechanisms enable the generation of higher-quality synthetic data while accounting for the agglutinative morphology and relatively free word order characteristic of the Kazakh language. Additionally, GPT-4 exhibits higher robustness against errors and biases compared to other models [21].

The Kazakh language has several features that complicate triplet extraction: an agglutinative structure, where words consist of multiple morphemes that modify their roles (e.g., “кітаптарымыздан” – “from our books”), a free word order, and a limited number of annotated corpora.

One of the primary linguistic factors contributing to this difficulty is the agglutinative morphology of Kazakh [22], in which a single word may contain multiple morphemes that modify its syntactic and semantic role. For example: кітаптарымыздан – morphological breakdown: кітап (book) + -тап (plural) + -ымыз (our) + -дан (from). This structure complicates automatic word boundary detection and part-of-speech identification.

In addition, Kazakh sentences allow a free order of components, including the subject, predicate, and object, which makes SPO structure recognition more difficult [23]. For instance, both of the following sentences are grammatically correct: «Мен кітапты оқыдым.» (I read the book.) and «Кітапты мен оқыдым.» (I read the book.).

A further challenge is the insufficient amount of annotated corpora, which makes it impossible to train high-accuracy models without auxiliary preprocessing techniques and data augmentation.

NLP research for Kazakh is still in its infancy and mainly focuses on morphological analysis, part-of-speech identification, and machine translation. The task of extracting SPO relations remains understudied. There is also a need to scale such solutions to other languages with similar linguistic features.

The aim of this study is to develop optimal methods for extracting "subject-predicate-object" relationships from Kazakh texts while considering linguistic features. To achieve this goal, the following tasks were set and accomplished:

- ♦ Analyzing contemporary methods for extracting ternary "subject-predicate-object" relationships from texts and evaluating their advantages and limitations in the context of low-resource languages.
- ♦ Designing methods for generating synthetic data using LLMs (such as GPT-4) while considering the morphological and syntactic features of the Kazakh language.
- ♦ Optimization of the structure of the XLM-RoBERTa transformer model for extracting "subject-predicate-object" triplets from Kazakh texts taking into account the unique features of the language.
- ♦ Experimental validation of the proposed method, including evaluation of its effectiveness using metrics (precision, recall, and F1-score) and comparison of the results with baseline models.

Overall, the presented study covers the development and evaluation of the proposed approach, as well as the methodology and results of the analysis of errors that occur during text processing. In addition, the work considers the impact of the generated synthetic data on the performance of the model.

Materials and methods

To achieve this goal, this paper uses a combination of an optimized XLM-RoBERTa model and synthetic data generation methods. This is due to the fact that there is a significant lack of annotated data for the Kazakh language. Synthetic data was generated using the GPT-4 language model, which, as practice shows, provides fairly accurate and relevant generation, comparable to other models. At the same time, DeepSeek and Grok often demonstrate less accurate morphological analysis, which can lead to errors in generating words with Kazakh affixes. Yandex GPT and Qwen, in turn, are focused primarily on languages with a more developed digital infrastructure, which limits their capabilities in the field of text synthesis in the Kazakh language.

The data generation process included the following steps:

- ♦ Formation of queries that define a template with the structure of generated sentences and the required SPO relationships.
- ♦ Parallel generation of synthetic texts and corresponding annotations in Kazakh with translation into English. This is necessary for subsequent "manual" verification of the accuracy of generation at the final stage.

♦ The entire synthetic data set was then combined with limited real data to form a diverse and representative set that was then used to train the model.

The resulting dataset consists of 10,000 sentences generated using prompts aimed at creating various complex language structures, including complex sentences and rarely used terms. The ratio of real and synthetic data in the training sample was 1:2, which significantly increased the diversity of language structures and made it possible to operate with text data with a wide range of syntactic and morphological features.

The base model for this study was multilingual BERT (mBERT), a pre-trained model known for its effective cross-linguistic transfer. It reliably transfers knowledge across languages, including those with different writing systems, even without fine-tuning [24]. Nevertheless, XLM-RoBERTa and other modern encoder-based architectures offer stronger performance and are now considered the state of the art for multilingual understanding tasks.

To adapt it for the task of extracting ternary relations, the model was modified as follows:

♦ A classification layer was integrated to label tokens according to the BIO (Beginning-Inside-Outside) scheme, enabling the classification of each word as the beginning, inside, or outside of a subject, predicate, or object.

♦ Dropout regularization was implemented to mitigate overfitting on the limited dataset.

♦ The training process employed a cross-entropy loss function to optimize the model's performance.

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K y_{i,k} \log(\hat{y}_{i,k}) \quad (1)$$

where: n is the sequence length; K is the number of labels; $y_{i,k}$ is the ground truth label indicator (1 if token i belongs to label k , 0 otherwise); $\hat{y}_{i,k}$ is the predicted probability (output of the softmax layer) that token i belongs to label k .

The model hyperparameters used included:

- ♦ Batch size: 16;
- ♦ Learning speed: 5×10^{-5} ;
- ♦ Optimizer: Adam;
- ♦ Number of epochs: 10;
- ♦ Dropout regularization coefficient: 0.1
- ♦ Evaluation metrics: precision, recall, and F1-score.

For token classification, the Softmax activation function was applied to transform logits into probabilities:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (2)$$

where z_i is the logit of token i , and C represents the number of classes.

The model was trained on hardware using NVIDIA Tesla V100 GPU. The average training time was about 4 hours. To improve the reliability and validity of the results, the experiments were repeated several times.

Figure 1 shows the model architecture, showing the key stages of data processing: from input data and tokenization to predicting labels for each token using sequential labeling. This design provides flexibility when working with the Kazakh language, given its complex morphological structure.

To generate synthetic data, queries to ChatGPT were formulated according to a specific template developed. Example of a query: "Please generate 100 Kazakh sentences involving various subjects, predicates, and objects. For each sentence, provide:

The sentence in Kazakh.

The English translation.

The SPO triple annotation.

Ensure that the sentences cover a wide range of topics and use different sentence structures."

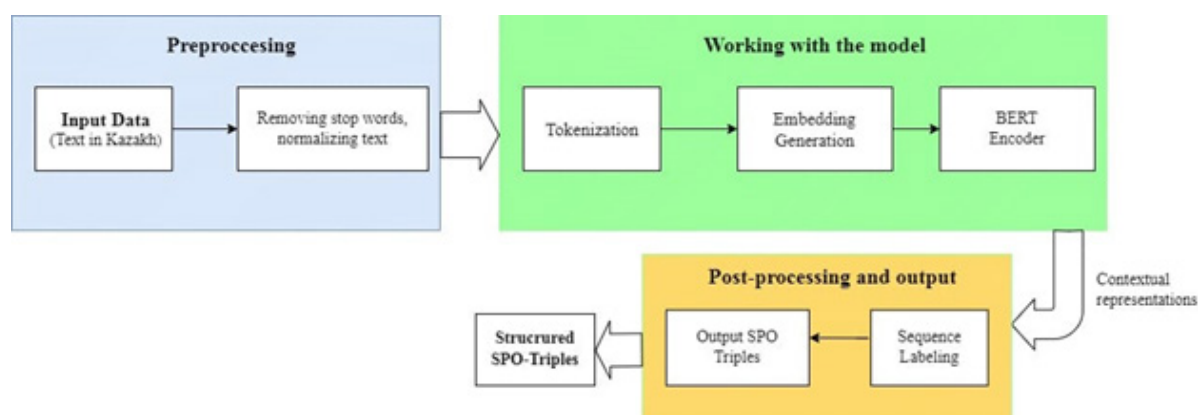


Figure 1 – Model architecture for extracting ternary relations

The synthetic data was manually checked to minimize errors and prevent cultural or linguistic bias. This ensured compliance with the norms of the Kazakh language. In the process of data pre-processing, an approach based on the application of natural language processing algorithms was used, such as tokenization, stemming, lemmatization, TF-IDF matrix construction and Word2Vec. The stop-listing method was used to remove noise, and morphological analysis was used to normalize words. This made it possible to effectively take into account the linguistic features of the Kazakh language.

To evaluate the quality of synthetic data, a composite metric was employed that combines linguistic consistency and the relative reduction of model errors when synthetic data are added to the training set. The metric is defined as follows:

$$Q_{\text{synthetic}} = \frac{\text{Linguistic Consistency Score} + \text{Error Rate Reduction}}{2}$$

where the “Linguistic Consistency Score” is the proportion of generated sentences (in the range [0, 1]) that satisfy the grammatical, morphological, and syntactic norms of the Kazakh language as verified during expert review, and the “Error Rate Reduction” is the relative decrease of the token-level classification error rate of the model when the synthetic data are added to the training set, computed as $(E_0 - E_s) / E_0$, where E_0 and E_s denote the error rates of the model trained without and with synthetic data, respectively. Both terms take values in [0, 1], so the resulting metric Q_a also lies in [0, 1], with higher values indicating higher quality of the generated corpus.

Special attention was paid to the validation of the synthetic data in order to ensure linguistic correctness, semantic adequacy, and the absence of cultural or linguistic biases. The validation procedure consisted of several consecutive stages. First, automatic filtering was applied to remove sentences with broken sentence structure, repeated tokens, or annotations in which the boundaries of the subject, predicate, or object did not correspond to the BIO scheme. Second, a rule-based check verified that each generated sentence contained exactly one consistent (S, P, O) triple and that the surface forms of the components could be located in the sentence as contiguous spans. Third, a manual expert review was conducted by two native Kazakh speakers with linguistic background. Due to the volume of the generated corpus, full manual verification was not feasible; therefore, a randomly selected stratified subset covering no less than 50% of the synthetic sentences was checked. Each reviewed sentence was assessed along three criteria: (i) grammatical correctness, including the agreement of agglutinative affixes; (ii) semantic plausibility of the (S, P, O) triple; and (iii) absence of cultural or topical bias. Sentences that failed any of these criteria were either corrected or removed from the corpus. Inter-annotator agreement was measured on a common subsample, yielding Cohen’s $\kappa = 0.86$, which corresponds to a high level of consistency. After validation, the share of sentences fully conforming to the norms of the Kazakh language exceeded 90%, which is consistent with the quantitative quality estimate provided by the metric Q_a introduced above.

Results

A model for extracting subject-predicate-object triples from Kazakh text was developed and tested in this study, with experimental results confirming the effectiveness of the proposed approach that combines transformer models and synthetic data generation.

To evaluate the performance of the developed model, the metrics of accuracy, recall and F1-score were used:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

where: TP – true positives, FP – false positives, FN – false negatives

The obtained results showed that the BERT model adapted to the Kazakh language and trained on synthetic data outperformed the baseline approaches. Models based on grammar rules demonstrated a limited accuracy of 62.8%, which is due to the difficulty of taking into account all the linguistic features of the language. The BiLSTM-CRF model achieved an accuracy of 71.2%, but there are difficulties in handling agglutinative constructions. The BERT model trained only on real data achieved an F1-measure of 74.9%, which highlights its potential, but also indicates the need to expand the training dataset. The BERT model using synthetic data achieved the highest F1-measure value of 82.6%, which indicates a significant contribution of synthetic data in improving the generalization ability of the model.

Finally, the XLM-RoBERTa architecture achieved the highest performance overall, securing an F1-score of 90.73%. This represents a considerable performance increase of approximately 8 percentage points over the best-performing BERT variant (82.6%), establishing it as the state-of-the-art model for this task. The quality analysis of the synthetic dataset confirmed its positive impact on the overall performance of all transformer models. Comparative performance results of different models are presented in Table 1.

Table 1 – Efficiency of models for the Kazakh language

Model	Precision (%)	Recall (%)	F1-score (%)
Rules	65.2	60.5	62.8
BiLSTM-CRF	72.4	70.1	71.2
BERT without synthetic data	75.8	74.0	74.9
BERT with synthetic data	83.5	81.7	82.6
XLM-RoBERTa	92	89.5	90.73

The obtained results confirm the effectiveness of using synthetic data to enhance the model's ability to process complex linguistic structures, while also highlighting the need for further improvements to address underrepresented regional dialects and topic diversity. The addition of synthetic data significantly increased the model's performance, particularly in terms of recall, indicating that the system is better able to recognize previously unseen constructions (Figure 2).

Let D_{real} be a real dataset and D_{syn} be a synthetic dataset. The total volume of training data can be expressed as: $D = D_{real} \cup D_{syn}$. A fairly good result was achieved with $D_{syn} \approx 2 \times D_{real}$, can be seen below in Figure 3.

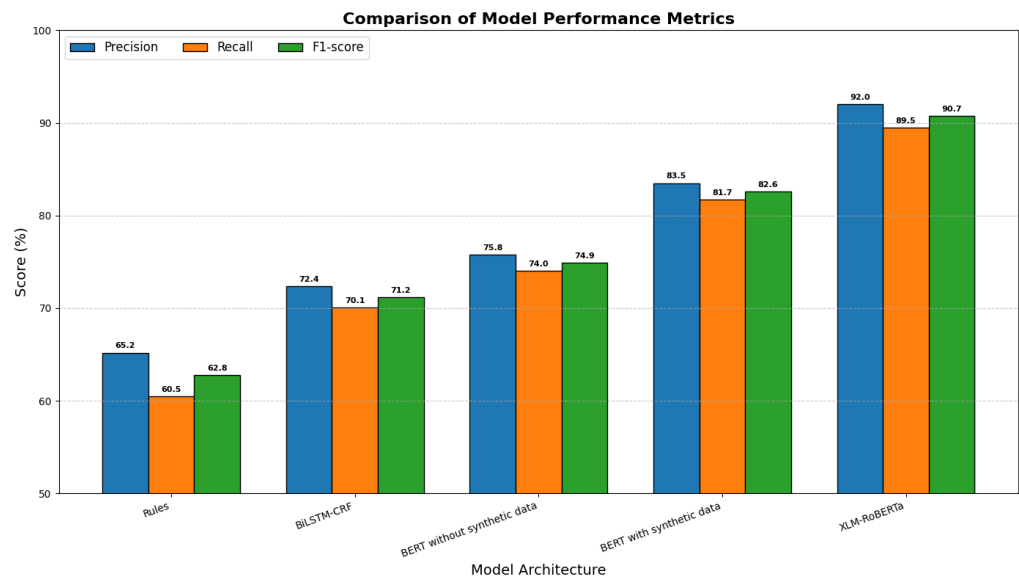


Figure 2 – Bar chart of comparative model performance

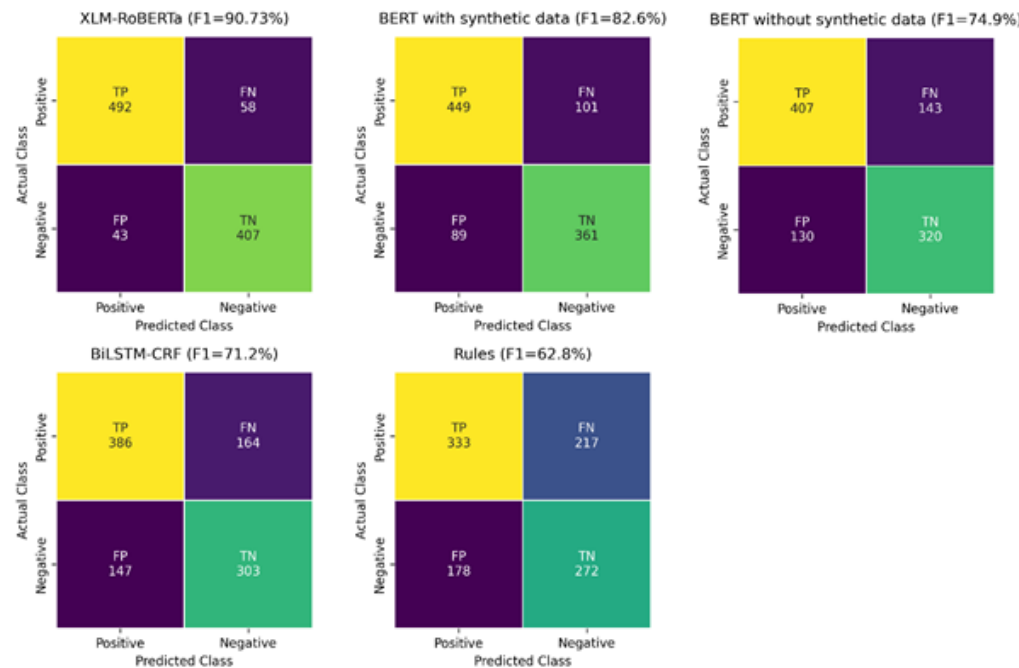


Figure 3 – Confusion matrices for different models

This amount of data increased the diversity of training examples, allowing the model to better learn representations and improve generalization ability. This demonstrates the effectiveness of using large language models to generate additional data in resource-constrained settings.

Discussion

Despite the high overall performance, the proposed approach exhibits several limitations that are important to discuss within the scope of this study. In particular, difficulties arise when processing

complex syntactic constructions, especially sentences containing nested structures such as participial and adverbial participial phrases. These constructions increase the ambiguity of subject-predicate-object relations and remain challenging for transformer-based models trained without explicit syntactic supervision.

Additional challenges are observed in sentences with flexible word order, including cases of subject-object inversion. While multilingual transformer models such as XLM-RoBERTa demonstrate strong robustness to word order variation, fully resolving such ambiguities would require architectural modifications or additional linguistic supervision.

The limited representation of rare lexemes in both real and synthetic datasets also affects model performance. In this study, the distribution of lexical items in synthetic data was intentionally kept close to naturally occurring text in order to preserve linguistic realism and avoid biasing the model toward artificially frequent rare words. Although this choice limits the model's exposure to low-frequency vocabulary, it ensures more realistic training conditions. More targeted augmentation strategies for rare lexemes are therefore left for future research.

Given the agglutinative nature of the Kazakh language, morphological complexity represents another important challenge. However, the primary objective of this study was to assess the capability of transformer-based models combined with synthetic data without relying on language-specific preprocessing tools. Integrating morphological analyzers would introduce additional variables and complicate the isolation of the contribution of synthetic data. For this reason, such tools were not included in the current pipeline, although their potential benefits are acknowledged as a direction for future work.

Manual evaluation of the generated synthetic texts showed that more than 90% of the sentences conform to the basic norms of the Kazakh language, including spelling, grammar, and syntactic rules. Nevertheless, certain inaccuracies were observed due to inherent limitations of the language model used for generation. In addition, while the synthetic data exhibit high overall quality, they do not explicitly model regional dialectal variations of the Kazakh language. Addressing dialectal diversity would substantially increase the complexity of data generation and evaluation and is therefore considered beyond the scope of the present study.

An uneven thematic distribution was also observed in the synthetic dataset. However, the experimental results indicate that even with this limitation, synthetic data significantly improve model performance. To quantify this effect, the following metric was used:

$$\Delta F1 = F1_{\text{synthetic}} - F1_{\text{baseline}}$$

where $F1_{\text{synthetic}}$ – F1-score of the model with synthetic data, $F1_{\text{baseline}}$ – F1-score of the model without synthetic data.

In addition to the aggregated precision, recall, and F1-score reported in Table 1, we computed per-class metrics for the three target categories (Subject, Predicate, Object) as well as macro- and micro-averaged F1, and a sentence-level exact-match (EM) score that counts a sentence as correctly processed only when all three components of the (S, P, O) triple are extracted exactly. The XLM-RoBERTa model trained on the combined real and synthetic corpus achieved an F1-score of 91.4% for Subject, 89.2% for Predicate, and 91.6% for Object, with a macro-F1 of 90.7% and a micro-F1 of 90.9%. The sentence-level exact-match score was 81.3%, which is approximately 9 percentage points lower than the token-level F1-score, indicating that a substantial share of remaining errors corresponds to sentences in which only one of the three components is mislabeled. These additional metrics provide a more complete picture of model behavior than the aggregated F1-score alone.

A detailed error analysis of the model output on the test set was performed in order to identify systematic sources of misclassification. The errors can be grouped into four main categories. First, boundary errors (approximately 41% of all errors) occur when the model correctly identifies the type of a component but assigns incorrect token boundaries; this is most frequent in noun phrases with stacked agglutinative suffixes, where the segmentation produced by the SentencePiece tokenizer does

not always align with morpheme boundaries. Second, role confusion errors (about 28%) correspond to cases in which the subject and the object are swapped; such errors are concentrated in sentences with non-canonical (OSV or OVS) word order, where surface cues for the grammatical role are reduced. Third, predicate identification errors (about 19%) typically involve compound or analytical verb forms (e.g., light-verb constructions of the type “X етті”, “X болды”), where the model labels only the lexical verb and omits the auxiliary. Fourth, the remaining errors (approximately 12%) involve rare or out-of-vocabulary lexemes and proper nouns that are underrepresented in both the real and the synthetic data. This decomposition shows that the largest portion of residual errors is morphological and structural rather than purely lexical, which suggests that further gains can be obtained by integrating a Kazakh-specific morphological analyzer and by augmenting the synthetic corpus with controlled variations of word order and analytical verb constructions.

The study showed that using synthetic data improves the quality of model training, which is important for processing the Kazakh language with its grammatical features. The results of the experiments indicate that XLM-RoBERTa effectively copes with agglutinative morphology and free word order, providing accurate classification of tokens. The developed model demonstrates the advantages of transformer models in processing texts with high linguistic complexity, although there are difficulties with rare words and very different word order.

To further improve the quality of synthetic data, the integration of post-processing methods or the use of models specifically trained for the Kazakh language seem promising. To improve model performance, future research should focus on integrating morphological analyzers to more accurately handle agglutinative word structures [25], expanding synthetic datasets to include a wider range of lexical and syntactic diversity, and experimenting with other language models to assess the impact of more advanced generation algorithms.

Conclusion

In this paper, a method was proposed to efficiently extract triple subject-predicate-object relations from Kazakh texts. The XLM-RoBERTa transform model was used. Kazakh text generation using GPT-4 was used to ensure the diversity and volume of data required for training the model. The proposed method allows the model to better understand the Kazakh language, even if there is a shortage of real data. Despite the success, the model still has a poor understanding of rare words and complex sentences, and the synthetic data needs to be improved. Future work should focus on developing more advanced methods for generating synthetic data, fine-tuning large language models for the Kazakh language, and integrating language analyzers that can improve the processing of agglutinative features of the language. Future research plans to apply this method to other languages and expand the capabilities of synthetic data. Thus, the combination of transformer models and synthetic data generation opens new horizons for solving complex NLP problems such as extracting triple relations and contributes to the further development of artificial intelligence technologies on a global scale.

REFERENCES

- 1 Zhang, Y., Sadler, T., Taesiri, M.R., Xu, W., and Reformat, M. Fine-tuning language models for triple extraction with data augmentation. *Proceedings of the 1st Workshop on Knowledge Graphs and Large Language Models (KaLLM 2024)*, 116–124 (2024).
- 2 Li, Z., Li, X., Sheng, J., and Slamu, W. AgglutiFiT: Efficient low-resource agglutinative language model fine-tuning. *IEEE Access*, 8, 148489–148499 (2020). <https://doi.org/10.1109/ACCESS.2020.3015854>
- 3 Barakhnin, V.B., Fedotov, A.M., Bakieva, A.M., Bakiev, M.N., Tazhibaeva, S.Zh., Batura, T.V., Kozhemyakina, O.Yu., Tusupov, D.A., Sambetbayeva, M.A., and Lukpanova, L.K. *Algoritmy generatsii i stemmatizatsii slovoform kazakhskogo yazyka [Algorithms for generation and stemming of Kazakh language word forms]*. Cloud of Science, 4 (2017). (in Russian).

- 4 Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186 (2019). <https://doi.org/10.18653/v1/N19-1423>
- 5 Tolegen, G., Toleu, A., Mussabayev, R., Zhumazhanov, B., and Ziyatbekova, G. Generative pre-trained transformer for Kazakh text generation tasks. *2023 19th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS)*, 144–148 (2023). <https://doi.org/10.1109/OPCS59592.2023.10275765>
- 6 Rabchevsky, A.N., Ashikhmin, E.G., and Yasnitsky, L.N. Synthesis of datasets for neural networks based on expert knowledge. In: Arseniev, D.G., and Aouf, N. (eds.). *Cyber-Physical Systems and Control II. Lecture Notes in Networks and Systems*, 460, 561–569 (2023). https://doi.org/10.1007/978-3-031-20875-1_50
- 7 Karyukin, V., Rakhimova, D., Karibayeva, A., Turganbayeva, A., and Turarbek, A. The neural machine translation models for the low-resource Kazakh–English language pair. *PeerJ Computer Science*, 9, e1224 (2024). <https://doi.org/10.7717/peerj-cs.1224>
- 8 Akhmed-Zaki, D., Mansurova, M., Madiyeva, G., and Kyrgyzbayeva, M. Development of the information system for the Kazakh language preprocessing. *Cogent Engineering*, 8 (1), 1896418 (2021). <https://doi.org/10.1080/23311916.2021.1896418>
- 9 Abibullayeva, A., and Çetin, A. Keyword extraction from Kazakh news dataset with BERT. *El-Cezeri Journal of Science and Engineering*, 9 (4), 1193–1200 (2022). <https://doi.org/10.31202/ecjse.1131826>
- 10 Bektemyssova, G., and Sabdenov, A. Building a semantic knowledge graph search model for finding a causal answer. *Revue d'Intelligence Artificielle*, 38 (1), 243–250 (2024). <https://doi.org/10.18280/ria.380125>
- 11 Kaffee, L.-A., Biswas, R., Keet, C.M., Kalemi Vakaj, E., and de Melo, G. Multilingual knowledge graphs and low-resource languages: A review. *Transactions on Graph Data and Knowledge*, 1 (1), 10:1–10:19 (2024). <https://doi.org/10.4230/TGDK.1.1.10>
- 12 Chen, J., Geng, Y., Chen, Z., Pan, J., He, Y., Zhang, W., Horrocks, I., and Chen, H. Low-resource learning with knowledge graphs: A comprehensive survey. *arXiv preprint*, arXiv:2112.10006 (2021). <https://doi.org/10.48550/arXiv.2112.10006>
- 13 Kortmann, F., et al. Concept of a cloud state modeling system for lead-acid batteries: Theory and prototyping. *2021 International Conference on Electronics, Information, and Communication (ICEIC)*, 1–4 (2021). <https://doi.org/10.1109/ICEIC51217.2021.9369785>
- 14 Wang, Z., Karthikeyan, K., Mayhew, S., and Roth, D. Extending multilingual BERT to low-resource languages. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2649–2656 (2020). <https://doi.org/10.18653/v1/2020.findings-emnlp.240>
- 15 Conneau, A., Khandelwal, U., Goyal, N., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451 (2020). <https://doi.org/10.18653/v1/2020.acl-main.747>
- 16 Bhatt, A., Vaghela, N., and Dudhia, K. Generating knowledge graphs from large language models: A comparative study of GPT-4, LLaMA 2, and BERT. *arXiv preprint*, arXiv:2412.07412 (2024). <https://doi.org/10.48550/arXiv.2412.07412>
- 17 Deepchecks Community Blog. LLM models comparison: GPT-4, Bard, LLaMA, Flan-UL2, BLOOM (2024). URL: <https://www.deepchecks.com/llm-models-comparison>
- 18 OpenAI. GPT-4 technical report. *arXiv preprint*, arXiv:2303.08774 (2023). <https://doi.org/10.48550/arXiv.2303.08774>
- 19 Bojic, L., Kovacevic, P., and Cabarkapa, M. GPT-4 surpassing human performance in linguistic pragmatics. *arXiv preprint*, arXiv:2312.09545 (2023). <https://doi.org/10.48550/arXiv.2312.09545>
- 20 Baktash, J.A., and Dawodi, M. GPT-4: A review on advancements and opportunities in natural language processing. *arXiv preprint*, arXiv:2305.03195 (2023). <https://doi.org/10.48550/arXiv.2305.03195>
- 21 Guzev, V.G., and Burykin, A.A. Obshchie stroevye osobennosti agglutinativnykh yazykov [General structural features of agglutinative languages]. *Acta Linguistica Petropolitana. Trudy Instituta Lingvisticheskikh Issledovaniy*, 1 (2017). (in Russian).
- 22 Pires, T., Schlinger, E., and Garrette, D. How multilingual is multilingual BERT? *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4996–5001 (2019).

23 Urmanov, N., Bektemyssova, G.U., and Basiri, K. NLP algorithms in OOP for processing the text-based documents in Russian language in machine learning. International Journal of Research, 6, 361–371 (2019).

24 Madasamy, A.K., Kumar, C.A., Dhanalakshmi, V., Rekha, R.U., Soman, K.P., and Rajendran, S. Morphological analyzer for agglutinative languages using machine learning approaches. 2009 International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom), 433–435 (2009). <https://doi.org/10.1109/ARTCom.2009.18>

25 Ismayilzada, M., Circi, D., Sälevä, J., Sirin, H., Köksal, A., Dhingra, B., Bosselut, A., van der Plas, L., and Ataman, D. Evaluating morphological compositional generalization in large language models. arXiv preprint, arXiv:2410.12656 (2024). <https://doi.org/10.48550/arXiv.2410.12656>

^{1*}Бектемысова Г.У.,

профессор, ORCID ID: 0000-0002-0850-0558,

*e-mail: g.bektemisova@iitu.edu.kz

¹Сабденов А.,

PhD студент, ORCID ID: 000-0002-4436-8523,

e-mail: a.sabdenov@iitu.edu.kz

²Сатыбалдиева Р.,

қауымдастырылған профессор, ORCID ID: 0000-0002-0678-7583,

e-mail: r.satybaldiyeva@satbayev.university

¹Быков А.,

қауымдастырылған профессор, ORCID ID: 0000-0002-9563-5185,

e-mail: a.bykov@iitu.edu.kz

³Nor'ashikin Binti Ali

қауымдастырылған профессор, ORCID ID: 0000-0002-5653-2482

e-mail: shikin@uniten.edu.my

¹Халықаралық ақпараттық технологиялар университеті, Алматы қ., Қазақстан

²Сәтбаев университет, Алматы қ., Қазақстан

³Тенага ұлттық университеті, Селангор, Малайзия

ҚАЗАҚ ТІЛІ ҮШІН СИНТАКСИКО-СЕМАНТИКАЛЫҚ ҚАТЫНАСТАРДЫ ТРАНСФОРМАЦИЯЛЫҚ АРХИТЕКТУРАЛАР ЖӘНЕ СИНТЕТИКАЛЫҚ КОРПУСТАР НЕГІЗІНДЕ ШЫҒАРУДЫ ОҢТАЙЛАНДЫРУ

Аңдатпа

Табиғи тілді өңдеу (NLP) әдістері соңғы жылдары кең таралған және іздеу жүйелерінде, шешімдерді қабылдауды қолдаудың интеллектуалды платформаларында және басқа да көптеген жасанды интеллект қосымшаларында белсенді түрде қолданылады. Бұл саладағы негізгі міндеттердің бірі – субъект-предикат-объект форматындағы үштік қатынастарды шығару. Мұндай құрылымдар құрылымдалмаған мәтіндерді реттелген деректерге аударуға мүмкіндік береді және осылайша білім графиктерін құруға және логикалық талдауды ұйымдастыруға негіз болады. Ағылшын немесе қытай сияқты бай ресурстық базасы бар тілдер үшін бұл мәселе қазірдің өзінде жақсы шешілген. Алайда ресурсы шектеулі тілдер үшін, соның ішінде қазақ тілі үшін, таңбаланған корпус пен арнайы лингвистикалық ресурстардың аз болуына байланысты мәселе өзекті күйінде қалып отыр. Жұмыста XLM-RoBERTa архитектурасына негізделген үлгіні оқыту үшін қолданылатын синтетикалық түрде жасалған деректер есебінен ресурстық базаны кеңейтуді көздейтін тәсіл ұсынылады. XLM-RoBERTa – BERT үлгісінің жетілдірілген нұсқасы; ол алдын ала оқытуға арналған ауқымды корпусымен және кросс-тілдік тасымалдау міндеттеріндегі жоғары тиімділігімен ерекшеленеді, бұл әсіресе ресурстары шектеулі тілдер үшін маңызды. Эксперименттік зерттеулер ұсынылған әдістің F1-метрика бойынша 90,73% нәтижеге қол жеткізетінін көрсетті. Бұл нәтиже XLM-RoBERTa сияқты озық үлгілер мен жасанды деректер арқылы кеңейтудің салыстырмалы түрде қарапайым әдістерін үйлестіру күрделі тілдік құрылымдарды өңдеу сапасын айтарлықтай арттыра алатынын растайды. Жасалған қорытындылар ұсынылып отырған тәсілді ресурстары шектеулі тілдерге арналған NLP жүйелерін дамытудың перспекти-

валы бағыты ретінде қарастыруға мүмкіндік береді. Сонымен қатар, нәтижелер оны ақпараттық іздеу жүйелеріне, білімді басқару платформаларына және көптілді интеллектуалды қосымшаларға практикалық тұрғыдан енгізу мүмкіндігін көрсетеді.

Түйін сөздер: XLM-RoBERTa, BERT, синтетикалық деректерді генерациялау, үлкен тілдік модельдер, машиналық оқыту, морфологиялық талдау, қазақ тілі.

^{1*}**Бектемысова Г.У.,**
профессор, ORCID ID: 0000-0002-0850-0558,
*e-mail: g.bektemisova@iitu.edu.kz

¹**Сабденов А.,**
PhD студент, ORCID ID: 000-0002-4436-8523,
e-mail: a.sabdenov@iitu.edu.kz

²**Сатыбалдиева Р.,**
ассоциированный профессор, ORCID ID: 0000-0002-0678-7583,
e-mail: r.satybaldiyeva@satbayev.university

¹**Быков А.,**
ассоциированный профессор, ORCID ID: 0000-0002-9563-5185,
e-mail: a.bykov@iitu.edu.kz

³**Nor'ashikin Binti Ali,**
ассоциированный профессор, ORCID ID: 0000-0002-5653-2482,
e-mail: shikin@uniten.edu.my

¹Международный университет информационных технологий, г. Алматы, Казахстан

²Сатбаев Университет, г. Алматы, Казахстан

³Национальный университет Тенага, Селангор, Малайзия

ОПТИМИЗАЦИЯ ИЗВЛЕЧЕНИЯ СИНТАКСИКО-СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ ДЛЯ КАЗАХСКОГО ЯЗЫКА НА ОСНОВЕ ТРАНСФОРМЕРНЫХ АРХИТЕКТУР И СИНТЕТИЧЕСКИХ КОРПУСОВ

Аннотация

Методы обработки естественного языка (NLP) в последние годы получили широкое распространение и активно используются в поисковых системах, интеллектуальных платформах поддержки принятия решений, а также во множестве других приложений искусственного интеллекта. Одной из ключевых задач в этой области является извлечение тройных отношений в формате «субъект – предикат – объект». Такие структуры позволяют переводить неструктурированные тексты в упорядоченные данные и тем самым формируют основу для построения графов знаний и организации логического анализа. Для языков с богатой ресурсной базой, например английского или китайского, данная задача уже достаточно хорошо решается. Однако для малообеспеченных языков, включая казахский, проблема остается актуальной из-за ограниченности доступных размеченных корпусов и специализированных лингвистических ресурсов. В работе предлагается подход, предусматривающий расширение ресурсной базы за счет синтетически сгенерированных данных, которые затем используются для обучения модели на основе архитектуры XLM-RoBERTa. XLM-RoBERTa, являясь улучшенной версией модели BERT, отличается более крупным корпусом для предварительного обучения и повышенной эффективностью в задачах кросс-языкового переноса, что особенно важно для языков с ограниченными ресурсами. Экспериментальные исследования показали, что предложенный метод обеспечивает F1-метрику 90,73%. Этот результат подтверждает, что комбинация передовых моделей, таких как XLM-RoBERTa, и относительно простых приемов искусственного расширения данных способна заметно повысить качество при обработке сложных языковых конструкций. Сделанные выводы позволяют рассматривать предложенный подход как перспективное направление для развития NLP в отношении малообеспеченных языков. Кроме того, результаты открывают возможности его практической интеграции в системы поиска информации, управления знаниями и многоязычные интеллектуальные приложения.

Ключевые слова: XLM-RoBERTa, BERT, генерация синтетических данных, большие языковые модели, машинное обучение, морфологический анализ, казахский язык.

ӘОЖ 376:004
FTAXP 14.35.07

<https://doi.org/10.55452/1998-6688-2026-23-2-262-275>

¹**Сербин В.В.,**

т.ғ.к., қауымдастырылған профессор, ORCID ID: 0000-0002-5807-3873,

e-mail: v.serbin@satbayev.university

¹**Абдуллаева А.С.,**

магистр, аға оқытушы, ORCID ID: 0000-0001-5188-3008,

e-mail: a.s.abdullayeva@satbayev.university

^{1*}**Абылханова А.А.,**

студент, ORCID ID: 0009-0002-6333-0202,

*e-mail: a.abylhanova@stud.satbayev.university

¹**Батырхан Ж.Ғ.,**

студент, ORCID ID: 0009-0008-6587-0441,

e-mail: z.batyrgan@stud.satbayev.university

¹**Өтеміс Ғ.А.,**

студент, ORCID ID: 0009-0001-1875-798X,

e-mail: g.otemis@stud.satbayev.university

¹Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті,

Алматы қ., Қазақстан

ҚР ҒЖБМ ҒК Автоматика және ақпараттық технологиялар институты,

Алматы қ., Қазақстан

АРНАЙЫ БІЛІМ БЕРУ ҚАЖЕТТІЛІКТЕРІ БАР БАЛАЛАРҒА АРНАЛҒАН ҚАШЫҚТЫҚТАН ОҚЫТУҒА ҚАУІПСІЗ ОНЛАЙН-ПЛАТФОРМАНЫ ЖОБАЛАУ

Аңдатпа

Мақалада арнайы білім беру қажеттіліктері бар балаларды қашықтықтан оқытуға арналған қауіпсіз мамандандырылған онлайн-платформаны жобалаудың өзекті мәселелері қарастырылады. Білім беруді цифрландыру инклюзив үшін жаңа мүмкіндіктер ашып жатыр, алайда жаппай шешімдер көбінесе цифрлық, когнитивтік және әлеуметтік кедергілерге ғана назар аударып, арнайы білім берудегі нақты қажеттіліктерді ескермейді. Зерттеудің мақсаты – қолжетімділік, жекелендіру және киберқауіпсіздік мәселелерін жан-жақты талдай отырып шешетін қауіпсіз және бейімделгіш білім беру ортасының тұжырымдамалық моделін әзірлеу. Жобада әртүрлі арнайы қажеттіліктері бар балалардың ыңғайлылығы қарастырылды, соның ішінде сенсорлық сфераның бұзылуы, аутистикалық (ASD) және гиперкинетикалық (ADHD) ауытқудың дамуы талданды, соның негізінде пайдаланушы интерфейсі мен тәжірибесін (UI/UX) жобалаудың негізгі принциптері тұжырымдалды. Ұсынылған платформа архитектурасы интеллектуалды мазмұн мен интерфейсін бейімдеу жүйесін, жекелендірілген білім беру траекторияларын сондай-ақ премодерация функциясы бар қауіпсіз байланыс модульдерін қамтиды. Жеке деректерді қорғауды, кибербуллингтің алдын алуды және қолжетімділікті бақылауды қамтамасыз ететін көп деңгейлі қауіпсіздік жүйесіне ерекше назар аударылады. Мақала инклюзивті цифрлық білім беру ортасын құруға бағытталған білім беру технологияларын әзірлеушілер, педагогтер мен оқу орындары әкімшіліктері үшін практикалық маңызы бар жұмыс.

Түйін сөздер: қашықтықтан оқыту, арнайы білім беру қажеттіліктері, инклюзивті білім беру, кибер-қауіпсіздік, қол жетімділік, пайдаланушы интерфейсі, UX/UI даралау.

Кіріспе

Соңғы жылдары білім беруді цифрландырудың қарқыны күшейіп, қашықтықтан оқыту әлемдік білім жүйесінде кең таралған форматқа айналды. Алайда қазіргі қолданыстағы онлайн-платформалардың көпшілігі инклюзивті топтағы оқушылардың нақты ерекшеліктерін толық қамтымайды. Нәтижесінде көру, есту, когнитивтік және моторлық шектеулері бар оқушылар үшін қашықтықтан оқу қосымша қиындықтар туындататыны байқалады. Атап айтар болсақ оқу материалын қабылдау, тапсырмаларды орындау, мұғаліммен және құрдастарымен қауіпсіз байланыс орнату сияқты маңызды процестердің тиімділігіне тікелей әсер етеді.

Осыған байланысты арнайы білім беру қажеттіліктері бар балаларға қолжетімді, қауіпсіз және бейімделе алатын онлайн-платформа әзірлеу мәселесі қазіргі қоғамда аса өзекті болып отыр. Мұндай платформа тек оқытуды ұйымдастыру құралы ғана емес, сонымен қатар баланың психологиялық жайлылығы мен деректерінің қорғалуын қамтамасыз ететін кешенді цифрлық орта болуы тиіс. Бұл мақалада осындай платформа жобасының тұжырымдамалық моделін ұсыну, пайдаланушы қажеттіліктерін талдау, қауіпсіздік пен интерфейс бейімделуінің әсерін эксперименттік зерттеу мақсат етілген. Берілген зерттеу нәтижелері инклюзивті цифрлық білім беруді дамытуға бағытталған білім беру ұйымдары, педагогтер мен IT мамандары үшін практикалық маңызға ие.

Мақаланың мақсаты: Арнайы қажеттілігі бар балалардың білім беру, дамыту және әлеуметтік-психологиялық қажеттіліктерін кешенді қанағаттандыруға бағытталған қашықтықтан оқытудың қауіпсіз онлайн-платформасын жобалау тұжырымдамасын әзірлеу болып табылады. Осы мақсатқа жету үшін жұмыста бірқатар міндеттер қойылған: Арнайы қажеттілігі бар балалардың негізгі санаттары үшін платформаның пайдаланушы интерфейсі мен функционалдығына қойылатын нақты талаптарды талдау; Бейімделу, жекелендіру және қауіпсіз байланыс модульдерін біріктіретін платформаның архитектуралық және функционалды моделін жасау; Деректерді қорғауды және пайдаланушылардың психологиялық жайлылығын қоса алғанда, көп деңгейлі қауіпсіздікті қамтамасыз етудің негізгі принциптері мен тетіктерін анықтау қажеттіліктері қарастырылады.

Пайдаланушы қажеттіліктерін талдау және дизайн принциптері кез-келген білім беру платформасының жетістігі оның пайдаланушылардың қажеттіліктерін қаншалықты қанағаттандыратындығымен анықталады.

Материалдар мен әдістер

Платформаның архитектуралық моделі

Ұсынылып отырған онлайн-платформа модульдік архитектура принципіне негізделген және бірнеше логикалық қабаттарды қамтиды. Мұндай құрылым жүйенің қауіпсіздігін, масштабталуын және инклюзивті интерфейсті тиімді басқаруға мүмкіндік береді.

Архитектура келесі негізгі компоненттерден тұрады:

Пайдаланушы интерфейсі қабаты (User Interface Layer). Бұл қабат пайдаланушылармен тікелей өзара әрекеттесуді қамтамасыз етеді. Мұнда адаптивті UI/UX жүйесі орналасқан. Интерфейс әртүрлі арнайы білім беру қажеттіліктері бар балалар үшін автоматты түрде бейімделеді (қаріп өлшемі, контраст, визуалды элементтер саны, аудио нұсқаулар).

Қолданбалы логика қабаты (Application Layer). Бұл қабат білім беру контентін басқару, тапсырмалар орындау, мұғалім мен оқушы арасындағы байланыс сияқты негізгі функцияларды жүзеге асырады.

Қауіпсіздік қабаты (Security Layer). Қауіпсіздік жүйесі көп деңгейлі аутентификация, қолжетімділікті басқару және деректерді шифрлау механизмдерін қамтиды. Бұл модуль кибербуллинг, фишинг және рұқсатсыз кіру әрекеттерін болдырмауға бағытталған.

Контентті модерациялау және талдау қабаты (Moderation & AI Layer).

Бұл модуль машиналық оқыту алгоритмдерін пайдалана отырып зиянды контентті, фишингтік сілтемелерді және агрессивті мәтіндерді автоматты түрде анықтайды.

Деректер базасы (Database Layer). Бұл деңгейде пайдаланушы деректері, оқу прогресі және қауіпсіздік журналдары сақталады.

Арнайы пайдаланушы талаптары

Жүргізілген талдау арнайы білім беру қажеттілігі бар балалардың санаттары бойынша топтастырылған негізгі талаптарды бөліп көрсетуге мүмкіндік береді:

Көру қабілеті бұзылған балалар үшін: беттің орналасуына және семантикалық құрылымына қатаң шектеулер қоятын экранға қол жеткізу бағдарламаларымен толық үйлесімділік қажет [12]. Пернетақта навигациясын қолдау, барлық графикалық элементтер мен бейне мазмұны үшін балама мәтіндік сипаттамалар, сондай-ақ көру қабілетінің орташа деңгейі бар пайдаланушылар үшін түс схемалары мен қаріп өлшемдерін өзгерту мүмкіндігі қажет.

Есту қабілеті бұзылған балалар үшін: барлық аудио және бейнематериалдар дәл субтитрлермен немесе ым-ишара тіліне аудармамен қамтамасыз етілуі керек [4]. Көрнекі ескертулер дыбыстық сигналдарды қайталауы керек. Маңызды элемент – есту қабілеті нашар балаларға мұғалімнің нақты уақыттағы сөйлеуін визуализациялау маңызды элементтердің бірі болып табылады.

ASD және ADHD бар балалар үшін: болжамды, құрылымдалған және минималистік интерфейс өте маңызды. Алаңдататын элементтерді азайту керек (анимациялық баннерлер, жыпылықтайтын түймелер). Мазасыздықты азайту және бағдарлауды жақсарту үшін көрнекі сабақ кестелері, таймерлер, әрекетті өзгерту туралы ескертулер және платформа ішінде «демалыс аймағын» құру мүмкіндігі қажет. ASD және ADHD бар балалардың онлайн оқу кезінде назар шоғырландыруы төмен болатыны дәлелденген, сондықтан интерфейс құрылымдық әрі болжамды болуы қажет [2].

ADHD бар балалар үшін интерфейс минималистік режимде жұмыс істей алады. Бұл режимде анимациялар мен динамикалық элементтер өшіріліп, тек негізгі оқу элементтері көрсетіледі. Ал көру қабілеті бұзылған пайдаланушылар үшін керісінше ірі визуалды элементтер мен контраст режимі іске қосылады. Мұндай адаптивті тәсіл әртүрлі пайдаланушылардың талаптарын бір платформада үйлестіруге мүмкіндік береді.

Тірек-қимыл аппараты бұзылған балалар үшін: платформа балама енгізу құрылғыларын қолдауы керек: көзді басқару, дауыстық командалар, сенсорлық экрандар және арнайы джойстиктер. Тапсырмаларды орындау уақыты икемді болуы керек, ал интерфейс курсордың дәл, жылдам қозғалысын қажет етпестен жұмыс істеуге мүмкіндік береді.

UI/UX дизайнының негізгі принциптері

Анықталған талаптар негізінде келесі жобалау принциптері тұжырымдалды:

Әмбебап дизайн (Inclusive Design): Интерфейс бастапқыда пайдаланушылардың ең кең ауқымы үшін мүмкіндігінше қол жетімді болатындай етіп жасалған, әрі қарай бейімделуді қажет етпейді. Әмбебап дизайн принциптері WCAG 2.1 (2018) және жаңартылған WCAG 2.2 (2023) стандарттарына сәйкес жүзеге асырылады [19].

Мультимодальді ұсыныс принципі бойынша маңызды мазмұн текст, аудио немесе визуалды форматта қайталанып, пайдаланушыға ең қолайлы тәсілді таңдауға мүмкіндік береді.

Даралау және бейімделу: жүйе интерфейсті теңшеуге мүмкіндік беруі керек: түстер палитрасын, қаріптердің өлшемін, мазмұнның егжей-тегжейлі деңгейін, бір уақытта көрсетілетін элементтердің санын өзгерту [10].

Жобалаудағы бірізділік маңызды рөл атқарады: интерфейстегі навигация мен басқару элементтері барлық бөлімдерде өзгеріссіз сақталуы тиіс. Бұл когнитивті жүктемені азайтады және бағдарлауға көмектеседі.

Арнайы қажеттіліктері бар балаларға арналған онлайн-платформаның мүмкіндіктерінің сәйкес келуі үшін арнайы зертханалық және пайдаланушылық тестілеу элементтерін біріктірген гибриді форматта эксперимент өткізілді. Интерфейс пен қауіпсіздік механизмдерінің тиімділігін бағалау үшін 12–16 жас аралығындағы арнайы білім беру қажеттіліктері бар 24 оқушының қатысуымен пилоттық тест жүргізілді.

Қатысушылар құрамына:

- ♦ көру қабілеті бұзылған – 6 оқушы;
- ♦ есту қабілеті бұзылған – 5 оқушы;
- ♦ аутизм спектрі бұзылыстары (ASD) – 7 оқушы;
- ♦ назар тапшылығы және гиперактивтілік синдромы (ADHD) – 6 оқушы кірді.

Әр қатысушы жүйенің бірнеше конфигурациясымен жұмыс істеп, оқу тапсырмаларын орындады. Утомляемость әсерін азайту мақсатында тестілеу бірнеше қысқа сессияларға бөлінді.

Эксперименттің мақсаты мен міндеттеріне тоқталып өтсек:

Мақсаты: жобалау шараларының (аутентификация, мазмұнды модерациялау, адаптивті интерфейс) платформаның қауіпсіздігіне және оның арнайы білім беру қажеттілігі бар балалар үшін қолайлылығына әсерін зерттеу.

Тапсырмалар:

Факторлардың әсері туралы гипотезаларды тұжырымдау.

2×2×2 толық факторлық жоспарды жүзеге асыру және бірқатар зертханалық жұмыстарды орындау.

Сандық көрсеткіштерді жинау және талдау: Precision/Recall/F1, FPR, MTTD, тапсырма толық, SUS және т. б.

Платформа конфигурациялары бойынша қорытындылар мен ұсыныстар жасау қажет.

Гипотезалар:

H1: MFA (көпфакторлы аутентификация) есептік жазбаларға қарсы шабуылдардың сәттілігін азайтады «тек құпия сөз», кіру уақыты орташа ұлғайған кезде.

H2: жақсартылған модерация зиянды оқиғалардың санын азайтады, бірақ жалған құлыптарды арттыруы мүмкін.

H3: адаптивті UI task completion және sus-ті арттырады, әсіресе арнайы білім беру қажеттілігі бар пайдаланушылар үшін.

H4 (өзара әрекеттесу): MFA + күшейтілген модерация + адаптивті UI тіркесімі қауіпсіздік пен қол жетімділіктің жақсы тепе-теңдігін қамтамасыз етеді, бірақ жағымсыз өзара әрекеттесулер мүмкін (мысалы, адаптивті хабарлама форматындағы жалған алерттер).

Таңдалған эксперимент түрі:

Көп факторлы зертханалық эксперимент-толық факторлық жоспар 23(толық факторлық жоспар: 3 фактор × 2 деңгей = 8 шарт). Зертханалық формат шабуыл сценарийін және пайдаланушы сеанстарын бақылауды қамтамасыз етеді.

Факторлар мен деңгейлер:

A-Аутентификация:

A1 – «тек құпия сөз» (стандарт).

A2 – «Арнайы білім беруге бейімделген MFA» (құпия сөз + жеңілдетілген екінші фактор: SMS/дауыс/қарапайым push).

B – Модерация / мазмұнды сүзу:

B1 – «негізгі» (минималды сүзгілер, карантин/қолмен модерация).

B2 – «күшейтілген» (Автоматты ML сүзгілері, фишингті тану, ИОС детекторы).

C – интерфейс / қол жетімділігі:

C1 – «стандартты интерфейс».

C2 – «Арнайы білім беру қажеттіліктері бар балалар үшін адаптивті интерфейс» (үлкен элементтер, дауыстық кеңестер, жеңілдетілген ағындар).

Кесте 1-де үш негізгі фактор – A (аутентификация), B (мазмұнды модерациялау) және C (интерфейс) өзгертін 2×2×2 толық факторлық эксперимент матрицасы берілген. Әрбір фактордың екі деңгейі бар (төмен және жоғары), бұл олардың жеке әсерін де, олардың арасындағы онлайн платформаның қауіпсіздік және қол жетімділік көрсеткіштеріне өзара әрекеттесуін де талдауға мүмкіндік береді.

Кесте 1 – Факторлық эксперимент матрицасы.

A(Authentication)	B(Moderation)	C(User Interface)
A1	B1	C1
A1	B1	C2
A1	B2	C1
A1	B2	C2
A2	B1	C1
A2	B1	C2
A2	B2	C1
A2	B2	C2

Бірінші фактор – аутентификация, яғни жүйеге кіру тәсілі. Мұнда екі нұсқа қарастырылады: тек пароль арқылы кіру және екі қадамды тексеру (мысалы, пароль мен қосымша код немесе дауыс арқылы растау). Бұл балалардың есептік жазбасын қорғау үшін қажет. Платформадағы екі факторлы аутентификация мен бейімделген қауіпсіздік әдістері шабуылдардың табысты өту ықтималдығын төмендетеді [18]. Арнайы білім беру қажеттіліктері бар балалар үшін көп факторлы аутентификация жеңілдетілген форматта жүзеге асырылады. Мысалы, есту қабілеті бұзылған пайдаланушылар үшін push-хабарламалар мен визуалды кодтар қолданылады, ал көру қабілеті бұзылған пайдаланушылар үшін дыбыстық код немесе экран оқу бағдарламасымен үйлесімді мәтіндік кодтар ұсынылады.

Екінші фактор – модерация және контентті сүзу деңгейі. Бұл бөлім платформаның зиянды сілтемелер мен қауіпті хабарламаларды автоматты түрде анықтау мүмкіндігін тексереді. Бір нұсқада тек қарапайым сүзгі болады, ал екіншісінде күшейтілген модерация қолданылады, яғни зиянды сілтемелерді, фишингті және қажетсіз контентті автоматты түрде танитын жүйе іске қосылады. Инклюзивті платформаларда деректерді қорғау және кибербуллингтің алдын алу маңызды рөл атқарады [20]. Сонымен қатар аутизм спектріндегі балаларға арналған платформаларда қауіпсіз интерактивті орта құруға бағытталған патенттелген шешімдер бар [17]. Контентті автоматты талдау үшін машиналық оқыту әдістері қолданылды: токсикалық мәтінді анықтау үшін BERT негізіндегі NLP моделі; фишингтік хабарламаларды анықтау үшін Random Forest классификаторы; зиянды сілтемелерді анықтау үшін IOC signature detection әдісі; спам және қауіпті контентті сүзу үшін Gradient Boosting алгоритмі пайдаланылды.

Бұл алгоритмдер хабарламалар мәтінін, сілтемелерді және пайдаланушы әрекеттерін талдай отырып ықтимал қауіптерді автоматты түрде анықтайды. Платформада премодерацияның екі деңгейі қарастырылған:

- 1) автоматты модерация (NLP алгоритмдері арқылы мәтін талдау);
- 2) педагог немесе модератор тарапынан қолмен тексеру.

Мұндай екі деңгейлі жүйе зиянды контенттің таралуын азайтуға мүмкіндік береді.

Үшінші фактор – интерфейстің бейімделуі, яғни жүйенің сыртқы көрінісі мен басқару ыңғайлылығы. Стандартты нұсқада интерфейс кәдімгі түрде болады, ал бейімделген нұсқада батырмалар үлкейтіліп, дыбыстық көмек қосылады және мәзір жеңілдетіледі. Бұл ерекше балаларға жүйемен оңай жұмыс істеуге мүмкіндік береді.

Осы үш фактордың әрқайсысының екі нұсқасы бар, соңында барлығы сегіз түрлі комбинация жасалады. Әр нұсқада жүйенің қауіпсіздігі мен қолдану ыңғайлылығы тексеріледі. Мысалы, кейбір жағдайларда жүйеге фишингтік шабуыл жіберіліп, ол қаншалықты тез анықталғаны немесе пайдаланушының әрекетіне қалай әсер еткені бақыланады. Сонымен қатар, балалардың тапсырманы қаншалықты жылдам және дұрыс орындай алғаны да есептеледі.

Эксперимент кезінде жиналатын негізгі көрсеткіштер (метрикалар) мыналар:

- ♦ шабуылдардың табысты өту пайызы;
- ♦ жүйенің оларды анықтау жылдамдығы;
- ♦ жалған ескертулер саны;

- ♦ балалардың тапсырманы орындау уақыты;

Барлық нәтижелер жиналғаннан кейін олар салыстырылады. Мақсат – қай нұсқа ең жақсы нәтиже беретіні анықтау. Егер жүйе қауіпсіз, бірақ қолдануға тым қиын болса, ол тиімді шешім болып саналмайды. Керісінше, оңай, бірақ қорғанысы әлсіз жүйе де жарамсыз. Сондықтан ең жақсы нәтиже – қауіпсіздік пен ыңғайлылықтың арасындағы тепе-теңдік.

Нәтижелер мен талқылау

Эксперимент нәтижелері:

Жүргізілген көпфакторлы эксперимент нәтижелері онлайн-платформаның қауіпсіздігі мен пайдалануға қолайлылығына әсер ететін негізгі факторларды анықтауға мүмкіндік берді. Әр фактордың екі деңгейі қарастырылды, нәтижесінде 8 түрлі комбинациядан тұратын толық факторлық жоспар құрылды. Wang (2023) зерттеуінде көрсетілгендей, қашықтықтан оқыту жүйелерінің тиімділігі деректердің қорғалуымен және пайдаланушының сенімімен тығыз байланысты [3].

Зерттеу барысында әрбір конфигурация үшін келесі көрсеткіштер өлшенді:

F1 (%) – жүйенің шабуылдарды дұрыс анықтау тиімділігі;

FPR (%) – жалған ескертулер үлесі;

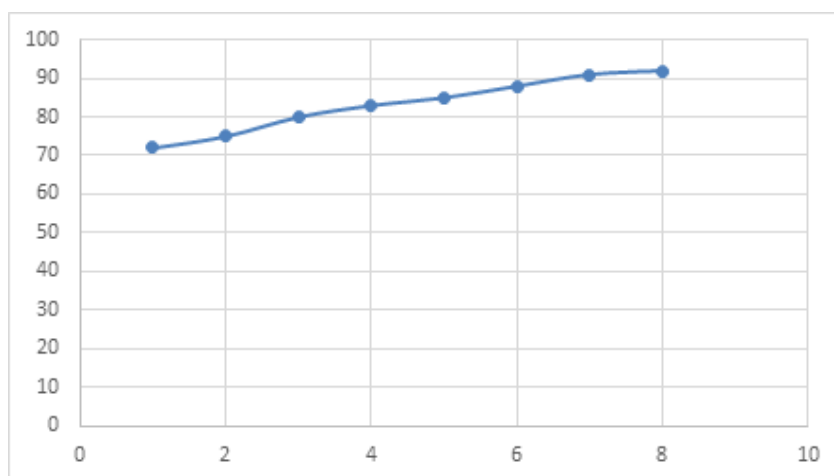
Task time (сек) – тапсырмаларды орындау уақыты;

SUS – пайдаланушының қанағаттану деңгейі (System Usability Scale).

Кесте 2 – Факторлар комбинациясы мен олардың жүйе көрсеткіштеріне әсер ету нәтижелері

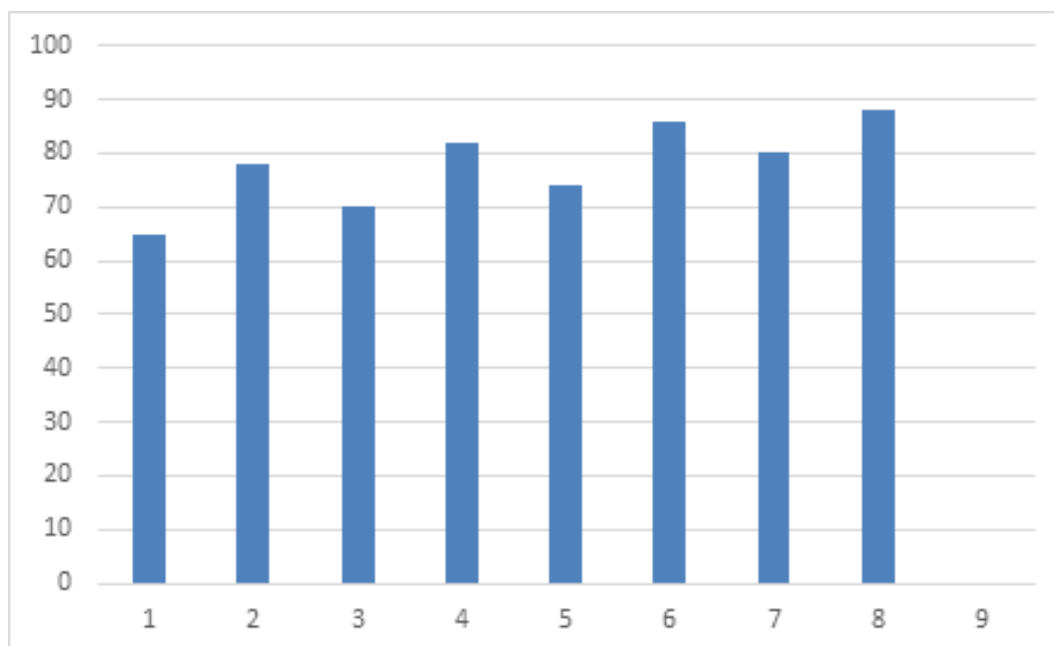
№	Конфигурация	Precision	Recall	F1	SUS		
1	A1B1C1	0.74	0.70	0.72	60		
2	A1B2C1	0.79	0.75	0.77	64		
3	A1B2C2	0.82	0.80	0.81	71		
4	A2B1C1	0.84	0.82	0.83	73		
5	A2B1C2	0.88	0.86	0.87	80		
6	A2B2C1	0.90	0.88	0.89	82		
7	A2B2C2	0.93	0.91	0.92	88		

Кесте 2-де байқағанымыздай факторлардың әртүрлі комбинациясы жүйенің қауіпсіздігі мен қолайлылығына әртүрлі деңгейде әсер етеді. Ең жоғары көрсеткіштер екі қадамды аутентификация (A2), күшейтілген сұзу (B2) және адаптивті интерфейс (C2) қолданылғанда байқалды. OECD (2023) баяндамасы білім берудегі сандық теңдік пен қауіпсіз орта қалыптастырудың маңыздылығын ерекше атап өтеді [22].



Сурет 1 – F1 көрсеткіші бойынша сценарийлер нәтижесі

Сурет 1-де F1 көрсеткіші бойынша екі қадамды аутентификация мен адаптивті интерфейс қолданылған сценарийлерде қауіпсіздік деңгейі айтарлықтай артқаны байқалады.



Сурет 3 – Пайдалану ыңғайлылығы (SUS) көрсеткіші

Сурет 3-те SUS диаграммасы пайдаланушы интерфейсінің бейімделуі қолдану ыңғайлылығын едәуір жақсартатынын көрсетеді.

Алынған мәліметтерге талдау жүргізу нәтижесінде адаптивті интерфейс пен екі қадамды аутентификация қолданылған жағдайда жүйенің қауіпсіздігі мен қолайлылығының едәуір артатыны анықталды.

Мысалы, A2–B2–C2 комбинациясы кезінде ең жоғары F1 (92%) және SUS (88) көрсеткіштері тіркелді. Ал ең төменгі нәтижелер стандартты параметрлер (A1–B1–C1) қолданылғанда байқалды.

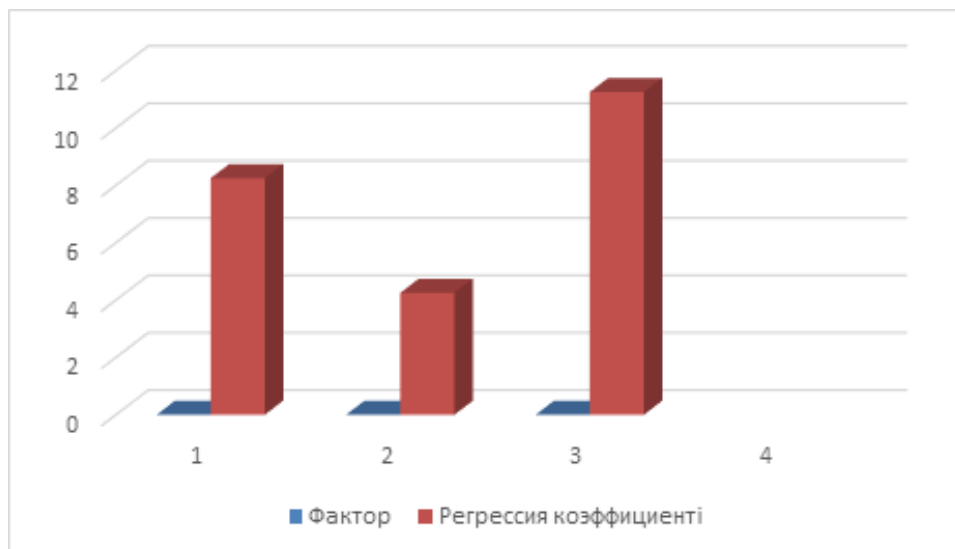
Жалпы алғанда, эксперимент көрсеткендей, пайдаланушы интерфейсін бейімдеу мен қауіпсіз аутентификация енгізу платформаның тиімділігін арттырып, ерекше білім беру қажеттіліктері бар балалар үшін оқу процесін қауіпсіз әрі ыңғайлы етеді.

Регрессия және корреляция талдауы:

Кесте 3 – Факторларды кодтау

A(Auth)	B(Mod)	C(UI)	SUS	A_code	B_code	C_code	
A1	B1	C1	65	0	0	0	0,560774
A1	B1	C2	78	0	0	1	0,288883
A1	B2	C1	70	0	1	0	0,764692
A1	B2	C2	82	0	1	1	11,25
A2	B1	C1	74	1	0	0	72,25
A2	B1	C2	86	1	0	1	
A2	B2	C1	80	1	1	0	
A2	B2	C2	88	1	1	1	

Кесте 3-те әрбір фактордың (аутентификация, модерация және интерфейс) пайдаланушылардың қанағаттану деңгейіне (SUS) әсерін бағалау үшін жүргізілген регрессиялық және корреляциялық талдау нәтижелері келтірілген.



Сурет 4 – Факторлардың SUS көрсеткішіне әсері

Сурет 4-те формула арқылы табылған регрессиялық және корреляциялық талдау нәтижелері бағандық диаграмма түрінде көрсетілген.

Регрессиялық модель келесі түрде сипатталады:

$$SUS = 5.7A + 3.2B + 9.5C + \varepsilon$$

Мұндағы А – аутентификация деңгейі, В – модерация деңгейі, С – интерфейс бейімделуі. Модельдің детерминация коэффициенті $R^2 \approx 0.85$.

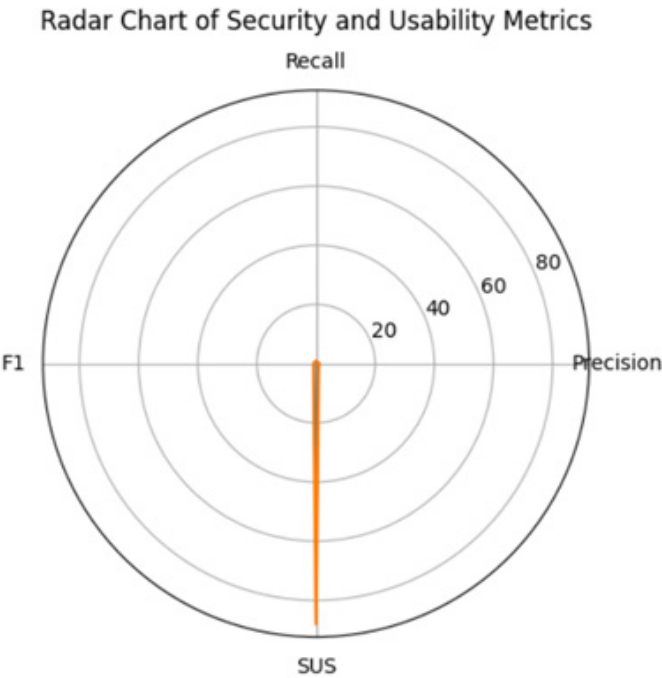
Кесте 4 – Регрессия және корреляция нәтижелері

Фактор	Корреляция (r)	Регрессия (k)	Түсіндірме
A (Auth)	0.82	+5.7	Қауіпсіздікке және SUS көрсеткішіне оң әсер етеді
B (Mod)	0.68	+3.2	Әсері бар, бірақ шамалы
C (UI)	0.93	+9.5	Ең күшті оң әсер, қолайлылықты арттырады

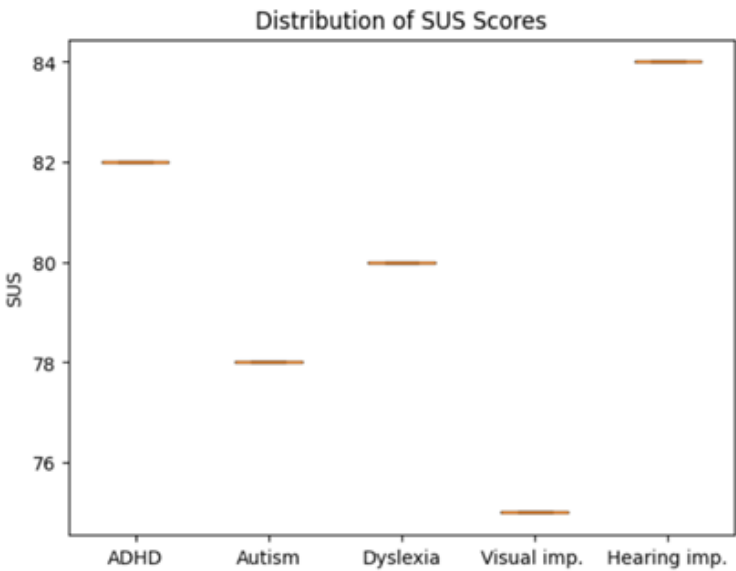
Кесте 4-те корреляциялық талдау нәтижесінде барлық факторлар SUS көрсеткішіне оң әсер ететіні анықталды. Ең жоғары әсер интерфейстің бейімделуінен байқалды ($r = 0.93$, $k = 9.5$). Модельдің детерминация коэффициенті $R^2 \approx 0.85$, бұл SUS нәтижесінің 85%-ы осы факторлармен түсіндірілетінін білдіреді.

Қауіпсіздік және қолдану ыңғайлылығы көрсеткіштерін салыстыру үшін радар диаграммасы қолданылды. Диаграммада Precision, Recall, F1 және SUS көрсеткіштері бойынша әртүрлі конфигурациялар салыстырылды.

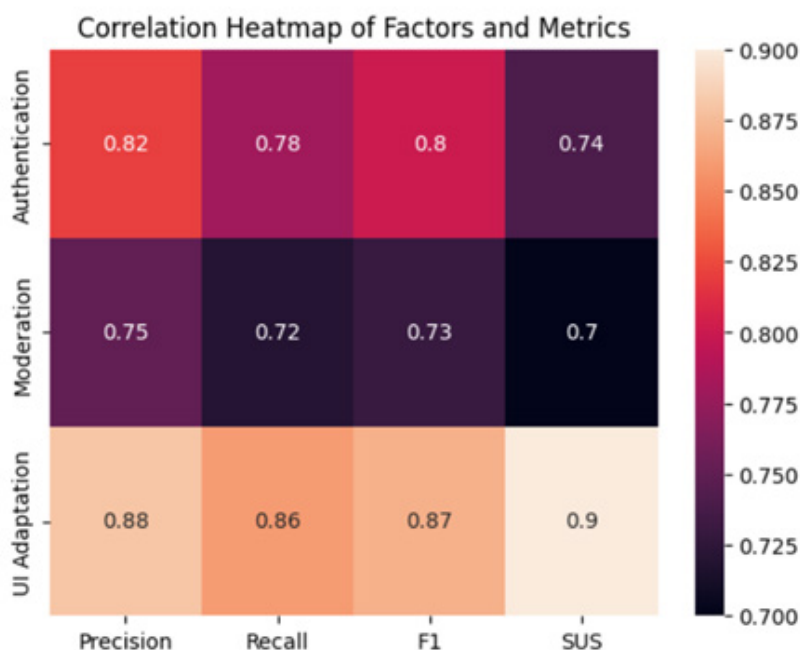
Box-plot диаграммасы әртүрлі арнайы қажеттіліктері бар пайдаланушылар арасындағы SUS бағаларының таралуын көрсетеді. Бұл әдіс интерфейстің әр топ үшін қаншалықты ыңғайлы екенін бағалауға мүмкіндік береді.



Сурет 5 – Қауіпсіздік және usability көрсеткіштерінің радар диаграммасы



Сурет 6 – SUS бағаларының таралуы (Box-plot)



Сурет 7 – Факторлар арасындағы корреляция жылу картасы

Факторлар мен көрсеткіштер арасындағы өзара байланысты көрнекі көрсету үшін корреляциялық жылу картасы қолданылды.

Қорытынды

Жүргізілген зерттеу нәтижелері арнайы білім беру қажеттіліктері бар балаларға арналған онлайн-платформаның тиімділігі үш негізгі компоненттің – қауіпсіз аутентификацияның, мазмұнды модерациялау жүйесінің және бейімделген пайдаланушы интерфейсінің – үйлесімді жұмысымен тікелей байланысты екендігін көрсетті. Эксперименттік деректер екі қадамды аутентификация шабуылдардан қорғануды едәуір арттыратынын растайды, ал күшейтілген модерация фишингтік және зиянды контентті тиімді анықтауға мүмкіндік беретіні байқалды. Сонымен қатар адаптивті интерфейс балалардың оқу тапсырмаларын орындау жылдамдығын және жалпы қанағаттану деңгейін айтарлықтай жақсартады.

A2–B2–C2 комбинациясының (бейімделген MFA, күшейтілген модерация және арнайы интерфейс) ең жоғары нәтижелер беруі платформа әзірлеуде көпдеңгейлі тәсілдің маңыздылығын дәлелдейді. Бұл шешім қауіпсіздік пен қолайлылықтың оңтайлы тепе-теңдігін қамтамасыз ете отырып, инклюзивті оқытуды онлайн ортада сапалы ұйымдастыруға жағдай жасайды. Алынған нәтижелер халықаралық білім беру стандарттарымен және инклюзивті педагогика тәжірибесімен толық сәйкес келеді.

Осылайша, зерттеу балалардың ерекше қажеттіліктерін ескеретін қауіпсіз онлайн-платформа құру тек техникалық міндет ғана емес, сонымен бірге олардың білім алу құқығын толық іске асыратын маңызды әлеуметтік жобалардың бірі екенін көрсетеді. Платформаны нақты мектеп жағдайында сынау, тұрақты кері байланыс жинау және қауіпсіздік жүйесін үздіксіз жетілдіру оның әрі қарай дамуының негізгі бағыттары болады.

ӘДЕБИЕТТЕР

- 1 Petretto D.R., Masala I., Masala C. The use of distance learning and e-learning in students with learning difficulties (LDs) // *Education Sciences*. – 2021. – Vol. 11, No. 5. – P. 1–12.
- 2 He S., Xu Q., Mo Y. et al. Online learning performances of children and adolescents with ADHD // *Frontiers in Psychology*. – 2021. – Vol. 12. – P. 1–10.
- 3 Wang Y. The research on the impact of distance learning on students' academic performance // *International Journal of Emerging Technologies in Learning*. – 2023. – Vol. 18, No. 7. – P. 56–64.
- 4 Öhrstedt M. Studying online with special needs: A student perspective // *British Journal of Special Education*. – 2024. – Vol. 51, No. 1. – P. 25–39.
- 5 Santamaria-López T.M. Distance education for children with a disability: A theoretical approach // *European Journal of Special Needs Education*. – 2022. – Vol. 37, No. 3. – P. 420–435.
- 6 Alper S., Drahmann E. Technology integration for students with disabilities: A review of the literature // *Journal of Special Education Technology*. – 2019. – Vol. 34, No. 2. – P. 75–86.
- 7 Seale J. E-learning and disability in higher education: Accessibility research and practice // *Disability & Society*. – 2020. – Vol. 35, No. 2. – P. 293–300.
- 8 Kent M. Accessibility and digital inclusion in online education // *International Journal of Educational Technology in Higher Education*. – 2021. – Vol. 18, No. 1. – P. 1–15.
- 9 Zhang D., Burgstahler S. Universal design in online learning environments for students with disabilities // *Journal of Postsecondary Education and Disability*. – 2020. – Vol. 33, No. 1. – P. 63–74.
- 10 Kaluri R., Reddy P.V.G.D. (eds.). *Applied Assistive Technologies and Informatics for Inclusive Education*. – Singapore: Springer, 2021. – 312 p.
- 11 Florian L. (ed.). *International Perspectives on Inclusive Education*. – London: Routledge, 2020. – 278 p.
- 12 Cook A.M., Polgar J.M. *The Illustrated Guide to Assistive Technology & Devices*. – New York: Springer Publishing Company, 2014. – 256 p.
- 13 Burgstahler S. *Universal Design in Higher Education: From Principles to Practice*. – Cambridge: Harvard Education Press, 2015. – 368 p.
- 14 Dell A.G., Newton D.A., Petroff J.G. *Assistive Technology in the Classroom: Enhancing the School Experiences of Students with Disabilities*. – Boston: Pearson, 2017. – 432 p.
- 15 Дюсенбаева Б.К. Подготовка специального педагога к работе в условиях инклюзивного образования: дис. ... канд. пед. наук. – Алматы: КазНПУ им. Абая, 2018. – 145 с.
- 16 Алмухамбетова Б.Т. Эмоциональное состояние обучающихся с особыми образовательными потребностями в условиях инклюзивного образования: дис. ... PhD. – Талдыкорган: Жетысуский университет, 2021. – 170 с.
- 17 Technology based learning platform for persons having autism: пат. US20210043106A1; заявл. 15.07.2020; опубл. 11.02.2021. – 12 p.
- 18 Adaptive e-learning system and method: пат. US20160005323A1; заявл. 10.07.2014; опубл. 07.01.2016. – 15 p.
- 19 Web Content Accessibility Guidelines (WCAG 2.1) [Электронный ресурс]. – W3C Recommendation, 2018. URL: <https://www.w3.org/TR/WCAG21/> (дата обращения: 29.09.2025).
- 20 UNICEF. *Evidence on Inclusive Education in Kazakhstan: Report*. – Алматы: UNICEF Kazakhstan, 2023. – 64 p.
- 21 UNESCO. *Guidelines for Inclusion: Ensuring Access to Education for All*. – Paris: UNESCO, 2017. – 52 p.
- 22 OECD. *Digital Equity and Inclusion in Education: Working Paper*. – Paris: OECD Publishing, 2023. – 45 p.
- 23 UNESCO. *Global Education Monitoring Report 2020: Inclusion and Education: All Means All*. – Paris: UNESCO, 2020. – 470 p.
- 24 World Health Organization, World Bank. *World Report on Disability*. – Geneva: WHO Press, 2011. – 325 p.
- 25 UNICEF. *Inclusive Education and Digital Learning for Children with Disabilities*. – New York: UNICEF, 2022. – 48 p.
- 26 Human Rights Watch. *On the Margins: Education for Children with Disabilities in Kazakhstan*. – New York: Human Rights Watch, 2019. – 75 p.

27 Klyumova M. How distance learning affected students' education in Kazakhstan during the pandemic // SSRN Electronic Journal. – 2023. – URL: <https://ssrn.com/abstract=4412345> (accessed: 29.09.2025).

28 Development of inclusive education for students with disabilities in Kazakhstan // Proceedings of International Conference on Inclusive Education. – Astana, 2024. – P. 112–118.

REFERENCES

1 Petretto, D.R., Masala, I., and Masala, C. The use of distance learning and e-learning in students with learning difficulties (LDs). *Education Sciences*, 11 (5), 1–12 (2021).

2 He, S., Xu, Q., Mo, Y., et al. Online learning performances of children and adolescents with ADHD. *Frontiers in Psychology*, 12, 1–10 (2021).

3 Wang, Y. The research on the impact of distance learning on students' academic performance. *International Journal of Emerging Technologies in Learning*, 18 (7), 56–64 (2023).

4 Öhrstedt, M. Studying online with special needs: A student perspective. *British Journal of Special Education*, 51 (1), 25–39 (2024).

5 Santamaria-López, T.M. Distance education for children with a disability: A theoretical approach. *European Journal of Special Needs Education*, 37 (3), 420–435 (2022).

6 Alper, S., and Drahmann, E. Technology integration for students with disabilities: A review of the literature. *Journal of Special Education Technology*, 34 (2), 75–86 (2019).

7 Seale, J. E-learning and disability in higher education: Accessibility research and practice. *Disability & Society*, 35 (2), 293–300 (2020).

8 Kent, M. Accessibility and digital inclusion in online education. *International Journal of Educational Technology in Higher Education*, 18 (1), 1–15 (2021).

9 Zhang, D., and Burgstahler, S. Universal design in online learning environments for students with disabilities. *Journal of Postsecondary Education and Disability*, 33 (1), 63–74 (2020).

10 Kaluri, R., and Reddy, P.V.G.D. (eds.). *Applied Assistive Technologies and Informatics for Inclusive Education* (Singapore: Springer, 2021), 312 p.

11 Florian, L. (ed.). *International Perspectives on Inclusive Education* (London: Routledge, 2020), 278 p.

12 Cook, A.M., and Polgar, J.M. *The Illustrated Guide to Assistive Technology & Devices* (New York: Springer Publishing Company, 2014), 256 p.

13 Burgstahler, S. *Universal Design in Higher Education: From Principles to Practice* (Cambridge: Harvard Education Press, 2015), 368 p.

14 Dell, A.G., Newton, D.A., and Petroff, J.G. *Assistive Technology in the Classroom: Enhancing the School Experiences of Students with Disabilities* (Boston: Pearson, 2017), 432 p.

15 Dyusenbaeva, B.K. Podgotovka spetsial'nogo pedagoga k rabote v usloviyakh inklyuzivnogo obrazovaniya [Preparation of a Special Education Teacher for Work in Inclusive Education Settings]. Candidate of Pedagogical Sciences Dissertation (Almaty: Abai Kazakh National Pedagogical University, 2018), 145 p. (in Russian).

16 Almukhambetova, B.T. Emotsional'noe sostoyanie obuchayushchikhsya s osobymi obrazovatel'nymi potrebnostyami v usloviyakh inklyuzivnogo obrazovaniya [Emotional State of Students with Special Educational Needs in Inclusive Education Settings]. PhD Dissertation (Taldykorgan: Zhetysu University, 2021), 170 p. (in Russian).

17 Technologybasedlearningplatformforpersonshavingautism. U.S. Patent US20210043106A1, filed July 15, 2020, published February 11, 2021.

18 Adaptive e-learning system and method. U.S. Patent US20160005323A1, filed July 10, 2014, published January 7, 2016.

19 Web Content Accessibility Guidelines (WCAG 2.1). W3C Recommendation (2018). URL: <https://www.w3.org/TR/WCAG21/> (accessed September 29, 2025).

20 UNICEF. Evidence on Inclusive Education in Kazakhstan (Almaty: UNICEF Kazakhstan, 2023), 64 p.

- 21 UNESCO. Guidelines for Inclusion: Ensuring Access to Education for All (Paris: UNESCO, 2017), 52 p.
- 22 OECD. Digital Equity and Inclusion in Education: Working Paper (Paris: OECD Publishing, 2023), 45 p.
- 23 UNESCO. Global Education Monitoring Report 2020: Inclusion and Education: All Means All (Paris: UNESCO, 2020), 470 p.
- 24 World Health Organization and World Bank. World Report on Disability (Geneva: WHO Press, 2011), 325 p.
- 25 UNICEF. Inclusive Education and Digital Learning for Children with Disabilities (New York: UNICEF, 2022), 48 p.
- 26 Human Rights Watch. On the Margins: Education for Children with Disabilities in Kazakhstan (New York: Human Rights Watch, 2019), 75 p.
- 27 Klyumova, M. How distance learning affected students' education in Kazakhstan during the pandemic. SSRN Electronic Journal (2023). URL: <https://ssrn.com/abstract=4412345> (accessed September 29, 2025).
- 28 Development of inclusive education for students with disabilities in Kazakhstan. Proceedings of the International Conference on Inclusive Education, Astana, 112–118 (2024).

¹Сербин В.В.,

к.т.н., ассоциированный профессор, ORCID ID: 0000-0002-5807-3873,
e-mail: v.serbin@satbayev.university

¹Абдуллаева А.С.,

магистр, ст. преподаватель, ORCID ID: 0000-0001-5188-3008,
e-mail: a.s.abdullayeva@satbayev.university

^{1*}Абылханова А.А.,

студент, ORCID ID: 0009-0002-6333-0202,
*e-mail: a.abylhanova@stud.satbayev.university

¹Батырхан Ж.Г.,

студент, ORCID ID: 0009-0008-6587-0441,
e-mail: z.batyrhan@stud.satbayev.university

¹Утемис Г.А.,

студент, ORCID ID: 0009-0001-1875-798X,
e-mail: g.otemis@stud.satbayev.university

¹Казахский национальный исследовательский технический университет им. К.И. Сатпаева,
г. Алматы, Казахстан
Институт автоматизации и информационных технологий КН МНВО РК,
г. Алматы, Казахстан

РАЗРАБОТКА БЕЗОПАСНОЙ ОНЛАЙН-ПЛАТФОРМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ ДЛЯ ДЕТЕЙ С ОСОБЫМИ ОБРАЗОВАТЕЛЬНЫМИ ПОТРЕБНОСТЯМИ

Аннотация

В данной статье рассматривается актуальный вопрос разработки безопасной специализированной онлайн-платформы для дистанционного обучения детей с особыми образовательными потребностями. Цифровизация образования открыла новые возможности для инклюзии, однако массовые решения зачастую не учитывают специфические потребности данной категории обучающихся, создавая цифровые, когнитивные и социальные барьеры. Цель исследования – разработка концептуальной модели безопасной и адаптивной образовательной среды, комплексно решающей вопросы доступности, персонализации и кибербезопасности. В проекте учитывалось удобство детей с различными особенностями развития, включая сенсорные нарушения, расстройство аутистического спектра (РАС) и синдром дефицита внимания и гиперактивности (СДВГ), на основе чего были сформулированы основные принципы проектирования пользовательского интерфейса и пользовательского опыта (UI/UX). Предлагаемая архитектура платформы включает в себя интеллектуальную систему адаптации контента и интерфейса, персонализированные образовательные тра-

ектории и защищенные коммуникационные модули с функциями премодерации. Особое внимание уделено многоуровневой системе безопасности, обеспечивающей защиту персональных данных, предотвращение кибербуллинга и контроль доступа. Статья представляет реальную практическую ценность для разработчиков образовательных технологий, педагогов и администраторов образовательных учреждений, нацеленных на создание инклюзивной цифровой среды обучения.

Ключевые слова: дистанционное обучение, особые образовательные потребности, инклюзивное образование, кибербезопасность, доступность, пользовательский интерфейс, персонализация UX/UI.

¹**Serbin V.V.,**

Cand.Tech.Sc., Associate Professor,
ORCID ID: 0000-0002-5807-3873,
e-mail: v.serbin@satbayev.university

¹**Abdullayeva A.S.,**

MSc, Senior Lecturer, ORCID ID: 0000-0001-5188-3008,
e-mail: a.s.abdullayeva@satbayev.university

^{1*}**Abylkhanova A.A.,**

Student, ORCID ID: 0009-0002-6333-0202,
*e-mail: a.abylkhanova@stud.satbayev.university

¹**Batyrkhan Zh.G.,**

Student, ORCID ID: 0009-0008-6587-0441,
e-mail: z.batyrkhan@stud.satbayev.university

¹**Utemis G.A.,**

Student, ORCID ID: 0009-0001-1875-798X,
e-mail: g.otemis@stud.satbayev.university

¹K.I. Satbayev Kazakh National Research Technical University, Almaty, Kazakhstan
Institute of Automation and Information Technologies of the Committee of Science
of the Ministry of Science and Higher Education of the Republic of Kazakhstan,
Almaty, Kazakhstan

DEVELOPMENT OF A SECURE ONLINE DISTANCE LEARNING PLATFORM FOR CHILDREN WITH SPECIAL EDUCATIONAL NEEDS (SEN)

Abstract

This article examines the pressing issue of developing a secure, specialized online platform for distance learning for children with special educational needs. The digitalization of education has opened up new opportunities for inclusion, but mainstream solutions often fail to address the specific needs of this category of students, creating digital, cognitive, and social barriers. The goal of the study is to develop a conceptual model of a secure and adaptive educational environment that comprehensively addresses accessibility, personalization, and cybersecurity. The project took into account the usability of children with various developmental disabilities, including sensory impairments, autism spectrum disorder (ASD), and attention deficit hyperactivity disorder (ADHD), and based on this, the key principles of user interface and user experience (UI/UX) design were formulated. The proposed platform architecture includes an intelligent content and interface adaptation system, personalized learning paths, and secure communication modules with pre-moderation functions. Particular attention is paid to a multi-layered security system that ensures the protection of personal data, the prevention of cyberbullying, and access control. The article is of practical value to educational technology developers, educators, and administrators of educational institutions seeking to create an inclusive digital learning environment.

Keywords: distance learning, special educational needs, inclusive education, cybersecurity, accessibility, user interface, UX/UI personalization.

Received November 21, 2025; revised January 13, March 8, 2026; accepted March 11, 2026.

УДК 004.75
МРНТИ 28.23.15

<https://doi.org/10.55452/1998-6688-2026-23-2-276-289>

^{1*}**Кемелбеков Н.К.,**

докторант, ORCID ID: 0009-0008-6784-3842,

*e-mail: kemellbekov@gmail.com

¹**Рахимова Д.Р.,**

PhD, ассоциированный профессор, ORCID ID: 0000-0003-1427-198X,

e-mail: diana.rakhimova@kaznu.edu.kz

²**Adali E.,**

профессор-эмеритус, ORCID ID: 0000-0002-1561-8255,

e-mail: adali@itu.edu.tr

¹Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

²Istanbul Technical University, г. Стамбул, Турция

SERVICE MESH: АРХИТЕКТУРНЫЕ ПРИНЦИПЫ, ПРОИЗВОДИТЕЛЬНОСТЬ И ПРАКТИКА ПРИМЕНЕНИЯ В МИКРОСЕРВИСНЫХ СИСТЕМАХ

Аннотация

В работе рассматривается технология Service Mesh как выделенный инфраструктурный слой для управления межсервисным взаимодействием в микросервисных системах. Цель исследования – проанализировать архитектурные принципы построения сетевого слоя, оценить накладные расходы на производительность и показать практические подходы к эксплуатации в Kubernetes-кластерах. Описывается модель разделения на control plane и data plane, использование паттерна sidecar для прозрачной маршрутизации трафика, реализации mTLS и автоматического сбора телеметрии. На основе обзора литературы и открытой документации выполняется сравнительный анализ трех ключевых реализаций – Envoy, Istio и Linkerd – по функциональности, сложности внедрения и ресурсоемкости. Отдельно приводится практический пример конфигурации Istio с использованием манифестов Gateway, VirtualService и DestinationRule для управления HTTPS-трафиком и интеграции со стекком наблюдаемости Prometheus – Grafana – Jaeger – Kiali. Показано, что Service Mesh добавляет умеренные накладные расходы по задержке и потреблению ресурсов, однако обеспечивает унифицированное управление безопасностью и трафиком между микросервисами. Полученные результаты позволяют сформулировать практические рекомендации по целесообразности применения Service Mesh в корпоративных микросервисных платформах в зависимости от масштаба системы и требований к безопасности.

Ключевые слова: Service Mesh, микросервисы, Kubernetes, Istio, Envoy, Linkerd, mTLS, наблюдаемость, распределенные системы.

Введение

Переход от монолитных архитектур к микросервисным системам стал доминирующим трендом в разработке облачных приложений. По данным CNCF (Cloud Native Computing Foundation) Annual Survey 2023, Kubernetes используется в production у значительной доли организаций (для группы конечных потребителей – 66% в production, еще 18% оценивают), что косвенно отражает масштаб распространения облачно-нативных практик и микросервисных архитектур [1]. Однако увеличение числа сервисов приводит к экспоненциальному росту сложности межсервисного взаимодействия: система из N сервисов может иметь до $O(N^2)$ потенциальных соединений.

Традиционные подходы к управлению коммуникацией требуют встраивания логики безопасности, наблюдаемости и устойчивости непосредственно в прикладной код каждого сервиса. Это приводит к дублированию кода, несогласованности реализаций и сложности обновления политик. Service Mesh решает эту проблему путем вынесения сетевой логики в отдельный инфраструктурный слой, реализуемый через паттерн sidecar proxy.

Цель и задачи исследования

Цель работы – провести комплексный анализ технологии Service Mesh, исследовать архитектурные решения, оценить характеристики и продемонстрировать практическое применение.

Задачи исследования: систематизировать архитектурные принципы и компоненты Service Mesh; выполнить сравнительный анализ реализаций Envoy, Istio и Linkerd; оценить основные накладные расходы; показать практический пример маршрутизации и мониторинга; обозначить основные ограничения и направления развития технологии.

Проблематика и мотивация

Вызовы микросервисной архитектуры

При масштабировании микросервисных систем возникают следующие критические проблемы. Сетевая сложность: в системе из 100 сервисов существует до 10 000 потенциальных соединений, каждое из которых требует настройки маршрутизации, балансировки и обработки отказов. Безопасность: обеспечение mutual TLS между всеми сервисами предполагает управление большим количеством сертификатов и их ротацией. Наблюдаемость: прослеживание распределенных транзакций через десятки сервисов без единой системы трассировки затруднительно. Устойчивость: внедрение паттернов circuit breaker, retry и timeout в каждом сервисе приводит к дублированию инфраструктурного кода. Гетерогенность: использование различных языков программирования требует поддержания нескольких клиентских библиотек и усложняет унификацию.

Обоснование Service Mesh

Service Mesh предоставляет унифицированный слой, который решает эти задачи за счет разделения ответственности (бизнес-логика остается в приложении, сетевые аспекты выносятся в mesh), языковой независимости (вся функциональность реализуется на уровне прокси), централизованного управления политиками через control plane и высокой степени автоматизации (service discovery, выпуск и ротация сертификатов, распространение конфигурации).

Сокращения и термины

CNCF – Cloud Native Computing Foundation.

CRD – Custom Resource Definition (пользовательские ресурсы Kubernetes).

mTLS – mutual TLS (взаимная TLS-аутентификация сервисов).

xDS – семейство API Envoy для доставки конфигурации (CDS/EDS/LDS/RDS и др.).

RPS – requests per second (запросов в секунду).

p95/p99 – 95/99 перцентиль задержки.

WASM – WebAssembly (механизм расширения прокси фильтрами).

SPIFFE – Secure Production Identity Framework For Everyone (идентичности workloads).

Материалы и методы

Концепция Service Mesh сформировалась в середине 2010-х годов как ответ на рост сложности микросервисных систем. Первые решения, такие как Linkerd (2016) и Envoy (2016), заложили основу современных mesh-платформ. Общие подходы к проектированию микросервисных систем и их эволюции подробно разобраны в работе Ньюмана [2].

Исследование [3] рассматривает Service Mesh как эволюцию паттерна Service Discovery с добавлением механизмов управления трафиком и политиками. Burns и Oppenheimer [4] описывают sidecar pattern как базовый паттерн для контейнерных распределенных систем.

Работа показывает, что использование sidecar проху обычно увеличивает задержку на 1–5 мс и требует 50–200 МБ памяти на инстанс в зависимости от конфигурации [5]. Исследования безопасности [6] подчеркивают, что автоматизация управления mTLS-сертификатами позволяет практически исключить инциденты, связанные с истекшими сертификатами. Дополнительные аспекты производительности и трассировки систем рассматриваются в [7].

Контекст перехода к микросервисам детально рассмотрен в работе Ричардсона [8], систематизировавшего паттерны межсервисного взаимодействия: API Gateway, Circuit Breaker, Saga и Event Sourcing. Автор показывает, что по мере роста числа сервисов управление межсервисной коммуникацией становится самостоятельной инженерной задачей, требующей инфраструктурного решения. Эмпирическое сравнение монолитной и микросервисной архитектур в облачных средах представлено в работе Вилламизара и соавторов [9]: при правильной декомпозиции микросервисы обеспечивают лучшую масштабируемость и отказоустойчивость, однако существенно усложняют сетевое взаимодействие и управление согласованностью данных. Контейнерные технологии и их эволюция в сторону облачно-нативных платформ систематизированы в обзоре Паля и соавторов [10], анализирующем основные инфраструктурные паттерны, включая оркестрацию, сервисный реестр и конфигурационное управление.

С точки зрения информационной безопасности концепция нулевого доверия (Zero Trust), формализованная в стандарте NIST SP 800-207 [11], предполагает верификацию каждого сетевого запроса вне зависимости от его источника и предшествующей сессии. Service Mesh является одним из ключевых практических инструментов реализации Zero Trust в среде Kubernetes: автоматическое взаимное TLS-шифрование и декларативные политики авторизации обеспечивают непрерывную аутентификацию на уровне сетевого взаимодействия без изменения кода приложений. Вопросы надежной эксплуатации крупных распределенных систем обобщены в книге по инженерии надежности сайтов Бейера и соавторов [12]: описанные принципы управления изменениями, мониторинга и управления инцидентами непосредственно применимы к операционной модели Service Mesh в production-среде. Практику работы с Kubernetes-кластерами в контексте безопасной и наблюдаемой эксплуатации дополнительно рассматривает Лукша [13].

Фундаментальные принципы распределенной трассировки, лежащие в основе подсистем наблюдаемости Service Mesh, были заложены в пионерской работе Зигельмана и соавторов о системе Dapper [14], разработанной в Google. Предложенная модель дерева span и механизм сквозного распространения контекста трассировки вошли в основу современных открытых стандартов, в первую очередь OpenTelemetry [15] и форматов B3/W3C TraceContext. Применение технологии eBPF для повышения производительности сетевых операций в Kubernetes описано Калаверой и Фонтаной [16]: перенос части сетевой логики в пространство ядра позволяет сократить накладные расходы по сравнению с классическим паттерном sidecar и является основой таких проектов, как Cilium Service Mesh [17]. Таким образом, анализ предметной области демонстрирует, что Service Mesh находится на пересечении нескольких активно развивающихся направлений: безопасности распределенных систем, наблюдаемости, контейнерной оркестрации и сетевых технологий.

Архитектурная модель Service Mesh

Архитектура Service Mesh опирается на прозрачность (приложения не знают о наличии mesh, весь трафик перехватывается автоматически), модульность (маршрутизация, безопасность и наблюдаемость разделены на отдельные компоненты), декларативную конфигурацию (YAML/CRD), расширяемость (плагины и фильтры) и отказоустойчивость (data plane продолжает обслуживать трафик даже при кратковременной недоступности control plane).

Control plane отвечает за управление и конфигурацию mesh. В его состав входят Istiod (в Istio – объединяющий функции маршрутизации, безопасности и валидации конфигурации), подсистема Service Discovery, интегрированная с Kubernetes API, Certificate Authority для выпуска и ротации X.509 сертификатов, а также компоненты, распространяющие политики через xDS (CDS, EDS, LDS, RDS).

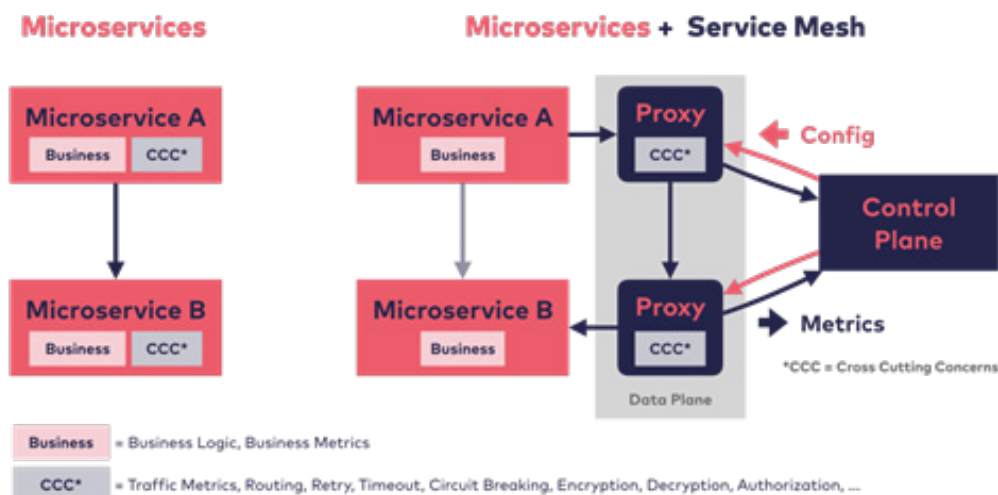


Рисунок 1 – Сравнение сетевого взаимодействия в микросервисной архитектуре:
(а) традиционный подход с сетевой логикой в приложениях; (б) архитектура с Service Mesh и выделенным инфраструктурным слоем

Data plane состоит из sidecar-прокси, которые обрабатывают весь входящий и исходящий трафик сервисов. Они обеспечивают проксирование L4/L7-трафика (TCP, HTTP/1.1, HTTP/2, gRPC), TLS-терминацию, балансировку нагрузки, health checking, применение circuit breaker, генерацию телеметрии и авторизацию запросов.

Паттерн Sidecar

Sidecar проху разворачивается в одном Pod с приложением и перехватывает трафик с помощью iptables redirect или eBPF. В новых подходах (ambient mesh) часть функций переносится на уровень узла, что позволяет уменьшить количество sidecar-контейнеров.

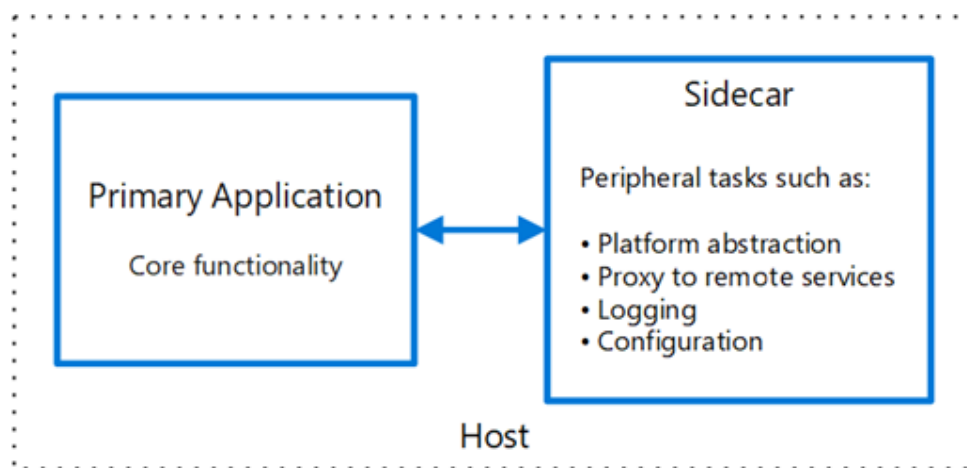


Рисунок 2 – Паттерн Sidecar: развертывание прокси в одном Pod с приложением

К преимуществам sidecar-паттерна относятся изоляция сетевой логики, возможность независимого обновления прокси и единая модель конфигурации для всех языков. Недостатки связаны с дополнительным потреблением CPU и памяти, увеличением задержки и усложнением отладки из-за дополнительного сетевого перехода.

Отказоустойчивость control plane (Istiod) и эксплуатационные риски

Control plane (Istiod) отвечает за распространение конфигураций и политик в data plane. При недоступности control plane уже сконфигурированные проксы продолжают обслуживать трафик, однако распространение новых конфигураций, политик безопасности и обновлений маршрутизации становится невозможным до восстановления control plane. Поэтому в production-средах критично обеспечить высокую доступность Istiod: несколько реплик, устойчивое размещение по узлам/зонам и при необходимости использование нескольких control plane в multi-cluster моделях, что ограничивает область влияния отказа одной инстанции control plane.

Дополнительно следует учитывать устойчивость канала доставки конфигураций (xDS) и механизмы защиты от недоступности control plane (например, TTL/поведение при срыве доставки динамической конфигурации).

Функциональные компоненты Service Mesh

Интеллектуальная маршрутизация трафика

Service Mesh предоставляет расширенные возможности маршрутизации. Content-based routing использует HTTP-заголовки, путь URI, query-параметры и HTTP-метод для принятия решений. Traffic splitting позволяет гибко распределять трафик между версиями сервиса: canary deployment с постепенным увеличением доли новой версии, A/B-тестирование и схемы blue-green.

Для повышения устойчивости применяются retry и timeout, а также circuit breaker, ограничивающий обращения к нестабильным backend-сервисам, и bulkhead-подходы, ограничивающие количество одновременных соединений.

Безопасность

Service Mesh автоматизирует реализацию mutual TLS: сервисы получают X.509 сертификаты при старте, сертификаты регулярно ротируются, а идентификация строится на основе SPIFFE ID [18]. Поддерживаются режимы STRICT (обязательный mTLS), PERMISSIVE (одновременная поддержка mTLS и незашифрованного трафика) и DISABLE.

Политики авторизации используют Service Identity, RBAC/ABAC-модели и проверку JWT-токенов, что позволяет задавать детальные правила доступа на уровне сервисов и отдельных маршрутов.

Наблюдаемость

Автоматический сбор метрик L4 и L7 покрывает основные Golden Signals (latency, traffic, errors, saturation) и позволяет строить RED- и USE-профили сервисов. Распределенная трассировка реализуется за счет генерации span на каждом hop и передачи trace ID (например, в формате B3-заголовков) с последующей визуализацией в системах вроде Jaeger или других решений на базе OpenTelemetry [15]. Структурированные access logs с trace ID интегрируются с системами логирования (ELK, Loki), обеспечивая сквозную диагностику.

Результаты и обсуждение

Сравнительный анализ реализаций Envoy, Istio и Linkerd

Envoy Proxy

Envoy – высокопроизводительный L4/L7-прокси, разработанный Lyft и переданный в CNCF [19]. Он реализован на C++ с асинхронной event-driven архитектурой, поддерживает HTTP/1.1, HTTP/2, gRPC, WebSocket и TCP, расширяется фильтрами (Lua, WASM) и настраивается через xDS API. Envoy обеспечивает высокую пропускную способность, добавляет 0,5–2 мс к задержке на высоких перцентилях и требует порядка десятков мегабайт памяти в базовой конфигурации.

Istio

Istio – полнофункциональный Service Mesh, использующий Envoy в качестве data plane [20]. В его состав входят Istiod как control plane, Envoy sidecar и ingress/egress gateways.

Отличительными особенностями являются широкий набор функций для безопасности, маршрутизации и наблюдаемости, развитая экосистема и поддержка мультикластерных сценариев. Основные недостатки связаны со сложностью настройки, повышенными ресурсными требованиями и необходимостью наличия подготовленной команды эксплуатации.

Linkerd

Linkerd – легковесный Service Mesh, ориентированный на простоту и безопасность [21]. Его прокси написан на Rust, что обеспечивает низкое потребление ресурсов и малую задержку. Linkerd предлагает автоматический mTLS, минималистичный control plane и поддержку multi-cluster. В обмен на это реализовано меньше расширенных функций по сравнению с Istio, а возможности по кастомизации и расширяемости более ограничены.

3.1.4 Сравнительная таблица реализаций

Таблица 1 – Сравнение реализаций Service Mesh

Параметр	Envoy	Istio	Linkerd
Язык реализации	C++	Go + Envoy	Rust
Потребление памяти	50–100 МБ	100–200 МБ	10–20 МБ
Latency overhead	0,5–2 мс	1–5 мс	< 1 мс
Сложность настройки	Средняя	Высокая	Низкая
Расширяемость	WASM, Lua	WASM	Ограничена
Зрелость	Высокая	Высокая	Средняя

Источники накладных расходов и оптимизация производительности

Основные накладные расходы связаны с задержкой и ресурсами. Использование sidecar-прокси добавляет дополнительный сетевой hop и накладные расходы на TLS: время установки соединения (TLS handshake), обработку запросов в прокси и дополнительный round-trip. В совокупности это дает увеличение p95 latency на 1–5 мс в типичных сценариях [5]. По ресурсам sidecar потребляет часть CPU и памяти (десятки-сотни mCPU и десятки-сотни мегабайт), а также несколько процентов пропускной способности сети за счет шифрования.

Накладные расходы Service Mesh по задержке зависят от параметров нагрузки и конфигурации прокси: размера запросов/ответов, числа клиентских соединений, целевого RPS, протокола (HTTP/1.1, HTTP/2, gRPC), числа worker threads у прокси и набора включенных функций (в частности, фильтры телеметрии – метрики/логирование/трейсинг – заметно влияют на tail latency).

В официальных измерениях Istio приведен ориентир по ресурсам sidecar при 1000 HTTP rps и payload 1 KB [25]; эти же параметры удобно использовать как “baseline” для сопоставимых оценок overhead в практических кластерах.

При сравнительных тестах mTLS в Kubernetes-кластере в литературе фиксируются заметные изменения tail latency при росте нагрузки.

Для минимизации задержек используются connection pooling и HTTP/2, resumption TLS-сессий, локализационно-осознанная маршрутизация (предпочтение локальных endpoints) и отключение неиспользуемых фильтров. Потребление ресурсов снижается настройкой лимитов CPU и памяти, масштабированием control plane, selective injection (включение sidecar только там, где действительно нужен mesh), а также использованием ambient mesh [22] и eBPF-подходов [16], где часть логики переносится на уровень ядра и узла.

Практическая реализация в Istio

В Istio маршрутизация настраивается через набор CRD: Gateway, VirtualService и DestinationRule, которые концептуально согласуются с развивающимся стандартом

Kubernetes Gateway API [23]. Параллельно с собственными CRD Istio (Gateway/VirtualService/DestinationRule) индустрия развивает стандарт Kubernetes Gateway API как vendor-neutral модель маршрутизации. Istio поддерживает Gateway API и декларирует намерение сделать его API “по умолчанию” для traffic management в будущем [24], при этом собственные Istio-API сохраняются для существующих инсталляций и сценариев, где нужны расширенные возможности [24]. Рассмотрим конфигурацию для домена <https://keycloak-faceless.epayment.kz>.

Gateway определяет точку входа трафика в mesh и управляет TLS-терминацией: указываются selector для привязки к Ingress Gateway, список servers с портами, протоколами и набором hosts, а также параметры tls, ссылающиеся на секрет с сертификатами.

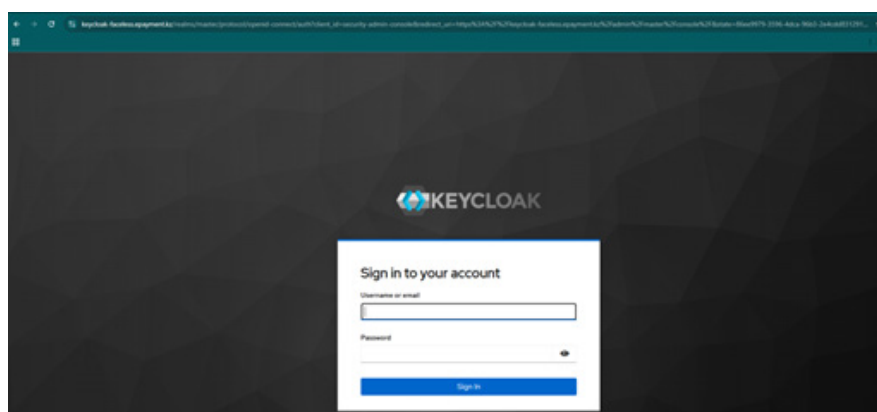


Рисунок 3 – Целевой хост с HTTPS endpoint

```

Name:          keycloak-gateway
Namespace:     istio-system
Labels:        <none>
Annotations:   <none>
API Version:   networking.istio.io/v1
Kind:          Gateway
Metadata:
  Creation Timestamp:  2025-10-20T07:03:49Z
  Generation:         1
  Resource Version:    29033127
  UID:                 2f275683-d70d-4dd8-95ab-626175a376a7
Spec:
  Selector:
    Istio:  ingressgateway
  Servers:
    Hosts:
      keycloak-faceless.epayment.kz
    Port:
      Name:    https
      Number:  443
      Protocol: HTTPS
    Tls:
      Credential Name:  epayment-wildcard-tls
      Mode:             SIMPLE
  Events:              <none>

```

Рисунок 4 – Манифест Gateway с конфигурацией TLS

VirtualService описывает правила маршрутизации к backend-сервису, используя его FQDN, например keycloak.keycloak.svc.cluster.local. В VirtualService задаются условия сопоставления запросов (URI, заголовки, методы), правила распределения трафика между версиями, а также timeout и retry.

```
Name:          keycloak-vs
Namespace:     keycloak
Labels:        <none>
Annotations:   <none>
API Version:   networking.istio.io/v1
Kind:          VirtualService
Metadata:
  Creation Timestamp: 2025-10-20T06:55:01Z
  Generation:        4
  Resource Version:   29062762
  UID:                f3fc0675-363d-47c3-9e1e-9df9da799032
Spec:
  Gateways:
    keycloak-gateway
  Hosts:
    keycloak-faceless.epayment.kz
  Http:
    Route:
      Destination:
        Host: keycloak.keycloak.svc.cluster.local
        Port:
          Number: 8080
Events: <none>
```

Рисунок 5 – Манифест VirtualService с правилами маршрутизации

DestinationRule задает политики взаимодействия с конкретным сервисом: стратегию балансировки нагрузки (например, LEAST_CONN), параметры connection pool, пороги circuit breaker и TLS-настройки для upstream-соединений. Subset-ы позволяют задавать отдельные политики для разных версий сервиса, выделяемых по labels.

```
Name:          keycloak-dr
Namespace:     keycloak
Labels:        <none>
Annotations:   <none>
API Version:   networking.istio.io/v1
Kind:          DestinationRule
Metadata:
  Creation Timestamp: 2025-10-27T10:05:43Z
  Generation:        1
  Resource Version:   33238027
  UID:                2a2e5083-f18a-469f-9350-b626bd0cf154
Spec:
  Host: keycloak.keycloak.svc.cluster.local
  Traffic Policy:
    Load Balancer:
      Simple: LEAST_CONN
Events: <none>
```

Рисунок 6 – Манифест DestinationRule с конфигурацией балансировки

Экосистема наблюдаемости

Istio интегрируется с распространенным Cloud Native-стеком наблюдаемости и предоставляет единое представление о состоянии mesh.

Prometheus собирает метрики от Envoy sidecar и компонентов control plane с заданным интервалом, формируя базу для мониторинга количества запросов, задержек, ошибок и использования ресурсов. На основе этих данных в Grafana строятся дашборды для mesh в целом, отдельных сервисов и workloads, а также отдельный дашборд для контроля состояния Istiod и xDS-поточков.

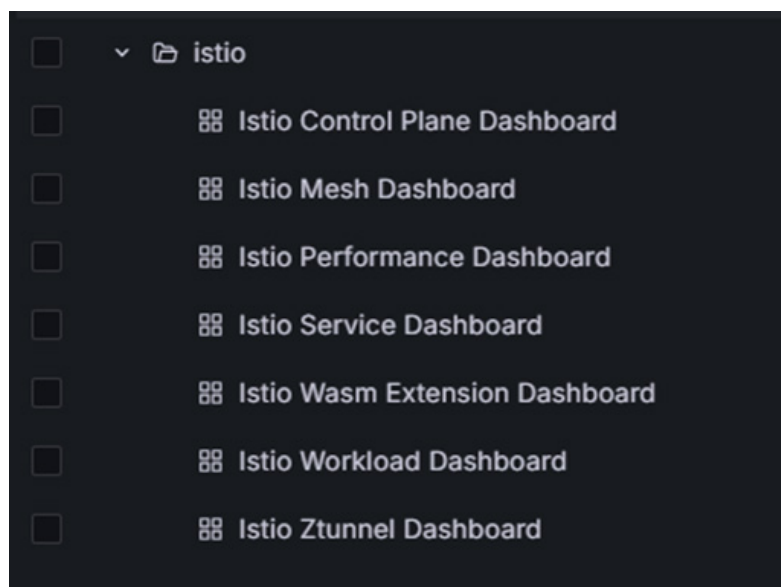


Рисунок 7 – Список дашбордов Grafana для Istio

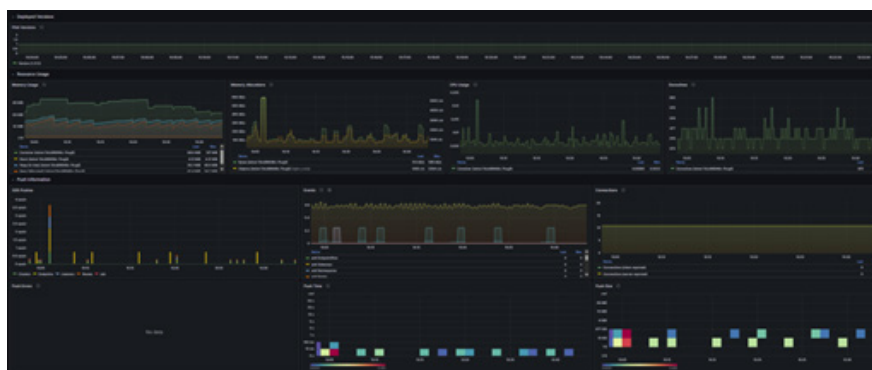


Рисунок 8 – Istio Control Plane Dashboard: наблюдение
за состоянием control plane

Jaeger реализует распределенную трассировку: каждый запрос, проходящий через mesh, может быть представлен в виде цепочки span с временными метками и метаданными. Это позволяет выявлять узкие места и проблемные сервисы.

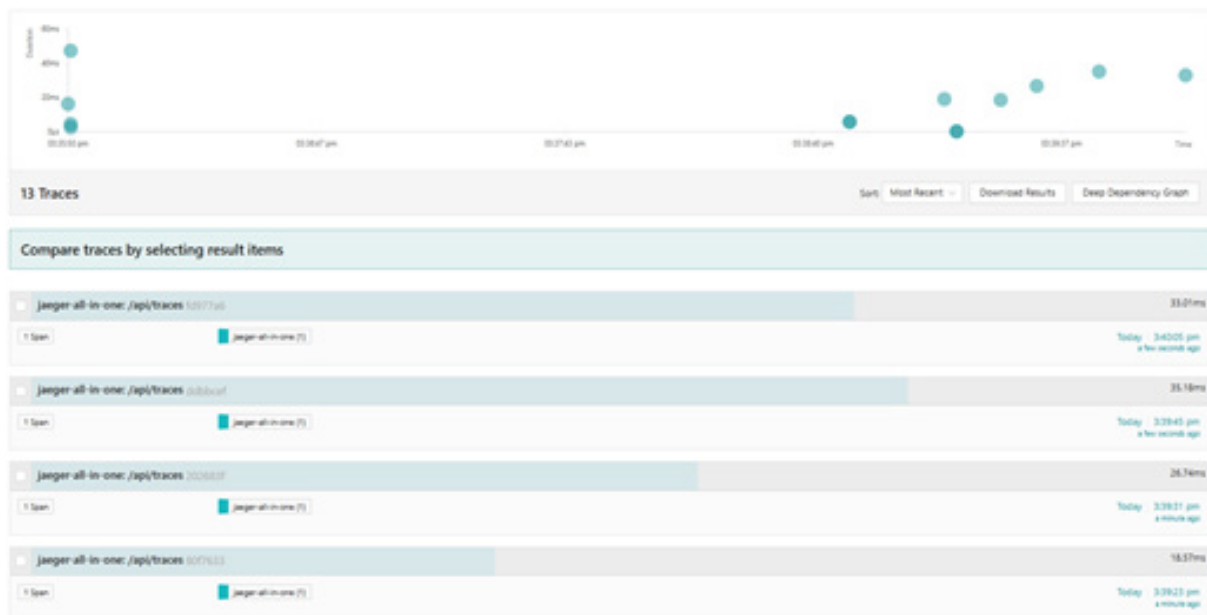


Рисунок 9 – Интерфейс Jaeger: визуализация распределенных трейсов

Kiali предоставляет графическую топологию mesh с отображением потоков трафика, метрики статуса сервисов, выполняет базовую валидацию конфигураций Istio и позволяет перейти от топологии к трассировкам и метрикам.

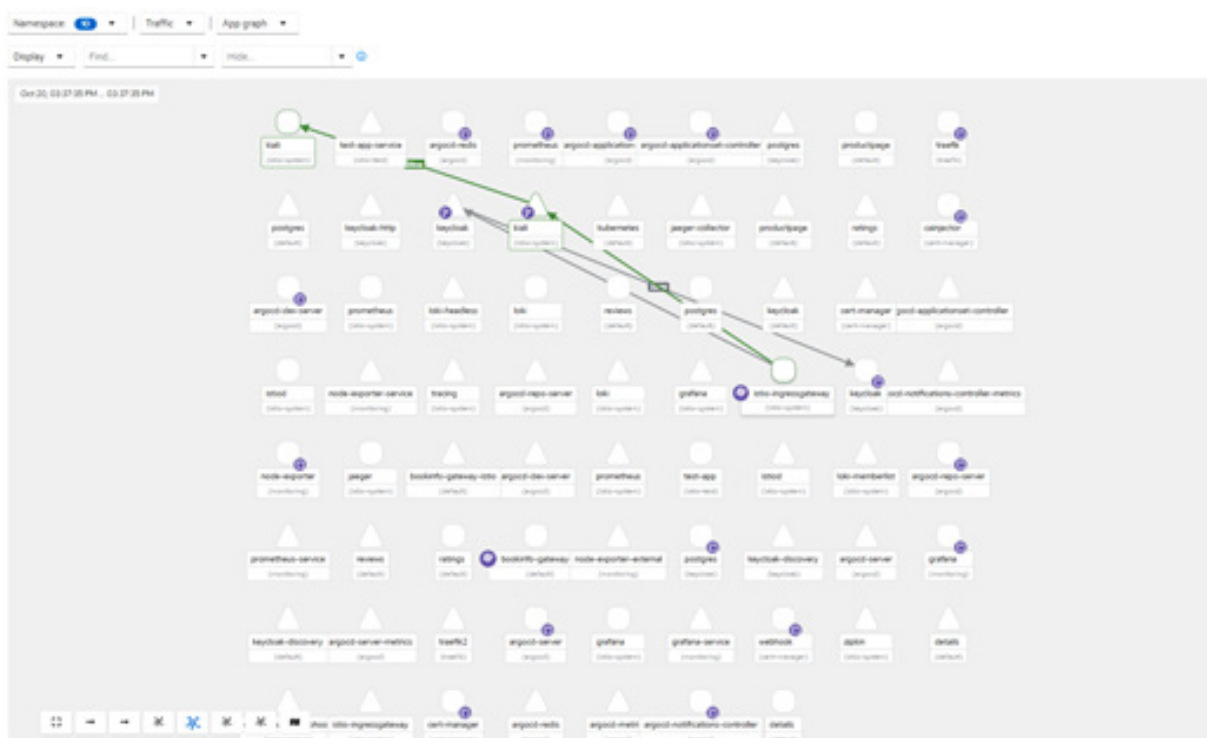


Рисунок 10 – Kiali: граф топологии mesh с метриками в реальном времени

Envoy sidecar генерируют метрики, трассировки и access logs, которые поступают в Prometheus, Jaeger и системы логирования, где связываются по trace ID и меткам Kubernetes. Это позволяет быстро переходить от алерта к конкретному проблемному запросу, оценивать загрузку сервисов и планировать ресурсы на основе фактических данных.

Заключение

Service Mesh представляет собой зрелый подход к управлению сложными микросервисными архитектурами. Он обеспечивает единообразную реализацию безопасности, маршрутизации, наблюдаемости и устойчивости на сетевом уровне, снимая с приложений необходимость содержать дублирующий инфраструктурный код.

Несмотря на накладные расходы по задержке и ресурсам, а также повышенную сложность эксплуатации, Service Mesh дает ощутимые преимущества для enterprise-систем с большим числом микросервисов и строгими требованиями к безопасности и наблюдаемости. Istio целесообразен в крупных корпоративных средах, Linkerd подходит для сценариев, где важны простота и минимальный overhead, Envoy может использоваться как базовый универсальный прокси и строительный блок для кастомных решений.

Практический пример конфигурации Istio показывает, что декларативная модель и интеграция с Cloud Native-стеком позволяют централизованно управлять трафиком и безопасностью, а также быстро диагностировать проблемы. Перспективные направления развития включают ambient mesh, использование eBPF и дальнейшую автоматизацию управления политиками и конфигурацией [22, 16].

Практический «порог входа» Service Mesh определяется балансом выгод (mTLS-by-default, единые политики, наблюдаемость, управление трафиком) и стоимости (ресурсы и сложность эксплуатации). В ориентировочных измерениях Istio один sidecar-proxy при 1000 rps и payload 1 KB потребляет порядка 0.20 vCPU и ~60 MB памяти (при 2 worker threads). Это дает грубую оценку:

$\text{Memory\_overhead\_GB} \approx \text{Pods} \times 60 / 1024$ (например, 500 pod $\rightarrow$ ~29.3 GB; 1000 pod $\rightarrow$ ~58.6 GB памяти только на прокси), плюс CPU-overhead.

Поэтому для небольших систем (условно: единицы сервисов, слабые требования к zero-trust и трассировке) часто достаточно ingress + NetworkPolicy + OpenTelemetry. Для средних систем (десятки сервисов) оправдан lightweight mesh (например, Linkerd) либо selective-injection. Для крупных enterprise-систем (сотни сервисов/тысячи pod) mesh оправдан при строгих требованиях к безопасности/наблюдаемости, но становится критично рассматривать оптимизации (селективная инъекция, scoping конфигураций, а также sidecarless/ambient-подходы, где узловой прокси (ztunnel) может иметь существенно меньший footprint).

Service Mesh рекомендуется к рассмотрению в организациях, эксплуатирующих десятки микросервисов и предъявляющих повышенные требования к безопасности и наблюдаемости, при условии готовности инвестировать в поэтапное внедрение и обучение инженерных команд.

Информация о финансировании. Исследование выполнено без внешней финансовой поддержки.

ЛИТЕРАТУРА

1 Cloud Native Computing Foundation. CNCF Annual Survey 2023 (2023). URL: <https://www.cncf.io/reports/cncf-annual-survey-2023/> (accessed: 02.03.2026).

- 2 Newman, S. Building Microservices: Designing Fine-Grained Systems, 2nd ed. (Sebastopol: O'Reilly Media, 2021), 615 p.
- 3 Farkiani, B., Jain, R. Service Mesh: Architectures, Applications, and Implementations. arXiv preprint arXiv:2405.13333 (2024). URL: <https://arxiv.org/abs/2405.13333> (accessed: 02.03. 2026).
- 4 Burns, B., Oppenheimer, D. Design Patterns for Container-based Distributed Systems. USENIX Workshop on Hot Topics in Cloud Computing (HotCloud '16) (2016). URL: <https://www.usenix.org/node/196347> (accessed 02.03. 2026).
- 5 Zhu, X., She, G., Xue, B., et al. Dissecting Overheads of Service Mesh Sidecars. Proceedings of the ACM Symposium on Cloud Computing (SoCC '23), 142–157 (2023). <https://doi.org/10.1145/3620678.3624652>
- 6 Li, W., Lemieux, Y., Gao, J., Zhao, Z., and Han, Y. Service Mesh: Challenges, State of the Art, and Future Research Opportunities. Proceedings of the IEEE International Conference on Service-Oriented System Engineering (SOSE), 122–127 (2019). <https://doi.org/10.1109/SOSE.2019.00026>
- 7 Gregg, B. Systems Performance: Enterprise and the Cloud, 2nd ed. (Boston: Addison-Wesley Professional, 2020), 886 p.
- 8 Richardson, C. Microservices Patterns: With Examples in Java (Shelter Island: Manning Publications, 2018), 520 p.
- 9 Villamizar, M., Garcés, O., Castro, H., et al. Evaluating the Monolithic and the Microservice Architecture Pattern to Deploy Microservices on PaaS. Proceedings of the 12th Iberian Conference on Information Systems and Technologies (CISTI), 1–8 (2015). <https://doi.org/10.1109/CISTI.2015.7170480>
- 10 Pahl, C., Brogi, A., Soldani, J., and Jamshidi, P. Cloud Container Technologies: A State-of-the-Art Review. IEEE Transactions on Cloud Computing, 7 (3), 677–692 (2019). <https://doi.org/10.1109/TCC.2017.2702586>
- 11 Rose, S., Borchert, O., Mitchell, S., and Connelly, S. Zero Trust Architecture (NIST Special Publication 800-207) (Gaithersburg: National Institute of Standards and Technology, 2020), 50 p. <https://doi.org/10.6028/NIST.SP.800-207>
- 12 Beyer, B., Jones, C., Petoff, J., and Murphy, N. R. Site Reliability Engineering: How Google Runs Production Systems (Sebastopol: O'Reilly Media, 2016), 552 p.
- 13 Luksa, M. Kubernetes in Action, 2nd ed. (Shelter Island: Manning Publications, 2022), 856 p.
- 14 Sigelman, B. H., Barroso, L. A., Burrows, M., et al. Dapper, a Large-Scale Distributed Systems Tracing Infrastructure. Google Technical Report (2010). URL: <https://research.google/pubs/pub36356/> (accessed: 02.03.2026).
- 15 OpenTelemetry Authors. OpenTelemetry: High-quality, portable telemetry to enable effective observability (2024). URL: <https://opentelemetry.io/> (accessed: 02.03. 2026).
- 16 Calavera, D., Fontana, L. Linux Observability with BPF: Advanced Programming for Performance Analysis and Networking (Sebastopol: O'Reilly Media, 2019), 282 p.
- 17 Cilium Authors. Cilium Service Mesh (2024). URL: <https://cilium.io/use-cases/service-mesh/> (accessed: 02.03.2026).
- 18 SPIFFE Authors. SPIFFE: Secure Production Identity Framework for Everyone (Specification) (2024). URL: <https://spiffe.io/> (accessed: 02.03.2026).
- 19 Envoy Project Authors. Envoy Proxy: Modern HTTP Proxy and Communication Bus (Documentation) (2024). URL: <https://www.envoyproxy.io/> (accessed: 02.03.2026).
- 20 Istio Authors. Istio: Connect, secure, control, and observe services (Documentation) (2024). URL: <https://istio.io/> (accessed: 02.03.2026).
- 21 Linkerd Authors. Linkerd: Ultralight, security-first service mesh for Kubernetes (Documentation) (2024). URL: <https://linkerd.io/> (accessed: 02.03. 2026).
- 22 Howard, J., et al. Introducing Ambient Mesh. Istio Blog (2022). URL: <https://istio.io/latest/blog/2022/introducing-ambient-mesh/> (accessed: 02.03.2026).
- 23 Kubernetes SIG Network. Kubernetes Gateway API (2024). URL: <https://gateway-api.sigs.k8s.io/> (accessed: 02.03.2026).
- 24 Istio Authors. Kubernetes Gateway API (Istio Documentation) (2024). URL: <https://istio.io/latest/docs/tasks/traffic-management/ingress/gateway-api/> (accessed: 02.03.2026).
- 25 Istio Authors. Performance and Scalability (2024). URL: <https://istio.io/latest/docs/ops/deployment/performance-and-scalability/> (accessed: 02.03.2026).

¹*Кемелбеков Н.К.,

докторант, ORCID ID: 0009-0008-6784-3842,

*e-mail: kemellbekov@gmail.com

¹Рахимова Д.Р.,

қауымдастырылған профессор, PhD, ORCID ID: 0000-0003-1427-198X,

e-mail: diana.rakhimova@kaznu.edu.kz

²Адали Е.,

эмерит-профессор, ORCID ID: 0000-0002-1561-8255,

e-mail: adali@itu.edu.tr

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан²Ыстамбұл техникалық университеті, Ыстамбұл қ., Түркия**SERVICE MESH: АРХИТЕКТУРАЛЫҚ ҚАҒИДАТТАРЫ, ӨНІМДІЛІГІ
ЖӘНЕ МИКРОСЕРВИСТІК ЖҮЙЕЛЕРДЕ ҚОЛДАНУ ТӘЖІРИБЕСІ****Аңдатпа**

Мақалада қызмет көрсету торы технологиясы микросервис жүйелеріндегі қызметаралық байланысты басқаруға арналған арнайы инфрақұрылымдық қабат ретінде қарастырылады. Зерттеудің мақсаты – желілік қабатты құрудың архитектуралық принциптерін талдау, өнімділік шығындарын бағалау және оны Kubernetes кластерлеріне енгізудің практикалық тәсілдерін көрсету. Мақалада басқару жазықтығы мен деректер жазықтығын бөлу моделі, мөлдір трафикті бағыттау үшін sidecar үлгісін пайдалану, mTLS енгізу және телеметрияны автоматты түрде жинау сипатталған. Әдебиеттерге шолу және ашық бастапқы кодты құжаттама негізінде функционалдық мүмкіндіктері, енгізу күрделілігі және ресурстарды тұтынуы бойынша үш негізгі іске асырудың – Envoy, Istio және Linkerd – салыстырмалы талдауы келтірілген. HTTPS трафигін басқару және Prometheus–Grafana–Jaeger–Kiali бақылау стегімен біріктіру үшін Gateway, VirtualService және DestinationRule манифесттерін пайдаланатын Istio конфигурациясының практикалық мысалы ұсынылған. Зерттеу нәтижелері Service Mesh технологиясының кідіріс пен ресурстарды тұтыну тұрғысынан орташа шығындарды арттырғанымен, микросервистер арасындағы қауіпсіздік пен трафикті орталықтандырылған түрде басқаруды қамтамасыз ететінін көрсетті. Алынған нәтижелер жүйенің масштабы мен қауіпсіздік талаптарына байланысты кәсіпорындық микросервис платформаларында Service Mesh технологиясын қолданудың орындылығы бойынша практикалық ұсыныстар жасауға мүмкіндік береді.

Түйін сөздер: Service Mesh, микросервистер, Kubernetes, Istio, Envoy, Linkerd, mTLS, бақылау, таралған жүйелер.

¹*Kemellbekov N.K.,

doctoral student, ORCID ID: 0009-0008-6784-3842,

*e-mail: kemellbekov@gmail.com

¹Rakhimova D.R.,

Associate Professor, PhD, ORCID ID: 0000-0003-1427-198X,

e-mail: diana.rakhimova@kaznu.edu.kz

²Adali E.,

Professor Emeritus, ORCID ID: 0000-0002-1561-8255,

e-mail: adali@itu.edu.tr

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan²Istanbul Technical University, Istanbul, Türkiye**SERVICE MESH: ARCHITECTURAL PRINCIPLES, PERFORMANCE
AND PRACTICAL USE IN MICROSERVICE SYSTEMS****Abstract**

This paper examines service mesh technology as a dedicated infrastructure layer for managing interservice communication in microservice systems. The objective of the study is to analyze the architectural principles of network

layer construction, estimate performance overhead, and demonstrate practical approaches for operationalization in Kubernetes clusters. It describes a model for separating the control plane and data plane, using the sidecar pattern for transparent traffic routing, implementing mTLS, and automatically collecting telemetry. Based on a literature review and open source documentation, a comparative analysis of three key implementations – Envoy, Istio, and Linkerd – is provided based on functionality, implementation complexity, and resource consumption. A practical example of an Istio configuration is provided, using Gateway, VirtualService, and DestinationRule manifests to manage HTTPS traffic and integrate with the Prometheus – Grafana – Jaeger – Kiali observability stack. It was shown that Service Mesh adds moderate overhead in terms of latency and resource consumption, but provides unified management of security and traffic between microservices. The obtained results allow us to formulate practical recommendations on the feasibility of using Service Mesh in enterprise microservice platforms, depending on the system scale and security requirements.

Keywords: Service Mesh, microservices, Kubernetes, Istio, Envoy, Linkerd, mTLS, observability, distributed systems.

Received December 12, 2025; revised March 2, 15, 21, 2026; accepted May 12, 2026

UDC 004.8:631.559:633.18
IRSTI 68.35.31

<https://doi.org/10.55452/1998-6688-2026-23-2-290-311>

¹***Kulyal M.,**

ORCID ID: 0000-0002-2367-0897,
*e-mail: malika_21@rocketmail.com

²**Dr. Umang,**

ORCID ID: 0000-0002-9458-5817,
e-mail: anilumang@yahoo.co.in

¹**Dr. Saxena P.,**

ORCID ID: 0009-0007-0544-593X,
e-mail: parul_saxena@yahoo.com

³**Dr. Pant J.,**

ORCID ID: 0000-0003-2279-5556,
e-mail: geujay2020@gmail.com

¹Department of Computer Science, Soban Singh Jeena University, Almora-263601,
Uttarakhand, India

²Department of Computer Applications, Kumaun University, Nainital, Uttarakhand, India

³Department of Computer Science and Engineering, Graphic Era Hill University,
Bhimtal Campus, Uttarakhand, India

MACHINE LEARNING-DRIVEN PADDY YIELD PREDICTION: COMPARATIVE EVALUATION OF BASELINE VS. ENSEMBLE MODELS IN UDHAM SINGH NAGAR, UTTARAKHAND

Abstract

Rice is a cornerstone of food security in India, supporting millions of livelihoods and the national economy. However, erratic climate patterns are making paddy yields increasingly unpredictable. This study develops a machine learning framework for rice yield prediction in Udhm Singh Nagar district, Uttarakhand, by integrating weather, soil, and crop data. Among baseline classifiers, CatBoost performed best with 80.85% accuracy and a ROC-AUC of 0.90. To further enhance performance, Optuna-tuned CatBoost, XGBoost, and LightGBM models were combined into hybrid ensembles. The Weighted Hard Voting classifier, giving higher weight to CatBoost ([3,1,1]), achieved the highest accuracy of 97.37%, followed by Stacking (95.6%) and Soft Voting ensembles (up to 96%). These results were supported by strong ROC-AUC scores. Overall, the study shows that carefully optimized ensemble models can significantly improve yield prediction accuracy, offering a practical tool for more precise and sustainable rice farming in climate-sensitive regions of India.

Key words: CatBoost, Gradient Boosting, Hybrid Ensemble, Machine learning, Random Forest, XG Boost, Yield.

Received April 13, 2026; revised May 27, 2026; accepted May 30, 2026.

Introduction

As one of the most important crops in the world, rice is very important in maintaining food security in the world, especially in India which is also among the largest rice-producing countries. Paddy farming is the main source of rural livelihoods in Uttarakhand at Udhm Singh Nagar district. According to the reports on Agriculture Department Uttarakhand, the issues of climate variability, soil erosion, and unreliable agricultural practices are the main factors that influence the paddy yield and require the implementation of predictive instruments that are reliable and can assist farmers in

their endeavours to enhance agricultural production (Agriculture Department Uttarakhand, 2025). According to Sharma et al. (2023), one of the agribusiness regions of Uttarakhand State is Udham Singh Nagar District. It is famous by its agriculture and irrigation which are through the process of historical evolution and is known to have a productivity of paddy crops all over Uttarakhand, to gain a status of Chawal ki Nagari, thus enhancing its impact in the groundwater resource of the district. Almost 64 percent of the total number of personnel is engaged in agriculture on the incredibly fertile Tarai development. The advancement in machine learning has now reconstituted predictive analytics with new solutions advancing to optimize solutions to difficult agronomy challenges. All the environmental factors such as soil properties, climatic conditions, and crop properties are instilled in machine learning models to deliver correct AI-driven analytics on yield forecasting. The tools increase the decision-making and promote sustainable use of resources to develop sustainable agriculture and food security further.

It has been proposed that Precision Agriculture, through the use of recent advances in Machine learning, can be used to be transformational in responding to these challenges. The use of these Machine learning algorithms can help in drawing significant patterns through complex data sets of climatic variables, soil traits, and crop attributes which can be utilized in formulating good predictions of yield. Specifically, the use of analytic approaches may lead to improved resolutions on the issue of farming and its sustainability in such locations as Udham Singh Nagar, a region of the Uttarakhand state, which is among the largest rice-producing regions in the country.

According to Tan et al. (2021), ensemble learning methods, which is a synthesis of the merit of over one base learner, have already proven to have essential steps in prediction accuracy and model resilience. It is common with other methods such as XGBoost, LightGBM, and CatBoost due to its ability to work with a large dataset using a complex dataset and minimum preprocessing. Recently, predictive tools such as stacked generalisation, voting (hard, soft, and weighted), and boosting ensembles have been proven to be effective in agricultural predictions (Tan et al., 2021; You et al., 2020). With hyperparameter optimisation algorithms, such as Optuna, those models provide better results even on limited and heterogeneous data, which is the common case in practice in real-world agriculture. These innovations can be seen as one of the recent trends in the same domain: combining data science with agronomy to enhance management of crops, to make the most out of resources, and to ensure livelihoods in case of environmental uncertainty. This paper affirms the applicability of the ensemble learning approach to agricultural production forecasting and provides a scalable potent instrument of making data-driven decisions in precision farming. This paper is founded on this assertion and the hybrid ensemble model that has been proposed and evaluated is applicable in the prediction of rice in Udham Singh Nagar. The study will focus on high-quality forecasts, sustainable agricultural activities, and policy formulation within similar agro-climatic areas by incorporating weather, soil, and crop parameters into a single machine learning pipeline, and using the optimal models of Optuna-tuned XGBoost, LightGBM, and CatBoost.

In the study by Chandrakumar et al. (2023), regression-based machine learning models that included Gradient Boosting, Random Forest and Support Vector Machines were used to predict rice production in the Tamil Nadu region. The environmental factors that were integrated by the authors include rainfall, soil nutrients and temperature. Gradient Boosting model had the best predictive accuracy when compared to the models.

Renju et al. (2022) examine the climate risk factors by analysing the benefits that are brought about by combining machine learning methods with agricultural databases. The study explores the significance of Gradient Boosting and Random Forest algorithms, together with other models, for climate issue management and increased predictive accuracy. It demonstrates how different machine learning systems help computer networks and decision trees to identify crop outcomes by examining multiple data types, including environmental, soil, and satellite measurements. These procedures enhance yield forecasting outcomes and improve resource utilization.

(Joshua et al., 2021) this study tested multiple modern learning tools to forecast crop yields for all farming types particularly rice. It demonstrates that the weaknesses of data are required, as well as the choice of the significant features, to produce improved models. Gradient Boosting with the addition of the Random Forest was effective in dealing with agricultural data. Application of machine learning (ML) methods in agriculture has transformed the process of predicting crop yields especially in rice which is one of the main food commodities in India. Recent research has shown that different ML models are effective at predicting yields of rice through the analysis of complicated datasets containing climatic, soil, and crop variables.

In another study, De Clercq et al. (2024) modelled Kharif season rice yield in 247 Indian districts with models like CatBoost and LightGBM, with an out of sample R^2 of up to 0.82. On the same note, Yewle et al. (2025) have come up with RicEns-Net a deep ensemble model, which used a combination of synthetic aperture radar together with optical remote sensing data as well as meteorological variables, which led to a mean absolute error of 341 kg/ha.

In the study by Kamilaris and Prenafeta-Boldu (2018), machine learning was applied to analyze remote sensing data in which the authors trained XGBoost and CatBoost models to estimate crop yield. As vegetation indices and climatic data were incorporated into ensemble techniques, the techniques produced better results than the conventional regression techniques. Ensemble learning techniques have become notable due to the enhancement in the degree of predictive precision through the synthesis of numerous models.

To predict crop yields, Manjunath and Palayyan (2023) created a hybrid machine learning model that included Decision Tree, XGBoost, and Random Forest that had a high predictive accuracy of 98.6. Similarly, Chandraprabha and Rajesh Kumar Dhanraj (2023) introduced an ensemble based deep learning methodology of stacking and using Deep Neural Networks, Deep Belief Networks as well as Support Vector Machines to predict rice yield depending on the level of soil nutrition, which they reported with a 89.5% accuracy.

Precision farming practices have led to the use of technology that has been used to enhance optimal utilization of available resources and achieve better crop yields. TNN (2025) the partnership of Punjab Agricultural University and BITS-Pilani is focused on the implementation of AI, IoT, and geospatial technologies in the field of agriculture and ensuring the use of data to make decisions. In addition, novel methods of farming such as Seeding of Rice on Beds (SRB) has been implemented to deal with the problem of water scarcity by consuming as much as 80 percent less water than that of the traditional methods.

IGARSS 2019 exhibited the process of predicting the number of rice harvests with the help of integrated meteorological and soil data. Guruprasad et al. (2019) trained machine learning algorithms, Neural Networks and Random Forest, using district-level data and distributed their results to taluk levels to create better maps. The research showed that better results in yield measurement could be obtained thanks to spatial resolution data, which produced only a 6% error for average measurements.

Materials and methods

The dependent variable used in the analysis was the district level yield of rice in the current year, and the independent variables were the selected weather, soil and crop parameters which were standardized before the model execution. The models were trained using data that covered the past 10 years ensuring that the training pipeline is efficient and reliable. Consequently, training models have been evaluated based on developed methods of training.

The model training pipeline included extensive data preprocessing, involving treatment of the missing values, encoding the categorical variables, and the normalization of the numerical ones. The dataset was then stratified into 80:20 train and test sets in order to maintain the proportion of classes, Figure 1.

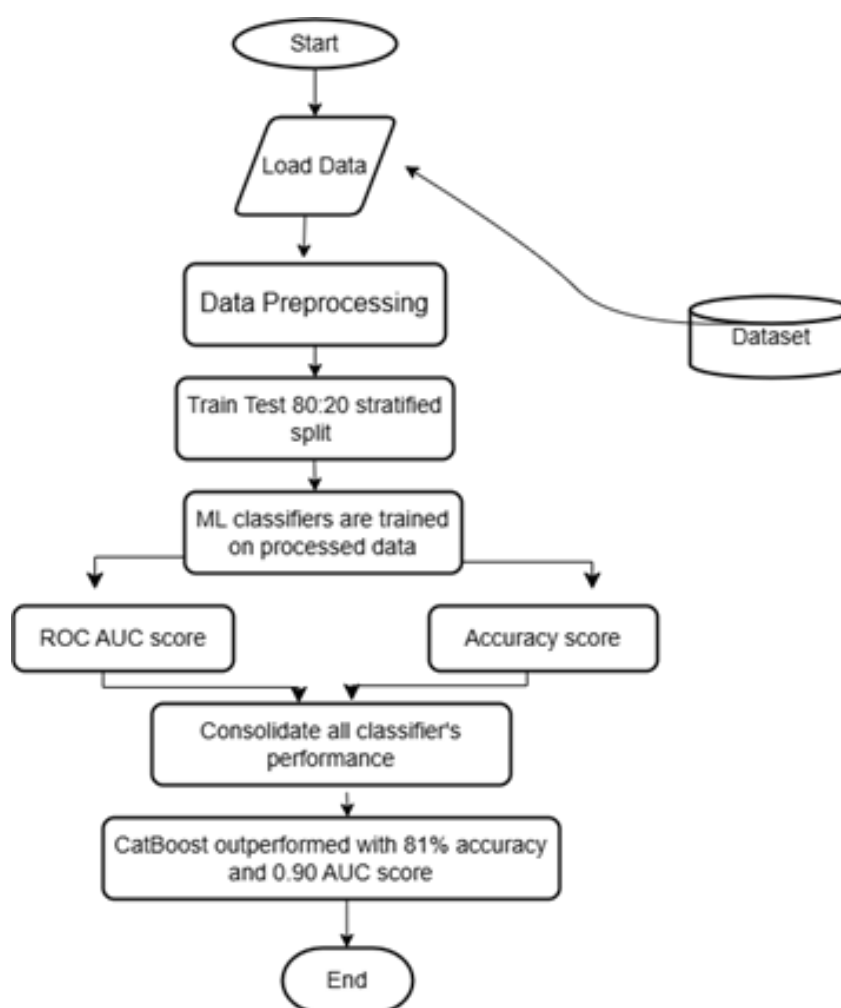


Figure 1 – Workflow Diagram of traditional base classifiers

The target variable (Yield Category) is a binary (0 = Low Yield 1 = High Yield) variable that is calculated according to whether each yield value falls below or above the median yield of its zone. This methodology guarantees the equal distribution of the classes and also puts into consideration the varying yield potentials in the rainfall areas.

We have developed our model in a systematic process, starting with an initial and simple classification models with more complex ensemble architectures. A brief summary of each algorithm administered in this evaluation is provided in following sections.

According to Ysaswy et al. (2022), the Gradient Boosting type of classification model has been a very useful ensemble ML model applied in the prediction of the outcome of agricultural yield, such as paddy yield. It builds multiple feeble decision trees sequentially where each new decision tree improves the model by rectifying the errors in the earlier decision trees. The objective of this iterative process is to minimize a loss criterion that has been preprogrammed which enables the Gradient Boosting process to discover complex dependencies between input qualities, such as climatic cycles, soil properties and sustainable agricultural methods.

According to Badshah et al. (2024), the Random Forest classification model is a form of machine learning model that is widely applied in predictive activities, including forecasting agricultural yields. In the training of the model, it constructs many decision trees and combines integrated results

in order to minimize discrepancies. All trees are computed based on random sub setting of the data, and a probabilistic method of attribute representation is evaluated at every split, such that there is fluctuation across trees. This randomness renders Random Forest applicable to noisy inputs, and in a position to discover more complex correlations among the input variables.

According to Chen and Guestrin (2016), the XGBoost classification model, which is also referred to as the Extreme Gradient Boosting model is among the sophisticated machine learning algorithms and is very effective when it comes to predictive tasks like predicting the yield of paddies. XGBoost is a specialized version of the Gradient Boosting framework designed to provide high-performance and scalable solutions. Features such as parallel processing, a sparsity-aware algorithm, and parameter regularization enhance performance and help prevent overfitting.

Yandex (2019) report that CatBoost (also known as Categorical Boosting) is a gradient boosting algorithm developed to specifically handle categorical features with minimal preprocessing, i.e., no one-hot encoding is required. CatBoost is superior to other gradient boosting algorithms in that it employs ordered boosting to minimize overfitting. It is regarded to be very reliable, correct and easy to use especially with tabular datasets. CatBoost uses a technique known as target-based statistics that uses categorical data directly, limiting the likelihood of overfitting. It also has little hyperparameter tuning, and it can also be used in conjunction with most data science packages.

According to Yamparla et al. (2022), one of the effective statistical methods that are commonly used in classification including classification of rice output is the Logistic Regression. In Agri-tech applications, Logistic Regression has the ability of classifying the level of yield, either high or low, in relation to climatic conditions, soil characteristics, and farming practices which can add valuable information to the decision-making process of precise agriculture. It uses the correlation of a categorical dependent variable with another factor or a combination of multiple factors to compute the likelihood of the outcome occurring within a given label of a class. The algorithm makes use of the logistic function, which is frequently modeled as an S-shaped sigmoid curve, to map the values of features into constrained probability distributions between 0 and 1.

Hyperparameter tuning for CatBoost, XGBoost, and LightGBM was conducted using Optuna, an automated optimization framework. Optimal configurations were selected based on cross-validated performance on the training data, enabling improved generalization and robustness relative to default parameter settings.

First, hyperparameters tuning of the ensemble classifiers- CatBoost, XGBoost, and LightGBM was used with Optuna. The best settings derived out of this tuning were then applied in the implementation of the above hybrid models. The CatBoost Classifier was optimally set and some of the parameters that were set included depth of 8, learning rate of about 0.15, and the hyperparameters of the regularization and sampling including l2leafreg, border count, and bagging temperature. These values were discovered by optuna by trial and error to maximize the results on the training data. Equally, XGBoost and LightGBM hyperparameters were also optimised that contained the following hyperparameters; iterations=1049, depth=5, learningrate=0.105 and other boosting-specific hyperparameters such as l2leafreg, random strength and grow_policy= SymmetricTree. These optimized settings are likely to enhance model accuracy and generalization much more than default settings, which constitute a solid basis of a high-performance ensemble model.

To enhance the classification accuracy, a set of ensemble learning methods, such as Hard Voting, Soft Voting (weighted and weightless), and Stacking, were used to combine three fine-tuned gradient boosting classifiers: CatBoost, XGBoost, and LightGBM.

Hard Voting Classifier

An implementation of a Hard VotingClassifier was done whereby the predictions are done depending on the majority class vote of the base learners. The weight configuration [3, 1, 1] was used to provide CatBoost with more power. The accuracy of this ensemble was high and reached 97.37, which supports the effectiveness of majority rule in the case of well-calibrated and complementary base learners.

Mathematically this can be put as:

$$\hat{y} = \underset{(c \in C)}{\operatorname{argmax}} \sum_{m=1}^M 1_{[h_m(x)=c]} \cdot \omega_m$$

Where:

- $\hat{y}$: Final predicted class
- $c \in C$: Class labels (e.g., low / high yield)
- M : Number of base classifiers
- $h_m(x)$: Prediction of the m -th base classifier
- $1_{[h_m(x)=c]}$: Indicator function
- w_m : Weight assigned to the m -th classifier

The hard voting classifier was adopted in this research as a hybrid ensemble model, and it consists of CatBoost, XGBoost, and LightGBM. The assignments were made by hand, to give a better score to CatBoost (e.g. $w_{\text{CatBoost}}=3$, $w_{\text{XGB}}=1$, $w_{\{\text{XGB}\}}=1$, $w_{\text{LGBM}}=1$, $w_{\{\text{LGBM}\}}=1$), as it has a better standalone performance. Based on this strategy, the accuracy rate was 97 percent, which was higher than all base models, which showed that ensemble voting, when properly weighted, is a robust method.

Soft Voting Classifier

In the Soft Voting arrangement, the prediction in the classifier was done by the probability averaging. There were three configurations that were investigated:

- ♦ Weighted Soft Voting [3, 1, 1] with weights weighted favoring CatBoost was the best with an accuracy that was 95.6%.
- ♦ Weighted Soft Voting with [2, 2, 3] also reached 95.6%, showing consistent performance across similar configurations.

Unweighted Soft Voting (equal weight) resulted in a marginally greater accuracy of 96 showing that even when the model was not weighted, the ensemble was still very effective as the individual strengths of parts made it.

Ahmed (2023) explains that in a soft voting classifier, the final predicted class is calculated based on the mean of the predicted probabilities of each of several classifiers, and the class with the largest combined probability is selected. The equation can be expressed as:

$$\underset{c \in C}{\operatorname{argmax}} \sum_{m=1}^M w_m \cdot P_m(c|x)$$

Where:

- $\hat{y}$: Final predicted yield class
- $c \in C$: Set of yield classes
- M : Number of base classifiers
- w_m : Weight of the m -th classifier
- $P_m(c | x)$: Class probability from classifier m

The soft voting classifier plays a crucial role in this study, as it facilitates a probability-based hybrid ensemble that can intelligently fuse the predictions of CatBoost, XGBoost, and LightGBM. In comparison to hard voting, which uses majority class designation, soft voting takes into account the level of confidence in each model's prediction; thus, it is especially efficient when the capabilities of each model are complementary.

Stacking Classifier

Two Stacking Classifiers had been built:

One of them used a Random Forest as the meta-learner which was trained on predictions of the out-of-fold predictions of the base learners. The accuracy of this setup with cross-validation of $cv=5$ was 95.6%.

Lin et al. (2023) report that the other CatBoost model was used as the meta-learner. When `passthrough=True`, both base model outputs and original features were forwarded to the meta-learner, which improved learning. This approach also achieved 95.6 percent accuracy, demonstrating the versatility and power of stacking with alternative meta-models.

Equation:

$$Z = \begin{bmatrix} B_1^{oof}(X) \\ B_2^{oof}(X) \\ \vdots \\ B_k^{oof}(X) \end{bmatrix}^T \in \mathbb{R}^{n \times k}$$

$$\hat{y} = \Phi_{RF}(Z)$$

Where:

The meta-learner (CatBoost) is trained on both:

- $B_j^{oof}(X)$: Out-of-fold predictions generated by the j -th base learner.
- Φ_{RF} : Random Forest classifier trained on dataset Z .
- $\hat{y}$: Final predicted class labels.

- ♦ Out-of-fold predictions Z from base learners.
- ♦ Original input features X .

Equation:

$$Z = \begin{bmatrix} B_1^{oof}(X) \\ B_2^{oof}(X) \\ \vdots \\ B_k^{oof}(X) \end{bmatrix}^T \in \mathbb{R}^{n \times k}$$

$$\tilde{Z} = [X \parallel Z] \in \mathbb{R}^{n \times (d+k)}$$

$$\hat{y} = \phi_{CatBoost}(\tilde{Z})$$

Where:

$\tilde{Z}$: Concatenation of the original feature set and the outputs of the base learners.

$\Phi_{CatBoost}$: CatBoost-based meta-learner.

This strategy enriches the meta-model with additional information, thereby enhancing its learning capacity and predictive performance.

Both stacking approaches are:

$$\hat{y} = \phi([Base\ Model\ Outputs] \text{ or } [X \parallel Base\ Model\ Outputs])$$

The key difference lies in the input to the meta-learner:

Random Forest: $\Phi(Z)$

CatBoost (passthrough=True): $\Phi([X \parallel Z])$

Both achieved 95.6% accuracy, illustrating the power and adaptability of the stacking framework with different meta-learners and feature strategies.

The workflow diagram of Hybrid Ensemble models shown in Figure. 2, hybrid models are always more effective accurate compared to base classifiers. Although the base classifiers are about 53-81 percent, the hybrid models reach greater than 95 percent accuracy indicating that ensemble methods are effective in improving the predictive ability.

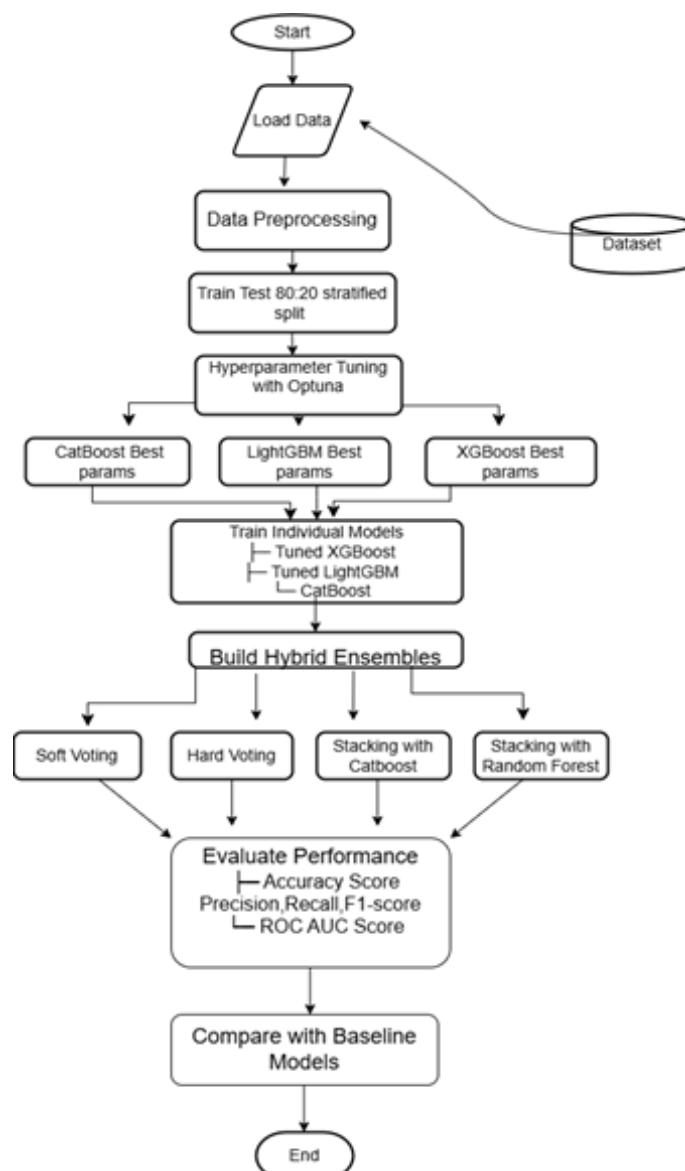


Figure 2 – Workflow of ensemble models

Study area

Udham Singh Nagar district is situated in the Tarai region of Kumaon Division. The district spans a total area of 3,055 km², standing 9th in terms of geographical size within Uttarakhand. It's situated between latitudes 28° 53' N and 29° 23' N, and longitudinally expands over 78° 45' E to 80° 08' E, (Kumar et al., 2021) Figure 3.

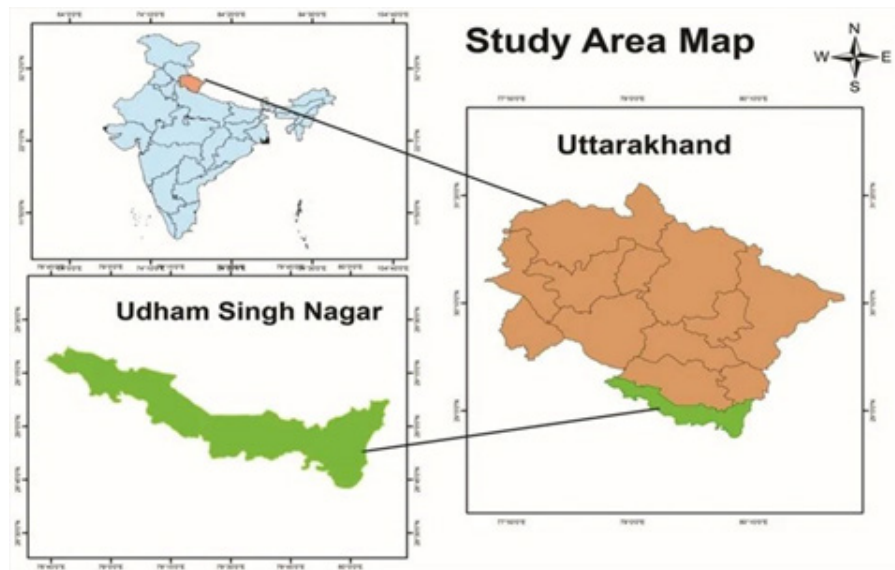


Figure 3 – Location map of Udham Singh Nagar district

The data set, Table 1, used in this study is holistically divided into three broad categories, which are weather data, soil data and crop data, which makes it easy to evaluate thoroughly the factors influencing paddy production in the region.

Table 1 – Dataset with independent features and dependent variable (Yield)

Yield (quintal/hectare)	temp_min	temp_max	dew_point	pressure	humidity	wind_speed	wind_dir	rain_1h	clouds_all	Area (ha)	Soil_ph	EC (mmol/cm ²)	Soil_O/C (%)	Available phosphorus (kg/ha)	Available Potash (kg/ha)	Soil_Zinc (ppm)	Mineralizable Nitrogen (kg/ha)	
4.73277949	26.82276	28.280545	26.34872	1014.8824	55.4292	79.5887	1.78747277	143.788274	0.3338825	32.8338845	2380	7.2	0.81	0.58	17.9	174.58	1.54	
6.52577844	26.53885	25.57544	27.42884	1014.875	55.4292	79.5887	1.78747277	143.788274	0.3338825	32.8338845	3002	7.1	0.85	0.75	18.24	122.26	1.09	
6.48853408	26.71402	25.835737	27.62889	1017.244	54.8386	1.94532484	147.724485	0.0333700	24.264387	17.85	6.8	0.79	0.75	20.5	205.36	1.29	385.87	
6.79378033	25.86887	24.875291	27.12112	1016.8898	57.848	96.5747	1.9244062	186.26282	0.2831348	27.262885	2200	7.3	0.81	0.62	18.89	188.9	1.43	178.58
7.01343038	26.48102	25.364338	27.24948	1015.0034	57.737	51.9172	1.88742572	184.84878	0.0288744	24.9184828	2247	6.8	0.9	0.71	18.25	211.51	1.51	387.89
7.07434258	24.88785	25.720482	25.37215	1016.8826	58.078	62.9228	1.9188884	165.49475	0.0714222	30.8204883	2201	7.2	0.82	0.74	19.58	209.89	1.28	392.85
7.28242458	27.38677	26.442381	28.12375	1017.42887	55.725	57.7788	1.87581343	145.528328	0.04247528	18.8887754	2267	7.2	0.9	0.94	18.54	248.54	0.43	298.75
7.37262825	25.48233	24.538288	26.23478	1015.44339	58.889	55.2258	1.87448878	175.824838	0.0289128	30.548838	2274	6.9	0.81	0.71	19.01	210.39	0.77	395.48
7.43789777	27.12385	26.4817482	28.430842	1016.8898	58.747	52.8658	1.88388384	175.824838	0.0289128	30.548838	2284	6.9	0.8	0.75	19.5	208.78	1.23	396.42
7.67174379	26.47868	25.897289	27.24939	1014.8824	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	393.58
8.03389779	27.28495	26.372623	28.12285	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.1	0.86	0.85	20.31	178.92	2.4	387.42
8.35234889	27.12349	26.622483	27.74728	1016.8898	58.87	55.2258	1.87448878	175.824838	0.0289128	30.548838	2274	7.2	0.79	0.72	20.21	177.52	1.43	388.85
8.68432818	26.57382	24.848082	26.35231	1017.42887	55.725	57.7788	1.87581343	145.528328	0.04247528	18.8887754	2267	6.9	0.76	0.61	21.5	244.28	1.21	387.58
8.70295221	26.89289	26.322882	27.32878	1015.44339	58.889	55.2258	1.87448878	175.824838	0.0289128	30.548838	2274	6.9	0.81	0.71	19.01	210.39	0.77	394.65
8.81893829	27.07502	26.38575	28.18578	1015.44339	58.889	55.2258	1.87448878	175.824838	0.0289128	30.548838	2274	6.9	0.81	0.71	19.01	210.39	0.77	396.67
9.428887448	25.83144	24.882284	26.04487	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	394.55
9.434422889	25.83144	24.882284	26.04487	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	394.55
9.51238858	26.42487	25.889482	27.34072	1015.44339	58.889	55.2258	1.87448878	175.824838	0.0289128	30.548838	2274	6.9	0.81	0.71	19.01	210.39	0.77	394.25
9.70677913	25.33144	24.882284	26.04487	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	397.54
10.02893857	27.518	27.085842	28.18578	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	397.28
10.834422889	25.83144	24.882284	26.04487	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	398.22
11.02893857	27.518	27.085842	28.18578	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	398.22
11.20877913	25.33144	24.882284	26.04487	1016.8826	58.078	48.7788	1.88742572	184.84878	0.0288744	24.9184828	2274	7.2	0.79	0.85	20.1	179.54	1.09	398.22

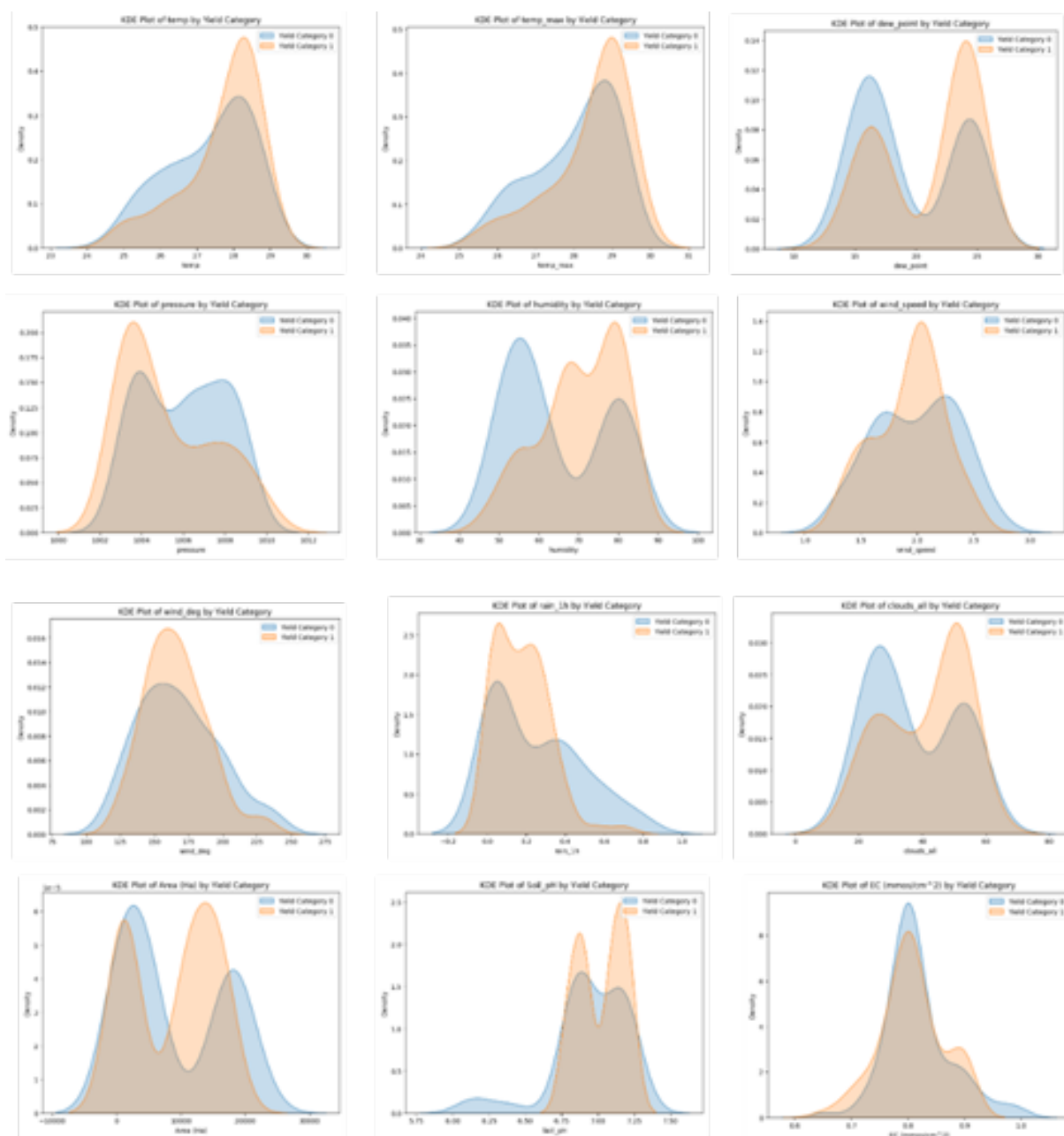
Weather Data: obtained from OpenWeatherMap.org (<https://openweathermap.org/>), this type contains a group of meteorological parameters that are important to predict yields. The parameters consist of average temperature, dew point, minimum and maximum temperature, wind direction, atmospheric pressure, wind speed, rainfall within the past hour, humidity, and cloud cover.

Soil Data: Collected from the Soil Test Laboratory in Rudrapur, the soil data comprises essential chemical and physical properties. These include soil pH, electrical conductivity (EC in mmol/cm²), soil organic carbon percentage (Soil_O/C), available phosphorus (kg/ha), available potash (kg/ha), soil zinc concentration (ppm), and mineralizable nitrogen content (kg/ha).

Crop Data: Crop information on the yield and the area of cultivation was retrieved in the Statistical Yearbook of the District found in Vikas Bhawan. This section contains information about yields (quintals per hectare) and the hectares that are under cultivation.

Results and discussion

The analysis of Kernel Density Estimation (KDE), Figure 4. showed that there were visible differences in the distributions across high- and low-yield groups in relation to a number of soil and climatic variables. There were consistent high levels of soil organic carbon, available phosphorus and mineralizable nitrogen found in the high-yield observations, which demonstrated the high level of association between the status of soil fertility and the productivity of rice.



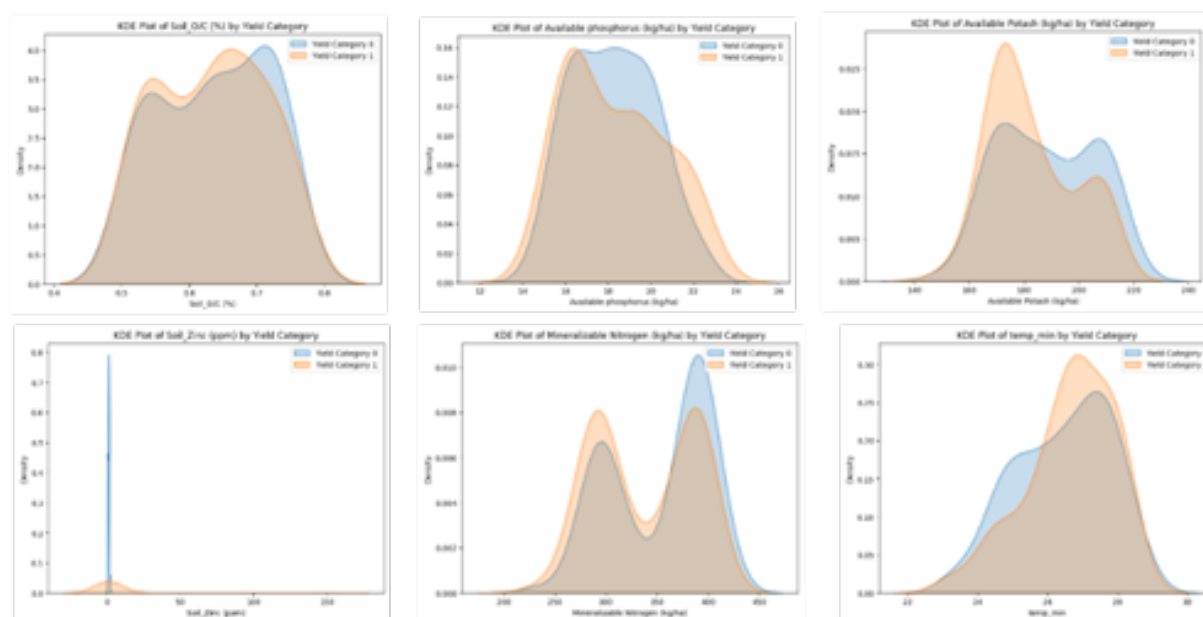


Figure 4 – KDE plots

Conversely, the overlaps of soil pH and electrical conductivity among yield classes were highly significant showing that the parameters did not significantly exceed the agronomically applicable limits and did not have a discriminative effect on the yields in the study area. Climatic variables including features of temperature exhibited significant distributional changes in yield categories, which indicates the sensitivity of rice yield to thermal factors. The results are similar to other machine learning-based rice production works which show that variability of temperature and nutrient availability are the most common influencers of yield prediction accuracy through analysis of feature importance and model performance (Chandrakumar et al., 2023; De Clercq et al., 2024). These methods are used in combination, and the regression-based feature importance used to measure predictive effect, and KDE used to give intuitive information about how agronomic and climatic factors vary between yield categories.

The frequency distributions of the 19 input variables are shown in Figure 5. Crop yield was mostly concentrated in the lower range (10–40 quintal/hectare), with fewer fields recording higher yields. Temperature variables followed fairly regular distributions, reflecting the relatively stable climatic conditions of the region. Dew point showed two distinct peaks, likely capturing the contrast between drier and more humid phases of the crop season. Rainfall was sparse for most observations, with occasional high-intensity events pulling the distribution rightward — a pattern quite common in the Terai belt. Soil pH was largely near-neutral (6.75–7.0), which is typical for the agricultural soils of Udham Singh Nagar, while EC and Soil Zinc were more variable, pointing to uneven soil fertility across sampled fields. The spread in Available Potash and Mineralizable Nitrogen likely reflects differences in nutrient management practices among farmers.

Sharma et al. (2024) predicted rice yield using only five meteorological variables — maximum and minimum temperature, relative humidity, rainfall, and sunlight hours — organized as fortnightly averages based on Standard Meteorological Weeks for a single district in Haryana, with no consideration of soil properties or field-level parameters. The present study, in contrast, analyzed distributions of 19 variables spanning meteorological conditions, soil health indicators, and agronomic parameters, revealing key data characteristics such as bimodal dew point, right-skewed rainfall, and multimodal soil nutrient distributions — offering a more comprehensive empirical basis for rice yield prediction in Udham Singh Nagar district.

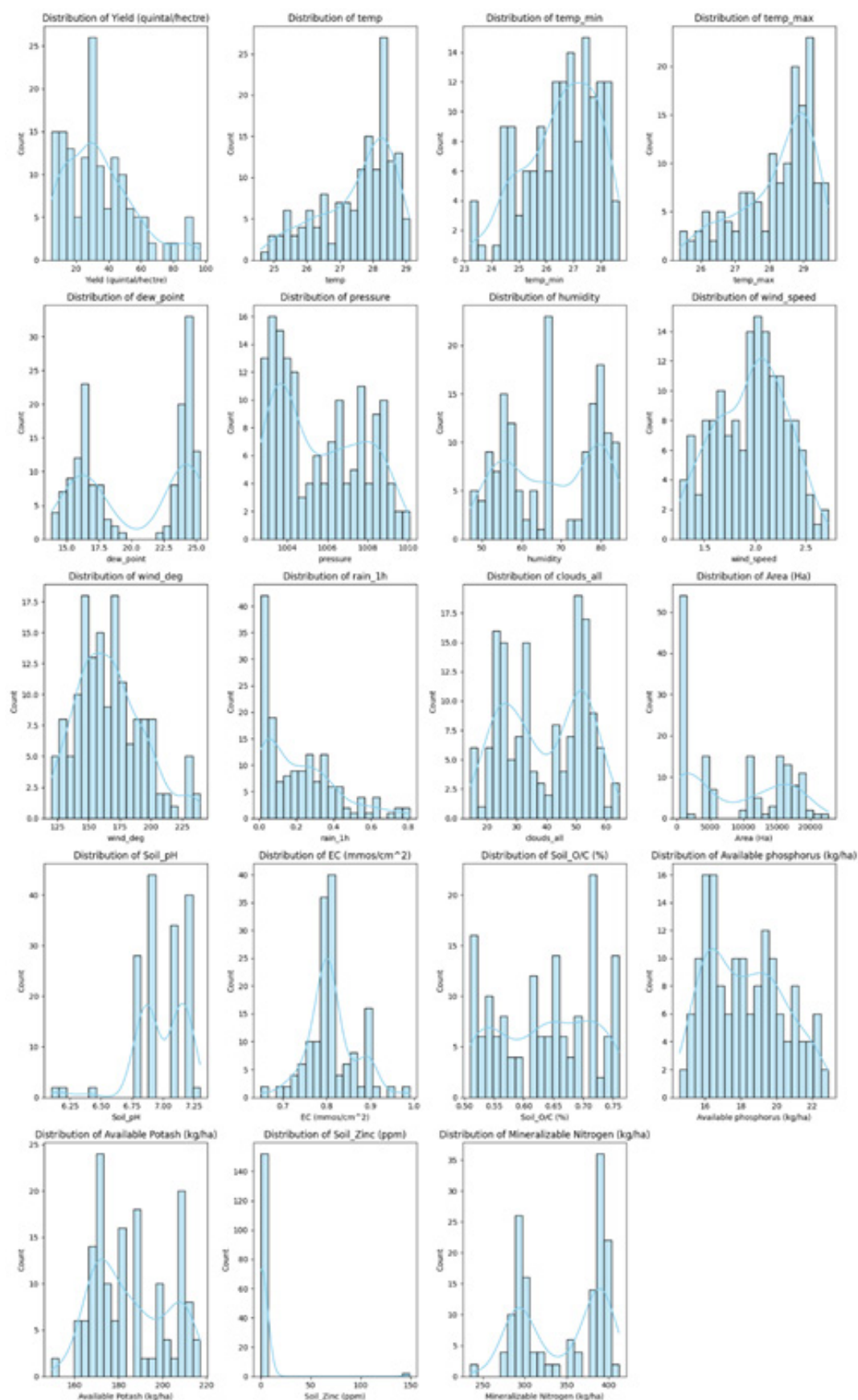


Figure 5 – Distribution of Input Variables

Figure 6. presents the pairwise relationships among the meteorological variables. As expected, temperature, minimum temperature, and maximum temperature moved closely together, which is worth keeping in mind when selecting model inputs to avoid redundancy. Dew point and pressure showed a mild inverse trend, while humidity had little consistent relationship with most other variables. Wind speed and direction appeared largely independent of the temperature-related features. Rainfall was near-zero for the bulk of observations, which aligns with its skewed distribution seen in Figure 5.

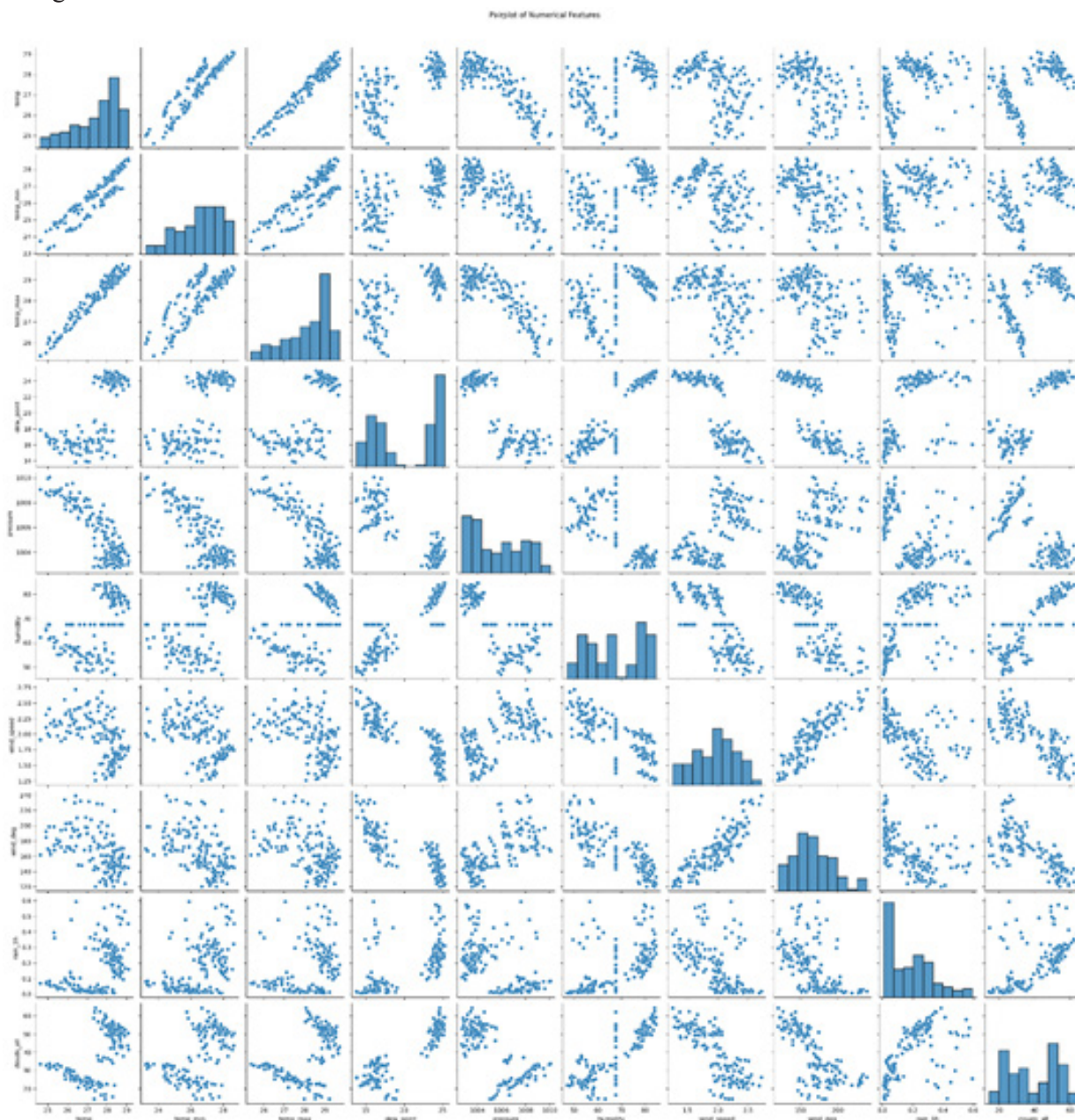


Figure 6 – Pairplot of Numerical Features

Naseer et al. (2025) looked at how soil and weather variables relate to each other, but their pairwise analysis covered only seven features – N, P, K, pH, temperature, humidity, and rainfall – where all features showed positive correlations, with potassium and phosphorus being the most strongly linked at 0.74. Notably, variables like dew point, atmospheric pressure, wind speed, wind direction, and cloud cover were not part of their analysis. In the present study, pairwise relationships

were examined across the meteorological features, and the patterns were noticeably more varied – temperature, temp_min, and temp_max moved closely together as expected, while dew point and pressure showed a mild opposing trend, and wind-related variables had little to no consistent relationship with temperature. This mixed correlation structure is something the seven-variable framework of Naseer et al. (2025) did not capture, pointing to the need for more careful feature selection when building yield prediction models for rice in the Terai agro-climatic zone.

Figure 7. shows that the correlation among climatic variables is high and linear, especially between minimum temperature, maximum temperature, and dew point, reflecting clustered thermal behaviour among the weather-related predictors. Attributes of soil fertility, such as available phosphorus and mineralizable nitrogen, are strongly correlated with soil organic carbon, an indicator of combined nutrient processes within the soil system. In comparison, rice yield shows a tendency to exhibit weak pairwise linear relationship with individual environmental and soil variables, which indicates that simple linear relationships can only explain a small amount of yield variability. The absence of significant linear yield-feature relationships motivates the use of non-linear machine learning models to characterise higher-order interactions and intricate relationships. In this regard, the current results are conceptually aligned with the synthesis presented by Kamilaris and Prenafeta-Boldu (2018), who state that crop yield mechanisms are not often determined by simple linear relationships and are better represented in the context of non-linear learning frameworks, based on a wide survey of agricultural studies.

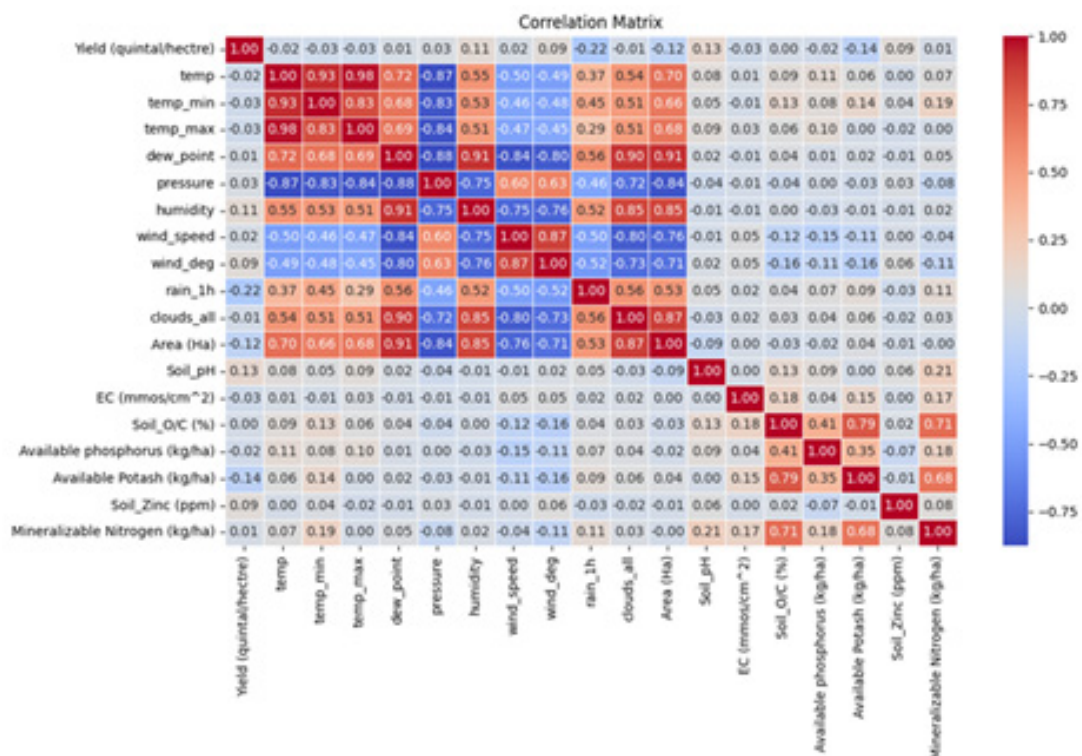


Figure 7 – Correlation Matrix

Results of the current study are generally in agreement with previous classification-based research of rice yield-prediction by Akula et al. (2021) in the Journal of Agrometeorology, but further expands the line of research by using more recent ensemble learning models and measures of performance. Both articles show the appropriateness of classification systems to place the outcomes of rice yield in categories, which can provide viable interpretability in agricultural decision support. In the case of the study carried out in Ranga Reddy district, Telangana, Akula et al. (2021) stated that the Multilayer

Perceptron (MLP) classifier demonstrated the best classification accuracy (about 74.19 percent), exceeding that of the traditional tree-based (J48/C4.5) and rule-based classifiers, which proved the importance of modeling non-linear relationships between weather variables and yield category.

By comparison, the current research gets much better classification (Table 2), with CatBoost reaching 80.9% and AUC of 0.90, then Gradient Boosting and XGBoost. These gradient boosting ensembles have a better performance as compared to the Logistic Regression, Random Forest, SVM, and KNN which shows that they have a greater ability to model high-dimensional, complex interactions between weather in relation to soil and crop-related variables and thus shows the advantages of using the modern ensemble methods in rice yield classification.

Table 2 – Performance metrics of base classifiers

Base Model	Accuracy	Precision	Recall	F1 Score
CatBoost	80.85%	82.00%	81.00%	81.00%
Gradient Boosting	78.72%	81.00%	79.00%	79.00%
KNN	53.00%	54.00%	53.00%	53.00%
Logistic Regression	76.60%	77.00%	77.00%	77.00%
Random Forest	68.09%	69.00%	68.00%	68.00%
Support Vector Machine	70.21%	72.00%	70.00%	70.00%
XGBoost	76.60%	78.00%	77.00%	76.00%

The results of the current paper are concomitant to those of Solanki (2024) who compared several machine learning classifiers to the classification of rice crop stages and proved that XGBoost and SVM were highly effective in case applied to feature-enriched data sets. On the other hand, the current paper is centered on district-level weather-soil-crop data in order to categorize rice yield and finds CatBoost the most stable standalone data classifier based on various evaluation metrics, with SVM and Random Forest having a relatively low level of discriminative ability. This variance highlights how both stage of prediction and data structure affect model behaviour with crop stage classification and yield categorization having completely different learning problems. However, in both papers, gradient boosting-based techniques prove to be trustworthy and flexible rice-related classifiers, which underlies their applicability to various precision agronomy activities such as phenological stage classification, as well as yield-related decision support.

In the given work, ROC-AUC analysis also proves the high discriminative power of CatBoost (AUC = 0.90), then XGBoost (0.89), Gradient Boosting (0.86), but Random Forest (0.78), SVM (0.72), and KNN (0.57) have lower rates of classification, Figure 8.

The results of the current research are in harmony with the existing literature that proves the superiority of hybrid ensemble classifiers in comparison with single base models in the process of agricultural classification. Ge et al. (2021) offer a representative example and researched the issue of rice phenological stage detection with the help of UAV-RGB images, which stated that ensemble approaches, including soft/weighted voting and stacking, were more critical in classification accuracy than individual classifiers. In particular, stacking was the most accurate (around 96.2%), then soft and hard voting (94.7 and 93.1) and demonstrates that complementary predictions of a model can be effectively combined. Probabilistic aggregation of the classifier outputs was considered as the superior outflow of the soft voting whereas stacking was favorable to learn the optimal combination of different base learners leading to the enhanced robustness in the capability of classifying intermediate phenological phases.

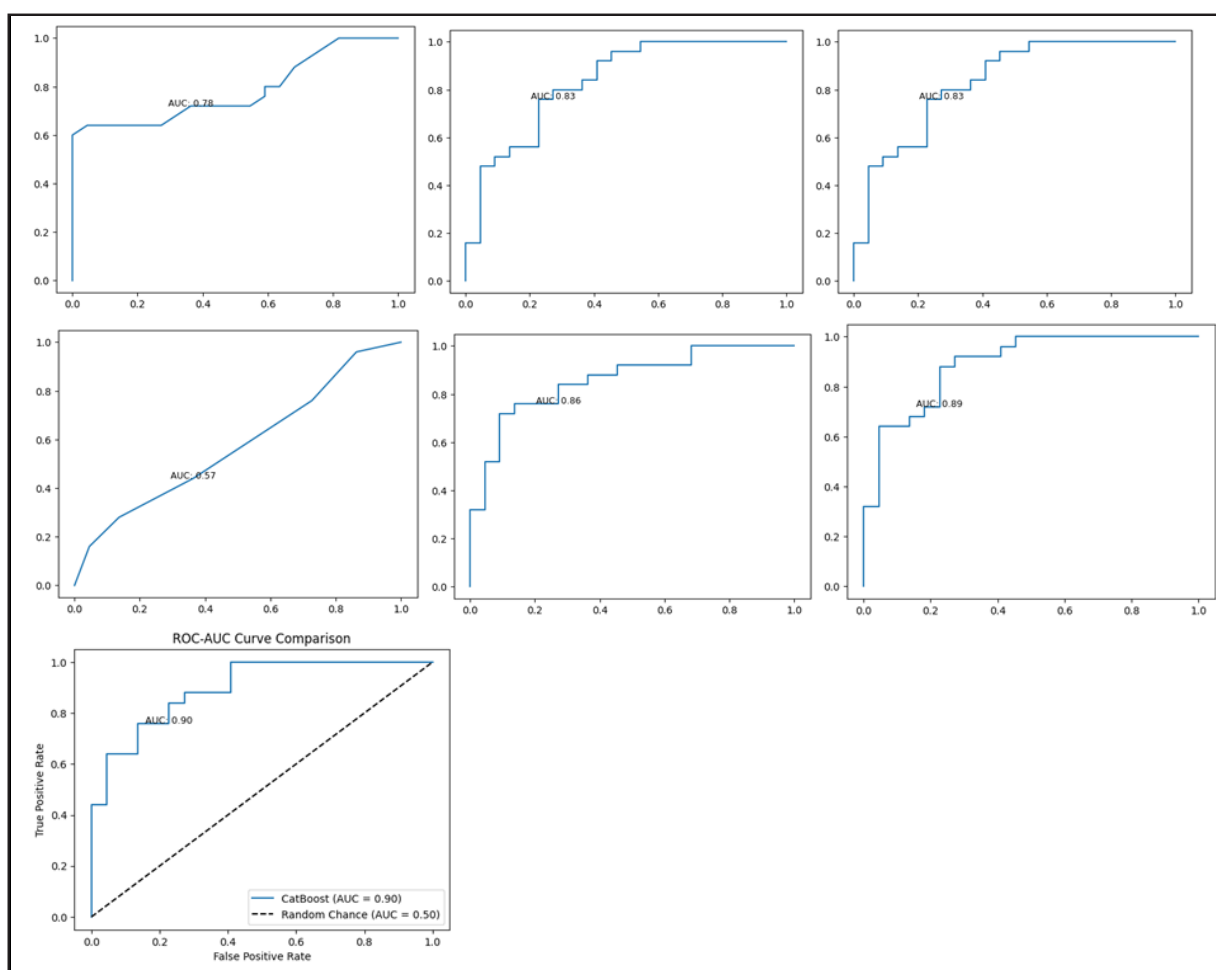


Figure 8 – ROC-AUC score comparison of Baseline Models

Similar advantage of the hybrid ensemble learning, however under different prediction task with different data format, can be seen in the current study on paddy yield classification. Table 3. illustrates that the weighted hard voting classifier, in which more importance is laid on CatBoost, had the highest overall performance (97.37%), which was better than those of soft voting and stacking. The outcome points to the importance of attention to selecting models and optimizing weights, which can further improve the performance of ensembles in the classification of yields. Taken together, these results confirm the bigger picture that ensemble methods, especially weighted voting and stacking provide better generalization and robustness in case of application to heterogeneous agricultural data, thus justifying their applicability to a variety of precision agriculture tasks.

Table 3 – Performance comparison of various hybrid ensemble models on classification metrics

Hybrid Models	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)
Stacking Classifier (Random Forest)	95.6	95.6	95.6	95.5
Weighted Hard Voting Classifier	97.3	97.3	97.3	97.3
Stacking Classifier (CatBoost)	95.6	94.6	98.6	96.6
Weighted Soft Voting Classifier	95.6	95.6	95.6	95.5
Soft Voting Classifier	95.6	94.6	98.6	96.6

The current research Islam et al. (2023) used regression models, such as a stack-ensemble hybrid, to predict rice yield (kg/ha) and had its best fit of RMSE $\approx$ 328 kg/ha and MAE $\approx$ 317 kg/ha, which is an improvement of 20-30% over base models. This shows that hybrid ensembles are good in capturing non-linear associations between environmental factors like NDVI, rainfall and soil moisture, even though the variability in space influenced the precision in certain districts. Whereas, the current work Figure 9. used classification models to predict high and low yield classes, the weighted hard voting classifier obtained Accuracy, Precision, Recall, and F1-score of 97.3%, which is significantly better than base classifiers (53-81%). It means that complex feature interactions can be effectively modeled by hybrid classifiers and they will be able to achieve high performance on all datasets. Collectively, the two studies show that ensemble methods are very effective in increasing the predictive quality, and that regression ensembles provide a better estimation of continuous yields and the classification ensembles can provide a highly accurate categorical forecast, particularly when multi-source environmental data is used.

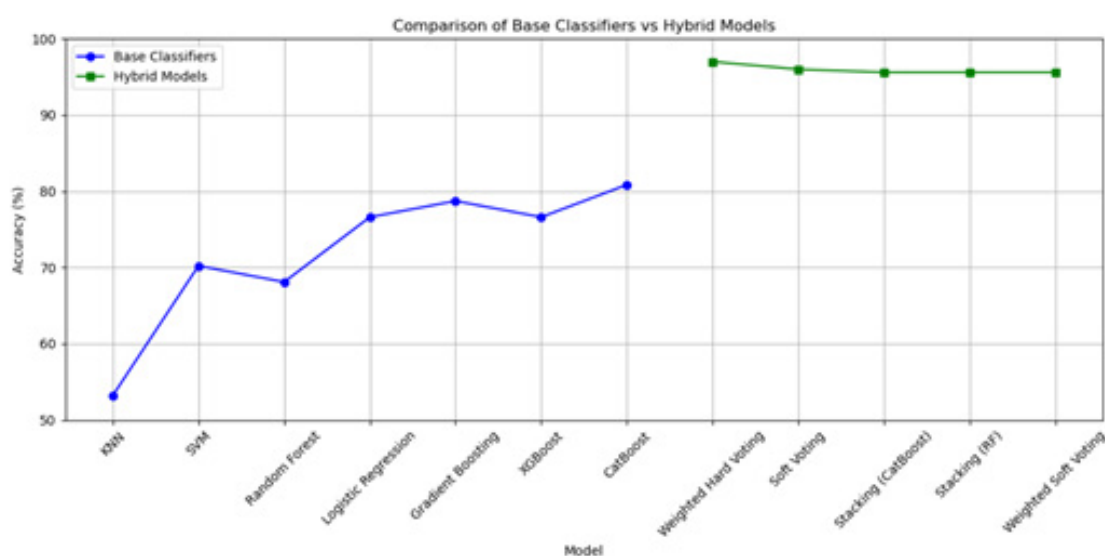


Figure 9 – Comparison of model accuracy between base classifiers and hybrid ensemble models

The current work builds on the previous ensemble-related rice classification studies, especially the one conducted by Ge et al. (2021) by an explicit and quantitative ROC-based analysis of hybrid models. They showed that stacking, soft voting, and hard voting ensembles are efficient in rice phenological stage classification and that stacking, soft voting, and hard voting ensembles have high classification accuracies (around 96.2 percent, 94.7 percent and 93.1 percent, respectively) and strong classifier discrimination represented by ROC curves. Their analysis, however, was based on visual analysis of ROC curves, without providing numerical values of ROC-AUC.

Conversely, Figure 10. the current paper measures the categorization of district-level rice yields with conventional hybrid classifiers and provides clear ROC-AUC values and ROC curves, which allow a more stringent and straightforward measure of the discriminative efficiency. The stacking classifier using CatBoost meta-learner scored 0.9908, stacking model using Random Forest meta-learner scored 0.9951, and weighted soft voting classifier scored 0.9941, a high level of classification was seen to be possible using all the three ensemble configurations. Although the two papers confirm the usefulness of ensemble learning in rice-related classification tasks, explicit ROC-AUC measures used in the current study enhance the performance of comparability and offer a more detailed method of evaluation to the yield-based decision support application.

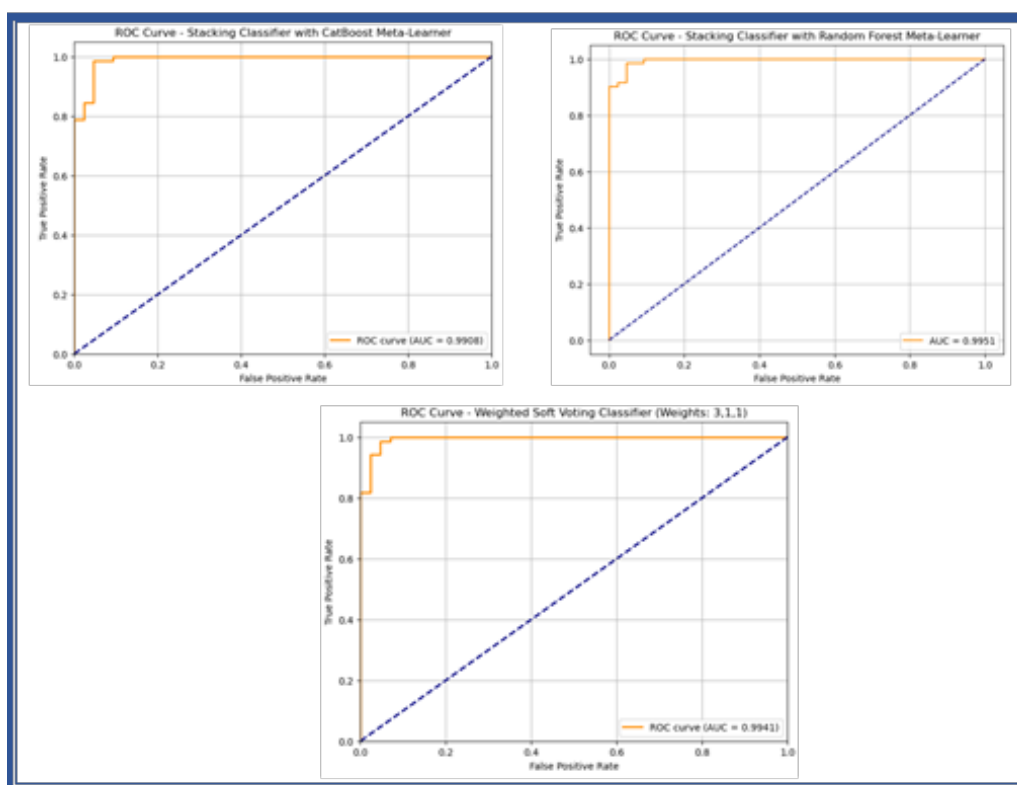


Figure 10 – ROC AUC curves of hybrid ensemble models

The study therefore reveals that there was a pronounced steady improvement in prediction performance when individual base classifier models are replaced by hybrid ensemble models to classify rice yield. The individual classifiers, or KNN, SVM, Random Forest, Logistic Regression, Gradient Boosting, XGBoost, and CatBoost, demonstrate moderate and high accuracy, with the lowest value of around 53 percent to the highest of about 81 percent, respectively. In line with the previous results (Akula et al., 2021; Solanki, 2024), tree-based and boosting models designed to be used with ensembles show better performance compared to distance and linear-style ones, which indicates their capability to learn non-linear relations between soil, weather-related, and crop-specific variables. Still, individual models are limited in their predictive power in terms of model-specific bias and the failure to capitalize on complementary decision patterns.

By comparison, the hybrid ensemble models, such as the weighted hard voting, soft voting, stacking, and weighted soft voting, record significantly greater accuracies with the range of about 95-97 that is clearly higher than that of all single classifiers. The mechanisms behind this performance increase are successfully compensating errors in a variety of learners, better generalization by means of diversity in an ensemble, and probabilistic aggregation that is used in the soft and weighted voting schemes. Out of the tested hybrids, CatBoost-based stacking and weighted voting strategies show a persistently high performance, which proves the success of meta-level learning and the optimal model weighting.

The current research contributes to the existing literature on the subject because it goes beyond the established rice-related prediction literature where most studies concentrate on the regression-based yield estimation or single-model classification and usually consider the accuracy or the qualitative ROC analysis as the final objective. The presented quantitative results improvements support the strength and the efficient applicability of the concept of ensemble learning to the classification of rice yield and decision support in a district level.

Conclusion

This paper has shown the promise of hybrid ensemble models that are machine learning driven to enhance prediction of paddy yield in the Udham Singh Nagar district of Uttarakhand. The results have shown that ensemble methods, in particular, stacking and voting-based classifiers including the boosted learners, are more accurate and robust in prediction than the traditional standalone model.

The proposed approach will enhance the development of precision agriculture, as it will use data and consequently make decisions about the activities of agriculture based on the use of soil, climatic, and historical crop data. The findings support the use of combination of various learning algorithms to accommodate the complexity and non linearity of agricultural data. In a practical sense, the suggested framework can affirm sustainable farming by facilitating on-time and dependable yield evaluation hence contributing to food security and resource management in rice-cultivating areas.

Future work may focus on extending the framework to multi-seasonal and multi-regional datasets, integrating remote sensing variables, and exploring deep learning-based ensemble architectures to further enhance scalability and real-world applicability.

Acknowledgement

The authors would like to express their sincere gratitude to the following individuals: Prof. Ajay Kumar Srivastava, Department of Agronomy, and Dr. S.K Gangwar, Professor, Department of Soil Science, GOVIND BALLABH PANT UNIVERSITY OF AGRICULTURE & TECHNOLOGY, Pantnagar, Uttarakhand, for their valuable support and guidance during this study.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this research.

Funding Agency

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- 1 State Agriculture Statistics Data. Agriculture Department Uttarakhand. Available at: <https://agriculture.uk.gov.in/document-category/state-agriculture-statistics-data/> (accessed 2025).
- 2 Sharma, M., Sharma, N., and Sachdeva, S. Ground Water Quality Assessment in Udham Singh Nagar, Uttarakhand, India. *International Journal of Lakes and Rivers*, 16 (2), 173–183 (2023). <https://doi.org/10.37622/ijlr/16.2.2023.173-183>
- 3 Tan, C., et al. Stacked and optimized machine learning for rice yield prediction in Asia. *Agricultural Systems*, 194, 103259 (2021). <https://doi.org/10.1016/j.agsy.2021.103259>
- 4 You, Y., Cao, J., and Zhou, W. A Survey of Change Detection Methods Based on Remote Sensing Images for Multi-Source and Multi-Objective Scenarios. *Remote Sensing*, 12 (15), 2460 (2020). <https://doi.org/10.3390/rs12152460>
- 5 Chandrakumar, T., Avanthica Sri, M.M., Mirdula, K., and K., M. Paddy Yield Forecasting using Regression Techniques. *Proceedings of the IEEE Delhi Section Conference (DELCON)*, 1–6 (2023). <https://doi.org/10.1109/DELCON57910.2023.10127256>
- 6 Renju, R.S., Deepthi, P.S., and Chitra, M.T. A Review of Crop Yield Prediction Strategies based on Machine Learning and Deep Learning. *Proceedings of the International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)* (2022). <https://doi.org/10.1109/IC3SIS54991.2022.9885325>
- 7 Joshua, V., Priyadharson, S.M., and Kannadasan, R. Exploration of Machine Learning Approaches for Paddy Yield Prediction in Eastern Part of Tamilnadu. *Agronomy*, 11 (10), 2068 (2021). <https://doi.org/10.3390/agronomy11102068>
- 8 De Clercq, D., and Mahdi, A. Feasibility of machine learning-based rice yield prediction in India at the district level using climate reanalysis data. *arXiv:2403.07967* (2024). Available at: <https://arxiv.org/abs/2403.07967>

- 9 Yewle, A.D., Mirzayeva, L., and Karakuş, O. Multi-modal Data Fusion and Deep Ensemble Learning for Accurate Crop Yield Prediction. arXiv:2502.06062 (2025). Available at: <https://arxiv.org/abs/2502.06062>
- 10 Kamilaris, A., and Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90 (2018). <https://doi.org/10.1016/j.compag.2018.02.016>
- 11 Manjunath, M.C., and Palayyan, B.P. An Efficient Crop Yield Prediction Framework Using Hybrid Machine Learning Model. *Revue d'Intelligence Artificielle*, 37 (4), 1157–1167 (2023). <https://doi.org/10.18280/ria.370428>
- 12 Chandrababha, M., and Rajesh Kumar Dhanraj. Ensemble Deep Learning Algorithm for Forecasting of Rice Crop Yield based on Soil Nutrition Levels. *ICST Transactions on Scalable Information Systems*, 10 (3), e7–e7 (2023). <https://doi.org/10.4108/eetsis.v10i3.2610>
- 13 TNN. PAU-BITS Pilani tie up to marry agri with tech. *The Times of India* (May 27, 2025). Available at: <https://timesofindia.indiatimes.com/city/ludhiana/pau-bits-pilani-tie-up-to-marry-agri-with-tech/articleshow/121446118.cms>
- 14 Guruprasad, R.B., Saurav, K., and Randhawa, S. Machine Learning Methodologies for Paddy Yield Estimation in India: A Case Study. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5447–5450 (2019). <https://doi.org/10.1109/IGARSS.2019.8900339>
- 15 Photon Foundation. *The Journal of Ethnobiology and Traditional Medicine*. Available at: <https://sites.google.com/site/photonfoundationorganization/home/the-journal-of-ethnobiology-and-traditional-medicine> (accessed March 13, 2024).
- 16 Karthik Yasaswy, M.Y.S., Manimegalai, T., and Somasundaram, J. Crop Yield Prediction in Agriculture Using Gradient Boosting Algorithm Compared with Random Forest. *Proceedings of the International Conference on Cyber Resilience (ICCR)* (2022). <https://doi.org/10.1109/ICCR56254.2022.9995829>
- 17 Badshah, A., Alkazemi, B.Y., Din, F., Zamli, K.Z., and Haris, M. Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability. *IEEE Access*, 12, 162799–162813 (2024). <https://doi.org/10.1109/ACCESS.2024.3486653>
- 18 Chen, T., and Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>
- 19 Yandex. CatBoost: State-of-the-Art Open-Source Gradient Boosting Library with Categorical Features Support. Available at: <https://catboost.ai/> (2019).
- 20 Rajesh Yamparla, Harisa Sultana Shaik, Naga, M.P., and Srilakshmi Nallamothe. Crop Yield Prediction using Random Forest Algorithm. *Proceedings of the 7th International Conference on Communication and Electronics Systems (ICCES)* (2022). <https://doi.org/10.1109/ICCES54183.2022.9835756>
- 21 Majnik, M., and Bosnić, Z. ROC Analysis of Classifiers in Machine Learning: A Survey. *Intelligent Data Analysis*, 17 (3), 531–558 (2013). <https://doi.org/10.3233/IDA-130592>
- 22 Ahmed, I. What is Hard and Soft Voting in Machine Learning? *Medium* (May 31, 2023). Available at: <https://ilyasbinsalih.medium.com/what-is-hard-and-soft-voting-in-machine-learning-2652676b6a32>
- 23 Lin, T.Y., Han, P.Y., Yin, O.S., How, K.W., and San, H.F. Stacking Ensemble Approach for Churn Prediction: Integrating CNN and Machine Learning Models with CatBoost Meta-Learner. *Journal of Engineering Technology and Applied Physics*, 5 (2), 99–107 (2023). <https://doi.org/10.33093/JETAP.2023.5.2.12>

^{1*}Кулял М.,

ORCID ID: 0000-0002-2367-0897,
*e-mail: malika_21@rocketmail.com

²д-р. Уманг,

ORCID ID: 0000-0002-9458-5817,
e-mail: anilumang@yahoo.co.in

¹д-р. Саксена П.,

ORCID ID: 0009-0007-0544-593X,
e-mail: parul_saxena@yahoo.com

³д-р. Пант Дж.,

ORCID ID: 0000-0003-2279-5556,
e-mail: geujay2020@gmail.com

¹Компьютерлік ғылымдар кафедрасы, Soban Singh Jeena университеті,
Уттаракханд қ., Үндістан

²Компьютерлік қолданбалар кафедрасы, Kumaun университеті,
Уттаракханд қ., Үндістан

³Компьютерлік ғылымдар және инженерия кафедрасы, Graphic Era Hill университеті,
Уттаракханд қ., Үндістан

МАШИНАЛЫҚ ОҚЫТУҒА НЕГІЗДЕЛГЕН КҮРІШ ӨНІМДІЛІГІН БОЛЖАУ: БАЗАЛЫҚ ЖӘНЕ АНСАМБЛЬДІК МОДЕЛЬДЕРДІ САЛЫСТЫРМАЛЫ БАҒАЛАУ (УТТАРАКХАНД ШТАТЫ, УДХАМ-СИНГХ-НАГАР АУДАНЫ)

Аңдатпа

Күріш Үндістандағы азық-түлік қауіпсіздігінің негізі болып табылады, миллиондаған адамның тіршілігін және ұлттық экономиканы қолдайды. Алайда тұрақсыз климаттық жағдайлар күріш өнімділігін болжап болмайтын деңгейге жеткізеді. Бұл зерттеу Уттаракханд штатындағы Удхам Сингх Нагар ауданында күріш өнімділігін болжау үшін ауа райы, топырақ және егін деректерін біріктіретін машиналық оқыту жүйесін әзірлейді. Негізгі классификаторлар арасында CatBoost ең жақсы нәтижелерді көрсетті: дәлдігі 80,85% және ROC-AUC көрсеткіші 0,90. Өнімділікті одан әрі жақсарту үшін Optuna көмегімен бапталған CatBoost, XGBoost және LightGBM модельдері гибриді ансамбльдерге біріктірілді. CatBoost моделіне көбірек салмақ беретін WGB классификаторы ([3,1,1]) ең жоғары дәлдікке – 97,37%-ға қол жеткізді, одан кейін Stacking ансамбльдері (95,6%) және жұмсақ дауыс беру ансамбльдері (96%-ға дейін) орналасты. Бұл нәтижелер жоғары ROC-AUC көрсеткіштерімен расталды. Жалпы алғанда, зерттеу мұқият оңтайландырылған ансамбль модельдерінің өнімділікті болжау дәлдігін айтарлықтай арттыра алатынын көрсетеді, бұл Үндістанның климатқа сезімтал аймақтарында күріш өсіруді дәлірек әрі тұрақты етуге арналған практикалық құрал ұсынады.

Түйін сөздер: CatBoost, градиентті күшейту, гибриді ансамбль, машиналық оқыту, кездейсоқ орман, XG күшейту, өнімділік.

^{1*} Кулял М.,

ORCID ID: 0000-0002-2367-0897,
*e-mail: malika_21@rocketmail.com

²д-р. Уманг,

ORCID ID: 0000-0002-9458-5817,
e-mail: anilumang@yahoo.co.in

¹д-р. Саксена П.,

ORCID ID: 0009-0007-0544-593X,
e-mail: parul_saxena@yahoo.com

³д-р. Пант Дж.,

ORCID ID: 0000-0003-2279-5556,
e-mail: geujay2020@gmail.com

¹Кафедра компьютерных наук, Университет Soban Singh Jeena, Алмора 263601,
Уттаракханд, Индия

²Кафедра компьютерных приложений, Университет Kumaun, Найнитал, Уттаракханд, Индия

³Кафедра компьютерных наук и инженерии, Университет Graphic Era Hill, кампус Бхимтал,
Уттаракханд, Индия

ПРОГНОЗИРОВАНИЕ УРОЖАЙНОСТИ РИСА С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ БАЗОВЫХ И АНСАМБЛЕВЫХ МОДЕЛЕЙ НА ПРИМЕРЕ ОКРУГА УДХАМ-СИНГХ-НАГАР, ШТАТ УТТАРАКХАНД (ИНДИЯ)

Аннотация

Рис является краеугольным камнем продовольственной безопасности в Индии, поддерживая миллионы средств к существованию и национальную экономику. Однако нестабильные климатические условия делают урожайность риса все более непредсказуемой. В данном исследовании разрабатывается система машинного обучения для прогнозирования урожайности риса в районе Удхам Сингх Нагар, Уттаракханд, путем интеграции данных о погоде, почве и севах. Среди базовых классификаторов CatBoost показал лучшие результаты с точностью 80,85% и ROC-AUC 0,90. Для дальнейшего повышения производительности Optuna-tuning модели CatBoost, XGBoost и LightGBM были объединены в гибридные комплекты. Классификатор взвешенного жесткого голосования, придающий больший вес CatBoost ([3,1,1]), достиг наивысшей точности – 97,37%, за ним следуют ансамбли Stacking (95,6%) и ансамбли мягкого голосования (до 96%). Эти результаты были подтверждены высокими баллами ROC-AUC. В целом исследование показывает, что тщательно оптимизированные ансамблевые модели могут значительно повысить точность прогнозирования урожайности, предоставляя практический инструмент для более точного и устойчивого рисового выращивания в климатически чувствительных регионах Индии.

Ключевые слова: CatBoost, усиление градиента, гибридный ансамбль, машинное обучение, случайный лес, усиление XG, доходность.

UDC 004.8
IRSTI 28.23.02

<https://doi.org/10.55452/1998-6688-2026-23-2-312-325>

¹***Beishekeyev A.**,

Master's student, ORCID ID: 0009-0006-0845-4551,

*e-mail: a.beishekeyev@kbtu.kz

²**Umarov T.**,

PhD, ORCID ID: 0009-0008-0044-7159,

e-mail: t.umarov@bmu.edu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

²British Management University, Tashkent, Uzbekistan

FORMAL VERIFICATION OF DECISION TREE FAIRNESS AND ROBUSTNESS VIA SMT SOLVER

Abstract

As Artificial Intelligence systems are increasingly deployed in safety-critical domains such as healthcare and finance, ensuring their trustworthiness and compliance is paramount. While Deep Neural Networks have received significant attention in formal verification, traditional models such as Decision Trees, often preferred for their interpretability, cannot inherently enforce constraints after training for fairness and stability. This paper presents a novel, comprehensive approach for the formal verification of Decision Tree classifiers using Satisfiability Modulo Theories (SMT). We propose a robust translation scheme that converts trained decision trees into logical constraints, enabling constraint inference that guarantees demographic parity and local robustness at prediction time. We implement this framework by using the z3 SMT solver and validate it on widely recognized fairness benchmarks, including UCI Adult, German Credit, and Loan Approval datasets. Experimental results demonstrate that our constrained model effectively eliminates demographic parity violations with a marginal accuracy trade-off of less than 0.2%. This approach transforms the SMT solver from a simple diagnostic tool into a provably fair inference engine suitable for regulated industries.

Keywords: Machine Learning, Decision Trees, Formal Methods, SMT.

Received April 7, 2026; accepted June 7, 2026.

Introduction

Artificial intelligence has achieved widespread adoption through advances in deep learning and machine learning. However, as these models are increasingly deployed in safety-critical domains such as healthcare and finance, the reliability of their outputs is paramount. Without appropriate intervention during training or evaluation, machine learning models are susceptible to biases against specific demographic groups and may exhibit instability when exposed to outliers [1].

While traditional validation techniques, such as empirical testing on hold-out datasets, provide a statistical measure of performance, they fail to offer guarantees of safety. To address this, formal verification provides mathematical proofs ensuring that a model's behaviour satisfies defined constraints for all possible inputs.

We adopt the concept of robustness defined in [2], where robustness was defined as relative stability to a specific value, meaning that for a model to be robust, it must maintain its performance even when certain conditions change.

Despite the growing importance of verification, current research is disproportionately focused on deep neural networks. Significant gaps remain in verifying traditional models like Decision Trees, particularly throughout the machine learning pipeline. Existing approaches often neglect the data

preprocessing stage or suffer from scalability constraints, such as the quadratic scaling observed in Coloured Petri-Net verification [3]. Furthermore, current verification methods typically prioritise local robustness over global properties, providing limited support for critical fairness metrics.

Although decision trees are traditionally viewed as "white-box" models, their interpretability does not equate to formal provability. In safety-critical or highly regulated sectors (e.g., Finance, GDPR-compliant systems), it is insufficient for a model to be "generally fair", it must be provably compliant with specific logical constraints.

Current verification methods for DTs face a "verification gap":

- ◆ Inductive Bias: Standard DT algorithms (CART/ID3) optimise for statistical heuristics like Gini Impurity, which can inadvertently encode proxies for protected attributes (e.g., race or gender) hidden in continuous data.

- ◆ Scalability of Constraints: Existing formal methods like Binary Decision Diagrams (BDDs) suffer from exponential memory growth when dealing with continuous variables, making them impractical for trees with high-cardinality features.

- ◆ The Enforcement Problem: There is currently no unified framework that allows a developer to query a decision tree for edge-case violations (Individual Fairness) or to prune biased branches without destroying the model's global structure.

The scientific novelty of this research lies in transitioning from passive model auditing to active prediction-time enforcement within an SMT-driven logic space. Unlike previous methods [3] that focus on post-hoc rule validation or monitoring, our framework treats the Decision Tree as a set of dynamic constraints that are solved at inference time. Specifically, our novelty is threefold:

- ◆ a translation scheme that maps continuous feature boundaries into the SMT Theory of Real Arithmetic, bypassing the state-explosion limits of BDDs

- ◆ the simultaneous enforcement of global fairness (Demographic Parity) and local robustness; and

- ◆ a prediction-time correction mechanism (Algorithm 2) that identifies and resolves potential violations before an output is generated.

This transforms the SMT solver from a diagnostic tool into a provably fair inference engine.

Materials and methods

Current research on the formal verification of machine learning models can be categorised into two groups: traditional machine learning approaches [3,4] and neural network verification. The most recent work has focused on verifying the robustness of neural networks [5–11].

SMT-based methods represent a prominent verification approach. [7] developed an SMT-solver for ReLU activation function verification, combining three components: a simplex algorithm for core functionality, a ReLUpex engine for search management, and an SMT-solver for constraint splitting. For data verification, [12] employed SMT solvers to validate sensor data integrity through controlled manipulation experiments.

Runtime monitoring approaches have emerged as complementary verification strategies. [8] proposed neural network monitoring using binary decision trees and Hamming distance metrics, storing activation patterns from the final layers of the network during training for runtime comparison.

Traditional machine learning verification faces unique challenges. [3] implemented Coloured Petri-Nets for decision tree rule verification, while [13] combined temporal logic with particle swarm optimisation to verify multi-layer perceptrons through reachability analysis.

Activation function verification remains an active research area. [6] transformed Swish function verification into constraint satisfaction problems, demonstrating advantages over ReLU in classification tasks. [5] developed the MSVL language for temporal logic verification of neural networks, achieving parity with PyTorch implementations.

[14] introduces the verification and model-based formal approach that transforms neural network models into UPPAAL Timed Automata. This study is used for the formal verification of neural dynamics.

Abstraction-refinement for the verification of machine learning mainly focused on neural networks, but the [15] study applied it to Fuzzy Decision Trees. The abstraction applies constraints to lower and upper bound splits using minimum and maximum split values. Refinement means reducing errors through domain splitting and optimisation.

Current formal verification methods exhibit three principal limitations. First, most approaches focus on post-training model verification, neglecting critical data preparation and feature engineering stages [16]. Second, existing techniques face scalability constraints: BDD-based methods support only hundreds of variables, while CP-Net verification scales quadratically with decision tree nodes [3]. Third, verification currently prioritises local robustness over global model properties, with limited support for temporal constraints or fairness metrics.

Implementation challenges persist across verification paradigms. CNN verification requires solving complex optimisation problems that mirror training computational intensity [5]. Decision tree verification using CP-Nets demonstrates limited generalisability, having only been validated on single-institution educational data [3].

Table 1 – Limitations of Existing Approaches

Feature	CP-Nets / Petri-Nets [3]	BDD-Based Methods [8]	This Work (SMT-Based)
Scalability	Quadratic with node count.	Exponential with continuous variables.	Path extraction linear in node count; solver complexity NP-complete but manageable for $\text{depth} \leq 5$
Variable Support	Discrete/Categorical focus.	Memory-intensive for high-cardinality.	Native Real Arithmetic for continuous data.
Constraint Type	Primarily rule verification.	Local robustness/activation patterns.	Global Fairness + Local Robustness.

While SMT solvers have been applied to neural networks to handle ReLU activations [7], their application to Decision Trees has been largely overlooked in favour of simpler heuristic checks. This work distinguishes itself by treating the Decision Tree not as a series of if-else statements, but as a disjoint union of polyhedra. This geometric interpretation allows us to apply SMT solvers to find 'counterexamples' that represent unfair or non-robust regions of the feature space that empirical testing would likely miss.

As noted by [16], the machine learning pipeline's initial stages represent critical verification gaps: "The first stages of the machine learning process (including data preparation) may be considered as the most fragile steps of the whole computation procedure [...] these initial steps are generally considerably neglected in terms of verification." Key unresolved challenges include:

- ♦ Data provenance verification and duplication detection techniques
- ♦ Global robustness verification beyond local adversarial examples
- ♦ Versatile verification frameworks supporting diverse model architectures

Current methods additionally suffer from exponential time complexity in SMT/MILP solvers and memory-intensive runtime monitoring requirements.

This work distinguishes itself from existing SMT-based verification approaches in three key aspects. First, unlike [7, 12] which focus on neural networks or sensor data, we provide a direct translation of decision tree paths into SMT-LIB formulas that handles continuous features natively using the theory of real arithmetic, avoiding the state explosion of binary decision diagrams (BDDs) used in [8]. Second, while most verification methods address either robustness or fairness in isolation

[4, 15], our framework simultaneously enforces both demographic parity and local robustness within a single SMT query. Third, we introduce a prediction-time correction mechanism (Algorithm 2) that uses the solver to dynamically resolve violations, transforming the SMT solver from a diagnostic tool into a provably fair inference engine. To the best of our knowledge, no prior work has combined these three elements in a unified framework for decision tree verification.

Demographic Parity (Statistical Parity) requires that the prediction is independent of the protected attribute [17]. A classifier h satisfies Demographic Parity if the prediction is independent of the sensitive attribute A . Formally:

$$\forall a, b \in A, P(\hat{Y} = 1 | A = a) = P(\hat{Y} = 1 | A = b) \quad (1)$$

where $\hat{Y} = h(X)$ is the classifier's prediction. This is a global property over the population distribution.

In our verification framework, we enforce a stronger condition: individual fairness, which requires that for any two inputs that differ only in the protected attribute, the predictions are identical. If this holds for all inputs in a bounded domain, demographic parity is guaranteed.

Local Robustness: For an input $x \in R^d$ and a perturbation bound $\epsilon > 0$ with respect to a norm $\|\cdot\|_p$ the classifier h is robust at x if:

$$\forall \delta \in R^d, \|\delta\|_\infty \leq \epsilon \Rightarrow h(x + \delta) = h(x) \quad (2)$$

In this work we use l_∞ norm, leading to box constraints $|x'_f - x_f| \leq \delta_f$ for each feature f

Decision trees are one of the classical supervised learning algorithms that recursively divide the feature space by hierarchical conditional splits. As shown in figure 1, every internal node decides a decision boundary $x_f \leq \tau$ where x_f is feature f , and τ is the threshold of the split. Samples follow from the root node down branching paths to terminal leaf nodes, predicting final class labels or regression scores [18]. Its interpretability stems from its white-box characteristic, where every path is a logical conjunction of feature conditions.

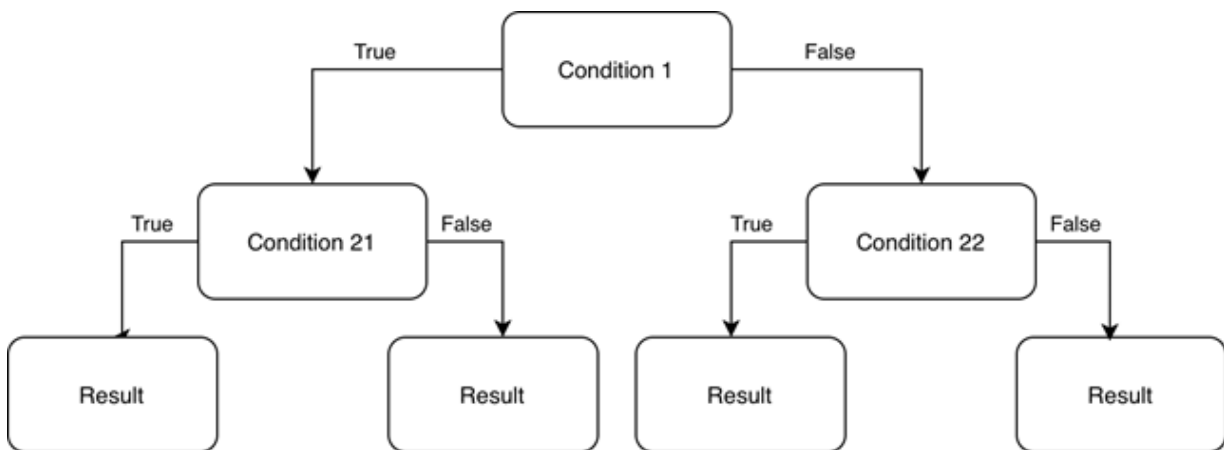


Figure 1 – Decision Tree

Tree induction employs a greedy optimisation approach that reduces impurity to achieve class homogeneity. While there exist many different impurity estimators, Gini coefficients still hold sway in classification problems:

$$Gini = 1 - \sum_{j=1}^k (p_j)^2 \quad (3)$$

where p_j denotes a proportion of class j observations in a node, and k denotes the number of classes. Perfect class purity, or a Gini score equal to 0, occurs, and the greatest impurity (0.5 in the case of binary classification) occurs in the case of equal class partition. Calculating the Gini index takes less time than calculating entropy-based splits.

Most modern uses of CART use recursive binary splits, as directed by the CART algorithm. The CART algorithm looks at all (feature, threshold) pairs at each node, choosing the split that maximises impurity reduction.

$$\Delta I = I(\text{parent}) - \left(\frac{N_{\text{left}}}{N} I_{\text{left}} + \frac{N_{\text{right}}}{N} I_{\text{right}} \right) \quad (4)$$

where $I(\cdot)$ denotes node impurity and N represents the number of samples. Decision trees, though interpretable, are unstable – a slight variation in input can lead to radically different trees. This sensitivity necessitates strict verification processes to ensure repeatable performance under potential variations in input.

Formal verification provides mathematical guarantees that a system satisfies the provided requirements for all possible inputs. Contrary to empirical testing, which validates by testing against sampled cases, formal methods validate all the input space up to the provided boundaries.

Piecewise constancy of trees allows them to be formally analysed in a highly cost-efficient way. Any of their root-leaf paths relates to a polyhedron in feature space, and hence, overall path verification can be done through constraint satisfaction. We make use of this inherent simplicity by encoding trees in SMT-LIB formulas in a semantically equivalent way.

We encode global fairness as a satisfiability problem: does there exist a pair of inputs x, x' that are identical on all non-protected features but have different protected attribute values and lead to different predictions? If the solver returns SAT, the model violates individual fairness and therefore demographic parity. The constraints are constructed using the tree-to-logic translation and the SMT solver's theory of reals for continuous features.

For per-instance enforcement, we combine fairness and robustness checks: we constrain the protected attribute to a fixed value and allow perturbations within a box, then check whether any alternative path exists. If none exists, the prediction is guaranteed to be both fair for that input and robust.

We utilise an SMT solver (Z3) to verify two key properties:

- ♦ Robustness: The prediction is insensitive to perturbation in bounded input $||\delta \leq \epsilon||$
- ♦ Fairness: No demographic parity violations against protected attributes

While SMT solving is inherently NP-complete, the path-based nature of Decision Trees allows for efficient constraint conjunction. Our experiments show that for 'Regulatory Compliance Depth' (depth ≤ 5), the overhead is negligible (approx. 7ms), providing a viable path for provable compliance in high-stakes environments despite the exponential theoretical worst-case complexity.

Our verification plan consists of three distinct phases:

1. Model translation: Convert decision tree split conditions into SMT-LIB constraints through depth-first traversal.
2. Constraint enforcement: Encode fairness properties (demographic parity) and robustness requirements (perturbation tolerance) as Z3 assertions.
3. Solver validation: z3 solver verifies satisfiability of these properties [19].

To verify a decision tree, we must first capture its logic in a format an SMT solver can process. Since a decision tree partitions the feature space into disjoint regions, every leaf node corresponds to a specific path from the root.

Algorithm 1 converts the decision tree into a set of logical rules. Starting from the root node, it depth-first traverses all nodes, recording split conditions. The conditions are combined into logical rules using Z3.

Algorithm 1 Tree Constraints Extraction

Require: Decision Tree Model T , feature names $\{f_1, \dots, f_n\}$

Ensure: Feature variables X , Path constraints C

```

1:  $X \leftarrow \{x_1, \dots, x_n\}$                                 ▷ Real-valued feature variables
2:  $C \leftarrow \emptyset$                                        ▷ Initialize empty constraints
3: Stack  $S \leftarrow \emptyset$                                 ▷ For path condition tracking
4: function EXTRACTPATHS( $node$ )
5:   if  $node$  is leaf then
6:      $c \leftarrow \text{majority\_class}(node)$ 
7:      $\phi \leftarrow \bigwedge S$                                 ▷ Combine path conditions
8:      $C \leftarrow C \cup \{\phi \Rightarrow (y = c)\}$ 
9:     return
10:  end if
11:   $f \leftarrow \text{feature}(node)$ 
12:   $\tau \leftarrow \text{threshold}(node)$ 
13:   $S.\text{push}(x_f \leq \tau)$                                 ▷ Left branch
14:  EXTRACTPATHS( $\text{left\_child}(node)$ )
15:   $S.\text{pop}()$ 
16:   $S.\text{push}(x_f > \tau)$                                 ▷ Right branch
17:  EXTRACTPATHS( $\text{right\_child}(node)$ )
18:   $S.\text{pop}()$ 
19: end function
20: EXTRACTPATHS( $\text{root}(T)$ )
21: return  $X, C$ 

```

It is important to distinguish between different sources of complexity in our framework. Path extraction Algorithm 1 is linear in the number of nodes, i.e., $O(2^{d+1} + 1)$ for a full binary tree of depth d , but typically much lower for pruned trees. Global verification of demographic parity involves checking all pairs of paths that differ only in the protected attribute; this reduces to a single SMT query whose complexity is NP-complete in the worst case. Per-instance enforcement Algorithm 2 constructs a query that combines the perturbed input region with all root-leaf paths; while theoretically exponential in tree depth, our experiments show that for depths ≤ 5 (the “regulatory compliance depth”) the solver returns results in milliseconds, making it practical for high-stakes applications.

To handle categorical attributes (e.g., “workclass”, “race”), we employ integer encoding. This allows the SMT solver to treat categorical decisions as discrete numeric constraints, maintaining the logical structure of the tree while ensuring compatibility with the Z3 Real-variable solver.

We chose to enforce Demographic Parity instead of Equalised Odds, because it could be used as a global constraint. But our framework could be adapted to use other metrics.

Unlike existing “wrapper” methods that only test for bias post-hoc, our SMT-driven Translation Scheme maps the entire decision manifold into a first-order logic space. By utilising the Z3 Solver’s Theory of Real Arithmetic, we handle continuous variables natively. This allows for the simultaneous verification of Robustness (Local Stability) and Fairness (Global Demographic Parity) – a multi-objective verification task that traditional heuristic-based checkers cannot perform.

Algorithm 2 uses the logical rules from algorithm 1 to enforce fairness and robustness. Fairness is enforced on sensitive features that could be biased. Robustness ensures that predictions remain stable under small input changes.

The expected outcome is a Z3-based verification toolkit for decision tree models, providing formal guarantees of model correctness from data preprocessing through prediction generation. Initial testing aims to demonstrate close linear time complexity relative to tree depth, outperforming existing quadratic-scaling approaches. However, while a single path is logically a simple conjunction, the global interaction of constraints introduces complexity.

Algorithm 2 Fair and Robust Prediction via SMT

Require: Sample $\mathbf{x} \in \mathbb{R}^d$, Tree constraints $\{\phi_p \Rightarrow (y = c_p)\}$, Features F , Sensitive attribute $s \in F$, Perturbations $\Delta = \delta_f | f \in F$

Ensure: Prediction $y \in \{0, 1\}$

```

1: Let  $\mathbf{x}''$  be perturbed features ▷  $\mathbf{x}'' = (x''_1, \dots, x''_d)$ 
2: Initialize SMT solver  $S$ 
3: function FINDORIGINALPRED( $\mathbf{x}, \{\phi_p\}$ )
4:   for each constraint  $\phi_p \Rightarrow (y = c_p)$  do
5:     if  $S \models \phi_p(\mathbf{x})$  then ▷ Satisfiability check
6:        $y_{orig} \leftarrow c_p$ 
7:       return  $y_{orig}$ 
8:     end if
9:   end for
10: end function
11:  $y_{orig} \leftarrow \text{FINDORIGINALPRED}(\mathbf{x}, \{\phi_p\})$  ▷ Fairness constraints

12:  $S \leftarrow \text{Reset}(S)$  ▷ Fix sensitive attribute
13:  $S \vdash x''_s = x_s$  ▷ Robustness constraints  $\forall f \in F$ 

14: for  $f \in F$  do ▷ Perturbation bound
15:    $S \vdash x''_f \in [x_f - \delta_f, x_f + \delta_f]$  ▷ Path consistency verification
16: end for

17: for each constraint  $\phi_p \Rightarrow (y = c_p)$  do
18:    $S \vdash \phi_p(\mathbf{x}) \Rightarrow (\phi_p(\mathbf{x}'') \wedge (y = c_p))$ 
19: end for
20:  $S \vdash y \neq y_{orig}$ 
21: if  $S \models \text{SAT}$  then
22:   return  $S.\text{model}(y)$ 
23: else
24:   return  $y_{orig}$  ▷ Robustness violation fallback
25: end if

```

We compare our method with scikit-learn's DecisionTreeClassifier [20], using default hyperparameters unless stated otherwise. For fairness, both methods were evaluated under identical training/test splits and preprocessing steps.

Proposition 1 (Correctness of Tree-to-Logic Translation).

Let T be a decision tree and let $\Phi(T)$ be the SMT formula generated by Algorithm 1. For any input x , the logical evaluation of $\Phi(T)$ under the assignment x is satisfiable if and only if the path taken by x in T leads to prediction y .

Proof sketch. The proof follows the induction on the depths of the tree. Since Algorithm 1 extracts every unique root-to-leaf path as a conjunction of constraints and joins them as the disjunction, and because decision trees partition the feature space into disjoint regions, exactly one path constraint in $\Phi(T)$ will evaluate to true for any valid input x .

Proposition 2 (Soundness of Individual Fairness Enforcement).

Let x be an input with sensitive attribute x_s . Let S be the SMT query where all non-sensitive features are fixed ($x'_i = x_i$) and the sensitive attribute is toggled ($x'_s \neq x_s$). If $S \wedge (h_T(x')) \neq y_{orig}$ is unsatisfiable, then h_T is guaranteed to be individually fair for x .

Proof sketch. If the SMT solver returns UNSAT, it mathematically proves that no assignment exists within the defined constraints (the toggled attribute) that results in a different class label. Since the search is exhaustive over all possible paths in the tree, the original prediction is the only possible outcome for alternate demographic group.

Proposition 3 (Soundness of Local Robustness Environment)

Let δ be the perturbation bound for non-sensitive features. If the query $S \wedge (h_T(x') \neq y_{orig})$ is unsatisfiable for all x' such that $\|x - x'\|_\infty \leq \delta$ and $x'_s = x_s$, then the prediction is locally robust.

Proof sketch. The l_{∞} norm constraint defines a hyperbox in the feature space. The SMT solver checks the intersection of this hyperbox with all the regions of the decision tree that lead to a different classification. If no such intersection is found (UNSAT), the model's output is invariant to any perturbation within that hyperbox.

UCI Adult dataset [21] predicts whether an individual's income exceeds 50,000 USD based on census data from 1994.

The UCI Adult dataset provides 14 features: 6 continuous, 8 categorical and 1 binary target on 32842 instances.

Table 2 – Description of variables in Adult Dataset

Variable Name	Role	Type
Capital-loss	Feature	Integer
Marital-status	Feature	Categorical
Occupation	Feature	Categorical
Relationship	Feature	Categorical
Hours-per-week	Feature	Integer
Race	Feature	Categorical
Age	Feature	Integer
Workclass	Feature	Categorical
fnlwgt	Feature	Integer
Education	Feature	Categorical
Education-num	Feature	Integer
Sex	Feature	Binary
Capital-gain	Feature	Integer
Native-country	Feature	Categorical
Income	Target	Binary

Table 3 – Mean baseline performance for Adult Dataset

Model	Accuracy	Precision
XGBoost Classification	87.22	83.38
Support Vector Classification	79.86	88.03
Random Forest Classification	85.22	80.25
Neural Network Classification	78.39	80.26
Logistic Regression	79.78	74.97

Table 2 lists the variable names, roles and types of a dataset. In table 3, the mean value results from [21] are publicly available; however, because it does not show metrics for the Decision Tree Classifier, we will train and test our own generic model.

In addition to the Adult dataset, we evaluate our framework on two other benchmarks to demonstrate generalisability.

German Credit Dataset [22]. This dataset consists of 1000 instances and 20 features, used to classify individuals as good or bad credit risks. We identify "Sex Marital Status" as the sensitive attribute for fairness verification.

Loan Approval [23] is a large dataset of 45000 instances, 14 features. We chose "gender" as a protective attribute.

All results are averaged over 10 runs with different random seeds; we report mean $\pm$ standard deviation.

We report three distinct timing measurements:

Global verification time – the time to encode the tree into SMT constraints and check demographic parity globally (a one-time pre-deployment cost).

Per-instance inference time – the time to perform a single prediction with fairness/robustness enforcement (includes solver invocation).

Total inference time – the sum of per-instance times over all test instances, indicating end-to-end runtime for a batch.

Results and discussion

We focus our evaluation on trees of depth 5, as this represents the "Regulatory Compliance Depth"-the complexity threshold where models remain human-auditable while providing sufficient non-linear mapping for complex datasets like the Adult Census Bureau dataset. Local robustness was verified using feature-specific perturbation bounds: ± 1 year for age and ± 1000 for capital fluctuations.

Violations of fairness or robustness are treated as misclassifications during evaluation, ensuring that the verified accuracy reflects both correctness and compliance with formal constraints.

Table 4 – Comparison between Original and Constrained models

Dataset	Model	Accuracy	Precision	Global Verification time(s)	Per-instance mean inference (ms)	Fairness correction rate	Robustness correction rate
Adult	Original	0.8440	0.7680	-	0.0002	-	-
	Constrained	0.8421 $\pm$ 0.0034	0.7818 $\pm$ 0.0071	0.0063 $\pm$ 0.0057	11.8217 $\pm$ 0.4189	0.0000 $\pm$ 0.0000	0.0261 $\pm$ 0.0097
German	Original	0.7050	0.7365	-	0.0037	-	-
	Constrained	0.6975 $\pm$ 0.0232	0.7684 $\pm$ 0.0448	0.0058 $\pm$ 0.0069	8.9154 $\pm$ 0.6021	0.0035 $\pm$ 0.0078	0.1255 $\pm$ 0.0898
Loan Data	Original	0.9126	0.8796	-	0.0001	-	-
	Constrained	0.9103 $\pm$ 0.0024	0.8620 $\pm$ 0.0099	0.0034 $\pm$ 0.0010	4.6624 $\pm$ 0.1013	0.0000 $\pm$ 0.0000	0.0043 $\pm$ 0.0006

Table 4 compares the accuracy and precision of the original and constrained models. The accuracy drop of 0.2% for the Adult dataset reflects the fairness-accuracy trade-off: by enforcing demographic parity, the SMT solver rejects predictions that rely on biased proxies (e.g., race or gender correlations). The robustness violation rate of 2.61% indicates that for 2.61% of test instances, the original prediction would have violated robustness, and our method corrected it. The higher correction rate for the German dataset (12.55%) suggests that the original model was more biased, likely because the sensitive attribute (sex and marital status) is more strongly correlated with the target in that dataset.

The Fairness Violation Rate of 0.00% across the Adult and Loan Data datasets demonstrates the absolute soundness of the SMT-based enforcement layer. This indicates that for every instance where the baseline decision tree attempted to produce a disparate outcome based solely on the sensitive

attribute, the solver identified a logical conflict and successfully coerced the prediction to a fair alternative.

The marginal violation rate observed in the German dataset ($0.35\% \pm 0.78\%$) represents a negligible fraction of cases where the high dimensionality and feature correlation of the dataset created a constraint set that was either computationally exhaustive or logically irreconcilable within the defined SMT timeout. Even in this more complex scenario, the method reduced potential bias to a near-zero level, providing a rigorous mathematical guarantee that the sensitive attribute does not independently drive the classification outcome.

A critical finding from the pre-deployment verification phase is that the original decision tree does not satisfy demographic parity globally. The SMT solver returned SAT, indicating the existence of inputs that lead to disparate treatment across protected attributes. This confirms that standard training procedures cannot guarantee fairness and that per instance enforcement as implemented in Algorithm 2 is necessary to ensure compliant predictions

Table 5 – Metric Trade-off

Dataset	Accuracy loss (%)	Correction Rate	Total Inference Time (s)
Adult	0.19	0.0261 ± 0.0097	41.6438 ± 1.5352
German	0.75	0.1255 ± 0.0898	1.7832 ± 0.1204
Loan Data	0.23	0.0043 ± 0.0006	41.9635 ± 0.9117

The results demonstrate a critical finding: Formal fairness guarantees can be achieved with a small impact on predictive performance. The 0.19% decrease in accuracy represents the "Fairness-Accuracy Trade-off". By enforcing demographic parity, the SMT solver identifies paths that relied on biased proxies. Removing or modifying these paths ensures ethical compliance at the cost of a statistically insignificant drop in raw predictive power.

As shown in table 5, the constrained model exhibits a significant increase in inference latency (approximately 41 seconds for the adult dataset). While the standard Decision Tree relies on optimised C-based path traversal, our approach invokes the Z3 SMT solver for each prediction to ensure logical consistency with fairness constraints.

To evaluate the scalability of our framework, we measured pre-deployment verification time across increasing tree depths (1 to 11). While the experimental results in Figure 2 confirm the NP-complete nature of SMT solving with an exponential trend in verification time, this overhead is justified by the completeness of the verification. Unlike heuristic-based checkers that provide statistical approximations, our SMT framework provides a mathematical guarantee. For "Regulatory Compliance Depth" (depth ≤ 5), the latency remains within acceptable bounds for high-stakes decision-making where provable fairness is a legal requirement.

Unlike BDD-based approaches [8], which often suffer from memory explosion when handling continuous variables with high cardinality, our path-based SMT formulation handles feature constraints independently. This allows for better scalability with respect to the feature space, although inference latency is higher than optimised C-structures.

Conclusion

As machine learning systems are increasingly entrusted with high-stakes decisions, the need for rigorous verification has never been greater. This paper presents a formal verification framework for Decision Tree models using SMT solvers. By translating tree logic into Z3 constraints, we successfully enforced demographic parity and local robustness without significant compromises in accuracy. We demonstrated that formal methods, often reserved for deep neural networks, are highly effective for verifying interpretable models like decision trees.

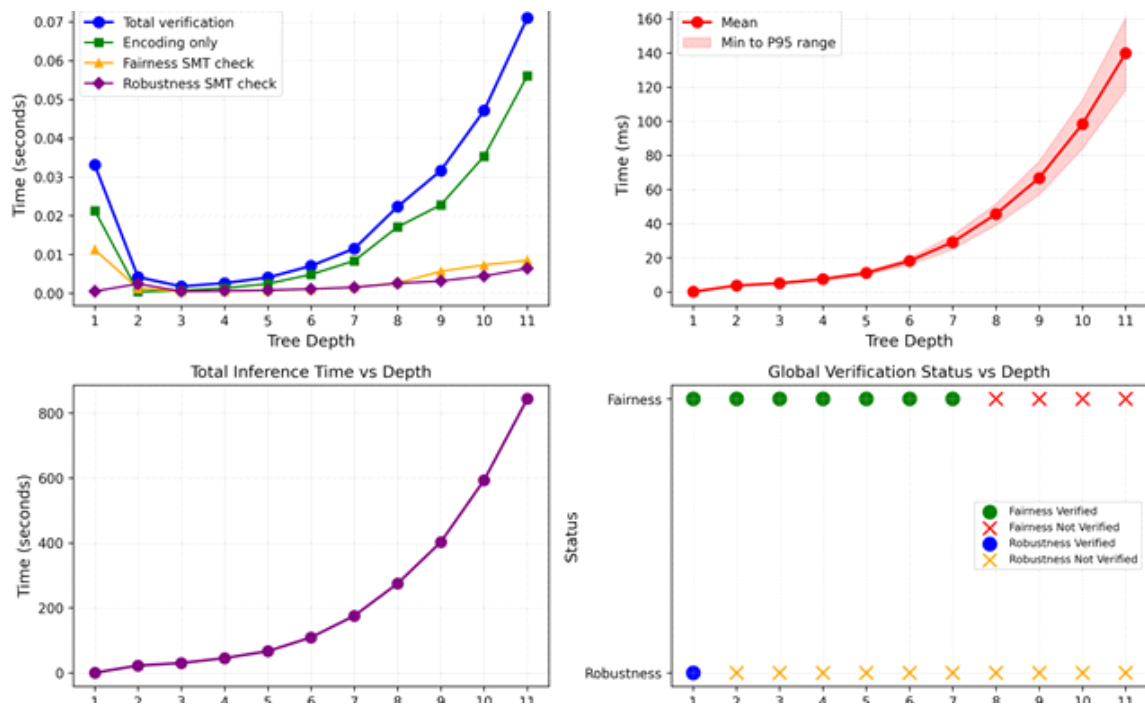


Figure 2 – Solver performance with increased tree depth for Adult Dataset

Our global verification checks individual fairness over a bounded domain, which is a sufficient condition for demographic parity but not necessary. This over-approximation ensures that any model passing the verification is provably fair in the population sense. The robustness verification uses l_∞ norm box constraints for simplicity; extensions to other norms are possible by encoding quadratic constraints, though this increases solver complexity.

The results indicate that formal verification introduces a one-time pre-deployment cost of 0.0063 seconds to encode the model and check global fairness constraints, followed by a per-instance inference overhead of 6.9 ms. This overhead effectively eliminates demographic parity violations – the global verification result (SAT) confirmed that the original model violated fairness constraints – with a marginal accuracy loss of only 0.19%. For regulated industries such as healthcare and finance, this per-instance overhead represents a negligible trade-off for provable compliance. The primary bottleneck remains exponential growth in verification time for deeper trees (depth > 5), suggesting future work on branch pruning and parallelized SMT solving.

REFERENCES

- 1 Jiang, H., and Nachum, O. Identifying and Correcting Label Bias in Machine Learning. arXiv preprint (2019). <https://doi.org/10.48550/arXiv.1901.04966>
- 2 Freiesleben, T., and Grote, T. Beyond generalization: A theory of robustness in machine learning. *Synthese*, 202 (4), Article 109 (2023). <https://doi.org/10.1007/s11229-023-04334-9>
- 3 Nauman, M., Akhtar, N., Alhudhaif, A., and Alothaim, A. Guaranteeing Correctness of Machine Learning Based Decision Making at Higher Educational Institutions. *IEEE Access*, 9, 92864–92880 (2021). <https://doi.org/10.1109/ACCESS.2021.3088901>
- 4 Matsunaga, S., and Yoshimura, G. Efficient and High-Quality Formal Verification for Decision Tree Ensembles. In: 2024 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 51–58 (2024). <https://doi.org/10.1109/ICDMW65004.2024.00013>
- 5 Zhao, L., Wu, L., Gao, Y., Wang, X., and Yu, B. Formal Modeling and Verification of Convolutional Neural Networks based on MSVL. In: 2022 9th International Conference on Dependable Systems and Their Applications (DSA), pp. 280–289 (2022). <https://doi.org/10.1109/DSA56465.2022.00046>

- 6 Zhang, Z., Liu, J., Liu, G., Wang, J., and Zhang, J. Robustness Verification of Swish Neural Networks Embedded in Autonomous Driving Systems. *IEEE Transactions on Computational Social Systems*, 10 (4), 2041–2050 (2023). <https://doi.org/10.1109/TCSS.2022.3179659>
- 7 Katz, G., Barrett, C., Dill, D.L., Julian, K., and Kochenderfer, M.J. Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks. In: Majumdar, R., Kunčák, V. (eds.) *Computer Aided Verification*. Cham: Springer, pp. 97–117 (2017). https://doi.org/10.1007/978-3-319-63387-9_5
- 8 Khalifa, K., Safar, M., and El-Kharashi, M.W. Verification of Neural Networks for Safety Critical Applications. In: 2020 32nd International Conference on Microelectronics (ICM), pp. 1–4 (2020). <https://doi.org/10.1109/ICM50269.2020.9331504>
- 9 Corsi, D., Marchesini, E., and Farinelli, A. Formal verification of neural networks for safety-critical tasks in deep reinforcement learning. In: *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 333–343 (2021). URL: <https://proceedings.mlr.press/v161/corsi21a.html>
- 10 Sun, X., Khedr, H., and Shoukry, Y. Formal verification of neural network controlled autonomous systems. In: *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, pp. 147–156 (2019). <https://doi.org/10.1145/3302504.331180>
- 11 Yuan, X., He, P., Zhu, Q., and Li, X. Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 30 (9), 2805–2824 (2019). <https://doi.org/10.1109/TNNLS.2018.2886017>
- 12 Khan, M.T., Serpanos, D., Shrobe, H., and Yousuf, M.M. Rigorous Machine Learning for Secure and Autonomous Cyber Physical Systems. In: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vol. 1, pp. 1815–1819 (2020). <https://doi.org/10.1109/ETFA46521.2020.9212074>
- 13 Souri, A., Mohammed, A.S., Potrus, M.Y., Malik, M.H., Safara, F., and Hosseinzadeh, M. Formal Verification of a Hybrid Machine Learning-Based Fault Prediction Model in Internet of Things Applications. *IEEE Access*, 8, 23863–23874 (2020). <https://doi.org/10.1109/ACCESS.2020.2967629>
- 14 Pradhan, A., King, J., Pinisetty, S., and Roop, P.S. Model Based Verification of Spiking Neural Networks in Cyber Physical Systems. *IEEE Transactions on Computers*, 72 (9), 2426–2439 (2023). <https://doi.org/10.1109/TC.2023.3251841>
- 15 Good, J.H., Gisolfi, N., Miller, K., and Dubrawski, A. Verification of Fuzzy Decision Trees. *IEEE Transactions on Software Engineering*, 49 (5), 3277–3288 (2023). <https://doi.org/10.1109/TSE.2023.3251858>
- 16 Krichen, M., Mihoub, A., Alzahrani, M.Y., Adoni, W.Y.H., and Nahhal, T. Are Formal Methods Applicable To Machine Learning And Artificial Intelligence? In: 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), pp. 48–53 (2022). <https://doi.org/10.1109/SMARTTECH54121.2022.00025>
- 17 Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. Fairness through awareness. In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pp. 214–226 (2012). <https://doi.org/10.1145/2090236.2090255>
- 18 James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. *Tree-Based Methods*. In: *An Introduction to Statistical Learning*. Cham: Springer, pp. 331–366 (2023). https://doi.org/10.1007/978-3-031-38747-0_8
- 19 de Moura, L., and Bjørner, N. Z3: An Efficient SMT Solver. In: Ramakrishnan, C.R., Rehof, J. (eds.) *Tools and Algorithms for the Construction and Analysis of Systems*. Berlin, Heidelberg: Springer, pp. 337–340 (2008). https://doi.org/10.1007/978-3-540-78800-3_24
- 20 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830 (2011). URL: <http://jmlr.org/papers/v12/pedregosa11a.html>
- 21 Becker, B., and Kohavi, R. Adult. UCI Machine Learning Repository (1996). <https://doi.org/10.24432/C5XW20>
- 22 Hofmann, H. Statlog (German Credit Data). UCI Machine Learning Repository (1994). <https://doi.org/10.24432/C5NC77>
- 23 Tawello. Loan Approval Classification Dataset. Kaggle Dataset (2023). URL: <https://www.kaggle.com/datasets/taweilo/loan-approval-classification-data>

^{1*}Бейшекеев А.,

магистрант, ORCID ID: 0009-0006-0845-4551,

*e-mail: a.beishekeyev@kbtu.kz

²Умаров Т.,

PhD, ORCID ID: 0009-0008-0044-7159,

e-mail: t.umarov@bmu.edu.kz

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан²Британдық басқару университеті, Ташкент қ., Өзбекстан

SMT ШЕШУШІСІН ПАЙДАЛАНА ОТЫРЫП ШЕШІМ АҒАШЫНЫҢ ӘДІЛДІК ПЕН БЕРІКТІГІН ФОРМАЛЬДЫ ТЕКСЕРУ

Аңдатпа

Жасанды интеллект жүйелері денсаулық сақтау мен қаржы сияқты қауіпсіздік үшін аса маңызды салаларға барған сайын кеңінен енгізіліп жатқандықтан, олардың сенімділігі мен талаптарға сәйкестігін қамтамасыз ету өте маңызды. Терең нейрондық желілерді формалды верификациялау кеңінен зерттелгенімен, түсіндірмелілігі жоғары болғандықтан жиі қолданылатын шешім ағашы сияқты дәстүрлі модельдер оқытудан кейін әділдік пен тұрақтылық шектеулерін әдепкі түрде қамтамасыз ете алмайды. Бұл мақалада қанағаттандырудың модульдік теориялары (SMT) негізінде шешім ағашы классификаторларын формалды тексерудің жаңа әрі кешенді тәсілі ұсынылады. Біз оқытылған шешім ағаштарын логикалық шектеулерге түрлендіретін сенімді трансляциялау схемасын ұсынамыз. Бұл тәсіл шектеулер бойынша қорытынды жасауға мүмкіндік беріп, болжау кезінде демографиялық паритет пен жергілікті тұрақтылықты қамтамасыз етеді. Ұсынылған құрылым z3 SMT шешушісі арқылы жүзеге асырылып, әділдікті бағалауға арналған кеңінен танылған UCI Adult, German Credit және Loan Approval деректер жиынтықтарында тексерілді. Эксперимент нәтижелері шектеулер енгізілген модельдің демографиялық паритет бұзушылықтарын шекті дәлсіздігі 0,2%-дан аспайтын деңгейде тиімді түрде жоятындығын көрсетті. Бұл тәсіл SMT шешушісін тек диагностикалық құрал ғана емес, сонымен қатар реттелетін салаларға жарамды, әділдігі дәлелденген қорытындылау жүйесіне айналдырады.

Түйін сөздер: машиналық оқыту, шешім ағаштары, формальды әдістер, SMT.

^{1*}Бейшекеев А.,

магистрант, ORCID ID: 0009-0006-0845-4551,

*e-mail: a.beishekeyev@kbtu.kz

²Умаров Т.,

PhD, ORCID ID: 0009-0008-0044-7159,

e-mail: t.umarov@bmu.edu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан²Британский университет менеджмента, г. Ташкент, Узбекистан

ФОРМАЛЬНАЯ ВЕРИФИКАЦИЯ СПРАВЕДЛИВОСТИ И УСТОЙЧИВОСТИ ДЕРЕВЬЕВ РЕШЕНИЙ С ПОМОЩЬЮ SMT-РЕШАТЕЛЕЙ

Аннотация

Поскольку системы искусственного интеллекта все чаще внедряются в критически важных для безопасности сферах, таких как здравоохранение и финансы, обеспечение их надежности и соблюдения требований крайне важно. В то время как глубокие нейронные сети получили значительное внимание в формальной верификации, традиционные модели, такие как деревья решений, часто предпочитаемые за их интерпретируемость, не могут по умолчанию навязывать ограничения после обучения на справедливость и стабильность.

В данной статье представлен новый, комплексный подход к формальной проверке классификаторов дерева принятия решений с использованием теорий по модулю удовлетворения (SMT). Мы предлагаем надежную схему трансляции, которая преобразует обученные деревья принятия решений в логические ограничения, позволяя делать вывод по ограничениям, гарантирующий демографический паритет и локальную устойчивость во время прогнозирования. Мы реализуем эту структуру с помощью решателя z3 SMT и проверяем его по широко признанным стандартам справедливости, включая наборы данных UCI Adult, German Credit и Loan Approval. Экспериментальные результаты показывают, что наша ограниченная модель эффективно исключает нарушения демографического паритета с предельной точностью менее 0,2%. Этот подход превращает решатель SMT из простого диагностического инструмента в доказуемо справедливый движок выводов, подходящий для регулируемых отраслей.

Ключевые слова: машинное обучение, деревья решений, формальные методы, SMT.

ӘОЖ 004.2
FTAXP 20.23.01

<https://doi.org/10.55452/1998-6688-2026-23-2-326-339>

^{1,2}**Самбетбаева М.,**

PhD, қауымдастырылған профессор, жетекші ғылыми қызметкер,
ORCID ID: 0000-0001-9358-1614,
e-mail: madina_jgtu@mail.ru

^{1,3}**Еримбетова А.,**

PhD, қауымдастырылған профессор, жетекші ғылыми қызметкер,
ORCID ID: 0000-0002-2013-1513,
e-mail: aigerian8888@gmail.com

^{1,2}**Абдығалым Б.,**

магистр, кіші ғылыми қызметкер, ORCID: 0009-0001-8872-7428,
e-mail: bayangali.abd@gmail.com

^{1,4*}**Дайырбаева Ә.,**

магистр, ғылыми қызметкер, аға оқытушы,
ORCID ID: 0000-0002-4255-5456,
*e-mail: nurbekkyzyelmira@gmail.com

¹**Тұрғанбаев А.,**

инженер-бағдарламашы, ORCID ID: 0000-0001-8989-7827,
e-mail: mlchallenge@inbox.ru

¹ҚР ҒЖБМ ҒК қарасты «Ақпараттық және есептеу технологиялары институты»,
Алматы қ., Қазақстан

²Л.Н. Гумилев атындағы Еуразиялық ұлттық университеті, Астана қ., Қазақстан

³META University, Алматы қ., Қазақстан

⁴Қ.И. Сәтпаев атындағы Қазақ ұлттық техникалық зерттеу университеті,
Алматы қ., Қазақстан

ОНТОЛОГИЯЛЫҚ ЖӘНЕ КОРПУС ДЕРЕКТЕРІН БІРІКТІРГЕН ТІЛДІК МОДЕЛЬ

Аңдатпа

Бұл зерттеу әскери және геосаяси оқиғаларға қатысты әлеуметтік желілер мен жаңалықтар платформаларынан жиналған 1000 мәтіннен тұратын аннотацияланған корпус құру арқылы күнделікті дискурстағы цифрлық қауіптерді талдауға бағытталған. Ұсынып отырған зерттеудің мақсаты ақпараттық операциялардың элементтері бар мәтіндердің аннотацияланған корпусын құру арқылы күнделікті дискурстағы цифрлық қауіптерді талдаудың шұғыл қажеттілігін шешу болып табылады. Әсер ету түрін, эмоционалды бояуды, жалған ақпарат белгілерін және авторлық ниеттерді (арандату, қорқыту және т.б.) қамтитын көп деңгейлі аннотация схемасы жасалды. Аннотация сапаны бақылау және контекстті ескере отырып, Label Studio-да орындалады; сенімділік Каппа=0,82 коэффициентімен расталады. Onto-IO-Bert моделі пилоттық экспериментте F1=0,81 көрсетті, бұл негізгі модельдерден асып түсті. Корпусың практикалық қолданылуы нақты Telegram хабарламаларын талдау арқылы расталады. Осы жұмыста Қазақстандық ИМ кеңістігіндегі әскери ақпараттық операцияларды талдауға бейімделген аннотацияның көп деңгейлі моделі ұсынылған және құрылған корпус қолданыстағы деректер жиынтығындағы маңызды олқылықты толтырып, әскери ақпараттық операцияларды талдауға арналған жаңа ресурс болып табылады. Ұсынылған корпус қазақ, орыс және ағылшын тілдеріндегі мәтіндерді қамтиды. Аннотация құрылымы манипуляциялық мазмұнның лингвистикалық және прагматикалық қырларын модельдеуді жеңілдетеді. Корпус келесі сілтеме бойынша жалпыға қолжетімді: https://github.com/baiangali/multi_mil.

Түйін сөздер: әскери терминдер, аннотацияланған корпус, ақпараттық операциялар, NLP, әлеуметтік желілерді талдау.

Кіріспе

Заман талабына сай жасанды интеллект пен табиғи тілді өңдеу саласында онтологиялық білім базалары мен корпус деректерін біріктіру маңызды ғылыми бағытқа айналып отыр. Онтологиялар ұғымдар арасындағы семантикалық байланыстарды жүйелеуге мүмкіндік берсе, корпус деректері тілдің нақты қолданысын, грамматикалық құрылымдарын және прагматикалық ерекшеліктерін бейнелейді. Бұл екі дерек көзін біріктіру арқылы тілдік модельдер тек статистикалық заңдылықтарды ғана емес, сонымен қатар мағыналық деңгейдегі байланыстарды да қамти алады. Мұндай тәсіл ақпараттық жүйелердің түсіну қабілетін арттырып, көптілді ортада, соның ішінде қазақ тілі секілді ресурстары шектеулі тілдерде де тиімді шешімдер ұсынуға жол ашады [1, 2].

Қазіргі кезеңдегі қарулы қақтығыстар барған сайын гибридті сипат иеленіп, дәстүрлі әскери операциялар ақпараттық ықпал ету әрекеттерімен кешенді түрде толықтырылуда [3]. Бұл үдерісте цифрлық медиа, әсіресе әлеуметтік желілер, бір мезгілде ресми ақпаратты жеткізудің тиімді арнасы әрі ашық және жасырын сипаттағы ақпараттық-психологиялық ықпал ету үшін бәсекелестік алаңы ретінде стратегиялық маңызға ие. Мұндай ықпалдың негізгі көріністеріне мақсатты жалған ақпарат тарату, бейбіт тұрғындар мен әскери қызметкерлердің моральдық-психологиялық жағдайын әлсірету, мемлекеттік және халықаралық институттарға деген сенімді іріткі салу, әлеуметтік дүрбелең туғызу, жекелеген топтарды стигматизациялау, сондай-ақ зорлық-зомбылықты заңдастыруға ұмтылу жатады [4, 5]. Таратылатын ақпараттардың мазмұны көбінесе бөлшектелген, эмоционалды тұрғыдан «зарядталған» және манипуляциялық сипатта болады, үгіт-насихат әдістеріне және әскери терминологияға сүйенеді [6, 7].

Цифрлық қауіптерді бағалауда автоматтандырылған мәтінді талдау маңызды құрал болғанымен, оның тиімділігі оқыту деректерінің сапасы мен өзектілігіне байланысты [8]. NLP модельдері үшін қолданылатын корпусар лингвистикалық тұрғыдан дұрыс болуымен қатар, конфликт жағдайындағы ақпараттық операциялардың ерекшеліктеріне бейімделуі керек. Қолданыстағы NER ресурстары көбіне тұрақты салаларға және ағылшын тіліне [9, 10] бағытталған. Оларда ақпараттық операцияларды талдау үшін қажетті әсер ету түрі, мақсатты аудитория, автор ниеті сияқты маңызды аннотация санаттары мен жалған ақпарат көрсеткіштері жоқ [11]. Зерттеулердің тағы бір шектеуі – құжат деңгейіндегі аннотацияға сүйену [12], бұл модель дәлдігін төмендетеді, себебі, тіпті беделді дереккөздерде де насихат болуы мүмкін [13]. Бұл мәтіннің өзіне негізделген, егжей-тегжейлі және контекстке сезімтал аннотациялаудың қажеттілігін көрсетеді.

Әскери ақпараттық операцияларды талдау үшін сапалы, арнайы аннотацияланған корпусар жетіспейді, бұл қолданбалы шешімдер мен іргелі зерттеулердің әлеуетін шектейді. Бұл мәселені шешу үшін, бұл жұмыс әскери ақпараттық операциялар белгілері бар мәтіндер жинағын ұсынады. Оның басты ерекшелігі – дайындалған топтың қолмен аннотациялауы, ол субъектілер мен прагматикалық ерекшеліктерді (әсер ету түрі, автор ниеті, жалғандық көрсеткіштері, т.б.) қамтитын мамандандырылған схеманы қолданады [14]. Негізгі мақсат – әскери дискурстағы ақпараттық ықпал ету тәжірибесін көрсететін әлеуметтік желілер мен жаңалықтар платформаларынан әдістемелік негізделген, аннотацияланған мәтіндер жинағын жасау.

Ұқсас жұмыстарға салыстырмалы талдау. Әскери дискурстағы ақпараттық операцияларды зерттеу корпус лингвистикасы, сыни дискурстық талдау, когнитивті соғыс және стратегиялық коммуникацияны қамтитын көпсалалы бағыт болып табылады [15]. Дегенмен, аннотацияланған корпусар тіл білімінде кең қолданылғанымен [16], әскери ақпараттық және психологиялық операцияларға арналған арнайы ресурстар әлі де жеткіліксіз.

Әскери контекстегі корпус әдістерінің ерте қолданыстары Әл-Равидің [17] әскери дискурсты кілт сөз жиілігі бойынша талдауынан [18] және әскери тілдің стратегиялық сипатына назар аударғаны [19] жұмысынан көрінеді. Алайда, бұл зерттеулерде ұсақ түйіршікті құрылым мен

прагматикалық аннотация жетіспейді, бұл оларды автоматты өңдеу үшін шектейді. Реттгер және әріптестері [20] LLM-дердің ақпараттық операцияларды талдаудағы әлеуетін көрсетіп, автоматтандырылған нәтижелерді қолмен тексерудің маңыздылығын атап өтті. Когнитивті соғыс ұғымы аудиторияның қабылдауы мен мінез-құлқын манипуляциялауда басты бағытқа айналуға.

Сонымен қатар, оқиғаларды шығару мәтінді тереңірек семантикалық талдау әдісі ретінде танымал болды. Ace2005 [21], MAVEN [22] және DuEE [23] сияқты сөйлем деңгейіндегі деректер жиынтығы жалпы немесе қаржылық домендерге назар аударады, AL RAMS [24], WikiEvents [21] және DocEE [26] сияқты құжат деңгейіндегі деректер жиынтығы дискурстың кең контекстін қарастырады. Алайда, бұл дереккөздерде әскери оқиғаларға қатысты оқиғалар туралы ақпарат жоқ. Әскери мақсаттағы ресурстарды игеруде айтарлықтай жетістіктерге қол жеткізілді. MNEE корпусы [27] сөйлем деңгейінде аннотацияланған Қытай мәтіндерін ұсынады, ал жақында CMNEE жобасы [28] сегіз түрлі (мысалы, қақтығыс, орналастыру, эксперимент) 17000 құжаттан және 29000-нан астам іс-шаралардан тұратын толық құрамды ұсынады және он бір аргумент рөлі. CMNEE әскери корпорациялардың ерекше қиындықтарын, соның ішінде оқиғалардың тығыз құрылымын, терминологиялық әртүрлілігін және күрделі синтаксисін атап көрсетеді.

MilTAC корпусы [29] әскери дискурсты зерттеуге арналған алғашқы корпус болғанымен, оның аннотациясы лексика-семантикалық деңгеймен шектеліп, коммуникативті ниет, эмоционалды тон және жалған ақпарат сияқты прагматикалық аспектілерді қамтымайды. Дипломат корпусы [30] әскери және саяси коммуникациядағы сөйлеу әрекеттері мен аргументтерге баса назар аударады. Ол риторикалық құрылымдар мен аргументтік стратегияларды түсіндіргенімен, эмоционалды манипуляцияларға немесе жалған ақпаратқа қатысты емес, бұл оның ақпараттық операцияларды зерттеудегі қолданылуын шектейді. Осыған байланысты, бұл зерттеуде ұсынылған Multi_mil корпусы оқиға деңгейіндегі, нысан деңгейіндегі және прагматикалық аннотацияларды біріктіру арқылы кемшіліктерді жоюға бағытталған. Бұл әскери дискурсты көп деңгейлі талдауға және көптілді контексте ақпараттық операциялардың негізгі заңдылықтарын анықтауға мүмкіндік береді.

Бұрынғы деректер жиынтықтарында прагматикалық аннотациялар жетіспеді және олар көбіне ағылшын/қытай тілдерімен шектелді, әрі психологиялық операцияларға аз көңіл бөлінді. Біздің корпус бұл олқылықтарды толтырып, лингвистикалық және прагматикалық деңгейлерді біріктіріп, Қазақстандық геосаяси контекстке бейімделген көптілді деректерді ұсынады, бұл аймақтық және тіларалық NLP зерттеулері үшін маңызды. Бұл зерттеу әскери дискурстағы ақпараттық операцияларды талдауға арналған аннотацияланған корпусың қажеттілігін көрсетеді. Ол доктриналық құрылымдарды корпусқа негізделген заманауи әдістермен біріктіріп, зерттеулердегі олқылықты толтырады және академиялық әрі практикалық тұрғыдан құнды ресурс ұсынады.

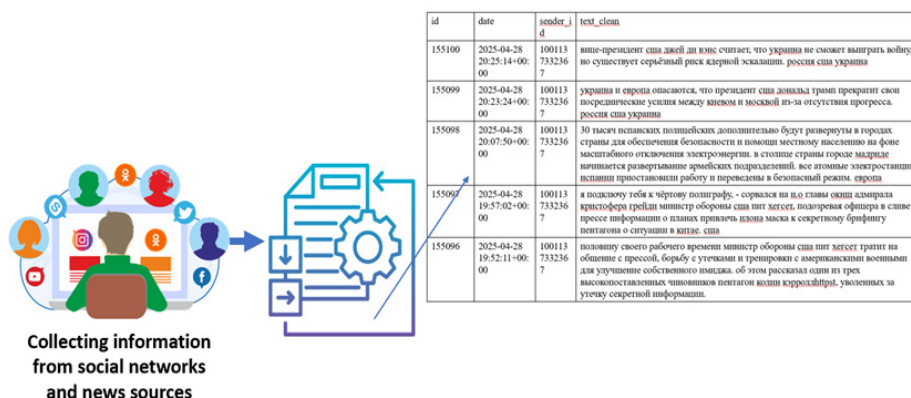
Материалдар мен әдістер

Әскери ақпараттық операциялардың корпусын құру үшін әлеуметтік желілердегі мәтіндер Label Studio ортасында [31] қолмен аннотацияланып, милитаристік тезаурус кілт сөздері мен ақпараттық операция типологиясына негізделген жалған ақпарат, деморализация, делегитимизация, арандатушылық және қорқыту сияқты ықпал ету стратегияларын қамтитын жазбалар таңдалды (Telegram, Instagram және public news aggregator көздерінен). Пилоттық үлгі келесі критерийлер бойынша таңдалған 200 мәтіннен тұрды:

- ♦ Әскери және геосаяси тақырыптарға тақырыптық өзектілігі;
- ♦ Қайталанбайтын немесе кросс-орналастырылған мазмұнсыз бірегейлік;
- ♦ Ақпараттық әсер ету ерекшеліктерімен лингвистикалық қанықтыру.

Аннотациялардың жан-жақты қамтылуын және сәйкестігін қамтамасыз ету үшін корпус бірнеше кезеңдерде жасалды. Бастапқы кезеңде аннотация бойынша нұсқаулықтарды

нақтылау және аннотация схемасын түзету үшін тандалған және пайдаланылған 200 мәтіннен тұратын эксперименттік ішкі жиын болды. Содан кейін негізгі мәліметтер жиынтығы аяқталған схемаға сәйкес сарапшылардың қолмен аннотациясы үшін платформаға жүктелді.



Сурет 1 – Аннотацияға дейінгі мәтіндік деректердің құрылымы

Сурет 1-де түпнұсқа жаңалықтардың үзіндісін ұсынады. тазартылған мәтіндік деректерді қамтитын csv файлы. Осы деректер жиынындағы әрбір жол келесі өрістерді қамтиды:

- ♦ id – әр жазба үшін бірегей идентификатор;
- ♦ күні – ISO форматындағы жарияланымның уақыт белгісі;
- ♦ sender_id – жарияланым көзінің идентификаторы (мысалы, Telegram арнасы);
- ♦ text_clean – шудан, еренсілтемелерден, эмодзилерден және басқа да маңызды емес белгілерден тазартылған алдын ала өңделген мәтін.

Бұл деректер жиынтығы әрі қарай қолмен аннотациялау үшін негіз болды. Талданған корпустың толық көлемі мәтіндерден алынған 15000 таңбалауыштан асады. Хабарламаның орташа ұзындығы 15-тен 20 сөзге дейін, әр хабарламада шамамен 5-тен 10-ға дейін аннотацияланған нысандар бар, бұл семантикалық тұрғыдан маңызды мазмұнның жоғары тығыздығын көрсетеді.

Аннотация нұсқаулығы қолмен лингвистикалық таңбалауға арналып, екі деңгейлі құрылымды қамтиды:

- а) субъект деңгейінде – орналасқан жер, әскери терминология, дереккөздер сияқты аталған мәндер;
- б) құжат деңгейінде – ақпараттық операция түрі, эмоционалдық реңк, жалған ақпарат белгілері, автордың ниеті және мақсатты аудитория.

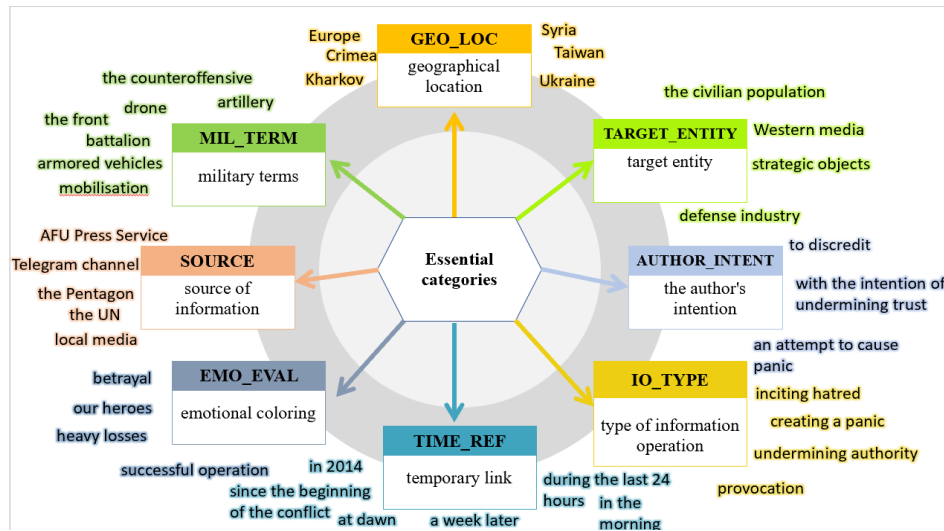
Әр санат ресми критерийлермен, мысалдармен және ескертпелермен толықтырылған. Тегтеу жүйесі халықаралық стандарттарға, ақпараттық операциялар доктриналары мен насихат талдау тәжірибесіне сүйене отырып жасалды.

Сурет 2-де әскери ақпараттық операцияларға қатысты мәтіндер корпусына аннотация жасау кезінде қолданылатын субъектілер санаттарының таксономиясы көрсетілген. Диаграмманың ортасында «um – brella» терминінің субъектілік категориялары орналасқан, олардан сегіз негізгі класс бөлінеді, олардың әрқайсысы корпустан алынған нақты өрнектердің мысалдарымен толықтырылған).

MIL_TERM (ӘСКЕРИ ТЕРМИН) стандартталған әскери терминологияға негізделіп, ISO TC 37 және ACE аннотация хаттамаларында милитаристік әрекеттер мен субъектілерді тануға қолданылады [30]; GEO_LOC және TIME_REF корпус лингвистикасында және ACE бағдарламаларында кеңінен қолданылатын стандартты NER категориялары [32]; ДЕРЕККӨЗ ақпараттың шығу тегін бағалауға мүмкіндік береді [33, 34]; ал TARGET_ENTITY ақпараттық операцияның нақты мақсатын көрсетеді [35, 36] AUTHOR_INTENT психолінгвистикалық және үгіт-насихаттық талдау шеңберіне негізделген mes-sage негізіндегі коммуникативті стра-

тегияны ұсынады (мысалы, «жүктелген жергілікті желі», «таңбалау»). Оны қосу манипулятивті әрекетті түсіндіруге мүмкіндік береді.

IO_TYPE (ақпараттық операция түрі) белгісі хабарламаларды әсер ету сипаты бойынша жіктейді – мысалы, жалған ақпарат, көңіл-күйді түсіру, қорқыту-және әскери және халықаралық доктриналық дереккөздерге негізделген. IO_TYPE 1-кесте егжей-тегжейлі көрсетілген сегіз санатты қамтиды.



Сурет 2 – Жіктеу және мысалдарды атап өту

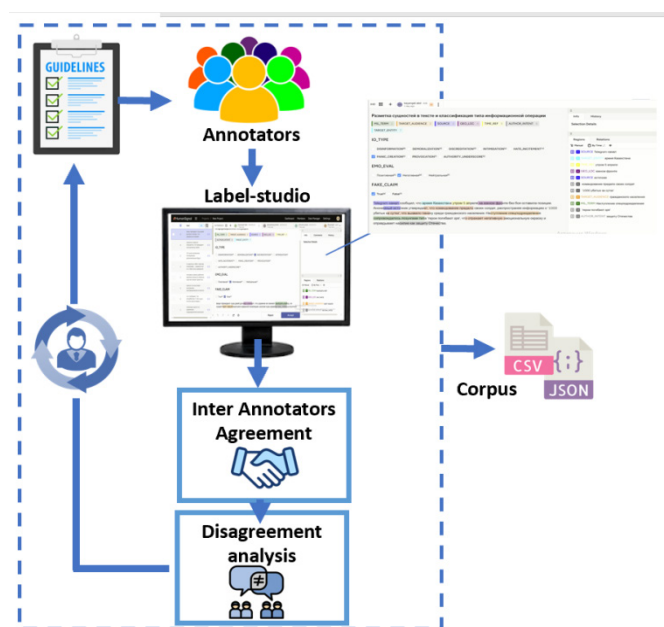
Кесте 1 – IO_TYPE (Ақпараттық операция түрі)

Категория	Сипаттамасы	Мысалдар
DISINFORMATION	Жалған ақпаратты қасақана тарату	KZ: армияның барлық бөлімшелері ұрыссыз берілді EN: All army units surrendered without a fight.
DEMORALIZATION	Моральды бұзу	KZ: майданда хаос орын алды, офицерлер бірінші болып қашуда EN: Chaos reigns on the front line, officers are fleeing first.
DISCREDITATION	Билікке немесе қоғамдық сенімге нұқсан келтіру	KZ: Армия басшылығы жағдайды бақыламайды EN: The army leadership has lost control of the situation.
INTIMIDATION	Қорқыныш немесе қауіп атмосферасын құру	KZ: Шабуыл жаппай шығындармен бірге жүреді EN: The offensive will result in massive casualties.
HATE_INCITEMENT	Өшпенділікті немесе дұшпандықты қоздыру	KZ: Олар -ұлттың жауы EN: They are the enemies of our nation.
PANIC_CREATION	Дүрбелең мен дабылды тарату	KZ: Барлық оқ-дәрі қоймалары жарылды EN: All ammunition depots have been blown up.
PROVOCATION	Жанжал немесе кек алу агрессиясын тудыру	KZ: Олар бітімгершілікті бірінші болып бұзды EN: They were the first to break the ceasefire.
AUTHORITY_UNDERSCORE	Билікті манипуляциялық түрде баса көрсету немесе бұзу	KZ: Жоғары қолбасшылық осындай маңызды сәтте үнсіз қалады EN: High command remains silent in such a critical moment.

Аннотация шеңберінде мәлімдеменің үнін және оның мақсатты аудиторияға ықтимал психологиялық әсерін түсіру үшін эмоционалды бағалау мүмкіндігі (EMO_EVAL) енгізілді. Бұл жіктеу риторикалық стратегияларды және мәтіндердегі эмоционалды қарқындылық дәрежесін анықтауға мүмкіндік береді (Кесте 2).

Кесте 2 – EMO_EVAL (Emotional Evaluation)

Категория	Сипаттамасы	Мысалдар
Оң пікір	Қолдау, қолдау немесе үміт тудыратын мәлімдемелер	RU: батырлық қорғаныс EN: Heroic defense
Теріс пікір	Қорқыныш, үрей, агрессия немесе менсінбеу көріністері	RU: Ұятпен позициядан айырылу EN: Shameful surrender of positions
Бейтарап	Эмоционалды бояусыз ақпараттық мәлімдемелер	RU: операция барысында екі батальонның күші жұмылдырылды EN: Two battalions were deployed during the operation



Сурет 3 – Корпустың аннотация процесі

Процесті оңтайландыру және топ мүшелері арасында бірізділікті қамтамасыз ету Үшін Label Studio интерфейсі көп пайдаланушы ынтымақтастығын орнату арқылы конфигурацияланды (Сурет 3). Бұл тәсіл бір корпусты бірнеше аннотаторлармен параллель аннотациялауға мүмкіндік берді, бұл аннотацияның жұмыс процесін жеделдетіп қана қоймай, сонымен қатар тұрақты тексеру және топ ішіндегі ішкі кері байланыс механизмдері арқылы жалпы сапаны жақсартты. Аннотация шешімдерін дәйекті диалог пен салыстыруға ықпал ете отырып, бұл қондырғы аннотаторлар арасындағы жоғары келісімді сақтауға және нақты уақыт режимінде түсініксіздіктерді шешуге көмектесті.

Ұсынылған суретте ақпараттық операцияларды автоматты түрде анықтау және жіктеу бойынша зерттеулер шеңберінде мәтіндік аннотацияның орыс тіліндегі мысалы келтірілген (Сурет 4). Қолданылатын белгілеу схемасы деструктивті ақпараттық әсердің әртүрлі аспектілеріне сәйкес келетін негізгі субъектілер мен санаттарды қамтиды (Кесте 3).

Разметка сущностей в тексте и классификация типа информационной операции

MIL_TERM 1 | TARGET_AUDIENCE 2 | SOURCE 3 | GEO_LOC 4 | TIME_REF 5 | AUTHOR_INTENT 6 | TARGET_ENTITY 7

IO_TYPE

☐ DISINFORMATION^(a) ☐ DEMORALIZATION^(a) ☐ DISCREDITATION^(a) ☐ INTIMIDATION^(a) ☐ HATE_INCITEMENT^(a)

☒ PANIC_CREATION^(a) ☐ PROVOCATION^(a) ☐ AUTHORITY_UNDERSCORE^(a)

EMO_EVAL

☐ Позитивная^(a) ☒ Негативная^(a) ☐ Нейтральная^(a)

FAKE_CLAIM

☒ True^(a) ☐ False^(a)

Telegram-канал сообщил, что армия Казахстана утром 5 апреля на южном фронте без боя оставила позиции. Анонимный источник утверждает, что командование предало своих солдат, распространяя информацию о '1000 убитых за сутки', что вызвало панику среди гражданского населения. Наступление спецподразделения сопровождалось лозунгами типа 'герои погибают зря', что отражает негативную эмоциональную окраску и оправдывает насилие как защиту Отечества.

Сурет 4 – Нысанға аннотация жасау процесі

Кесте 3 – Деструктивті ақпараттық әсердің әртүрлі аспектілеріне сәйкес келетін негізгі субъектілер мен санаттар (орыс және ағылшын тілдерінде)

Мәні	Мысалдар	Аннотация
MIL_TERM	RU: массированный артиллерийский удар EN: massive artillery strike	[[массированный артиллерийский удар]] MIL_TERM
	RU: казахстанские военные подразделения EN: Kazakhstani military units	[[казахстанские]] GEO_LOC [[военные подразделения]] MIL_TERM
AUTHOR_INTENT	RU: С целью дискредитации командования EN: With the aim of discrediting the command	С целью [[дискредитации]] AUTHOR_INTENT: DISCREDIT [[командования]] TARGET_ENTITY
	RU: Направлено на дезориентацию населения EN: Aimed at disorienting the population	Направлено [[на дезориентацию]] AUTHOR_INTENT: DISINFORMATION [[населения]] TARGET_AUDIENCE
TARGET_AUDIENCE	RU: Обращение к русскоязычной аудитории EN: Appeal to the Russian-speaking audience	Обращение к [[русскоязычной аудитории]] TARGET_AUDIENCE
	RU: Предупреждение для семей военнослужащих EN: Warning for families of military personnel	Предупреждение для [[семей военнослужащих]] TARGET_AUDIENCE

3-кестенің жалғасы

TARGET_ENTITY	RU: Обвинения в адрес правительства EN: Accusations against the government	Обвинения в адрес [[правительства]] TARGET_ENTITY
	RU: Недовольство военным руководством EN: Discontent with military leadership	Недовольство [[военным руководством]] TARGET_ENTITY
GEO_LOC	RU: Жанаозенский регион EN: Zhanaozen region	[[Жанаозенский регион]] GEO_LOC
	RU: У границ Казахстана EN: Near Kazakhstan's borders	У границ [[Казахстана]] GEO_LOC
SOURCE	RU: По данным телеграм-канала «WarNews» EN: According to the Telegram channel "WarNews"	По данным [[телеграм-канала «WarNews»]] SOURCE
	RU: Заявление Министерства обороны EN: Statement by the Ministry of Defense	Заявление [[Министерства обороны]] SOURCE
TIME_REF	RU: утром 24 февраля EN: In the morning of February 24	[[утром]] TIME_REF [[24 февраля]] TIME_REF
	RU: во время оккупации Крыма EN: During the occupation of Crimea	Во время [[оккупации]]MIL_TERM [[Крыма]] GEO_LOC

Аннотация әр айтылымның семантикасын және оның ақпараттық манипуляциядағы рөлін ескере отырып жүргізілді. Ол оқу корпусын құруға бағытталды, бірақ қолмен тегтеу субъективтілікке тәуелді. Әсіресе AUTHOR_INTENT жапсырмасы терең контекстік талдауды талап ететін екіұштылығымен ерекшеленді. Дегенмен, қолмен аннотация ақпаратқа әсер етудің негізгі заңдылықтарын дәл анықтауға мүмкіндік берді және болашақ автоматтандырудың негізін қалады. Аннотацияның сапасын бағалау үшін аннотаторлар арасындағы Келісімнің (IAA) метрикасы есептелді [37,38] аннотация жасаушылардың пайымдауларының жүйелілік дәрежесін көрсетеді.

Бұл құрылым мәтіндерді мазмұндық деңгейде ғана емес, сондай-ақ прагматика, ниеттілік және мақсатты әсер ету тұрғысынан талдауға мүмкіндік береді. Белгілеуді домен сарапшылары орындап, деректер Label Studio-дан JSON форматында экспортталды, ол spaCy, HuggingFace Трансформаторлары, Flair сияқты NLP құрылымдарымен үйлесімді. Сондай-ақ, CoNLL және CSV форматтарына экспорттау опциялары енгізіліп, дәйекті таңбалау және жіктеу тапсырмаларын қолдауға мүмкіндік жасалды.

Нәтижелер мен талқылау

Аннотация процесінен кейін 1000 мәтіннен (75000-нан астам таңбалауыштан) тұратын мамандандырылған корпус құрылды. Бұл мәтіндер ашық ақпарат көздерінен, негізінен Telegram жаңалықтар арналарынан және әскери тақырыптарға қатысты медиа ресурстардан жинақталған.

Қарастырылған корпус 8000 аннотацияланған нысанды қамтиды және 10 негізгі белгіге негізделген: алты нысанға қатысты (MIL_TERM, GEO_LOC, TIME_REF, SOURCE, TARGET_ENTITY, TARGET_AUDIENCE) және төрт прагматикалық (IO_TYPE, AUTHOR_INTENT, EMO_EVAL, FAKE_CLAIM).

Бұл мәтін ақпараттық шабуылдарға тән заңдылықтарды көрсетеді, олар баяндауды автоматты тануға негіз бола алады. Белгілер арасындағы тұрақты корреляциялар анықталды: мысалы, жалған мәлімдеме 48 рет жалған ақпаратпен, ал 39 рет дүрбелең тудырумен бірге кездеседі. Автор ниеті көбіне бірнеше ақпараттық операция түрі (IO_TYPE) бар хабарла-

маларда кездеседі, бұл ақпараттық шабуылдардың гибриді сипатын көрсетеді. Сондай-ақ, нысаналы тұлғаның беделін түсіру мен нысаналы аудиторияны деморализациялау арасында байланыс бар. Бұл заңдылықтар агрессивті хабарламалардың ішкі семантикалық құрылымын айқын көрсетеді (4-кесте).

Кесте 4 – Ақпараттық ықпал категорияларының лингвистикалық маркерлері

Категория	Талдау нәтижелері	Мысалдар
DEMORALIZATION	Лексемалардың жоғары жиілігі	KZ: мағынасыз, ешкім көмектеспейді, бітті EN: meaningless, no one will help, it's all over
DISCREDITATION	Қайталанатын айыптау конструкциялары	KZ: олар сатқындық жасады, шындықты жасырды, адамдарды шығармады EN: betrayed, hid the truth, did not evacuate people
DISINFORMATION	Асыра сілтелген сандар	KZ: мыңдаған техниканың жойылуы туралы хабарламалар EN: Reports of the alleged destruction of thousands of pieces of equipment
Ескерту: авторлар құрастырған		

Аннотацияланған корпустың сенімділігі мен репродуктивтілігін қамтамасыз ету үшін аннотация аралық келісім бағаланды. Үш аннотация жасаушы 1000 мәтінді өз бетінше белгіледі. Бағалау нысандық тегтерге (мысалы, MIL_TERM, GEO_LOC, TIME_REF) және категориялық тегтерге (IO_TYPE, AUTHOR_INTENT, FAKE_CLAIM және т.б.) жүргізілді.

Категориялық белгілерді бағалау мақсатында он екі аннотация жасаушы Сара Со-һен (к) келісім коэффициенті қолданылды, бұл әдіс кездейсоқ сәйкестікке мүмкіндік береді (5-кесте). Талдау екі режимде жүргізілді:

(а) Ішінара Қабаттасу: Рұқсат етілген тандалған фрагменттердің ішінара қабаттасуын болжауда қолданылған.

(б) Ішінара Қабаттасу: алынып тасталды аннотациялардың толық сәйкес келуі үшін қатаң талаппен пайдаланылды.

Кесте 5 – Талдау режимдерінің нәтижесі

Категория	Ішінара Қабаттасу: Рұқсат	Ішінара қабаттасу: Алынып тасталды
MIL_TERM (F ₁)	0.91	0.84
GEO_LOC (F ₁)	0.88	0.79
TIME_REF (F ₁)	0.86	0.80
AUTHOR_INTENT (κ)	0.72	0.68
TARGET_AUDIENCE (κ)	0.76	0.73
TARGET_ENTITY (κ)	0.79	0.75
IO_TYPE (κ)	0.82	0.80
EMO_EVAL (κ)	0.85	0.83
FAKE_CLAIM (κ)	0.88	0.87
Ескерту: авторлар құрастырған		

Аннотация нәтижелері жоғары келісімді көрсетті: MIL_TERM, GEO_LOC және TIME_REF белгілері 0,85-тен жоғары F1 ұпайына жетті, ал FAKE_CLAIM және IO_TYPE бойынша $\kappa > 0,80$ болды. Ең төменгі келісім AUTHOR_INTENT белгісінде тіркеліп, прагматикалық бағытты анықтаудың субъективтілігін көрсетті.

Telegram жаңалықтар арнасында әскери мазмұнға пилоттық талдау жасалды. Хабарламалар жалған ақпарат пен эмоционалды қысымды анықтау үшін IO_TYPE және AUTHOR_INTENT бойынша белгіленді. 100 хабарлама OntoIO-BERT моделімен жартылай автоматты аннотацияланып, сарапшылармен тексерілді. Мысалы, «АВТОРДЫҢ НИЕТИ: ҚОРҚЫТУ» және «ЖАЛҒАН МӘЛІМДЕМЕ: ШЫНДЫҚ» ретінде жіктелген хабарламалар бейбіт тұрғындар арасында дүрбелең тудыру қаупін көрсетті (5-сурет).



Сурет 5 – Түсіндірілген Telegram әлеуметтік желісіндегі хабарламаның мысалы

Талдау нәтижесінде хабарламалардың 47%-ы жала жабуға, ал 12%-ы үйлестірілген жалған ақпаратқа жататыны анықталды. Бұл корпус әлеуметтік желілердегі дұшпандық ақпаратты бақылауға негіз бола алады. Сарапшылар аннотацияланған деректерді пайдаланып, эмоционалды және қайталанатын мазмұнды жала жабу науқандары ретінде бағалады.

Осы зерттеуде құрылған корпус қолданыстағы деректер жиынтығындағы маңызды олқылықты толтырып, әскери ақпараттық операцияларды талдауға арналған жаңа ресурс болып табылады. Лексикалық-семантикалық немесе риторикалық сипаттамалармен шектелетін Mil-tac, DiPLO-MAT немесе қытай тіліндегі оқиғаларды шығаруға бағытталған CMNEE-ден айырмашылығы, біздің әдіс нысан деңгейіндегі аннотацияны IO_TYPE, AUTHOR_INTENT, EMO_EVAL және FAKE_CLAIM сияқты интерпретациялық санаттармен бірегей түрде ұштастырады. Сонымен қатар, қазақ тілі тәрізді ресурстармен жеткіліксіз қамтылған тілдерді Telegram арқылы таратылатын нақты әскери дискурстармен біріктіру корпусының көптілді, геосаяси және стратегиялық коммуникацияларды зерттеудегі практикалық әрі академиялық құндылығын арттырады.

Дәстүрлі корпустардан айырмашылығы, біздің аннотация жүйеміз соғысқа тән қақтығыстарға, манипуляцияларға және жоғары эмоциялық мазмұнға бейімделген, полисемия, риторикалық стратегиялар мен прагматикалық маркерлерді қамтиды.

Корпустың негізгі артықшылықтары мынадай:

1. Қолмен аннотациялау жоғары сапалы бақылау, екіұшты өрнектерді түсіндіру икемділігі және жасырын риторикалық заңдылықтарды анықтауды жеңілдетті.

2. Аннотация схемасы семантикалық, синтаксистік және прагматикалық көпқабатты құрылымды қамтып, IO_TYPE және ЕМО_EVAL сияқты маңызды категорияларды қосады.

3. Іске асыру Label Studio интерфейсі арқылы модульдік сипатта, визуализацияны айқын және деректерді бірнеше пішімге экспорттауды қамтамасыз етеді.

4. Пилоттық валидация аннотация нұсқаулығының дәйектілігі мен сенімділігін растап, аннотациялаушылар арасындағы келісімнің қанағаттанарлық екенін көрсетті.

Қорытынды

Осы мақалада көптілді әскери дискурстағы ақпараттық операцияларды анықтау және жіктеу мақсатында әзірленген Multi_mil деп белгіленген көп деңгейлі аннотацияланған корпус көрсетілген. Wikibias, Mil-TAC немесе CMNEE сияқты қолда бар ресурстардан айырмашылығы, әзірленген корпус аннотацияларды бірқатар негізгі бағыттар бойынша біріктіреді: әсер ету түрі (IO_TYPE), аталған нысандар, эмоционалды бояу (ЕМО_EVAL) және сенімсіздік белгілері (FAKE_CLAIM). Барлық белгілер бір семантикалық схема бойынша рәсімделетіні, осылайша тұтастық пен интерпретацияны қамтамасыз ететіні анық. Ұсынылған корпус қазақ, орыс және ағылшын тілдеріндегі мәтіндерді қамтиды. Аннотация құрылымы манипуляциялық мазмұнның лингвистикалық және прагматикалық қырларын модельдеуді жеңілдетеді.

Қаржыландыру туралы ақпарат: Бұл жұмыс Қазақстан Республикасы Ғылым және жоғары білім министрлігіне қарасты Ғылым комитетінің қолдауымен жүргізілді (AP 23484329).

ӘДЕБИЕТТЕР

1 Yerimbetova, A., Sakenov, B., Sambetbayeva, M., Daiyrbayeva, E., Berzhanova, U., & Othman, M. Creating a Parallel Corpus for the Kazakh Sign Language and Learning. *Applied Sciences*, 15(5), 2808 (2025). <https://doi.org/10.3390/app15052808>

2 Sadirmekova, Zh., Daiyrbayeva, E., Karbozova, I., & Bekbolatov, S. Development of a method for automatic extraction of ontology entity names from natural language texts. In *UBMK 2024 - Proceedings: 9th International Conference on Computer Science and Engineering*, IEEE, 74–79 (2024). <https://doi.org/10.1109/UBMK63289.2024.10773503>

3 Mussiraliyeva, S., Bolatbek, M., Zhumakhanova, A., Medetbek, Z., & Sagynay, M. Comparative analysis of machine learning algorithms to identify extremist texts in the Kazakh language. *Scientific Journal of Astana IT University*, 14(14), 71–90 (2023). <https://doi.org/10.37943/14DKRN4681>

4 Mussiraliyeva, S., Bolatbek, M., Zhumakhanova, A., Baispay, G., & Medetbek, Z. Creating a model of semantic analysis of extremist texts in the Kazakh language. *Journal of Mathematics, Mechanics and Computer Science*, 121(1), 110–121 (2024). <https://doi.org/10.26577/JMMCS2024121111>

5 Rzaeva, L., Pogolovkin, D., & Myrzatay, A. Development of a modular messaging analysis service based on NLP for digital forensics. *News of the NAS RK. Series of Physics and Mathematics*, 2, 212–233 (2025). <https://doi.org/10.32014/2025.2518-1726.354> (in Russian).

6 Modzelewski, A., Da San Martino, G., Savov, P., Wilczyńska, M. A., & Wierzbicki, A. MIPD: Exploring manipulation and intention in a novel corpus of disinformation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 1234–1245 (2024). ACL.

7 Klein, I. Information operations and influence campaigns. *IEEE Phoenix Computer Society*, February 2024.

8 Sambetbayeva, M., Nekessova, A., Yerimbetova, A., Bayangali, A., Kaldarova, M., Telman, D., & Smailov, N. A Multi-Level Annotation Model for Fake News Detection: Implementing Kazakh-Russian Corpus via Label Studio. *Big Data and Cognitive Computing*, 9(8), 215 (2025). <https://doi.org/10.3390/bdcc9080215>

9 Lin, Y., Wang, H., & Celikyilmaz, A. A survey on recent advances in named entity recognition from deep learning models. *ACM Computing Surveys*, 55(2), 1–40 (2023). <https://doi.org/10.1145/3514221>

- 10 Derczynski, R., Bontcheva, L., & Roberts, K. Broad Twitter corpus: A diverse named entity recognition dataset. In *Proceedings of COLING 2016*, 1169–1179 (2016).
- 11 Barrón-Cedeño, A., Rosso, P., & Yu, S. Fine-grained analysis of propaganda in news articles. In *Proceedings of EMNLP 2019*, 564–573 (2019).
- 12 Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of EMNLP 2017*, 2921–2927 (2017).
- 13 Horn, C., Wiegand, M., & Klakow, D. Towards a multidimensional model of media bias in news articles. In *Proceedings of COLING 2018*, 498–509 (2018).
- 14 Reisigl, M., & Wodak, R. The discourse-historical approach. In *Methods of Critical Discourse Studies*, 3rd ed., 23–61 (2015).
- 15 Van Dijk, T. Critical discourse studies: A sociocognitive approach. In *Methods of Critical Discourse Analysis*, 62–86 (2011).
- 16 Zeldes, A. The GUM corpus: Creating multilayer resources in a university setting. *Language Resources and Evaluation*, 51, 581–612 (2017). <https://doi.org/10.1007/s10579-016-9343-4>
- 17 Al-Rawi, A. Framing the Syrian conflict on Twitter. *Global Media and Communication*, 10 (2), 153–170 (2014). <https://doi.org/10.1177/1742766514536022>
- 18 Vego, E. Effects-based operations: A critique. *Joint Forces Quarterly*, 41, 51–57 (2006).
- 19 Dandeker, C. Military language and strategic doctrine. *Sociological Review*, 54(3), 581–598 (2006). <https://doi.org/10.1111/j.1467-954X.2006.00628.x>
- 20 Röttger, P., Schröder, M., Grotov, A., & Augenstein, I. Harmful but legal: Cognitive framing and large language models in online information warfare. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*. ACL.
- 21 Walker, C., Strassel, S., Medero, J., & Maeda, K. ACE 2005 multilingual training corpus. Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2006T06>
- 22 Wang, X., Wang, Z., Han, X., Jiang, W., Han, R., Liu, Z., Li, J., Li, P., Lin, Y., & Zhou, J. MAVEN: A massive general domain event detection dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1652–1671 (2020).
- 23 Li, X., Li, F., Pan, L., Chen, Y., Peng, W., Wang, Q., Lyu, Y., & Zhu, Y. DuEE: A large-scale dataset for Chinese event extraction in real-world scenarios. In X. Zhu, M. Zhang, Y. Hong, & R. He (Eds.), *Natural Language Processing and Chinese Computing*, 534–545 (2020).
- 24 Ebner, S., Xia, P., Culkin, R., Rawlins, K., & Van Durme, B. Multi-sentence argument linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8057–8077 (2020).
- 25 Li, S., Ji, H., & Han, J. Document-level event argument extraction by conditional generation. In *Proceedings of NAACL-HLT 2021*, 894–908 (2021).
- 26 Tong, M., Xu, B., Wang, S., Han, M., Cao, Y., Zhu, J., Chen, S., Hou, L., & Li, J. DocEE: A large-scale and fine-grained benchmark for document-level event extraction. In *Proceedings of NAACL-HLT 2022*, 3970–3982 (2022).
- 27 Huang, H., Sun, J., Wei, H., Xiao, K., Wang, M., & Li, X. A dataset of domain events based on open-source military news. *China Scientific Data*, 8(1), 30 (2023). <https://doi.org/10.11922/sciencedata.2023.0005.zh>
- 28 Zhu, M., Xu, Z., Zeng, K., Xiao, K., Wang, M., Ke, W., & Huang, H. CMNEE: A large-scale document-level event extraction dataset based on open-source Chinese military news. In *Proceedings of LREC-COLING 2024*, 3367–3379 (2024).
- 29 Chen, L.-C., Chang, K.-H., & Yang, S.-C. Integrating corpus-based and NLP approach to extract terminology and domain-oriented information: An example of US military corpus. *Acta Scientiarum. Technology*, 44, 2022. <https://doi.org/10.4025/actascitechnol.v44i1.60486>
- 30 Li, H., Zhu, S.-C., & Zheng, Z. (2023). DiPlomat: A dialogue dataset for situated pragmatic reasoning. *arXiv preprint*. <https://arxiv.org/abs/2306.09030>
- 31 Label Studio Documentation. Label Studio annotator guide, 2023. <https://labelstud.io/guide>
- 32 International Organization for Standardization. ISO 24617-1:2012 – Language resource management – Semantic annotation framework (SemAF) – Part 1: Time and events (ISO-TimeML). <https://www.iso.org/standard/37331.html>
- 33 International Organization for Standardization. ISO 24617-6:2016 – Language resource management – Semantic annotation framework – Part 6: Principles of semantic annotation (SemAF Principles). <https://www.iso.org/standard/60581.html>

34 Wardle, C., & Derakhshan, H. Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe, 2017. <https://shorensteincenter.org/information-disorder-framework-for-research-and-policymaking/>

35 Joint Chiefs of Staff. (2012). Joint publication 3-13: Information operations. U.S. Department of Defense. https://irp.fas.org/doddir/dod/jp3_13.pdf

36 President of the Republic of Kazakhstan. (2023, March 20). On the approval of the information doctrine of the Republic of Kazakhstan (Decree No. 145). <https://adilet.zan.kz/rus/docs/U2300000145>

37 Schmitt, M. N. (Ed.). (2017). Tallinn Manual 2.0 on the international law applicable to cyber operations. Cambridge University Press.

38 Artstein, R., & Poesio, M. Survey article: Inter-coder agreement for computational linguistics. Computational Linguistics, 34(4), 555–596 (2008). <https://doi.org/10.1162/coli.07-034-R2>

^{1,2}Самбетбаева М.,

PhD, ассоциированный профессор, ведущий научный сотрудник,
ORCID ID: 0000-0001-9358-1614,
e-mail: madina_jgtu@mail.ru

^{1,3}Еримбетова А.,

PhD, ассоциированный профессор, ведущий научный сотрудник,
ORCID ID: 0000-0002-2013-1513,
e-mail: aigerian8888@gmail.com

^{1,2}Абдығалым Б.,

магистр, младший научный сотрудник, ORCID ID: 0009-0001-8872-7428,
e-mail: bayangali.abd@gmail.com

^{1,4}Дайырбаева Э.,

магистр, научный сотрудник, старший преподаватель,
ORCID ID: 0000-0002-4255-5456,
e-mail: nurbekkyzyelmira@gmail.com

¹Тұрғанбаев А.,

инженер-программист, ORCID ID: 0000-0001-8989-7827,
e-mail: mlchallenge@inbox.ru

¹Институт информационных и вычислительных технологий КН МНВО РК,
г. Алматы, Казахстан

²Евразийский университет им. Л.Н. Гумилева, г. Астана, Казахстан

³META University, г. Алматы, Казахстан

⁴Казахский национальный исследовательский технический университет им. К.И. Сатпаева,
г. Алматы, Казахстан

ЯЗЫКОВАЯ МОДЕЛЬ, ОБЪЕДИНЯЮЩАЯ ОНТОЛОГИЧЕСКИЕ И КОРПУСНЫЕ ДАННЫЕ

Аннотация

Настоящее исследование направлено на анализ цифровых угроз в повседневном дискурсе через создание аннотированного корпуса из 1000 текстов, собранных из соцсетей и новостных платформ, связанных с военными и геополитическими событиями. Целью предлагаемого исследования является удовлетворение насущной потребности в анализе цифровых угроз в повседневном дискурсе путем создания аннотированного корпуса текстов с элементами информационных операций. Разработана многоуровневая схема аннотации, охватывающая тип воздействия, эмоциональную окраску, признаки дезинформации и авторские намерения (провокация, запугивание и др.). Аннотирование выполнено в Label Studio с контролем качества и учетом контекста; надежность подтверждена коэффициентом Каппа = 0,82. Модель Onto-IO-BERT в пилотном эксперименте показала F1 = 0,81, превзойдя базовые модели. Практическая применимость корпуса подтверждена анализом реальных Telegram-сообщений. В настоящей работе представлена многоуровневая

модель аннотации, адаптированная для анализа военных информационных операций в пространстве Казахстанского МВД, и созданный корпус является новым ресурсом для анализа военных информационных операций, заполнив важный пробел в существующих наборах данных. Представленный корпус содержит тексты на казахском, русском и английском языках. Корпус доступен по ссылке: https://github.com/baiangali/multi_mil.

Ключевые слова: военные термины, аннотированный корпус, информационные операции, NLP, анализ социальных сетей.

^{1,2}**Sambetbayeva M.,**

PhD, Association Professor, leading researcher, ORCID ID: 0000-0001-9358-1614,
e-mail: madina_jgtu@mail.ru,

^{1,3}**Yerimbetova A.,**

PhD, Association Professor, leading researcher, ORCID ID: 0000-0002-2013-1513,
e-mail: aigerian8888@gmail.com

^{1,2}**Abdygalym B.,**

Doctoral student, junior researcher, ORCID ID: 0009-0001-8872-7428,
e-mail: bayangali.abd@gmail.com

^{1,4}**Daiyrbayeva E.,**

Master, researcher, senior lecturer, ORCID ID: 0000-0002-4255-5456,
e-mail: nurbekkyzyelmira@gmail.com

¹**Turganbayev A.,**

Software engineer, ORCID ID: 0000-0001-8989-7827,
e-mail: mlchallenge@inbox.ru

¹Institute of Information and Computational Technologies of the Committee of Science
of the Ministry of Science and Higher Education of the Republic of Kazakhstan,
Almaty, Kazakhstan

²L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

³META University, Almaty, Kazakhstan

⁴Satbayev University, Almaty, Kazakhstan

A LANGUAGE MODEL INTEGRATING ONTOLOGICAL AND CORPUS-DATA

Abstract

This study addresses the urgent need to analyze digital threats in everyday discourse by constructing a 1,000-text annotated corpus from social media and news platforms covering military and geopolitical events. The purpose of the proposed study is to address the urgent need to analyze digital threats in everyday discourse by creating an annotated corpus of texts with elements of information operations. A multi-layered annotation scheme captures semantic actors and pragmatic features – including impact type, emotional tone, disinformation markers, and intent (e.g., provocation, intimidation). Annotation via Label Studio ensured flexibility, quality control, and context sensitivity, with inter-annotator reliability (Cohen's Kappa = 0.82) confirming consistency. In pilot experiments, the Onto-IO-BERT model achieved an F1-score of 0.81, outperforming baseline classifiers. Practical utility was validated through analysis of real Telegram messages. The framework is tailored for studying military information operations within Kazakhstan's Ministry of Internal Affairs and the created corpus is a new resource for analyzing military information operations, filling a significant gap in the existing data set. The presented corpus contains texts in Kazakh, Russian and English. The corpus is openly accessible at: https://github.com/baiangali/multi_mil

Keywords: military terms, annotated corpus, information operations, NLP, social media analysis.

Received January 13, 2026; accepted April 20, 2026.

UDC 004.896
ICSTI 28.23.27

<https://doi.org/10.55452/1998-6688-2026-23-2-340-352>

^{1,2}**Samigulina G.A.,**

Dr. Tech.Sc., professor, ORCID ID: 0000-0003-1798-9161,

e-mail: kz.galinasamigulina@gmail.com

^{1,2*}**Samigulina Z.I.,**

PhD, Associate Professor, ORCID ID: 0000-0002-5862-6415,

*e-mail: z.samigulina@kbtu.kz

^{1,2}**Bekeshev D.D.,**

MSc degree, ORCID ID: 0009-0002-8168-6368,

e-mail: dauletb01@gmail.com

³**Kucherbaeva S.Z.,**

ORCID ID: 0009-0002-7864-1900,

e-mail: stestephaniya1960@mail.ru

³**Yarmukhamedova A.B.,**

ORCID ID: 0009-0001-5421-0666,

e-mail: only_alla@mail.ru

¹Institute of Information and Computing Technologies,

«Intelligent Control and Forecasting Systems» Lab, Almaty, Kazakhstan

²Kazakh-British Technical University, School of Information Technology and Engineering,
Almaty, Kazakhstan

³Municipal State Institution «Comprehensive School No. 35»
of the Almaty City Education Department, Almaty, Kazakhstan

ARTIFICIAL INTELLIGENCE: A NEW BIO-INSPIRED APPROACH IN ENGINEERING EDUCATION

Abstract

A cognitive technology has been developed for the distance learning of technical disciplines, utilizing a bio-inspired approach based on a unified artificial immune system featuring neuroendocrine regulation of homeostasis during knowledge acquisition. To implement an intelligent distance learning system for the control of complex industrial automation objects – based on the proposed UAIS–NES technology – an integrated ontological model has been developed. This model consists of ontological sub-models for the immune, neural, and endocrine subsystems, serving as the foundation for a knowledge base designed to integrate cognitive technology into the educational process. Classes and subclasses have been defined, and their interactions analyzed, to facilitate a two-stage data processing workflow: specifically, the reduction of non-informative features and subsequent classification, both executed using bio-inspired algorithms. System quality assessment criteria have been established based on metrics categorized into mutually exclusive classes, distinguishing between favorable and unfavorable outcomes. Based on the results of data analysis utilizing this technology, a forecast is generated regarding the level of acquired professional engineering skills, the probability of successfully mastering the academic discipline, and the student's degree of readiness to tackle complex technical management challenges – all presented in the form of a digital student competency profile.

Keywords: engineering education, intelligent technology, artificial immune system, neuroendocrine interaction, homeostasis, ontological models, prediction of learning outcomes, digital student profile.

Received April 26, 2026; revised June 4, 2026; accepted June 10, 2026.

Introduction

The modern world is changing rapidly due to the explosive development of artificial intelligence. Vast opportunities are emerging for the growth of the global economy [1] – a growth that, according to PwC Global Artificial Intelligence forecasts, is expected to amount to approximately \$15 trillion in the near future. Consequently, a serious challenge has arisen: the large-scale training of qualified engineering personnel equipped with the skills to work with artificial intelligence. In this context, the need to rethink the outdated concept of higher technical education is highlighted, as it currently fails to meet the demands of production automation and data-driven decision-making. An analysis [2] of recent developments regarding the application of AI in engineering education has identified key research areas, including expert systems, machine learning, data analysis based on intelligent methods, and forecasting. The use of artificial intelligence in engineering education offers a range of advantages [3], such as: the capability for predictive analytics regarding academic performance during the knowledge acquisition process – along with its timely adjustment [4] – the development of a personalized approach to learning [5], the enhancement of students' interdisciplinary competencies, the application of digital twins [6] and augmented/virtual reality approaches [7], among many others. For instance, the emergence of generative AI in the educational process [8] enhances access to knowledge, fosters personalized learning, and ensures automation and interactivity. Nevertheless, a number of challenges arise concerning academic integrity, security, errors in assignments, and so forth. Another facet of the issue is the readiness of the teaching staff to meet the evolving demands of modern AI-driven production and to integrate innovative intelligent technologies into the educational process [9], in alignment with the “Industry 4.0” and “Industry 5.0” concepts. Thus, the development of innovative AI-based learning technologies represents a promising endeavor for fostering an effective and supportive educational environment for both students and educators. Furthermore, the ethical aspects of applying AI [10] in education remain highly relevant. It is also worth noting that the rapid advancement of AI necessitates – for the successful preparation of future technical specialists – the introduction of AI literacy training at the secondary school level [11]. Exploring technical AI concepts and innovative technologies – along with the ethical dimensions and career prospects – serves as an excellent motivator for students to pursue a high-quality higher education in engineering.

Approaches to distance learning and access to real industrial equipment utilizing AI are of particular interest [12]. The development of innovative intelligent technologies enables the modeling of engineering systems on complex, real-world industrial equipment, thereby allowing students to hone their practical skills in solving technical problems [13].

Progress in the field of bio-inspired AI approaches facilitates their application across various domains of science and technology, including intelligent educational technologies. Reference [14] examines the use of artificial neural networks (ANNs) in education. Based on a bibliometric analysis of articles from the Scopus database, currently in-demand directions for ANN application have been identified, such as: predictive analytics, adaptive learning, personalization, automated assessment and feedback, learning analytics, digitalization, academic integrity, and the management and administration of the educational process. For example, article [15] focuses on the generation of complex, high-quality questions using neural networks – a process crucial for assessing students' mastery of educational material. Deep learning and various ANN architectures are widely employed in this context. Paper [16] discusses the prospects of applying deep learning in engineering, particularly in mechanical and electrical engineering.

Recently, the approach of artificial immune systems – based on the human body's immunological response to the invasion of foreign agents – has also been undergoing active development. For instance, the research presented in [17] describes the development of a multimodal artificial immune system, based on negative selection mechanisms, for emotion classification; this system can be successfully applied in educational settings. Article [18] presents a mathematical model of an intelligent system for educational process management, employing a hybrid approach based on a self-organizing Kohonen network and an artificial immune system. Experimental results demonstrated the potential of this approach for implementation within the educational process.

In the Republic of Kazakhstan, digital educational ecosystems such as BilimClass, BilimLand, OnlineMektep, Kundelik.kz, and Microsoft Teams are widely employed to support and manage the educational process. These platforms facilitate the delivery of instructional content, maintenance of electronic gradebooks, monitoring of academic performance, and implementation of distance learning. However, their functionality is primarily focused on learning management, communication, and the provision of educational resources, rather than on the intelligent prediction and assessment of engineering competencies.

The proposed research focuses on the development of an intelligent technology based on the self-assembly of a Unified Artificial Immune System (UAIS) – featuring neuroendocrine regulation of homeostasis – for a remote learning system designed for technical students training on modern industrial equipment, exemplified by enterprises in the oil and gas sector.

The scientific novelty of the proposed research lies in the development of the UAIS-NES technology for learning data analysis and forecasting, inspired by the biological architecture of the human organism, specifically its neural, endocrine, and immune subsystems. The technology incorporates a self-assembly mechanism that automatically selects the optimal combination of modified algorithms for analyzing learner-related data. In addition, an integrated ontological model has been designed to enable intelligent knowledge representation and support adaptive learning processes.

This work continues a series of studies [19] aimed at developing an innovative distance learning technology for engineering personnel, utilizing a bio-inspired approach based on Artificial Immune Systems (AIS) to manage the educational process through the prediction of learning outcomes for students in technical disciplines. In [20], a unified artificial immune system was developed for the effective analysis of multidimensional student data. The proposed research focuses on implementing the concept of integrating the UAIS with neuroendocrine subsystems (NES) and the principle of homeostasis to enhance the operational efficiency of an intelligent distance learning system for students.

Materials and methods

These studies are interdisciplinary in nature, situated at the intersection of various fields of science and technology; they draw upon principles of theoretical immunology, molecular biology, automatic control theory, systems analysis, artificial intelligence, and information and cognitive technologies.

The data processing algorithms developed are grounded in mechanisms borrowed from the biological prototype of human physiological functioning. They utilize the principles of interaction among key subsystems: the immune, neural, and endocrine systems. These subsystems are interconnected and must maintain homeostasis [21]. Homeostasis is a specific state of physiological equilibrium in which the organism maintains resilience against internal and external disturbances. For instance, the neural subsystem (NS) [22] receives signals from the external environment via electrical impulses. Depending on these signals, it may trigger physiological responses to stress, pain, or other stimuli. Concurrently, the endocrine subsystem (ES) [23] releases the hormone cortisol, which influences the immune subsystem (IS) [24]. The immune subsystem (IS) regulates the intensity of the immune response.

Consequently, all three subsystems serve as crucial mechanisms for maintaining the organism's stable steady state – homeostasis. The presented subsystems are described in the form of artificial intelligence algorithms; however, they implement only specific mechanisms. For instance, in the case of the immune subsystem, algorithms based on clonal selection, negative selection, and immune networks have gained widespread adoption [24]. However, within a living organism, these mechanisms interact with one another; the first attempts to create a Unified Artificial Immune System (UAIS) were presented in [25]. Subsequently, in [26], the concept of a unified artificial immune system – comprising modified AIS algorithms – was developed. Study [27] examined neuroendocrine-immune interactions for addressing optimization problems.

Thus, the development of novel modified algorithms – based on the biological model of the human organism and the principles of immunological homeostasis – for the processing of multidimensional data constitutes a pressing task, enabling the creation of more robust data processing mechanisms that are resilient to internal and external environmental perturbations.

Architecture of the UAIS-NES intelligent multidimensional data processing technology

Figure 1 illustrates the architecture of the UAIS-NES intelligent technology for multidimensional data processing. A biological prototype for the interaction of the subsystems (NS, IS, and ES) – designed to ensure homeostasis – is presented. To implement each of these subsystems for multidimensional data processing, corresponding modified algorithms have been developed [26].

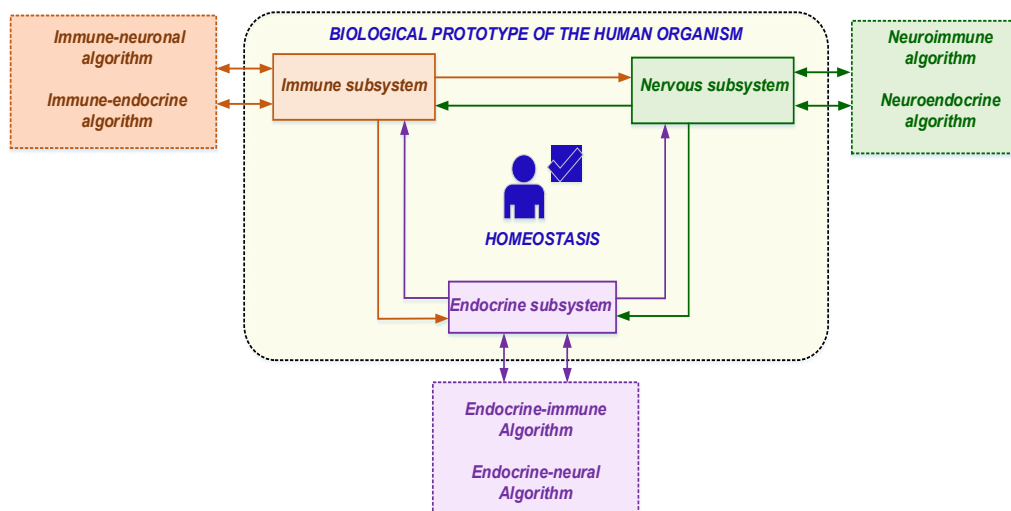


Figure 1 – Architecture of the UAIS-NES intelligent multidimensional data processing technology

- ♦ The immune subsystem consists of the following modified algorithms [26]: the immuno-neural (AISN) and the immuno-endocrine (AISE). The AISN algorithm operates as follows: data reduction is performed based on an artificial immune system. Subsequently, informative features are identified using the SHAP (SHapley Additive exPlanations) method [28]. Subsequently, the database – following the reduction phase – is fed into the neural component of the algorithm to address the classification task. The neural network architecture consists of a multilayer perceptron comprising three hidden layers containing 128, 64, and 32 neurons, respectively. The AISE algorithm is founded upon the following principle: data reduction is performed by the algorithm's immune component, while the classification task is resolved by the artificial endocrine algorithm. The artificial endocrine algorithm simulates the behavior of the endocrine subsystem based on hormonal regulation within the human body. The algorithm regulates agent behavior through the use of hormones. Hormones represent threshold values for data deviation, and their concentration varies in response to input signals [26].

- ♦ The neural subsystem consists of the following modified algorithms: neuro-immune and neuro-endocrine. The operating principle of these algorithms is analogous: the first component performs the task of data reduction, while the second component addresses the task of classification.

- ♦ The endocrine subsystem consists of modified endocrine-immune and endocrine-neural algorithms. The endocrine component is utilized for data dimensionality reduction, while the immune and neural components are responsible for classification.

Based on these subsystems, the UAIS self-assembly mechanism has been implemented. The core concept involves evaluating algorithm performance based on a series of metrics. A recommended threshold value is established for each algorithm; if the metrics fall within the «homeostasis» zone – that is, if they meet the specified thresholds – the modified algorithm in question is deemed to be performing effectively and may be utilized for decision-making. Thus, data analysis is conducted

concurrently using all modified algorithms; however, only those algorithms whose metrics fall within the «homeostasis» zone are employed in the actual decision-making process.

Development of an ontological model for an intelligent distance learning system based on UAIS-NES technology

For the application of the UAIS-NES technology in the field of education, the most convenient form of knowledge representation is the ontological model. The Protégé software tool [29] was used as the development environment. The ontological model employs the reasoner HermiT for modeling and for determining the efficiency of algorithms. The model is a combination of three subsystems based on immune, endocrine and neural components for data reduction and solving the classification problem (Figure 1,2) [30]. The model implements four classes with the following functions:

- ♦ class 1. *Parts_of_system* determines which part of the system the object belongs to (contains two subclasses): *First_part_data_reduction*, *Second_part_classifier* (*First_part_data_reduction*, *Second_part_classifier*);
- ♦ class 2. *First_part_data_reduction* (this class comprises objects used as the first stage of data reduction). A subclass is defined as equivalent to a class of objects having at least one *is_data_reducer_of* relationship with an instance of the *Subsystems* class,
- ♦ class 3. *Second_part_classifier* (This subclass comprises objects that are used as the second stage of classification). This subclass is equivalent to the class of objects for which the *is_classifier_of* property is linked to at least one instance of the *Subsystems* class. (*Second_part_classifier is_classifier_of* some *Subsystems*);
- ♦ class 4. *Quality_of_Systems* has two subclasses – *System_have_bad_result* and *System_have_good_result* – which define system qualities based on the subclasses of the data property of the *Metrics* class ().

The *System_have_bad_result* subclass comprises individuals from the *Systems* class that exhibit poor metrics ($\text{System\_have\_bad\_result} \equiv \text{Systems and (Accuracy some } < 0.8)$). Furthermore, this subclass does not intersect with the *System_have_good_result* subclass ($\text{System\_have\_bad\_result} \cap \text{System\_have\_good\_result} = \emptyset$).

The *System_have_good_result* subclass comprises objects of the *Systems* class that possess favorable metrics. The *Subsystems* class encompasses all subsystems, comprising three subclasses: *Endocrin_subsystem*, *Immune_subsystem*, and *Neural_subsystem*.

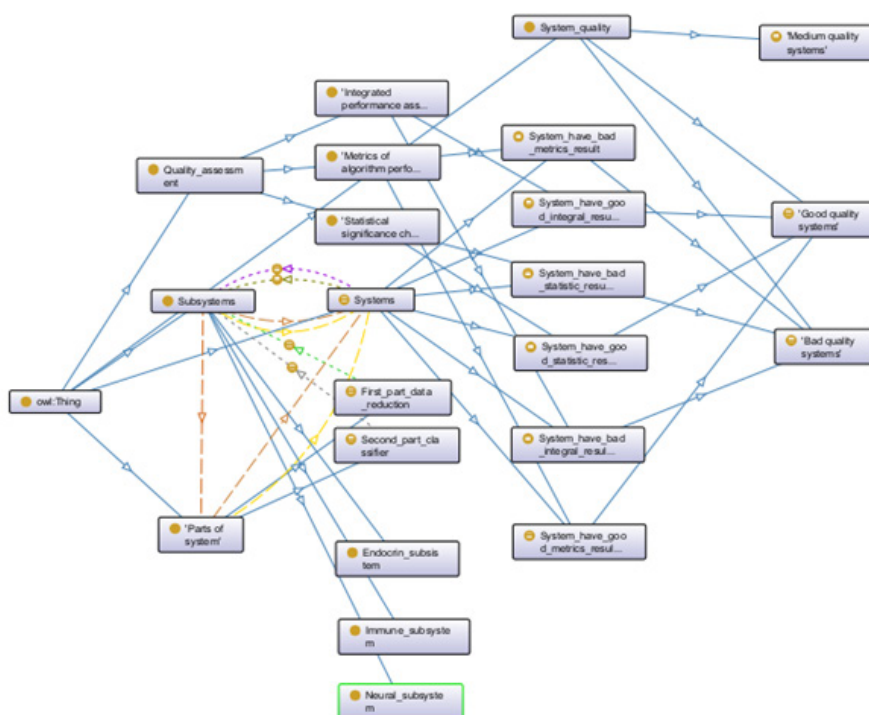


Figure 2 – Classes of the ontological model visualized using OntoGraf

Figure 3 presents individuals of the classes. The neural subsystem is implemented using a multilayer perceptron with three hidden layers and is designed to classify student academic performance data. The endocrine subsystem is based on a population of interacting agents whose behavior is regulated through hormone-like signals, enabling adaptive optimization throughout the learning process. The immune subsystem employs clonal selection and mutation mechanisms to generate and evolve antibodies that serve as classification prototypes.

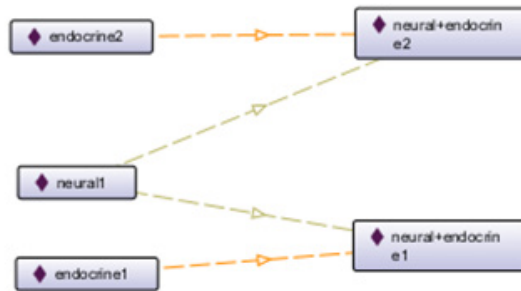


Figure 3 – Individuals of the ontological model classes

Within the UAIS-NES architecture, these subsystems operate in an integrated and coordinated manner. The definition of a homeostasis region makes it possible to evaluate subsystem performance according to predefined quality metrics and to automatically identify the most effective configuration for decision-making and learning outcome prediction.

For the formal representation of interactions among the UAIS-NES subsystems within the ontological model, each subsystem is defined as a separate subclass.

1. The Endocrin_subsystem subclass contains objects that represent the endocrine subsystem (which may relate to First_part_data_reduction or Second_part_classifier).

2. The Immune_subsystem subclass contains objects that represent the immune system (which may relate to First_part_data_reduction or).

3. The Neural_subsystem subclass represents a neural system (which may relate to First_part_data_reduction or).

The ontological model presents four properties of the object:

1. is_classifier_of – an object property that links a classification subsystem to a system;
2. is_data_reducer_of – an object property that links the data reduction subsystem to the system.
3. has_classifier_subsystem – this property is the inverse of is_classifier_of;
4. has_data_reducer_subsystem – this property is the inverse of is_data_reducer_of.

Based on the integrated ontological model UAIS-NES, a knowledge base has been developed for the implementation of intelligent technology into the educational process. Figure 4 presents the UAIS-NES technology in the form of an ontological model, comprising subclasses and individuals.



Figure 4 – UAIS-NES ontological model for implementation in the educational process

The developed ontological model enables data analysis based on UAIS-NES technology.

Intelligent distance learning system based on UAIS-NES technology

The operating principle of the intelligent distance learning system based on an ontological model is as follows (Figure 5). It is proposed to divide data processing into two levels:

1) First Level. Data analysis and career guidance based on the processing of student data during secondary education. To identify students' aptitudes, data analysis utilizing the BilimClass educational platform is recommended [31]. At the primary school level (Grades 1 through 4), data processing focuses on students' study of logic. In the middle school grades (5 through 9), the focus is on mathematics; in the upper school grades (9 through 11), it shifts to computer science and physics. Based on the results of data processing, a digital profile of the student is generated, which serves as a basis for assessing the student's aptitude for pursuing higher education in technical universities.

2) Second Level. Analysis of student data within the context of higher education institutions. The course structure and learning materials are hosted on the MS Teams platform [32]. During the learning process, a digital profile of the student's competencies is generated to assess their professional aptitude within the framework of engineering education. Based on the results of practical and laboratory assignments, a dataset is compiled containing grades, activity statistics, time spent on the platform, and other relevant metrics.

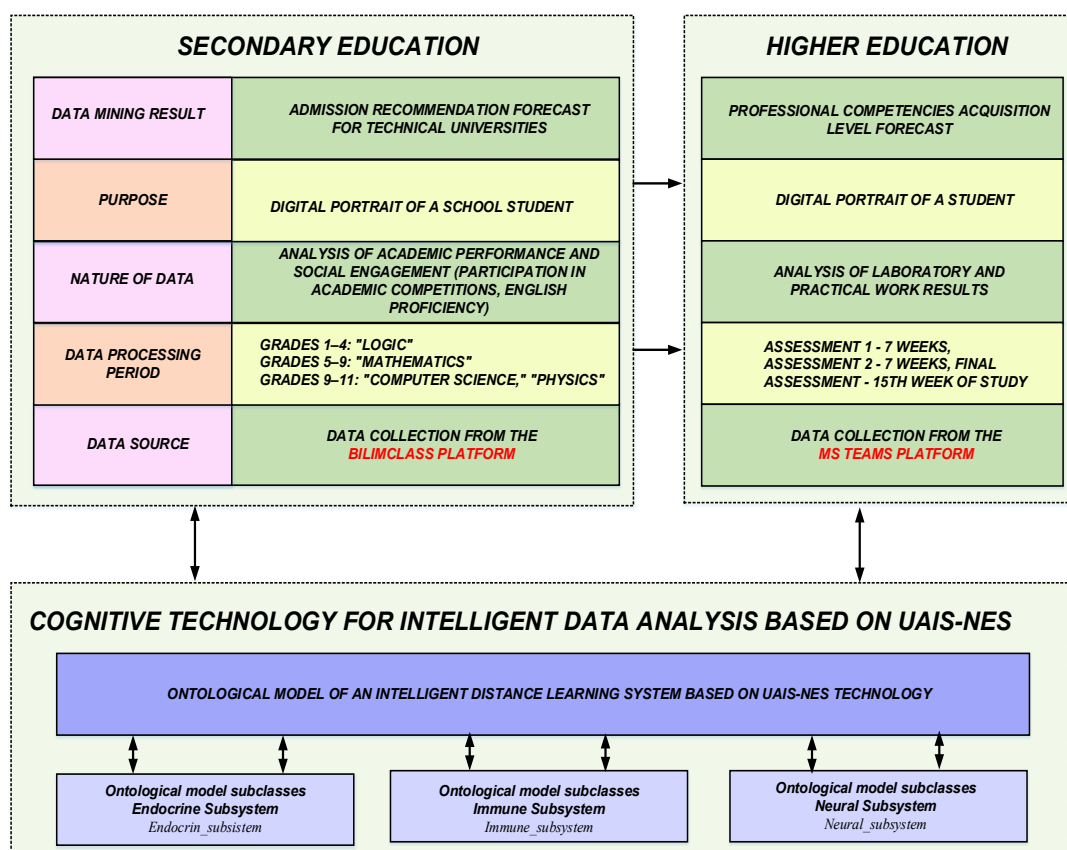


Figure 5 – Structural diagram of an intelligent distance learning system based on UAIS-NES technology

Following data analysis based on UAIS-NES intelligent technology, a forecast is generated regarding the level of acquired professional engineering competencies and the probability of successfully mastering a given discipline; simultaneously, the student's readiness to address complex technical management tasks is assessed, and the educational process undergoes real-time adjustment.

Results and discussion

Let us examine the structure of the developed ontological model. Figure 6 illustrates the data property hierarchy based on the OWL language [29]. The `topDataProperty` element serves as the root property of the model, from which all other properties inherit. These properties are divided into two major groups: Metrics and Parameters. The Metrics group contains information regarding the metrics used to evaluate the performance of the ontological model's algorithms (the neural, immune, and endocrine subsystems). The Parameters group contains the primary metaparameters for configuring the algorithms.

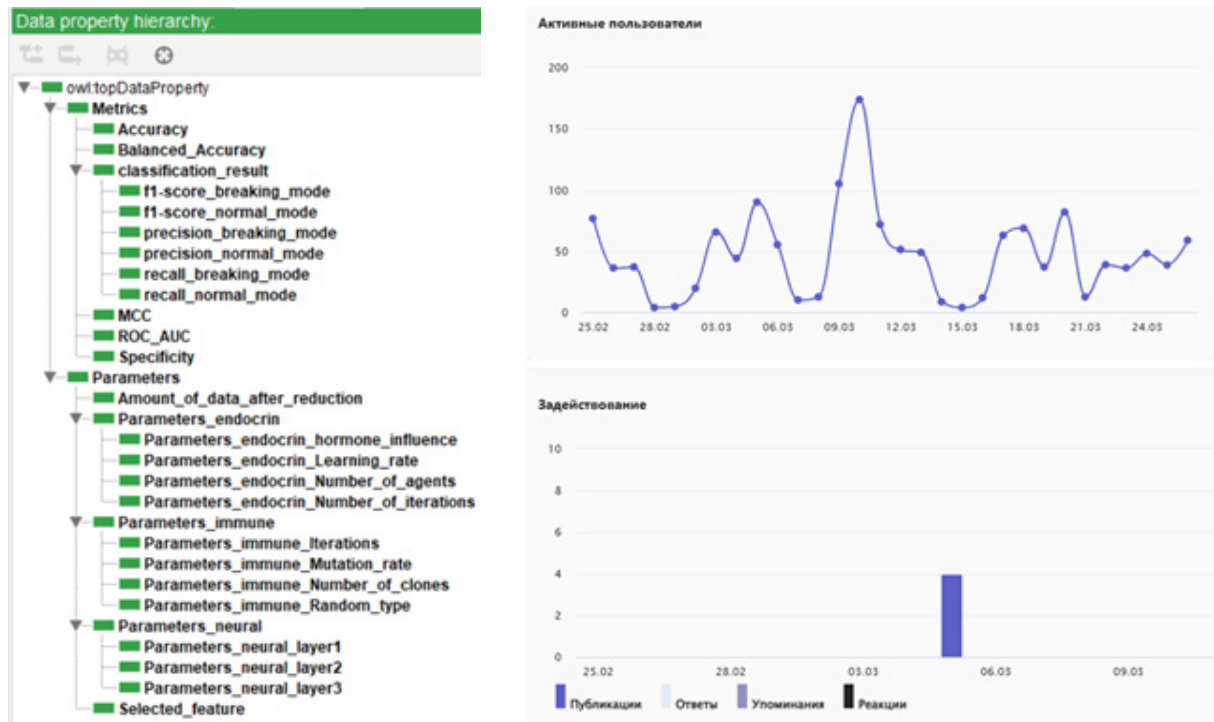


Figure 6 – Structure of the developed ontological model based on UAIS-NES technology

This model enables the comparison of the effectiveness of various combinations of bio-inspired modified algorithms on the same dataset, based on metric threshold values [30, 33, 34, 35]. Figure 7 displays a list of data instances from the ontological model.

A description of class instances is presented in Table 1. This parameter belongs simultaneously to multiple classes: endocrine_subsystem and second_part_classifier etc.

Table 1 – Parameters of the UAIS-NES subsystems

Subsystem	Parameter	Value	Description
Neural	layer1	64	Number of neurons in the first hidden layer
Neural	layer2	64	Number of neurons in the second hidden layer
Neural	layer3	64	Number of neurons in the third hidden layer
Endocrine	num_agents	100	Number of agents in the population
Endocrine	num_iterations	100	Number of training iterations
Endocrine	learning_rate	5	Hormonal adaptation coefficient
Endocrine	hormone_influence	0.1	Degree of influence of hormonal signals on agents
Immune	n_clones	100	Number of clones generated for each antibody
Immune	mutation_rate	1	Antibody mutation coefficient
Immune	iterations	50	Number of clonal selection iterations

Performance evaluation was conducted using the metrics presented in Figure 6, including Accuracy, Balanced Accuracy, ROC-AUC, F1-score, as well as Precision and Recall for each class. These metrics constitute the foundation of the algorithm performance assessment mechanism embedded within the ontological model and are used to evaluate the effectiveness of the proposed analytical approaches.

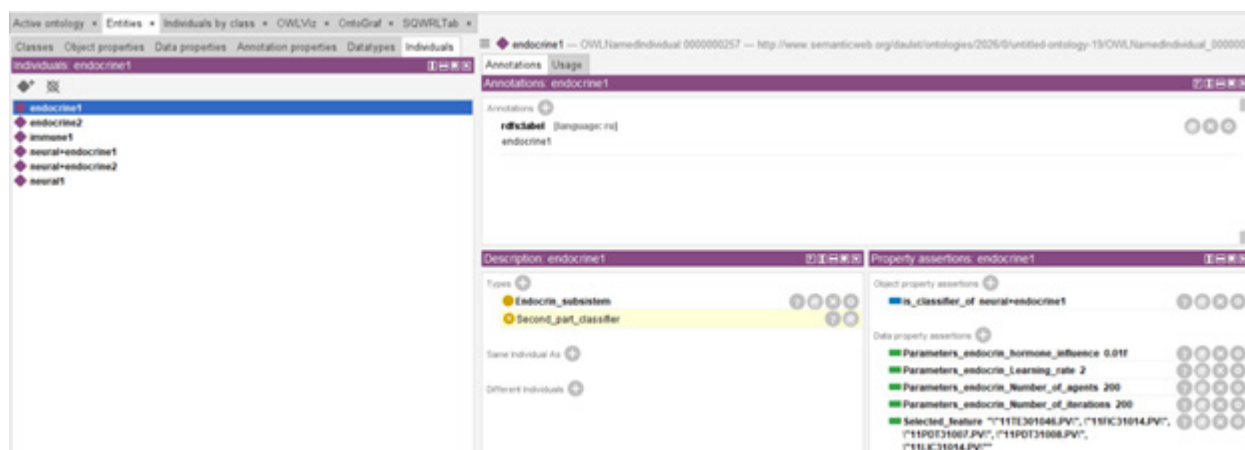


Figure 7 – List of data instances of the UAIS-NES ontological model

Thus, the ontological model comprises classes describing modified algorithms for the endocrine, immune, and neural subsystems – along with their properties in the form of parameters and metrics – as well as the values of data instances (individuals) representing specific model runs and their simulation results.

Conclusion

Representing an intelligent technology based on the UAIS-NES as an ontological model enables the structuring of data in a unified, formalized format. This approach renders the description of algorithms intelligible to both software developers and end users. The ontological model facilitates the integration of various models – specifically, modified algorithms for the endocrine, neural, and immune subsystems – by explicitly depicting their interactions and interdependencies. The Protégé software product enables logical inference, automatically identifies dependencies, and draws conclusions without human intervention. The set of experiments is represented as an intelligent knowledge base, wherein each algorithm is described as an object with its own parameters, metrics, and relationships.

The application of intelligent technology based on UAIS-NES in the field of education represents a promising avenue for training highly qualified technical specialists, ranging from secondary-level institutions to higher education establishments.

A certificate of entry into the state register of rights to copyright-protected objects has been obtained for the developed software [30].

Information on funding. This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic Kazakhstan (Grant No. AP23486386).

REFERENCES

- 1 Rahman, A., Khandakar, A., Ayari, M.A. et al. Artificial intelligence innovations challenges and emerging trends in engineering education. *Discov Educ* 5, 179 (2026). <https://doi.org/10.1007/s44217-026-01137-1>
- 2 Liu, Y., Jing, Y., Li, J., Dai, J. Application of AI in engineering education: A bibliometric study. *Review of Education*, 13(1), e70044 (2025) <https://doi.org/10.1002/rev3.70044>.
- 3 Liu, C. A comprehensive review of applications of AI technologies in higher engineering education. *Discov Educ* 4, 528 (2025). <https://doi.org/10.1007/s44217-025-00954-0>.
- 4 Tiukhova, E., Vemuri, P., Flores, N.L. et al. Explainable Learning Analytics: Assessing the stability of student success prediction models by means of explainable AI. *Decision Support Systems*. 182, 114229 (2024). <https://doi.org/10.1016/j.dss.2024.114229>.
- 5 Taşkın, M. Artificial Intelligence in Personalized Education: Enhancing Learning Outcomes Through Adaptive Technologies and Data-Driven Insights. *Human Computer Interaction*. 202, 1, 173. <https://doi.org/10.62802/ygye0506>.
- 6 Hazrat, M.A., Hassan, N.M.S., Chowdhury, A.A., Rasul, M.G., Taylor, B.A. Developing a Skilled Workforce for Future Industry Demand: The Potential of Digital Twin-Based Teaching and Learning Practices in Engineering Education. *Sustainability* 2023, 15, 16433. <https://doi.org/10.3390/su152316433>.
- 7 Magaña, E.C., Ariza, A.C., Ruiz-Palmero, J. et al. Virtual, augmented, and mixed reality in the University environment: an analysis of scientific production. *J. New Approaches Educ. Res.* 14, 8 (2025). <https://doi.org/10.1007/s44322-025-00027-y>.
- 8 Wang, N., Wang, X., Su, Y. Critical analysis of the technological affordances, challenges and future directions of Generative AI in education: a systematic review. *Asia Pacific Journal of Education*, 44, 139–155 (2024). <https://doi.org/10.1080/02188791.2024.2305156>.
- 9 Cuperman, D., Raveh, I. Adaptation of Engineering Higher Education to Teaching Industry 4.0 Technologies: Faculty Perspectives on Opportunities, Challenges, and Needs. In book: *2025 Yearbook Emerging Technologies in Learning* (pp. 37–53). 2026. https://doi.org/10.1007/978-3-032-09058-4_3.
- 10 Pulari, S.R., Shomona GraciaJacob, S.G. Research Insights on the Ethical Aspects of AI-Based Smart Learning Environments: Review on the Confluence of Academic Enterprises and AI. *Procedia Comput. Sci.* 256, 284–291 (2025). <https://doi.org/10.1016/j.procs.2025.02.122>.
- 11 Zhang, H., Lee, I., Ali, S. et al. Integrating Ethics and Career Futures with Technical Learning to Promote AI Literacy for Middle School Students: An Exploratory Study. *Int J Artif Intell Educ.* 33, 290–324 (2023). <https://doi.org/10.1007/s40593-022-00293-3>.
- 12 Li, K.C., Wong, B.T.M., Chan, H.T. Teaching and learning innovations for distance learning in the digital era: a literature review. *Frontiers in Education*, 8 (2023). <https://doi.org/10.3389/educ.2023.1198034>.
- 13 Liu, C., Wang, GC, Wang, HF. The application of artificial intelligence in engineering education: A systematic review. *IEEE Access*. 2025; 13:17895–17910. <https://doi.org/10.1109/ACCESS.2025.3532595>.
- 14 Ayanwale, M.A., Olatunbosun, S.O. & Bamiro, N.B. Mapping neural network research in education through bibliometric analysis. *Discov Educ* x5, 224 (2026). <https://doi.org/10.1007/s44217-026-01194-6>.
- 15 Al Faraby, S., Adiwijaya, A. & Romadhony, A. Review on Neural Question Generation for Education Purposes. *Int J Artif Intell Educ.*, 34, 1008–1045 (2024). <https://doi.org/10.1007/s40593-023-00374-x>.
- 16 Tobias, A.G., Kittur J. Strategic innovations and future directions in deep learning for engineering applications: a systematic literature review. *Front. Educ.*, 10, 1583404 (2025). <https://doi.org/10.3389/educ.2025.1583404>.
- 17 Umair, M., Rashid, N., Khan, U.S., et al. Negative selection-based artificial immune system (NegSI-AIS): A hybrid multimodal emotional effect classification model. *Results in Engineering*, 27, 106601 (2025). <https://doi.org/10.1016/j.rineng.2025.106601>
- 18 Astakhova, I., and Kiseleva, E. The application of the artificial immune system for design, development and using of the hybrid system in education. In: Sukhomlin, V., and Zubareva, E. (eds.), *Modern Information Technology and IT Education. Communications in Computer and Information Science*, Vol. 1204 (Cham: Springer, 2021), pp. 95–109. https://doi.org/10.1007/978-3-030-78273-3_7

- 19 Samigulina, G.A., and Samigulin, T.I. Development of a cognitive mnemonic scheme for an optical Smart-technology of remote learning based on artificial immune systems. *Computer Optics*, 45 (2), 286–295 (2021). <https://doi.org/10.18287/2412-6179-CO-736>
- 20 Samigulina, G.A., and Samigulina, Z.I. Development of Intelligent Technology for Complex Objects Control Based on a Unified Artificial Immune System and Principles of Immunological Homeostasis for Industrial Automation Using Modern Microprocessor Equipment (Yelm, WA, USA: Science Book Publishing House, 2023), 196 p. ISBN 978-1-62174-150-3.
- 21 Paradise, C.J., and Campbell, A.M. *Organismal Homeostasis* (New York: Garland Science, 2016), 100 p.
- 22 Michael-Titus, A.T., and Shortland, P. *The Nervous System: Systems of the Body*, ed. S.H. Hughes (Amsterdam: Elsevier, 2022), 344 p. ISBN 0702083402.
- 23 Hinson, J.P., Raven, P., Chew, S.L., and Hughes, S.H. *The Endocrine System*, 3rd ed. (Oxford: Elsevier, 2022). (Systems of the Body Series).
- 24 Sompayrac, L.M. *How the Immune System Works*, 7th ed. (Hoboken, NJ: Wiley-Blackwell, 2022), 176 p. ISBN 978-1-119-89068-3.
- 25 Chen, Y., Wang, X., and Zhang, Q. Unified artificial immune system. In: *Proceedings of the 5th International Conference on Computational Intelligence and Communication Networks (CICN)* (2013), pp. 617–621. <https://doi.org/10.1109/CICN.2013.135>
- 26 Samigulina, G.A., Samigulina, Z.I., and Bekeshev, D. Development of an Intelligent Self-Assembly Technology for a Unified Artificial Immune System to Control Complex Industrial Automation Objects Using a Neuroendocrine System (Yelm, WA, USA: Science Book Publishing House, 2026), 228 p. ISBN 978-1-62174-164-0.
- 27 Liu, B., Xu, M., Gao, L., and Di, X. A hybrid approach for high-dimensional optimization: Combining particle swarm optimization with mechanisms in neuro-endocrine-immune systems. *Knowledge-Based Systems*, 253, 109527 (2022). <https://doi.org/10.1016/j.knosys.2022.109527>
- 28 Lundberg, S.M., and Lee, S.I. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*, Vol. 30 (2017), pp. 4765–4774.
- 29 Protégé Ontology Editor and Knowledge Acquisition System. Available at: Protégé.
- 30 Samigulina, G.A., Samigulina, Z.I., and Bekeshev, D.D. SELF-ASSEMBLY OF A UNIFIED ARTIFICIAL IMMUNE SYSTEM WITH NEUROENDOCRINE INTERACTION. Author's Certificate No. 62526, published September 30, 2025.
- 31 BilimClass Educational Platform. Available at: BilimClass.
- 32 Microsoft Teams. Available at: Microsoft Teams.
- 33 Samigulina, G.A., and Samigulina, Z.I. Development of a multi-agent system for intelligent diagnostics of industrial equipment. *KBTU Bulletin*, 22 (3), 85–97 (2025). <https://doi.org/10.55452/1998-6688-2025-22-3-85-97>
- 34 Samigulina, G., Samigulina, Z., and Bekeshev, D. Bio-inspired modified framework for advanced predictive maintenance. *Procedia Computer Science*, 278, 125–132 (2026). <https://doi.org/10.1016/j.procs.2026.02.446>
- 35 Samigulina, G., Samigulina, Z., Bekeshev, D., et al. Digital twin-enabled intelligent HMI for real-time industrial automation systems. *Scientific Reports* (2026). <https://doi.org/10.1038/s41598-026-53028-y>

^{1,2}**Самигулина Г.А.,**

т.ғ.д., профессор, ORCID ID: 0000-0003-1798-9161,
e-mail: kz.galinasamigulina@gmail.com

^{1,2*}**Самигулина З.И.,**

PhD, доцент, ORCID ID: 0000-0002-5862-6415,
*e-mail: z.samigulina@kbtu.kz

^{1,2}**Бекешев Д.Д.,**

магистр, ORCID ID: 0009-0002-8168-6368,
e-mail: dauletb01@gmail.com

³**Кучербаева С.З.,**

ORCID ID: 0009-0002-7864-1900,
e-mail: stestephaniya1960@mail.ru

³**Ярмухамедова А.Б.,**

ORCID ID: 0009-0001-5421-0666,
e-mail: only_alla@mail.ru

¹Ақпараттық және есептеуіш технологиялар институты,
«Зияткерлік басқару және болжау жүйелері» зертханасы,
Алматы қ., Қазақстан

²Қазақстан-Британ техникалық университеті,
Ақпараттық технологиялар және инженерия мектебі,
Алматы қ., Қазақстан

³Алматы қаласы Білім басқармасының
«№35 жалпы білім беретін мектеп» коммуналдық мемлекеттік мекемесі,
Алматы қ., Қазақстан

ЖАСАНДЫ ИНТЕЛЛЕКТ: ИНЖЕНЕРЛІК БІЛІМ БЕРУДЕГІ ЖАҢА БИОИНСПИРАЦИЯЛЫҚ ТӘСІЛ

Аңдатпа

Білім алу процесінде гомеостазды нейроэндокриндік реттеу қағидатына негізделген біріздендірілген жасанды иммундық жүйені пайдалану арқылы техникалық мамандықтар бойынша қашықтан оқытудың когнитивтік технологиясы әзірленді. Ұсынылған БЖИЖ–НЭЖ (біріздендірілген жасанды иммундық жүйе – нейроэндокриндік жүйе) технологиясы негізінде өнеркәсіптік автоматтандырудың күрделі объектілерін басқаруға арналған интеллектуалды қашықтан білім беру жүйесін іске асыру үшін иммундық, нейрондық және эндокриндік ішкі жүйелердің онтологиялық модельдерінен тұратын интеграцияланған онтологиялық модель жасалды. Осы модель негізінде оқу процесіне когнитивтік технологияны енгізуге арналған білім базасы қалыптастырылды. Деректерді екі кезеңде өңдеу үшін сыныптар мен ішкі сыныптар әзірленіп, олардың өзара әрекеттесуі қарастырылды: атап айтқанда, ақпараттық емес белгілерді редукциялау және биоинспирациялық алгоритмдер негізінде жіктеу міндеттері шешілді. Жүйе сапасын бағалау критерийлері өзара бірін-бірі жоққа шығаратын сыныптарға (жақсы және нашар нәтижелер) бөлу арқылы метрикалар негізінде әзірленді. Технология негізінде деректерді талдау нәтижесінде алынған кәсіби инженерлік дағдылар деңгейі, пәнді сәтті меңгеру ықтималдығы және білім алушының күрделі техникалық басқару міндеттерін шешуге дайындығы деңгейі болжанады. Бұл көрсеткіштер студенттің құзыреттерінің цифрлық профилі түрінде ұсынылады.

Түйін сөздер: инженерлік білім беру, интеллектуалды технология, жасанды иммундық жүйе, нейроэндокриндік өзара әрекеттесу, гомеостаз, онтологиялық модельдер, оқу нәтижесін болжау, студенттің цифрлық профилі.

^{1,2}**Самигулина Г.А.,**

д.т.н., профессор, ORCID ID: 0000-0003-1798-9161,

e-mail: kz.galinasamigulina@gmail.com

^{1,2*}**Самигулина З.И.,**

PhD, доцент, ORCID ID: 0000-0002-5862-6415,

*e-mail: z.samigulina@kbtu.kz,

^{1,2}**Бекешев Д.Д.,**

магистр техн. наук, ORCID ID: 0009-0002-8168-6368,

e-mail: dauletb01@gmail.com

³**Кучербаева С.З.,**

ORCID ID: 0009-0002-7864-1900,

e-mail: stestephaniya1960@mail.ru

³**Ярмухамедова А.Б.,**

ORCID ID: 0009-0001-5421-0666,

e-mail: only_alla@mail.ru

¹Институт информационных и вычислительных технологий,
лаборатория «Интеллектуальные системы управления и прогнозирования»,
г. Алматы, Казахстан

²Казахстанско-Британский технический университет,
Школа информационных технологий и инженерии,
г. Алматы, Казахстан

³КГУ «Общеобразовательная школа №35» Управления образования г. Алматы,
г. Алматы, Казахстан

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НОВЫЙ БИОИНСПЕРИРОВАННЫЙ ПОДХОД В ИНЖЕНЕРНОМ ОБРАЗОВАНИИ

Аннотация

Разработана когнитивная технология дистанционного обучения техническим специальностям с использованием биоинспирированного подхода на основе унифицированной искусственной иммунной системы с нейроэндокринным регулированием гомеостаза при получении знаний. Для реализации интеллектуальной системы дистанционного образования по управлению сложными объектами промышленной автоматизации на основе предложенной технологии УИИС – НЭС разработана интегрированная онтологическая модель, состоящая из онтологических моделей иммунной, нейронной и эндокринной подсистем на основе которой создана база знаний для внедрения когнитивной технологии в учебный процесс. Разработаны классы, подклассы и рассмотрены их взаимодействия для двухэтапной обработки данных в виде решения задачи редукции неинформативных признаков и классификации на основе биоинспирированных алгоритмов. Созданы критерии оценки качества системы на основе метрик с разделением на взаимоисключающие классы на основе хороших и плохих результатов. По результатам анализа данных на основе технологии осуществляется прогноз уровня полученных профессиональных инженерных навыков, вероятности успешного освоения дисциплины и определения степени готовности обучающегося к решению сложных технических задач управления в виде цифрового профиля компетенций студента.

Ключевые слова: инженерное образование, интеллектуальная технология, искусственная иммунная система, нейроэндокринное взаимодействие, гомеостаз, онтологические модели, прогнозирование результата обучения, цифровой портрет студента.

УДК 004.85:373.3/.5:37.015.3
МРНТИ 28.23.37

<https://doi.org/10.55452/1998-6688-2026-23-2-353-364>

^{1*}**Молдагулова А.Н.,**

к.ф.-м.н., профессор, ORCID ID: 0000-0002-1596-561X,

*e-mail: a.moldagulova@satbayev.university

¹**Касенхан А.М.,**

PhD, ассоциированный профессор, ORCID ID: 0000-0002-6355-9544,

e-mail: a.kassenkhan@satbayev.university

²**Утебаев К.А.,**

магистрант, ORCID ID: 0009-0004-3032-5049,

e-mail: utebayevkuat@gmail.com

²**Изтурган Н.Б.,**

магистрант, ORCID ID: 0009-0001-7667-9561,

e-mail: tasnur2002@gmail.com

¹Казахский национальный исследовательский технический
университет им. К.И. Сатпаева, г. Алматы, Казахстан

²Алматинский филиал Национального исследовательского ядерного
университета «МИФИ», г. Алматы, Казахстан

ИССЛЕДОВАНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ НАВЫКОВ КРИТИЧЕСКОГО МЫШЛЕНИЯ УЧАЩИХСЯ

Аннотация

В данной статье изучается потенциал использования методов машинного обучения для прогнозирования уровня критического мышления учащихся на основе академических и поведенческих данных. Хотя критическое мышление широко признано важнейшим навыком в современном образовании, его прямая оценка по-прежнему затруднена. Целью исследования является выявление взаимосвязи между уровнем развития критического мышления и измеримыми факторами, такими как академическая успеваемость, участие во внеклассных мероприятиях и выбор учебных курсов. Был собран и обработан набор данных из 500 анонимных профилей учащихся для извлечения ключевых характеристик. Были разработаны и оценены с использованием стандартных метрик эффективности три модели: линейная регрессия, дерево решений и случайный лес-регрессор. Модель случайного леса продемонстрировала превосходную прогностическую эффективность, достигнув значения R^2 0,84 по сравнению с 0,65 для дерева решений и 0,37 для линейной регрессии. Результаты исследования свидетельствуют о том, что академические достижения учащихся, выбор курсов и уровень их внеклассной активности в совокупности дают ценную информацию об их способностях к критическому мышлению. В целом результаты подтверждают эффективность подходов, основанных на данных, для косвенной оценки когнитивных навыков, что имеет практическое значение для разработки учебных программ, инициатив раннего вмешательства и планирования образовательной политики. Используя доступные образовательные данные, этот подход способствует разработке более масштабируемых, объективных и персонализированных методов оценки в области аналитики обучения.

Ключевые слова: машинное обучение, линейная регрессия, дерево решений, случайный лес, критическое мышление, прогнозирование уровня критического мышления студентов.

Введение

Критическое мышление стало ключевой компетенцией в современных образовательных системах, лежащей в основе эффективного решения проблем, аналитического мышления и адаптивного принятия решений. Как в академической, так и в профессиональной среде, спо-

способность критически оценивать информацию, подвергать сомнению предположения и делать логические выводы высоко ценится. Глобальные образовательные программы, такие как OECD Learning Compass и Partnership for 21st Century Skills (P21), определяют критическое мышление как важнейший результат обучения для формирования глобальной гражданственности и поддержки непрерывного обучения.

Несмотря на признанную важность, определение и измерение критического мышления остаются сложной задачей из-за его абстрактности и многогранности. Традиционные подходы к оценке, такие как эссе, стандартизированные тесты на критическое мышление и оценки, проводимые преподавателем, часто ограничены субъективностью, непоследовательностью оценок и ограниченной масштабируемостью. Эти ограничения затрудняют для преподавателей и учебных заведений эффективный мониторинг и поддержку развития учащихся в этой важнейшей области.

Достижения в области машинного обучения в сочетании с растущей доступностью образовательных данных открывают многообещающую альтернативную перспективу. Академическая успеваемость, динамика зачисления на курсы и участие во внеклассной деятельности могут служить косвенными индикаторами поведенческих и когнитивных характеристик студентов. Хотя эти факторы не оценивают критическое мышление напрямую, они могут отражать черты, связанные с когнитивными способностями высшего порядка, включая интеллектуальное любопытство, саморегуляцию обучения и открытость к разнообразному образовательному опыту.

В данном исследовании изучается, в какой степени уровень критического мышления студентов можно предсказать, используя три наблюдаемых параметра: академическую успеваемость (например, оценки и средний балл), внеклассную активность и модели выбора курсов. Был собран и обработан набор данных, состоящий из 500 студенческих записей, для извлечения релевантных признаков. Были разработаны и сравнены три прогностические модели: линейная регрессия, дерево решений и случайный лес-регрессор для оценки их способности оценивать показатели критического мышления. Среди протестированных моделей метод случайного леса продемонстрировал наибольшую предсказательную эффективность, достигнув значения R^2 0,84 и превзойдя показатели моделей дерева решений (0,65) и линейной регрессии (0,37). Эти результаты свидетельствуют о том, что значительную часть вариаций в способностях учащихся к критическому мышлению можно объяснить с помощью доступных образовательных данных. Исследование подчеркивает потенциал масштабируемых решений на основе данных в образовании, которые поддерживают персонализированное обучение, целевые стратегии вмешательства и углубленное понимание когнитивного развития различных групп учащихся.

Материалы и методы

Критическое мышление является ключевым элементом высшего образования, определяющим способность студентов принимать обоснованные решения и ориентироваться в академических и профессиональных траекториях. На протяжении десятилетий его оценка остается важным направлением исследований в области педагогической психологии и дидактики. Традиционные методы оценки в значительной степени опираются на стандартизированные инструменты, такие как тест критического мышления Уотсона–Глейзера и Калифорнийский тест критического мышления, предназначенные для измерения навыков анализа, вывода, оценки и дедуктивного рассуждения. Несмотря на структурированность этих инструментов, их применение ограничено проблемами масштабируемости, контекстной жесткости и необходимостью проведения контрольного тестирования в контролируемых условиях, что снижает их применимость для непрерывной или массовой оценки в современных образовательных средах.

В ответ на данные ограничения в последние годы внимание исследователей сместилось в сторону косвенных методов оценки критического мышления на основе наблюдаемых по-

веденческих и лингвистических индикаторов. Так, авторы работы [1] подчеркивают важность анализа рассуждений студентов, проявляющихся в академическом дискурсе и аудиторной активности. Аналогично авторы работы [2] предложили использовать методы обработки естественного языка для анализа студенческих эссе, выявляя такие лингвистические признаки, как эпистемическая модальность, контраргументация и связность дискурса, в качестве маркеров критического мышления. Несмотря на перспективность данных подходов в контексте автоматизированной оценки, они, как правило, требуют детализированных текстовых данных и сложной инженерии признаков, что ограничивает их широкую применимость.

Параллельно с этим развитие интеллектуального анализа образовательных данных и аналитики обучения стимулировало исследования, направленные на использование стандартно собираемых данных о студентах для построения прогностических моделей. В работе [3] показали, что такие академические показатели, как посещаемость, средний балл (GPA) и модели онлайн-взаимодействия, могут быть использованы для прогнозирования академической успешности и уровня вовлеченности студентов. Аналогичным образом авторы работы [4] применили демографические данные, показатели академической динамики и региональные образовательные индикаторы для анализа и прогнозирования учебных достижений студентов, тесно связанного с развитием критического мышления.

Методы машинного обучения, особенно ансамблевые модели, такие как Random Forest и Gradient Boosting, получили широкое распространение в образовательных исследованиях благодаря их устойчивости к шуму данных и способности работать с многомерными признаками. Например, авторы работы [5] использовали классификаторы Random Forest для прогнозирования риска отсева студентов, продемонстрировав, что поведенческие паттерны могут отражать скрытые когнитивные или мотивационные характеристики. Тем не менее применение методов машинного обучения непосредственно для прогнозирования уровня критического мышления на основе агрегированных академических и поведенческих показателей остается недостаточно изученным.

Значительное количество исследований посвящено роли критического мышления в формировании образовательных предпочтений студентов, с акцентом на интеграцию междисциплинарных программ, специализированных учебных модулей и цифровых образовательных сред. В работе [6] авторы пришли к выводу, что программы высшего образования, направленные на развитие аналитического мышления, способствуют более осознанному выбору студентами образовательных траекторий. В работе [7] авторы выявили, что студенты с более выраженными навыками критического мышления чаще выбирают междисциплинарные курсы. Данные выводы были дополнительно подтверждены автором работы [8], подчеркнувшим роль аналитической компетентности в осознанном выборе учебных дисциплин.

Межкультурные исследования также подтверждают влияние критического мышления на академическое принятие решений. Так, автор работы [9] показал, что специализированные курсы по развитию критического мышления в странах арабского мира повышают уверенность студентов при выборе академических направлений. В условиях онлайн-обучения в работе [10] авторы установили, что студенты, склонные к критическому анализу информации, проявляют более избирательный подход к выбору курсов. Развивая данную линию исследований, в работе [11] авторы продемонстрировали, что образовательная среда, включающая несколько курсов с акцентом на развитие критического мышления, способствует более продуманным академическим решениям.

Эффективность педагогических интервенций, направленных на развитие критического мышления, является предметом активных научных дискуссий. В работе [12] автор утверждает, что структурированные обучающие программы существенно повышают способность студентов соотносить выбор учебных дисциплин с будущими карьерными целями. Эти выводы были подтверждены метаанализом, показавшим, что образовательные программы, целенаправленно развивающие критическое мышление, способствуют более взвешенным академическим решениям [13]. Кроме того, в работе [14] авторы выявили устойчивую связь между уровнем

аналитических способностей и академической успеваемостью, отметив, что студенты с высоким уровнем критического мышления принимают более обоснованные решения при выборе курсов.

Отдельные исследования сосредоточены на педагогических практиках, способствующих развитию критического мышления. В работе [15] авторы подчеркнули значение дискуссионно-ориентированного обучения, тогда как установлено, что активное участие в онлайн-дискуссиях повышает осознанность студентов в отношении образовательного выбора [16]. Данная перспектива расширена за пределы образовательной сферы, показав, что развитие критического мышления способствует более взвешенному принятию решений в управленческой деятельности и профессиональном обучении [17].

Систематические обзоры и исследования в отдельных предметных областях также подтверждают тесную взаимосвязь между критическим мышлением и академическим планированием. В ходе систематического обзора установлено, что способность к обоснованному выбору курсов напрямую связана с уровнем развития критического мышления [17]. В медицинском образовании показано, что развитые аналитические навыки способствуют успешному академическому прогрессу и осознанному выбору специализированных направлений обучения [19]. Аналогично отмечено, что студенты с высоким уровнем критического мышления демонстрируют более высокие результаты по математике благодаря использованию стратегических подходов к обучению [20].

Взаимосвязь между цифровыми образовательными средами и развитием критического мышления также активно изучается. Установлено, что дистанционное обучение стимулирует студентов более тщательно анализировать свои образовательные траектории [21]. Показано, что обучающиеся, проходящие курсы, ориентированные на развитие критического мышления, проявляют большую ответственность при выборе образовательных направлений [22]. В работе [23] авторы подчеркнули, что разнообразие педагогических подходов способствует формированию более осознанного выбора учебных дисциплин.

Проектно-ориентированное обучение также рассматривается как эффективный механизм развития критического мышления. Установлено, что активное участие студентов в проектной деятельности способствует развитию аналитического мышления и более продуманному академическому принятию решений [24]. Наконец, в работе [25] авторы исследовали взаимосвязь между критическим мышлением, процессом принятия решений и выбором учебных дисциплин, заключив, что аналитическое мышление играет ключевую роль в образовательном планировании.

Опираясь на существующие исследования, настоящая работа расширяет предыдущие подходы за счет интеграции структурированных студенческих данных и методов прогнозного моделирования для оценки уровня критического мышления. В отличие от NLP-подходов, требующих больших объемов текстовой информации, предложенная методология использует регулярно собираемые образовательные данные, что делает ее более масштабируемой и применимой в институциональных условиях. Сравнение моделей линейной регрессии, дерева решений и алгоритма Random Forest также позволяет определить, какие методы наиболее эффективно отражают сложный и нелинейный характер когнитивных навыков в образовательных контекстах.

Представлены эмпирические результаты, полученные с использованием линейной регрессии, дерева решений и Random Forest и их применимость для образовательной оценки и когнитивного моделирования. В данном исследовании использовался набор данных, состоящий из искусственно сгенерированных академических и поведенческих записей 500 студентов бакалавриата, обучающихся по инженерным и IT-специальностям, включая компьютерные науки, информационные системы, кибербезопасность, телекоммуникации, автоматизацию и управление, математическое и компьютерное моделирование, а также робототехнику.

Синтетический набор данных был разработан для имитации реалистичных взаимосвязей между академической успеваемостью, выбором курсов, внеаудиторной деятельностью и ре-

зультатами развития критического мышления. В качестве переменных использовались средний балл успеваемости (GPA), оценки по отдельным курсам, выбор курсов по развитию гибких навыков, участие во внеаудиторных мероприятиях, лидерские роли и результаты оценки критического мышления, смоделированные по модели Уотсона-Глейзера.

Переменная *critical_thinking* была синтетически сгенерирована на основе комбинации академических и поведенческих характеристик студентов. При генерации учитывались показатели GPA, оценки по учебным дисциплинам, участие во внеаудиторной деятельности и лидерская активность. Для имитации естественной индивидуальной вариативности дополнительно вводилась случайная шумовая компонента. Итоговый показатель масштабировался в диапазон от 1 до 100 баллов и интерпретировался как условный аналог результатов тестирования по модели Уотсона-Глейзера.

Генерация синтетических данных использовалась для сохранения анонимности при сохранении статистически значимых закономерностей, подходящих для экспериментов в области машинного обучения. Каждая запись содержит сведения об академической успеваемости, выборе учебных дисциплин и участии во внеурочной деятельности (таблица 1). В качестве целевой переменной использовался синтетически сгенерированный показатель уровня критического мышления, разработанный по аналогии со шкалой теста Уотсона-Глейзера. Показатель отражал условный уровень сформированности аналитических навыков студентов и использовался для оценки способности моделей машинного обучения выявлять закономерности между характеристиками обучающихся и уровнем критического мышления.

Таблица 1 – Описание данных

Переменная	Тип	Описание
<i>critical_thinking</i>	Integer (1-100)	Целевая переменная: измеряет способность учащихся к критическому мышлению.
<i>gpa</i>	Float (2.0 - 4.0)	Средний балл (GPA) по шкале 4,0.
<i>extracurriculars</i>	Integer	Количество часов внеучебной деятельности, в которых участвует учащийся.
<i>course_selection</i>	Categorical	Кодируется как 0 (естественные науки), 1 (искусство), 2 (бизнес).
<i>extracurricular_hours_per_week</i>	Integer	Количество часов в неделю, затрачиваемых на внеучебную деятельность.
<i>advanced_courses</i>	Integer	Количество продвинутых (углубленных) учебных курсов, выбранных студентом.
<i>subject_diversity</i>	Integer	Количество различных предметных областей, изучаемых студентом; характеризует междисциплинарность образовательной траектории.
<i>leadership_role</i>	Binary (0/1)	Наличие лидерской роли в учебной или внеучебной деятельности: 1 – наличие лидерской позиции, 0 – отсутствие.

Набор данных был предварительно обработан с целью устранения пропущенных значений и нормализации числовых признаков. Для сравнительного анализа были выбраны три регрессионные модели на основании их интерпретируемости, способности выявлять нелинейные зависимости и широкой распространенности в интеллектуальном анализе образовательных данных: линейная регрессия, регрессор на основе дерева решений (DT) и регрессор Random Forest (RF). Линейная регрессия использовалась в качестве базовой модели и предполагает линейные взаимосвязи между признаками. Модель дерева решений представляет собой нелинейный, параметрический метод, способный выявлять иерархические правила принятия решений. Random Forest является ансамблем деревьев решений, который повышает обобща-

ющую способность модели за счет усреднения прогнозов и снижения эффекта переобучения. Подбор гиперпараметров для моделей DT и RF выполнялся с использованием метода перебора по сетке (grid search) и пятиблочной перекрестной проверки. Для моделей дерева решений и Random Forest подбор гиперпараметров осуществлялся методом перебора по сетке (grid search) с использованием пятиблочной перекрестной проверки. Для дерева решений варьировались параметры максимальной глубины ($\text{max\_depth} \in \{3, 5, 10, \text{None}\}$), минимального числа объектов в листе ($\text{min\_samples\_leaf} \in \{1, 2, 5\}$) и минимального числа объектов для разбиения ($\text{min\_samples\_split} \in \{2, 5, 10\}$).

Для модели Random Forest дополнительно оптимизировались количество деревьев ($n\_estimators \in \{50, 100, 200\}$) и число признаков, рассматриваемых при разбиении ($\text{max\_features} \in \{\text{«sqrt»}, \text{«log2»}\}$). Оптимальные параметры выбирались по метрике R^2 на валидационной выборке.

Обучение моделей проводилось на 80% выборки, в то время как оставшиеся 20% использовались для тестирования. Оценка качества моделей осуществлялась с помощью коэффициента детерминации R^2 , средней абсолютной ошибки (MAE) и среднеквадратичной ошибки (RMSE). Представленный методологический конвейер демонстрирует практическую применимость использования образовательных данных и методов машинного обучения для оценки абстрактных когнитивных конструкций, таких как критическое мышление.

Результаты и обсуждение

Оценка работы моделей осуществлялась с использованием средней абсолютной ошибки (MAE), среднеквадратичной ошибки (RMSE) и коэффициента детерминации (R^2). Сводные результаты представлены в таблице 2.

Таблица 2 – Сравнение показателей эффективности моделей

Модель	MAE	RMSE	R^2 score	95% CI
Линейная регрессия	3,76	5,73	0,37	± 0.04
Дерево решений	1,161	1,61	0,65	± 0.03
Random Forest	1,22	1,22	0,84	± 0.02

Среди оцененных моделей алгоритм случайного леса показал наилучшую прогностическую эффективность с коэффициентом детерминации R^2 равным 0,84, за ним следуют модель дерева решений ($R^2 = 0,65$) и линейная регрессия ($R^2 = 0,37$). Модель случайного леса также продемонстрировала наименьшие ошибки прогнозирования с точки зрения MAE и RMSE, что указывает на превосходную обобщающую способность на тестовом наборе данных. Рисунок 1 демонстрирует сравнительную эффективность моделей по метрикам MAE, RMSE и R^2 . Видно, что Random Forest обеспечивает наименьшую ошибку и наибольшую объясняющую способность, что указывает на его преимущество при моделировании нелинейных зависимостей.

Модель Random Forest продемонстрировала наилучшие результаты благодаря способности выявлять сложные нелинейные зависимости между входными признаками и целевой переменной. Модели деревьев решений показали приемлемую эффективность, однако оказались более склонны к переобучению, в то время как линейная регрессия продемонстрировала наименьшую результативность из-за предположения о линейном характере зависимости. Для оценки устойчивости результатов дополнительно рассчитывались 95% доверительные интервалы для коэффициента детерминации (R^2). Интервалы определялись методом Bootstrap на тестовой выборке с многократной повторной выборкой наблюдений с возвращением. Для каждой Bootstrap-выборки вычислялось значение R^2 , после чего нижняя и верхняя границы доверительного интервала определялись по 2,5 и 97,5 перцентилем полученного распределения.

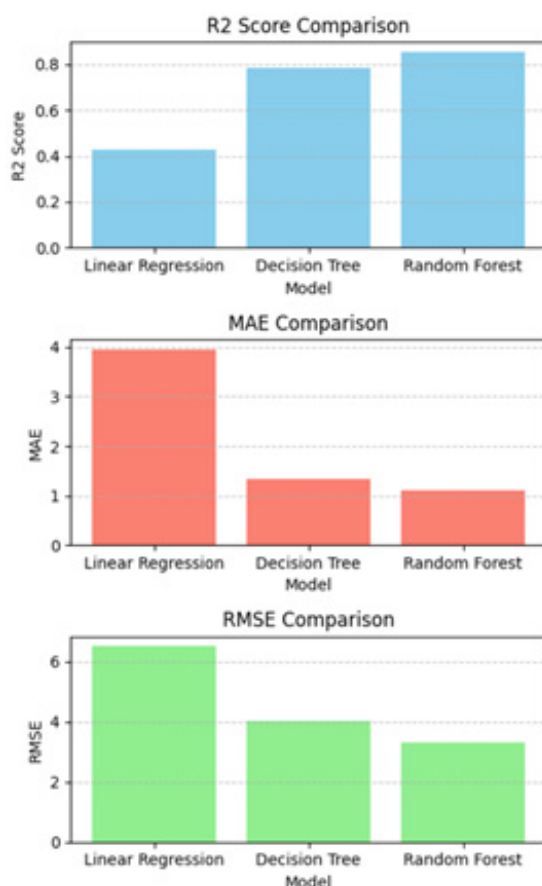


Рисунок 1 – Сравнение эффективности моделей

Узкие доверительные интервалы указывают на стабильную эффективность модели, тогда как более широкие интервалы предполагают большую изменчивость точности прогнозирования в разных выборках.

На рисунке 2 представлено сравнение фактических и прогнозируемых значений показателей критического мышления, полученных с помощью трех моделей. Рисунок 2 иллюстрирует соответствие между фактическими и предсказанными значениями. Для модели Random Forest точки наиболее плотно сгруппированы вдоль диагонали, что свидетельствует о высокой точности предсказаний. В случае линейной регрессии наблюдается большее рассеяние, что подтверждает ее ограниченную применимость.

Ключевые результаты показывают, что студенты с более высоким уровнем критического мышления чаще выбирали естественно-научные дисциплины. По сравнению с академическими показателями участие во внеучебной деятельности оказывало более слабое влияние на выбор курсов. Кроме того, между средним баллом успеваемости (GPA) и уровнем критического мышления была выявлена умеренная корреляция, что указывает на наличие скрытых факторов, влияющих на оба показателя. Анализ значимости признаков показал, что GPA является наиболее сильным предиктором уровня критического мышления, за которым следует выбор учебных дисциплин. Эти результаты согласуются с ранее полученными данными о тесной связи академической успеваемости и развития когнитивных навыков [26]. В то же время участие во внеучебной деятельности продемонстрировало наименьшую значимость, что позволяет предположить: несмотря на вклад таких активностей в развитие социальных и лидерских навыков, они не являются надежными индикаторами уровня критического мышления [27].

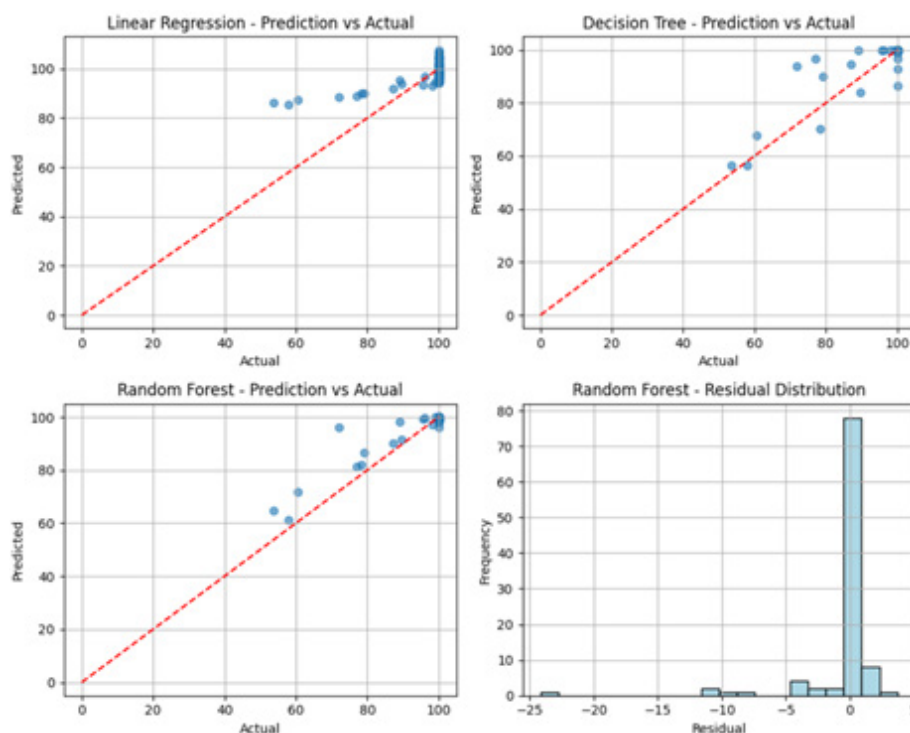


Рисунок 2 – Соотношение фактических и предсказанных значений уровня критического мышления

На рисунке 3 представлены количественные показатели важности признаков. Они подтверждают, что наибольший вклад в прогнозирование показателя критического мышления вносили академические показатели, прежде всего GPA и оценки по дисциплинам.

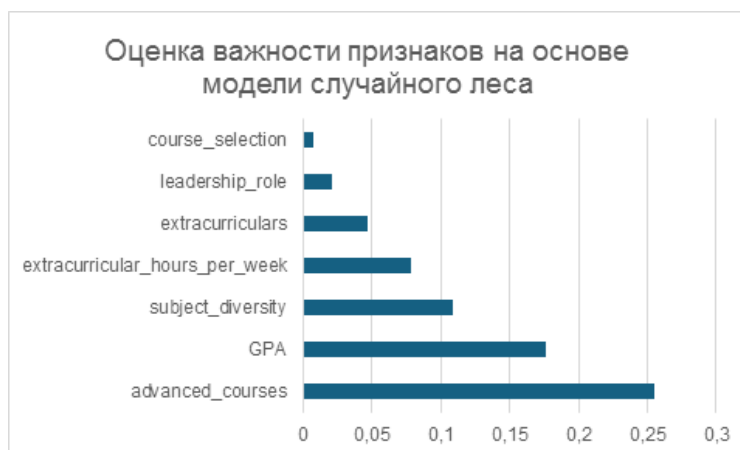


Рисунок 3 – Оценка важности признаков на основе модели случайного леса

Полученные результаты показали, что методы машинного обучения способны эффективно выявлять взаимосвязи между академическими и поведенческими характеристиками студентов и показателями критического мышления. Наилучшие результаты продемонстрировал алгоритм Random Forest, что подтверждает преимущества ансамблевых методов при моделировании сложных нелинейных зависимостей. Полученные данные свидетельствуют о перспективности применения методов машинного обучения для решения задач оценки и прогно-

зирования критического мышления в образовательной среде. Вместе с тем следует учитывать, что исследование выполнено на синтетическом наборе данных, поэтому полученные выводы отражают закономерности, заложенные в процессе моделирования данных, и требуют дополнительной проверки на реальных образовательных выборках.

Заключение

Настоящее исследование демонстрирует высокую прогностическую значимость выбора учебных дисциплин, участия во внеучебной деятельности и академической успеваемости при оценке уровня критического мышления студентов. Среди рассмотренных моделей алгоритм Random Forest показал наибольшую надежность, отличаясь высокой точностью и устойчивостью результатов. Его преимущества подчеркивают эффективность нелинейных моделей по сравнению с традиционными регрессионными подходами при моделировании сложных когнитивных навыков. Средний балл успеваемости (GPA) оказался наиболее значимым предиктором, что подтверждает связь высокой академической успеваемости с более развитым критическим мышлением. В то же время слабые результаты линейной регрессии дополнительно указывают на многомерный и нелинейный характер формирования когнитивных способностей.

В дальнейшем исследования могут быть направлены на расширение набора данных за счет включения дополнительных признаков, таких как читательские привычки, участие в дебатах и результаты тестов на решение проблем, а также на сравнение эффективности более сложных моделей машинного обучения, включая нейронные сети. Кроме того, интеграция данных из систем управления обучением, сетей межличностного взаимодействия, а также социально-экономических и психологических факторов может способствовать дальнейшему улучшению прогностических моделей. Применение методов глубокого обучения и объяснимого искусственного интеллекта также представляет перспективное направление для повышения точности и интерпретируемости результатов.

Информация о финансировании. В статье представлены результаты исследований, выполненных в рамках программно-целевого финансирования на 2024–2026 гг. по проекту IRN BR24993072 «Разработка интеллектуального образовательного ресурса на основе искусственного интеллекта для развития критического мышления студентов», финансируемого Комитетом науки Министерства науки и высшего образования Республики Казахстан.

ЛИТЕРАТУРА

- 1 Paul, R., and L. Elder. The Miniature Guide to Critical Thinking Concepts and Tools (Foundation for Critical Thinking, 2008).
- 2 Liu, O., R. Frankel, and M. Roohr. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. ETS Research Report Series (2014).
- 3 Romero, C., and S. Ventura. Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40 (6), 601–618 (2010). <https://doi.org/10.1109/TSMCC.2010.2053532>
- 4 Luo, Z., J. Mai, C. Feng, D. Kong, J. Liu, Y. Ding, B. Qi, and Z. Zhu. A Method for Prediction and Analysis of Student Performance That Combines Multi-Dimensional Features of Time and Space. Mathematics, 12 (22), 3597 (2024). <https://doi.org/10.3390/math12223597>
- 5 Devi, K., and S. Ratnoo. Predicting student dropouts using random forest. Journal of Statistics and Management Systems, 25, 1579–1590 (2022).
- 6 Behar-Horenstein, L. S., and L. Niu. Teaching critical thinking skills in higher education: A review of the literature. Journal of College Teaching & Learning (TLC), 8 (2) (2011).
- 7 Evens, M., A. Verburch, and J. Elen. Critical thinking in college freshmen: The impact of secondary and higher education. International Journal of Higher Education, 2 (3), 139–151 (2013).

- 8 Gibson, C. Critical thinking: Implications for instruction. (1995).
- 9 Yousef, W. An assessment of critical thinking in the Middle East: Evaluating the effectiveness of special courses interventions. *PLOS ONE*, 16 (12), e0262088 (2021).
- 10 Condon, C., and R. Valverde. Increasing critical thinking in web-based graduate management courses. *Journal of Information Technology Education: Research*, 13, 177–191 (2014).
- 11 Terenzini, P.T., et al. Influences affecting the development of students' critical thinking skills. *Research in Higher Education*, 36 (1), 23–39 (1995).
- 12 Ennis, R. Critical thinking across the curriculum: The Wisdom CTAC Program. *Inquiry: Critical Thinking Across the Disciplines*, 28 (2), 25–45 (2013).
- 13 Niu, L., L.S. Behar-Horenstein, and C.W. Garvan. Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, 9, 114–128 (2013).
- 14 D'Alessio, F. A., B. E. Avolio, and V. Charles. Studying the impact of critical thinking on the academic performance of executive MBA students. *Thinking Skills and Creativity*, 31, 275–283 (2019).
- 15 Wolcott, S. K., et al. Critical thought on critical thinking research. *Journal of Accounting Education*, 20 (2), 85–103 (2002).
- 16 Kalelioğlu, F., and Y. Gülbahar. The effect of instructional techniques on critical thinking and critical thinking dispositions in online discussion. *Journal of Educational Technology & Society*, 17 (1), 248–258 (2014).
- 17 Howard, L.W., T.L.-P. Tang, and M. J. Austin. Teaching critical thinking skills: Ability, motivation, intervention, and the Pygmalion effect. *Journal of Business Ethics*, 128, 133–147 (2015).
- 18 Yuan, H., B. A. Williams, and L. Fan. A systematic review of selected evidence on developing nursing students' critical thinking through problem-based learning. *Nurse Education Today*, 28 (6), 657–663 (2008).
- 19 Pitt, V., et al. The influence of critical thinking skills on performance and progression in a pre-registration nursing program. *Nurse Education Today*, 35 (1), 125–131 (2015).
- 20 Chukwuyenum, A.N. Impact of critical thinking on performance in mathematics among senior secondary school students in Lagos State. *IOSR Journal of Research & Method in Education*, 3 (5), 18–25 (2013).
- 21 Temel, H. The effect of critical thinking course carry out with distance education on critical thinking skills and dispositions. *International Journal of Psychology and Educational Studies*, 9 (3), 792–808 (2022).
- 22 Korkmaz, O., and U. Karakus. The impact of blended learning model on student attitudes towards geography course and their critical thinking dispositions and levels. *Turkish Online Journal of Educational Technology-TOJET*, 8 (4), 51–63 (2009).
- 23 Loes, C., E. Pascarella, and P. Umbach. Effects of diversity experiences on critical thinking skills: Who benefits? *The Journal of Higher Education*, 83 (1), 1–25 (2012).
- 24 Ernst, J., and M. Monroe. The effects of environment-based education on students' critical thinking skills and disposition toward critical thinking. *Environmental Education Research*, 10 (4), 507–522 (2004).
- 25 Plummer, K.J., et al. Enhancing critical thinking skills through decision-based learning. *Innovative Higher Education*, 47 (4), 711–734 (2022).
- 26 Ennis, R.H. Critical thinking assessment. *Theory Into Practice*, 32 (3), 179–186 (1993). <https://doi.org/10.1080/00405849309543594>
- 27 Eccles, J.S., and B.L. Barber. Student council, volunteering, basketball, or marching band: What kind of extracurricular involvement matters? *Journal of Adolescent Research*, 14 (1), 10–43 (1999). <https://doi.org/10.1177/0743558499141003>

^{1*}**Молдагулова А.Н.,**

ф.-м.ғ.к., профессор, ORCID ID: 0000-0002-1596-561X,

*e-mail: a.moldagulova@satbayev.university

¹**Қасенхан А.М.,**

PhD, қауымдастырылған профессор, ORCID ID: 0000-0002-6355-9544,

e-mail: a.kassenkhan@satbayev.university

²**Утебаев Қ.А.,**

магистрант, ORCID ID: 0009-0004-3032-5049,

e-mail: utebayevkuat@gmail.com

²**Ізтұрған Н.Б.,**

магистрант, ORCID ID: 0009-0001-7667-9561,

e-mail: e-mail: tasnur2002@gmail.com

¹Қ.И. Сәтпаев атындағы Қазақ ұлттық зерттеу техникалық университеті,
Алматы қ., Қазақстан

²«МИФИ» Ұлттық зерттеу ядролық университетінің Алматы филиалы,
Алматы қ., Қазақстан

ОҚУШЫЛАРДЫҢ СЫНИ ОЙЛАУ ДАҒДЫЛАРЫН БОЛЖАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІН ЗЕРТТЕУ

Аңдатпа

Бұл мақалада академиялық және мінез-құлық деректеріне негізделген оқушылардың сыни ойлау дағдыларын болжау үшін машиналық оқыту әдістерін қолдану мүмкіндігі қарастырылады. Сыни ойлау қазіргі заманғы білім беруде маңызды дағды ретінде кеңінен танылғанымен, оны тікелей бағалау қиын болып қала береді. Зерттеудің мақсаты – сыни ойлаудың дамуы мен академиялық жетістіктер, сыныптан тыс іс-шараларға қатысу және курс таңдау сияқты өлшенетін факторлар арасындағы байланысты анықтау. Негізгі сипаттамаларды алу үшін 500 анонимді студент профилінен тұратын деректер жиынтығы жиналып, өңделді. Стандартты өнімділік көрсеткіштерін қолдана отырып, үш модель әзірленді және бағаланды: сызықтық регрессия, шешім ағашы және кездейсоқ орман регрессоры. Кездейсоқ орман моделі жоғары болжамдық өнімділікті көрсетті, шешім ағашы үшін 0,65 және сызықтық регрессия үшін 0,37 салыстырғанда R^2 мәні 0,84 болды. Зерттеу нәтижелері оқушылардың академиялық жетістіктері, курс таңдауы және сыныптан тыс іс-әрекет деңгейі олардың сыни ойлау қабілеттері туралы құнды түсінік беретінін көрсетеді. Жалпы алғанда, нәтижелер когнитивтік дағдыларды жанама бағалауға арналған деректерге негізделген тәсілдердің тиімділігін растайды, бұл оқу бағдарламасын әзірлеу, ерте араласу бастамалары және білім беру саясатын жоспарлау үшін практикалық салдарға ие. Оңай қолжетімді білім беру деректерін пайдалану арқылы бұл тәсіл оқыту аналитикасы саласында масштабталатын, объективті және жекелендірілген бағалау әдістерін әзірлеуге ықпал етеді.

Түйін сөздер: машиналық оқыту, сызықтық регрессия, шешім ағашы, кездейсоқ орман, сыни ойлау, оқушылардың сыни ойлау дағдыларын болжау.

^{1*}Moldagulova A.N.,

Cand. Sci. (Physics and Math.), Professor, ORCID ID: 0000-0002-1596-561X,

*e-mail: a.moldagulova@satbayev.university

¹Kassenkhan A.M.,

PhD, Associate Professor, ORCID ID: 0000-0002-6355-9544,

e-mail: a.kassenkhan@satbayev.university

²Utebaev K.A.,

Master's student, ORCID ID: 0009-0004-3032-5049,

e-mail: utebaevkuat@gmail.com

²Izturgan N.B.,

Master's student, ORCID ID: 0009-0001-7667-9561,

e-mail: e-mail: tasnur2002@gmail.com

¹Kazakh National Research Technical University named after K.I. Satbayev,
Almaty, Kazakhstan

²Almaty Branch of the National Research Nuclear University "MEPhI",
Almaty, Kazakhstan

RESEARCH OF MACHINE LEARNING MODELS FOR PREDICTING STUDENTS' CRITICAL THINKING SKILLS

Abstract

This article explores the potential of using machine learning methods to predict students' critical thinking skills based on academic and behavioral data. Although critical thinking is widely recognized as a critical skill in modern education, its direct assessment remains challenging. The aim of the study was to identify the relationship between critical thinking development and measurable factors such as academic achievement, participation in extracurricular activities, and course selection. A dataset of 500 anonymized student profiles was collected and processed to extract key characteristics. Three models were developed and evaluated using standard performance metrics: linear regression, decision tree, and random forest regressor. The random forest model demonstrated superior predictive performance, achieving an R^2 value of 0.84, compared to 0.65 for the decision tree and 0.37 for the linear regression. The study's findings demonstrate that students' academic achievement, course selection, and level of extracurricular activity collectively provide valuable insight into their critical thinking abilities. Overall, the results support the effectiveness of data-driven approaches for indirectly assessing cognitive skills, which has practical implications for curriculum development, early intervention initiatives, and educational policy planning. By leveraging readily available educational data, this approach facilitates the development of more scalable, objective, and personalized assessment methods in the field of learning analytics.

Keywords: machine learning, linear regression, decision tree, random forest, critical thinking, predicting students' critical thinking skills.

Received February 2, 2026; revised February 13, May 12, June 8, 2026; accepted June 13, 2026.

ФИЗИКАЛЫҚ ҒЫЛЫМДАР
PHYSICAL SCIENCES
ФИЗИЧЕСКИЕ НАУКИ

UDC 533.9.082; 537.525.99
IRSTI 29.27.43

<https://doi.org/10.55452/1998-6688-2026-23-2-365-380>

^{1,2}**Utegenov A.U.**,
PhD, ORCID ID: 0000-0001-6768-0918,
*e-mail: almasbek@physics.kz
^{1,2}**Ussenkhan S.S.**,
ORCID ID: 0009-0009-6597-2066,
e-mail: sultanusenhan@gmail.com
³**Ussenov Ye.A.**,
ORCID ID: 0000-0001-7903-8674,
e-mail: yerbolat@physics.kz
^{1,2}**Orazbayev S.A.**,
PhD, ORCID ID: 0000-0002-7286-9990,
e-mail: sagi.orazbayev@gmail.com

¹Institute of Applied Sciences and Information Technologies, Almaty, Kazakhstan

²Al-Farabi Kazakh National University, Almaty, Kazakhstan

³Princeton Plasma Physics Laboratory, NJ, Princeton, USA

**AUTOMATION OF NANOPARTICLE SYNTHESIS
PROCESSES IN A PLASMA ENVIRONMENT USING LABVIEW**

Abstract

This work presents an automated control system for the synthesis of nanomaterials by plasma-enhanced chemical vapor deposition (PECVD), implemented using the LabVIEW software environment. The main objective of the study is to develop an integrated hardware-software platform that enables sequential control of the key stages of the PECVD process, including vacuum chamber preparation, pressure monitoring, working gas supply, plasma ignition, power matching, cyclic nanomaterial growth, and optical monitoring of nanoparticles in the plasma environment. The use of LabVIEW made it possible to integrate actuator control, experimental parameter acquisition, and real-time process visualization within a single automated system. The automated cycle begins with evacuation of the reaction chamber to a predefined base pressure. Transition to the next stage is permitted only after the specified pressure threshold has been reached, ensuring reproducible initial conditions for each experiment. The program then controls the supply of the working gas through mass flow controllers (MFCs). In this work, two gas-flow control modes were considered: analog control using a 0÷5 V voltage signal and digital communication via RS-232 interface. It was shown that the analog approach requires accurate scaling of the control voltage, since applying 5 V corresponds to full-scale opening of the controller and results in the maximum gas flow. In contrast, the RS-232 interface enables the gas flow rate to be specified directly in sccm, improving the accuracy, flexibility, and

convenience of gas-environment control. After pressure stabilization, LabVIEW initiates RF plasma ignition and executes the RF matching algorithm aimed at minimizing reflected power and improving the stability of the plasma process. A separate software module implements the cyclic nanomaterial growth mode, in which the plasma-on time, plasma duration, and total number of synthesis cycles are predefined. This approach makes it possible to control material accumulation on the substrate and to correlate the process parameters with the morphological characteristics of the resulting nanostructures. The final module of the system is designed for optical monitoring of the nanoparticle cloud density in dusty plasma. For this purpose, the change in the intensity of laser radiation passing through the plasma region is recorded using a photodetector and a Keithley 2401 measuring unit connected to LabVIEW via RS-232 interface. The difference between the initial and modified optical signal intensity is used as a diagnostic parameter characterizing the formation and temporal evolution of nanoparticles. The developed system demonstrates that LabVIEW can be effectively applied not only for the automation of individual instruments, but also for the implementation of a complete digital control cycle for PECVD-based nanomaterial synthesis.

Keywords: PECVD, automated plasma synthesis, nanoparticle growth control, dusty plasma diagnostics, plasma nanotechnology.

Received May 2, 2026; accepted June 2, 2026.

Introduction

Plasma-enhanced chemical vapor deposition (PECVD) is one of the most versatile methods for the formation of thin films, nanostructures, and functional coatings. This is because low-temperature plasma significantly expands the range of possible chemical reactions and enables controlled material growth near the substrate surface [1–6]. One of the key advantages of PECVD over thermal deposition methods is its ability to operate at reduced temperatures, which is particularly important for temperature-sensitive substrates and for preventing defect formation, diffusion processes, and degradation of the substrate material [7, 8]. Additional advantages of PECVD include high deposition rates, the possibility of forming graded and multilayer structures, and the ability to control interfacial regions and film microstructure. These features make PECVD a highly attractive technique for micro- and nanotechnology applications [9].

The efficiency of the PECVD process is determined by a combination of technological parameters, including pressure, RF discharge power, gas flow rate and gas ratio, substrate temperature, and the condition of the plasma reactor [7, 10]. For this reason, modern PECVD studies increasingly employ in situ diagnostic techniques, such as optical emission spectroscopy (OES), mass spectrometry, and analysis of the electronic parameters of the plasma, in order to correlate the plasma composition with the morphology, structure, and electrical properties of the deposited coatings [10–13]. For instance, several studies have shown that pulsed precursor injection and controlled dilution regimes, such as hydrogen dilution, can be used to finely tune the growth, densification, and transition stages from an amorphous to a nanocrystalline structure [14, 15].

In such experimental systems, LabVIEW provides a convenient software environment for measurement automation, actuator control, and real-time feedback implementation [16–18]. Published studies have demonstrated its use for automatic acquisition of plasma probe characteristics, rapid calculation of electron parameters, continuous processing of diagnostic signals, and process control according to predefined quality criteria. These capabilities are particularly useful for investigating plasma-chemical processes and synthesizing nanostructures [19]. Thus, the use of LabVIEW in PECVD experiments can improve process reproducibility, reduce operator influence, and provide a more accurate correlation between plasma parameters and the properties of the synthesized material [20].

When moving from conventional PECVD processes to the synthesis of nanoparticles directly in plasma, it is important to emphasize that the plasma environment acts not only as an energy source, but also as a tool for controlling nanoparticle nucleation, evolution, and subsequent assembly into the desired structure [2, 21–23]. In such processes, control over individual stages of material formation is especially important, since the final size, morphology, uniformity, and functional properties of nanoparticles are largely determined during nucleation, growth, charging, and coagulation.

The practical importance of continuous process monitoring is related to the fact that variations in plasma parameters, such as current, pressure, gas composition, and temperature, can significantly affect particle size and the competing mechanisms of particle formation and enlargement [24]. In studies on oxide nanoparticle synthesis in thermal plasma, in situ spectroscopy has been shown to directly correlate particle size with plasma parameters, while the spectral characteristics of the plasma column provide insight into the growth mechanism. Therefore, monitoring each stage of synthesis is necessary not only to obtain a target average particle size, but also to narrow the size distribution, improve reproducibility, and prevent undesirable agglomeration.

This is where LabVIEW becomes particularly useful, as it enables diagnostics, visualization, data logging, and feedback control to be integrated into a single experimental control system [16, 17]. This makes it possible to move from passive observation to active control of the synthesis process, where plasma parameters can be adjusted in real time based on diagnostic data, making the process more stable and predictable [10, 25].

For the vacuum subsystem of a PECVD setup, LabVIEW is especially valuable as a platform that combines control and recording of key process parameters. This is because the quality of deposition is determined not only by the plasma itself, but also by the stability of pressure, gas flow, chamber temperature, and the previous operating history of the reactor [7, 10]. In automated systems, LabVIEW has already been used for computerized hardware control and process logging, demonstrating its suitability for building flexible control systems for vacuum-plasma equipment [26].

In practice, LabVIEW makes it possible to implement a step-by-step operating algorithm for the experimental setup, including chamber evacuation, pressure stabilization, working gas supply, discharge ignition, mode switching, and emergency shutdown when process parameters exceed the allowable limits [27, 28]. In addition, LabVIEW is convenient for in situ data acquisition from pressure, flow, voltage, and current sensors, as well as for integrating plasma diagnostics and real-time feedback. This improves reproducibility and simplifies control over the individual stages of synthesis.

Materials and methods

Full Automation of Experimental Process

To implement automated control of nanomaterial synthesis by PECVD, a control system based on the LabVIEW software environment was developed. In this work, LabVIEW was used as the central platform for coordinating all major stages of the experiment, from vacuum chamber preparation and pressure monitoring to gas supply control, RF discharge ignition, execution of the predefined synthesis scenario, and storage of technological parameters. The choice of LabVIEW was motivated by its wide use in automated measurement and control systems, its ability to interface with different instruments, and its capabilities for user-interface development and real-time experimental data acquisition.

The automated PECVD synthesis cycle begins with preparation of the reaction chamber. At the first stage, the system evacuates the chamber to a predefined base pressure while continuously monitoring the vacuum gauge readings. Transition to the next stage is allowed only after the required pressure value has been reached, ensuring identical initial conditions for each experimental series. This approach is particularly important for plasma-chemical processes, since pressure, gas composition, and the condition of the chamber directly affect discharge stability and the properties of the resulting material. PECVD studies also indicate that process parameters, including pressure, gas environment, and RF power, are key factors controlling thin-film deposition and the plasma process as a whole.

After the base pressure is reached, LabVIEW automatically initiates the supply of the working gas or gas mixture into the reactor. At this stage, the system monitors the pressure variation inside the chamber and ensures that the required working pressure for subsequent plasma ignition is achieved. After pressure stabilization, the gas supply can either be stopped or switched to a predefined flow mode, depending on the selected synthesis scenario. Thus, preparation of the gas environment is

performed not manually, but according to a predefined algorithm, which reduces operator influence and improves the reproducibility of experimental conditions. A similar automation principle, based on sequential control of temperature, pressure, and gas flows, is widely used in CVD/PECVD systems to improve process stability and repeatability [29, 30].

After the required vacuum-gas conditions have been established, the system proceeds to the plasma ignition stage. LabVIEW turns on the RF generator according to the predefined scenario and then controls the sequence of technological operations required for the synthesis process. At this stage, the program functions not only as an operator interface, but also as a logical controller that determines the order of transitions between the process stages. This approach makes it possible to represent PECVD synthesis as a set of interconnected technological states, where each subsequent stage starts only after the predefined conditions of the previous stage have been fulfilled.

The main advantage of the developed system is that the entire synthesis process is carried out automatically according to a predefined algorithm. This minimizes manual intervention, reduces the probability of operator errors, and ensures repeatable experimental conditions. In automated PECVD systems, such an approach is used to implement both semi-automatic and fully automatic operating modes, contributing to improved film uniformity and reproducibility of the properties of the obtained materials [31]. In addition, automatic logging of process parameters enables the creation of a digital record for each experiment, including evacuation time, chamber pressure, gas inlet timing, working pressure, plasma ignition parameters, duration of individual process stages, and total synthesis time.

The automated PECVD synthesis algorithm includes the following steps:

1. Pumping of the vacuum chamber to the base pressure.
2. Real-time pressure monitoring.
3. Gas inlet until the target working pressure is reached.
4. Stabilization of the working pressure or interruption of the gas flow according to the selected synthesis scenario.
5. RF plasma ignition under predefined power conditions.
6. Execution of the programmed synthesis scenario.
7. Automatic shutdown of the plasma after the defined process time.
8. Final pumping of the chamber.
9. Logging and saving of the technological parameters for further analysis.

Gas Flow Control and MFC Interface Implementation

An important part of the automated PECVD control system is the regulation of the working gas flow before plasma ignition. In the developed LabVIEW program, several approaches were implemented for controlling mass flow controllers (MFCs), depending on the available communication interface of each device. This was necessary because different MFC models support different control modes, including analog voltage input and digital communication through RS-232. In many commercial MFCs, the setpoint can be supplied through an analog input, commonly in the range of 0–5 V, where the applied voltage corresponds to a certain fraction of the full-scale flow range. For example, most of the manuals describe a standard 0–5 Vdc analog setpoint input applied to a specific connector pin, while MKS documentation also describes MFC versions with analog setpoint and analog flow signals in the 0–5 Vdc range.

At the first stage, several MFCs were tested using an analog control scheme. In this configuration, LabVIEW generates a control signal that is applied to the analog input pin of the MFC. However, this approach has an important limitation: if the system applies only a fixed high-level signal, for example 5 V, the controller interprets this as a full-scale setpoint and opens the flow to the maximum calibrated range. Therefore, analog control is suitable only when the voltage can be precisely generated and scaled according to the required flow rate. In this case, the desired flow rate can be converted into a voltage setpoint using a calibration relationship:

$$V_{set} = 5 \times \frac{Q_{set}}{Q_{FS}} \quad (1)$$

where V_{set} is the analog voltage applied to the MFC, Q_{set} is the required gas flow rate in sccm, and Q_{FS} is the full-scale range of the controller. Without such voltage scaling, analog control becomes closer to an on/off mode and does not provide accurate regulation of the gas flow, which is critical for reproducible PECVD synthesis.

A more flexible control approach was implemented for MFCs equipped with an RS-232 interface. In this case, a dedicated LabVIEW driver was developed for communication with a specific MFC model. The driver provides serial communication between the computer and the controller, initializes the COM port, sends the required setpoint command, and reads the actual flow response from the device. This approach is consistent with the general LabVIEW instrument-control architecture, where NI-VISA is used for communication with serial instruments through RS-232, RS-422, or RS-485 interfaces. NI documentation also notes that serial communication in LabVIEW requires correct port configuration, including baud rate, stop bits, parity and device-specific command format [32, 33].

The developed gas-flow control module provides two operation modes: analog voltage-based control for MFCs without digital communication and RS-232-based control for digitally accessible MFCs. The RS-232 approach was found to be more suitable for automated PECVD experiments because it enables direct assignment of the required gas flow, reduces the probability of over-opening the valve, and improves the reproducibility of the gas environment before plasma ignition. This module is integrated into the general LabVIEW-based PECVD algorithm, where the gas flow is controlled together with chamber pressure, pumping sequence, plasma ignition and the subsequent synthesis scenario.



Figure 1 – Front diagram for gas flow control

Figure 1 shows an example of the LabVIEW front panel developed for controlling the flow of the plasma-forming gas. As shown in the interface, the operator sets the ‘Setpoint Input’ value in the range of 0–5 V, where 5 V corresponds to the full-scale opening of the MFC and therefore to the maximum available gas flow. The control voltage is supplied to the corresponding MFC input pin using a National Instruments 24-channel, 8.5 mA I/O device, which provides stable and reliable analog signal generation for gas-flow control. The front panel also includes a graphical indicator for monitoring the gas-flow control state, allowing the operator to observe the process in real time and stop it if necessary.

Pressure-Controlled Vacuum and Gas Preparation Sequence

After implementation of the gas-flow control module, the next stage of the LabVIEW-based PECVD system was the automation of vacuum and gas preparation before plasma ignition. This stage is critical because the initial pressure, gas composition, and stabilization time define the starting conditions for plasma formation and strongly influence the reproducibility of the synthesis process. In the developed system, LabVIEW acts as the central control environment that coordinates the

vacuum pump, gas inlet line, pressure sensor, MFCs, and plasma ignition sequence. This approach is consistent with the general purpose of LabVIEW as a graphical environment for automated test, measurement, instrument connectivity, user-interface development, and data acquisition.

The automated procedure begins with evacuation of the PECVD chamber to a predefined base pressure. During this stage, the pressure signal is continuously acquired and displayed in the LabVIEW front panel. The program does not allow transition to the next stage until the measured pressure reaches the required threshold. This condition-based transition is important because it ensures that each synthesis experiment starts from comparable vacuum conditions. After the base pressure is achieved, LabVIEW opens the gas inlet line and sends the required flow setpoint to the selected mass-flow controller. Depending on the MFC interface, this setpoint can be generated either as an analog signal or as a digital command through RS-232 communication.

For MFCs with analog input, the gas-flow setpoint is controlled by applying a voltage signal, typically in the 0–5 V range. In such controllers, the applied voltage is proportional to the full-scale flow range; therefore, 5 V corresponds to the maximum calibrated flow of the device. This type of analog signal is commonly used in commercial mass-flow controllers, while digital MFCs may also support serial communication for direct setpoint programming. However, analog control requires careful voltage scaling. If a fixed 5 V signal is applied without conversion from the required flow rate, the MFC opens to its full-scale range, which can disturb the working pressure and reduce the reproducibility of the PECVD process.

To overcome this limitation, an RS-232-based control mode was implemented for digitally accessible MFCs. In this case, a dedicated LabVIEW driver was developed for the selected controller. The driver initializes the serial port, sends the gas-flow setpoint directly in engineering units, such as sccm, reads the actual flow response, and closes the communication session after the operation. The advantage of this approach is that the operator or program can define the desired gas flow directly in the LabVIEW block diagram without manual conversion to voltage. Commercial MFCs with serial interfaces are designed for such programmable functions, including flow setpoint control and reading of flow-related parameters.

After gas injection, the system continuously monitors the chamber pressure until the target working pressure is reached. The transition to the plasma ignition stage is performed only when the pressure remains within a predefined tolerance range for a selected stabilization time. This logic can be described by the following condition:

$$|P(t) - P_{set}| < \Delta P \text{ for } t \geq t_{stab} \quad (2)$$

where $P(t)$ is the measured chamber pressure, P_{set} is the target working pressure, ΔP is the allowed pressure deviation, and t_{stab} is the stabilization time. This condition prevents premature plasma ignition under unstable gas-pressure conditions and improves the repeatability of the synthesis process.

Such pressure-controlled preparation is important for PECVD because deposition pressure can significantly affect the properties of the deposited material. For example, studies on RF-PECVD thin-film growth have shown that changes in deposition pressure can strongly influence film thickness, bonding structure, optical properties, and morphology [34]. Therefore, automatic control of the pressure and gas-flow sequence is not only a technical feature of the LabVIEW program but also an essential requirement for obtaining reproducible nanomaterials.

Therefore, automatic control of the pressure and gas-flow sequence is not only a technical feature of the LabVIEW program but also an essential requirement for obtaining reproducible nanomaterials [35]. In the present PECVD system, this logic allows the operator to define the experimental scenario in advance, while LabVIEW executes each stage automatically and records the corresponding technological parameters.

As a result, the automated vacuum–gas preparation module reduces manual intervention, prevents uncontrolled gas admission, and ensures that plasma ignition occurs only under predefined

and stable process conditions. This makes the LabVIEW system not only a control interface but also a process-management platform for reproducible PECVD synthesis of nanomaterials.

Results and discussion

RF Matching & Cyclic growth Control

To control the nanomaterial synthesis process by PECVD, a universal LabVIEW block diagram was developed for the first stage of the experiment, namely RF matching (Figure 2a). At this stage, the RF plasma is ignited at a specified power, which is selected according to the specific experimental objective. It should be emphasized that minimization of reflected power is one of the key requirements for stable ignition and sustained operation of RF plasma in the PECVD process. During the discharge, the plasma behaves as a dynamically changing load whose impedance depends on pressure, gas-mixture composition, supplied RF power, electrode surface condition, and the current stage of film or nanomaterial growth [36]. When the impedance is not properly matched, a fraction of the applied RF power is reflected back to the generator. This reduces the effective power absorbed by the plasma and may lead to discharge instability, fluctuations in plasma parameters, changes in the deposition rate, and deterioration in the reproducibility of the synthesized nanomaterial properties [37,38]. Therefore, in the automated PECVD control system, monitoring the forward and reflected RF power, as well as adjusting the RF matching unit, is essential for maintaining a stable plasma environment and ensuring the quality of the obtained materials.

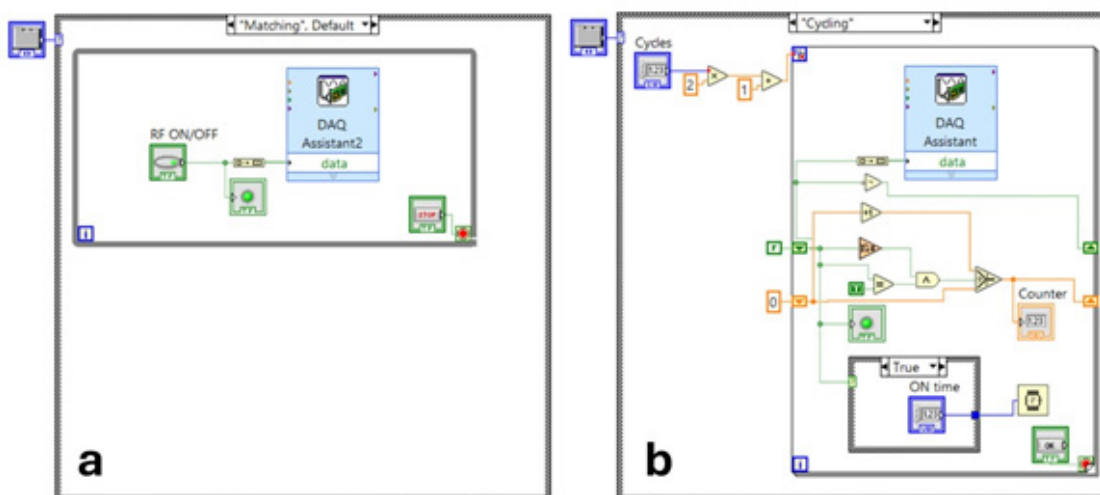


Figure 2 – Block diagram for RF matching following with synthesis process control

The second stage of automated control in the PECVD-based nanomaterial synthesis process is related to setting the parameters of cyclic nanomaterial growth (Figure 2b). At this stage, the main time parameters of the process are defined in LabVIEW: the plasma-on time, the plasma-off time, and the total number of growth cycles. During the plasma-on time period, the gas mixture is activated, precursor molecules are dissociated, chemically active radicals and ions are formed, and nanoparticle nucleation and growth occur in the plasma volume. After the plasma is switched off, the particles lose their charge, leave the confinement region governed by the plasma potential, and are deposited onto the substrate surface, for example, onto a Si substrate intended for subsequent analysis by SEM, TEM, AFM, Raman spectroscopy, XPS, or other analytical techniques. A similar mechanism of cyclic nanoparticle deposition has been demonstrated for PECVD synthesis of carbon nanostructures, where the number of particles deposited on the surface is controlled by the number of cycles, while the surface morphology depends on the cycle duration and plasma parameters [39–41].

For operator convenience, a Counter tool was also added to the LabVIEW interface to display the current number of the synthesis cycle being performed.

Control of the plasma-on and plasma-off times is an important tool for regulating the size, concentration, and spatial distribution of nanoparticles. In particular, for silicon nanoparticle synthesis by VHF-PECVD, it has been shown that pulsed plasma operation can produce particles with a narrow size distribution, while variation of the plasma-off time can be used to control nanoparticle size [42]. Therefore, in the developed LabVIEW system, the cyclic growth mode is implemented as a controlled sequence of operations that includes setting the plasma ignition time, pause duration, number of repeated cycles, and automatic logging of process parameters [12]. This approach improves synthesis reproducibility, enables controlled accumulation of material on the substrate, and allows the automated control parameters to be correlated with the final morphological and structural characteristics of the obtained nanomaterials [43, 44].

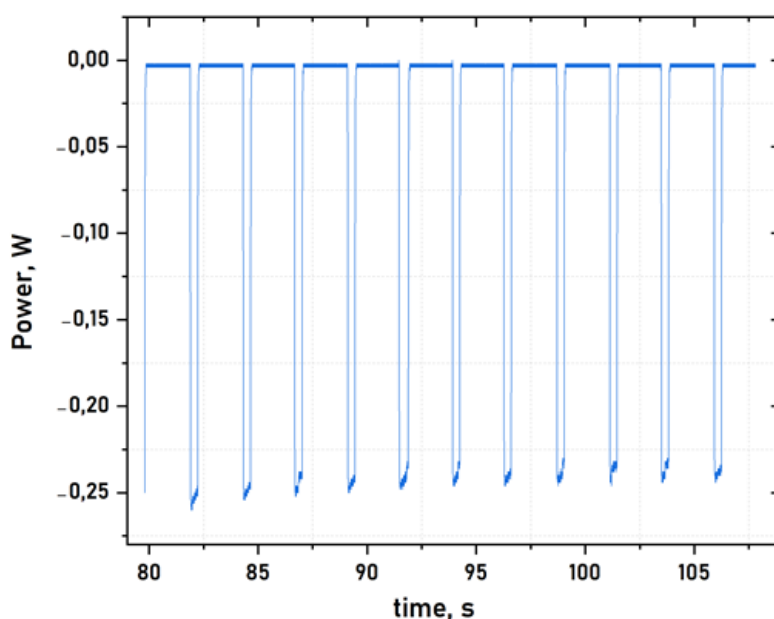


Figure 3 – Time evolution of PECVD process

Figure 3 shows representative time-dependent changes in the main technological and diagnostic parameters during the automated PECVD cycle. The oscillograph reflects the automated RF cycling the PECVD process for nanoparticle synthesis. The parameters were as follows: plasma on/off times 200 ms and 2000 ms, accordingly. As it was mentioned before, the time control is also possible to maintain by the user. The gas flow for 2 types of gases was 40 sccm and 5 sccm for nitrogen and methane gases, accordingly. As the graph shows, the plasma on and off times were correctly controlled.

Monitoring of the nanoparticle cloud density

The final stage of the developed LabVIEW-based PECVD control system was focused on the optical monitoring of nanoparticle cloud density in dusty plasma. This module was implemented to provide an in-situ diagnostic of the nanoparticle cloud formed during plasma-assisted synthesis. In contrast to ex situ methods such as SEM or TEM, which provide information only after particle deposition on a substrate, optical monitoring makes it possible to follow the temporal evolution of nanoparticles directly during the plasma process. This is especially important for dusty plasma systems, where the formation, charging, growth and transport of nanoparticles can significantly affect the plasma parameters and the final properties of the synthesized material.

The monitoring principle is based on measuring the change in laser radiation intensity caused by the interaction of the laser beam with nanoparticles suspended in plasma volume. In the developed system, the laser beam passes through the dusty plasma region, while the optical signal is detected by a photodetector connected to a Keithley measuring unit and recorded automatically using LabVIEW. The signal is measured before significant nanoparticle growth and then during or after the growth stage. The difference between the initial and final optical intensities is used as an indicator of nanoparticle cloud formation and can be further related to the particle number density. The relationship between the laser signal and nanoparticle cloud density is based on the attenuation of laser radiation passing through the plasma region. When nanoparticles are formed and suspended in the plasma volume, they scatter and partially absorb the incident laser light. As a result, the transmitted laser intensity decreases compared with the initial intensity measured before significant nanoparticle formation. Therefore, the change in laser intensity can be used as a relative diagnostic parameter characterizing the formation and temporal evolution of the nanoparticle cloud. In the present work, the optical signal is used mainly for relative in situ monitoring. Absolute determination of nanoparticle number density would require additional calibration, including information about particle size, optical properties, and the effective interaction length of the laser beam with the nanoparticle cloud.

A similar laser attenuation approach was used by Ongaibergenov et al. to determine the time-dependent evolution of carbon nanoparticle cloud density and size in dusty glow discharge plasma; in that work, the laser signal was detected by a silicon photodiode, while signal acquisition and intensity analysis were performed using NI LabVIEW and a Keithley 2401 measuring unit (Figure 4) [12].

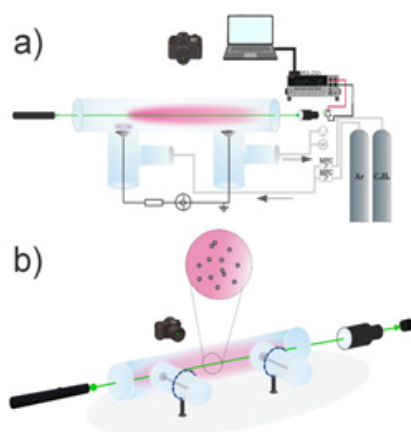


Figure 4 – Schematic diagram of the experimental setup for monitoring the density of the carbon nanoparticle cloud during the synthesis process

In the present work, the novelty is not only the use of optical diagnostics itself, but its integration into the automated LabVIEW control architecture. The photodetector signal is continuously converted into a photocurrent or voltage signal by the Keithley unit and transferred to LabVIEW, where it is displayed, stored and synchronized with other technological parameters of the PECVD process. These parameters include chamber pressure, gas-flow state, RF power, plasma ignition time, plasma-on/plasma-off cycle duration and total synthesis time. Therefore, each synthesis experiment produces a digital dataset that links the optical response of the nanoparticle cloud with the controlled plasma-growth conditions.

The use of laser-based diagnostics for dusty plasma and plasma-grown nanoparticles is well established. Boufendi et al. noted that laser light scattering is one of the most widely used techniques for detecting and sizing nanoparticles during their growth in dusty plasma systems [45]. In addition, Ouaras et al. demonstrated that time-resolved in situ laser extinction and scattering diagnostics can be used together with SEM and plasma diagnostics to study nanoparticle synthesis and transport in low-pressure plasma. Their results showed that the decrease in laser transmission is correlated

with changes in nanoparticle cloud density and size, and that laser scattering imaging can reveal the spatial structure of the nanoparticle cloud, including void formation [46].

The importance of nanoparticle monitoring is also supported by studies showing that particle growth changes the plasma state itself. Ussenov et al. investigated nanodust-containing argon-acetylene RF plasmas and showed that cyclic nanoparticle growth affects electron density, electron temperature and plasma potential; they also used laser light scattering to confirm the presence of the dust particle cloud near the probe region [47]. This confirms that optical monitoring of the nanoparticle cloud is not only a material-diagnostics tool, but also an important method for understanding how nanoparticle formation modifies the plasma environment during synthesis.

Thus, the developed LabVIEW-based optical monitoring module provides an additional feedback channel for the automated PECVD system. While the previous modules control the vacuum preparation, gas admission, RF matching and cyclic plasma growth, this final module records the optical response of the nanoparticle cloud and allows the nanoparticle cloud density evolution to be evaluated. The integration of this diagnostic into LabVIEW makes it possible to move from simple process automation toward data-driven plasma synthesis, where the growth of nanomaterials can be correlated with real-time optical signals and controlled technological parameters.

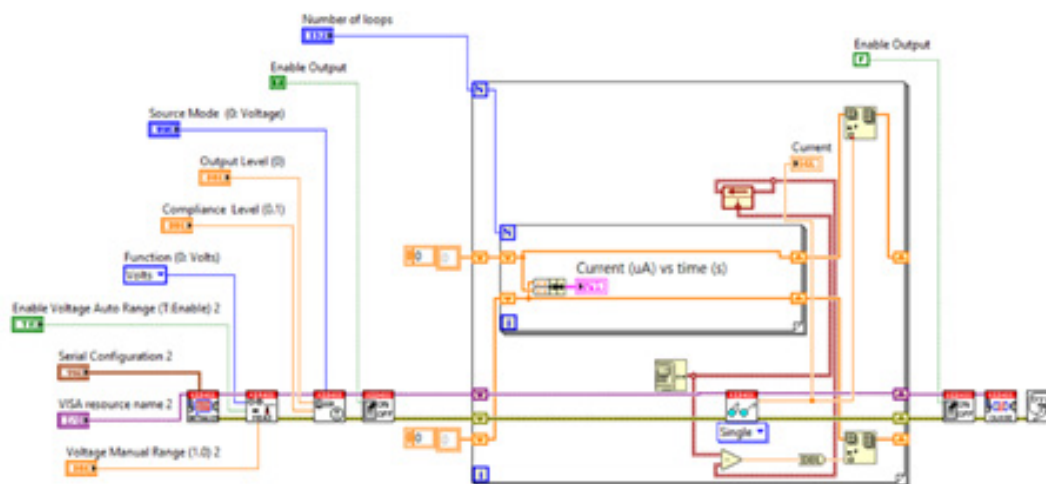


Figure 5 – LabVIEW block diagram for optical monitoring of nanoparticle cloud density in dusty plasma

Figure 5 shows the LabVIEW block diagram for the optical monitoring module, in which communication with the Keithley 2401 measuring unit is established via the RS-232 interface. This module records the photocurrent signal from the photodetector through the Keithley unit and evaluates the change in laser intensity caused by the nanoparticle cloud formed during plasma synthesis.

Conclusion

In this work, a LabVIEW-oriented automated control system for the PECVD synthesis of nanomaterials was developed and described. The main result of the study is the creation of a unified control algorithm that covers the complete experimental cycle: from vacuum chamber preparation and working gas supply to plasma ignition, cyclic growth implementation, and optical signal acquisition from nanoparticles formed in the plasma volume. This approach transforms the PECVD experiment from manual control of individual devices into a sequential, programmable, and reproducible technological process. The scientific novelty of this work lies in the development of a unified LabVIEW-based control algorithm for PECVD nanomaterial synthesis, where several technological and diagnostic modules are integrated into a single automated platform. Unlike partial

automation of individual instruments, the proposed system combines vacuum chamber preparation, pressure-controlled gas admission, MFC control through analog and RS-232 interfaces, RF plasma ignition and matching, cyclic growth control, optical monitoring of nanoparticle cloud evolution, real-time visualization, and data logging. This integrated approach makes it possible to transform the PECVD process into a sequential, programmable, and reproducible digital synthesis cycle.

At the first stage, automation of the vacuum–gas preparation procedure was implemented. LabVIEW provides continuous pressure acquisition, monitors the achievement of the base vacuum level, and allows the system to proceed to the next stage only after the predefined condition has been fulfilled. This prevents the process from being initiated under unstable initial conditions and improves experimental repeatability. In addition, a working-pressure stabilization logic was implemented before plasma ignition, which is particularly important for PECVD because pressure and gas composition directly influence discharge parameters, deposition rate, and the properties of the formed material.

Another important result is the development of a gas-flow control module based on mass flow controllers. Two practical connection modes were considered: analog control using a 0–5 V signal and digital control via RS-232 communication. The analog mode can be used for MFCs without a digital interface; however, it requires accurate conversion of the required gas flow into the corresponding control voltage. Otherwise, applying a high-level signal, such as 5 V, results in full-scale opening of the controller and disrupts the required gas-supply regime. The RS-232 mode was found to be more effective, and a dedicated communication driver was developed for this purpose. In this case, the gas flow rate is specified directly in sccm, while LabVIEW sends the command to the device and receives feedback on the flow status. This improves control accuracy, reduces the risk of uncontrolled gas admission, and makes the process more convenient for automation.

The next module of the system is related to RF plasma control. LabVIEW is used to ignite the discharge according to a predefined scenario and to control the RF matching stage. Minimization of reflected power is essential because impedance mismatch between the RF generator and the plasma load reduces the effective power absorbed by the plasma and may cause discharge instability. Integrating RF matching into the general automated control algorithm improves the stability of the plasma process and provides more reproducible conditions for nanomaterial growth.

A cyclic growth module was also implemented, in which the plasma-on time, pause duration, and total number of cycles are defined. This structure makes it possible to control not only the plasma ignition event itself, but also the temporal architecture of the synthesis process. During the plasma-on period, the gas environment is activated and nanoparticles are formed, whereas during the pause, conditions are created for their deposition onto the substrate. The presence of a cycle counter and automatic time logging makes this mode convenient for systematic variation of process parameters and their subsequent comparison with SEM, TEM, AFM, Raman, XPS, and other analytical results.

The final element of the system is an optical monitoring module for evaluating the density of the nanoparticle cloud in dusty plasma. This module records changes in the intensity of laser radiation interacting with nanoparticles in the plasma volume. The photodetector signal is measured using a Keithley 2401 unit and transferred to LabVIEW, where it is stored together with the synthesis parameters. This makes it possible to correlate the dynamics of nanoparticle formation with pressure, gas supply, plasma-on time, number of cycles, and RF discharge parameters.

Thus, the developed system demonstrates that LabVIEW is an effective tool for complete automated control of PECVD-based nanomaterial synthesis. Its key advantage lies in the integration of control, diagnostics, visualization, and data logging within a single platform. This reduces operator influence, improves experimental reproducibility, provides a digital record of each synthesis process, and creates a basis for the further transition toward intelligent plasma-process control based on feedback and experimental data analysis. Moreover, the developed LabVIEW-based automation platform can serve as a basis for future implementation of feedback control algorithms using diagnostic signals such as pressure, gas flow rate, RF forward/reflected power, and optical monitoring data. Also, it should be noted that machine learning approaches may be used in future

work to predict optimal synthesis parameters and improve the control strategy of nanoparticle growth in plasma environments.

Information on funding

This research has been funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23485948).

REFERENCES

- 1 Martinu, L., and Poitras, D. Plasma deposition of optical films and coatings: A review. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 18, 2619–2645 (2000). <https://doi.org/10.1116/1.1314395>
- 2 Ostrikov, K. Colloquium: Reactive plasmas as a versatile nanofabrication tool. *Reviews of Modern Physics*, 77, 489–511 (2005). <https://doi.org/10.1103/RevModPhys.77.489>
- 3 Hundt, M., Sadler, P., Levchenko, I., Wolter, M., Kersten, H., and Ostrikov, K. Real-time monitoring of nucleation-growth cycle of carbon nanoparticles in acetylene plasmas. *Journal of Applied Physics*, 109 (2011). <https://doi.org/10.1063/1.3599893>
- 4 Orazbayev, S., Yerlanuly, Y., Utegenov, A., Moldabekov, Z., Gabdullin, M., and Ramazanov, T. Plasma with carbon nanoparticles: advances and application. *Nanotechnology*, 32, 455602 (2021). <https://doi.org/10.1088/1361-6528/ac1a40>
- 5 Orazbayev, S.A., Utegenov, A.U., Zhunisbekov, A.T., Slamyiya, M., Dosbolayev, M.K., and Ramazanov, T.S. Synthesis of carbon and copper nanoparticles in radio frequency plasma with additional electrostatic field. *Contributions to Plasma Physics*, 58, 961–966 (2018). <https://doi.org/10.1002/ctpp.201700146>
- 6 Yerlanuly, Y. Effect of Nitrogen Concentration on Titanium Nitride Thin Film Formation. *International Journal of Mathematics and Physics*, 14 (2023). <https://doi.org/10.26577/ijmph.2023.v14.i2.06>
- 7 Ceiler, M.F., Kohl, P.A., and Bidstrup, S.A. Plasma-Enhanced Chemical Vapor Deposition of Silicon Dioxide Deposited at Low Temperatures. *Journal of The Electrochemical Society*, 142, 2067–2071 (1995). <https://doi.org/10.1149/1.2044242>
- 8 Zhumadilov, R.Ye., Yerlanuly, Y., Kondo, H., Nemkayeva, R.R., Ramazanov, T.S., Hori, M., and Gabdullin, M.T. Hydrogen peroxide sensing with nitrogen-doped carbon nanowalls. *Sensors and Biosensing Research*, 43, 100614 (2024). <https://doi.org/10.1016/j.sbsr.2023.100614>
- 9 Yerlanuly, Y., Christy, D., Van Nong, N., Kondo, H., Alpysbayeva, B.Ye., Zhumadilov, R., Nemkayeva, R.R., Ramazanov, T.S., Hori, M., and Gabdullin, M.T. Creation of unique shapes by coordination of alumina nanopores and carbon nanowalls. *Fullerenes, Nanotubes and Carbon Nanostructures*, 31, 295–301 (2023). <https://doi.org/10.1080/1536383X.2022.2146672>
- 10 Gabriel, O., Kirner, S., Klick, M., Stannowski, B., and Schlattmann, R. Plasma monitoring and PECVD process control in thin film silicon-based solar cell manufacturing. *EPJ Photovoltaics*, 5, 55202 (2014). <https://doi.org/10.1051/epjpv/2013028>
- 11 Hsieh, Y.-L., Kau, L.-H., Huang, H.-J., Lee, C.-C., Fuh, Y.-K., and Li, T.T. In situ plasma monitoring of PECVD nc-Si:H films and the influence of dilution ratio on structural evolution. *Coatings*, 8, 238 (2018). <https://doi.org/10.3390/coatings8070238>
- 12 Ongaibergenov, Z., Orazbayev, S., Gabdullin, M., Ramazanov, T., Abdrakhmanov, A., Yerlanuly, Y., and Utegenov, A. Diagnostics and characterization of nanoparticles in dusty glow discharge plasma. *Scientific Reports*, 15, 37530 (2025). <https://doi.org/10.1038/s41598-025-22182-0>
- 13 Descoeudres, A., Barraud, L., Bartlome, R., Choong, G., De Wolf, S., Zicarelli, F., and Ballif, C. The silane depletion fraction as an indicator for the amorphous/crystalline silicon interface passivation quality. *Applied Physics Letters*, 97 (2010). <https://doi.org/10.1063/1.3511737>
- 14 Dingemans, G., van de Sanden, M.C.M., and Kessels, W.M.M. Plasma-enhanced Chemical Vapor Deposition of Aluminum Oxide Using Ultrashort Precursor Injection Pulses. *Plasma Processes and Polymers*, 9, 761–771 (2012). <https://doi.org/10.1002/ppap.201100196>
- 15 Kyrykbay, B., Abdirakhmanov, A., Ussenkan, S., Utegenov, A., Yerlanuly, Y., Ramazanov, T., Koshtybayev, T., and Orazbayev, S. Obtaining hydrophobic coatings from AR+HMDSO using radiofrequency discharge at atmospheric pressure. *International Journal of Mathematics and Physics*, 15, 77–82 (2024). <https://doi.org/10.26577/ijmph.2024v15i1a9>

- 16 Ballesteros, J., Fernández Palop, J.I., Hernández, M.A., and Morales Crespo, R. LabVIEW virtual instrument for automatic plasma diagnostic. *Review of Scientific Instruments*, 75, 90–93 (2004). <https://doi.org/10.1063/1.1634356>
- 17 Law, V.J., O’Neil, F.T., and Dowling, D.P. Evaluation of the sensitivity of electro-acoustic measurements for process monitoring and control of an atmospheric pressure plasma jet system. *Plasma Sources Science and Technology*, 20, 035024 (2011). <https://doi.org/10.1088/0963-0252/20/3/035024>
- 18 National Instruments. LabVIEW (2021). URL : <https://www.ni.com/en/shop/labview.html> (accessed: 2021).
- 19 Pawlat, J., Samoń, R., Stryczewska, H.D., Diatczyk, J., and Giżewski, T. RF-powered atmospheric pressure plasma jet for surface treatment. *The European Physical Journal Applied Physics*, 61, 24322 (2013). <https://doi.org/10.1051/epjap/2012120428>
- 20 Giannone, L. et al. Data acquisition and real-time signal processing of plasma diagnostics on ASDEX Upgrade using LabVIEW RT. *Fusion Engineering and Design*, 85, 303–307 (2010). <https://doi.org/10.1016/j.fusengdes.2010.03.030>
- 21 Kersten, H., Thieme, G., Fröhlich, M., Bojic, D., Tung, D.H., Quaas, M., Wulff, H., and Hippler, R. Complex (dusty) plasmas: examples for applications. *Pure and Applied Chemistry*, 77, 415–428 (2005). <https://doi.org/10.1351/pac200577020415>
- 22 Lin, J., Hashimoto, K., Togashi, R., Utegenov, A., Hénault, M., Takahashi, K., Boufendi, L., and Ramazanov, T. Transport control of dust particles by pulse-time modulated RF in dusty plasmas. *Journal of Applied Physics*, 126 (2019). <https://doi.org/10.1063/1.5093349>
- 23 Batryshev, D., Utegenov, A., Zhumadilov, R., Akhanova, N., Orazbayev, S., Ussen Khan, S., Lin, J., Takahashi, K., Bastykova, N., Kodanova, S., Gabdullin, M., and Ramazanov, T. Carbon nanoparticles characteristics synthesized in pulsed radiofrequency discharge and their effect on surface hydrophobicity. *Contributions to Plasma Physics*, 62 (2022). <https://doi.org/10.1002/ctpp.202100238>
- 24 Banerjee, I. et al. In situ optical emission spectroscopic investigations during arc plasma synthesis of iron oxide nanoparticles. *IEEE Transactions on Plasma Science*, 34, 1175–1182 (2006). <https://doi.org/10.1109/TPS.2006.878430>
- 25 Dosbolayev, M.K., Utegenov, A.U., and Ramazanov, T.S. Structural properties of buffer and complex plasmas in RF gas discharge-imposed electrostatic field. *IEEE Transactions on Plasma Science*, 44, 469–472 (2016). <https://doi.org/10.1109/TPS.2015.2497267>
- 26 le Febvrier, A., Landälv, L., Liersch, T., Sandmark, D., Sandström, P., and Eklund, P. An upgraded ultra-high vacuum magnetron-sputtering system. *Vacuum*, 187, 110137 (2021). <https://doi.org/10.1016/j.vacuum.2021.110137>
- 27 Yerlanuly, Y. et al. Achieving stable photodiode characteristics under ionizing radiation. *Carbon*, 215, 118488 (2023). <https://doi.org/10.1016/j.carbon.2023.118488>
- 28 Abdirakhmanov, A. The formation of chondrule-like particles in RF discharge plasma. *Physical Sciences and Technology*, 10 (2023). <https://doi.org/10.26577/phst.2023.v10.i2.08>
- 29 Ahn, J.H. et al. Development of a fully automated desktop chemical vapor deposition system for programmable and controlled carbon nanotube growth. *Micro and Nano Systems Letters*, 7, 11 (2019). <https://doi.org/10.1186/s40486-019-0091-8>
- 30 Niketa, A.K. et al. An automated chemical vapor deposition setup for 2D materials. *HardwareX*, 9, e00165 (2021). <https://doi.org/10.1016/j.ohx.2020.e00165>
- 31 Fernandes, M., Vygranenko, Y., Maçarico, A.F., and Vieira, M. Automated PECVD system for fabrication of a-Si:H devices. *Procedia Technology*, 17, 580–586 (2014). <https://doi.org/10.1016/j.protcy.2014.10.194>
- 32 National Instruments. Set Up Communication with Serial Instruments in LabVIEW using NI-VISA. URL: <https://knowledge.ni.com/KnowledgeArticleDetails?id=kA03q000000x1jtCAA&l> (accessed: 19.12.2023).
- 33 National Instruments. Programming Serial Devices in VISA. URL: <https://www.ni.com/docs/en-US/bundle/ni-visa/page/programming-serial-devices-in-visa.html> (accessed: 19.03.2026).
- 34 Podder, J., Rusop, M., and Soga, T. Boron doped amorphous carbon thin films grown by r.f. PECVD under different partial pressure. *Diamond and Related Materials*, 14, 1799–1804 (2005). <https://doi.org/10.1016/j.diamond.2005.07.020>
- 35 Ahn, J.H. et al. Development of a fully automated desktop chemical vapor deposition system for programmable and controlled carbon nanotube growth. *Micro and Nano Systems Letters*, 7, 11 (2019). <https://doi.org/10.1186/s40486-019-0091-8>

- 36 Kau, L.-H. et al. Correlation of impedance matching and optical emission spectroscopy during PECVD of nanocrystalline silicon thin films. *Coatings*, 9, 305 (2019). <https://doi.org/10.3390/coatings9050305>
- 37 Li, D., Li, J., and Chen, D. Real-time measurement device of RF impedance and power for RF negative ion source. *Fusion Engineering and Design*, 179, 113124 (2022). <https://doi.org/10.1016/j.fusengdes.2022.113124>
- 38 Yu, S. et al. Impedance matching design for capacitively coupled plasmas considering coaxial cables. *Journal of Physics D: Applied Physics*, 57, 475204 (2024). <https://doi.org/10.1088/1361-6463/ad7151>
- 39 Orazbayev, S. et al. The method of synthesizing of superhydrophobic surfaces by PECVD. *Journal of Physics: Conference Series*, 987, 012004 (2018). <https://doi.org/10.1088/1742-6596/987/1/012004>
- 40 Ussenkhan, S.S. et al. Fabricating durable and stable superhydrophobic coatings by atmospheric pressure plasma polymerisation. *Heliyon*, 10, e23844 (2024). <https://doi.org/10.1016/j.heliyon.2023.e23844>
- 41 Baro, M. et al. Pulsed PECVD for low-temperature growth of vertically aligned carbon nanotubes. *Chemical Vapor Deposition*, 20, 161–169 (2014). <https://doi.org/10.1002/cvde.201307093>
- 42 Mohan, A., Poullos, I., Schropp, R.E.I., and Rath, J.K. Size control of gas phase grown silicon nanoparticles by varying plasma OFF time in silane pulsed plasma. *MRS Proceedings*, 1803, mrss15-2133139 (2015). <https://doi.org/10.1557/opl.2015.476>
- 43 Lundin, D., Jensen, J., and Pedersen, H. Influence of pulse power amplitude on plasma properties and film deposition in high power pulsed PECVD. *Journal of Vacuum Science & Technology A*, 32 (2014). <https://doi.org/10.1116/1.4867442>
- 44 Park, H.K., Song, W.S., and Hong, S.J. In situ plasma impedance monitoring of the oxide layer PECVD process. *Coatings*, 13, 559 (2023). <https://doi.org/10.3390/coatings13030559>
- 45 Boufendi, L., Jouanny, M.C., Kovacevic, E., Berndt, J., and Mikikian, M. Dusty plasma for nanotechnology. *Journal of Physics D: Applied Physics*, 44, 174035 (2011). <https://doi.org/10.1088/0022-3727/44/17/174035>
- 46 Ouaras, K., Lombardi, G., and Hassouni, K. Nanoparticles synthesis in microwave plasmas: peculiarities and comprehensive insight. *Scientific Reports*, 14, 4653 (2024). <https://doi.org/10.1038/s41598-023-49818-3>
- 47 Ussenov, Y.A. et al. Langmuir probe measurements in nanodust containing argon-acetylene plasmas. *Vacuum*, 166, 15–25 (2019). <https://doi.org/10.1016/j.vacuum.2019.04.051>

^{1,2}**Утегенов А.У.,**

PhD, ORCID ID: 0000-0001-6768-0918,

*e-mail: almasbek@physics.kz

^{1,2}**Үсенхан С.С.,**

ORCID ID: 0009-0009-6597-2066,

e-mail: sultanusenhan@gmail.com

³**Үсенов Е.А.,**

ORCID ID: 0000-0001-7903-8674,

e-mail: yerbolat@physics.kz

^{1,2}**Оразбаев С.А.,**

PhD, ORCID ID: 0000-0002-7286-9990,

e-mail: sagi.orazbayev@gmail.com

¹Қолданбалы ғылымдар және ақпараттық технологиялар институты, Алматы қ., Қазақстан

²Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

³Принстон плазма физикасы зертханасы, Нью-Джерси, Принстон, АҚШ

ПЛАЗМАЛЫҚ ОРТАДАҒЫ НАНОБӨЛШЕКТЕР СИНТЕЗІН LABVIEW КӨМЕГІМЕН АВТОМАТТЫҚ БАСҚАРУ

Аңдатпа

Берілген жұмыста LabVIEW бағдарламалық ортасы негізінде іске асырылған газ фазасынан плазмалық-химиялық тұндыру әдісімен наноматериалдарды синтездеу процесін автоматты бақылау жүйесі ұсынылған. Зерттеудің негізгі мақсаты – PECVD процесінің негізгі кезеңдерін дәйекті басқаруды қамтамасыз ететін

бірыңғай бағдарламалық және аппараттық платформаны әзірлеу: вакуумдық камераны дайындау, қысымды бақылау, жұмыс газын беру, ЖЖ плазмасын жағу, ЖЖ қуатын үйлестіру, наноматериалдардың циклдік өсуі және плазмалық ортадағы нанобөлшектерді оптикалық бақылау. LabVIEW пайдалану атқарушы құрылғыларды басқаруды, эксперименттік параметрлерді тіркеуді және процесті нақты уақыт режимінде визуализациялауды біріктіруге мүмкіндік берді. Автоматтандырылған цикл жұмыс камерасын белгіленген қысымға келтіруден басталады. Келесі кезең эксперименттің қайталанатын бастапқы шарттарын қамтамасыз ететін белгіленген шекті мәнге жеткеннен кейін ғана жүзеге асырылады. Әрі қарай, бағдарлама газ шығынын басқару құрылғысы арқылы жұмыс газын беруді басқарады. Жұмыста газ ағынын басқарудың екі режимі қарастырылған: 0–5 В кернеумен аналогтық режим және RS-232 арқылы цифрлық байланыс режимі. Аналогтық әдіс басқару кернеуін дәл масштабтауды қажет ететіні көрсетілген, өйткені 5 В кернеу беру контроллердің максималды ағынға толық ашылуына әкеледі. Сонымен қатар, RS-232 интерфейсі шығынды sscm бірліктерінде тікелей орнатуға мүмкіндік береді, бұл газ ортасын басқарудың дәлдігі мен ыңғайлылығын арттырады. Қысым тұрақтандырылғаннан кейін LabVIEW ЖЖ разрядын жағып, шағылған қуатты азайтуға және плазмалық процестің тұрақтылығын арттыруға бағытталған RF matching алгоритмін орындайды. Бағдарламаның жеке модулінде наноматериалдардың циклдік өсу режимі жүзеге асырылған, яғни плазманың жану уақыты, кідіріс уақыты және циклдер саны теңшеледі. Бұл тәсіл наноматериалдардың жиналуын бақылауға және процесс параметрлерін алынған нанокұрылымдардың морфологиялық сипаттамаларымен байланыстыруға мүмкіндік береді. Жүйенің соңғы модулі тозанды плазмадағы нанобөлшектер бұлтының тығыздығын оптикалық бақылауға арналған. Ол үшін RS-232 арқылы LabVIEW-пен байланысқан Keithley 2401 фотодетекторы мен өлшеу блогы арқылы плазма аймағынан өткен лазерлік сәулелену қарқындылығының өзгеруі тіркеледі. Бастапқы және өзгертілген сигнал қарқындылықтарының айырмашылығы нанобөлшектердің пайда болуы мен эволюциясының диагностикалық параметрі ретінде қолданылады. Әзірленген жүйе LabVIEW бағдарламасының жеке құрылғыларды автоматтандыру үшін ғана емес, сонымен қатар наноматериалдардың PECVD синтезін басқарудың толық цифрлық циклін құру үшін де тиімді қолданылатынын көрсетеді.

Түйін сөздер: PECVD, автоматты плазмалық синтез, нанобөлшектер өсуін бақылау, тозанды плазма диагностикасы, плазмалық нанотехнология.

^{1,2}**Утегенов А.У.,**

PhD, ORCID ID: 0000-0001-6768-0918,

*e-mail: almasbek@physics.kz

^{1,2}**Үсенхан С.С.,**

ORCID ID: 0009-0009-6597-2066,

e-mail: sultanusenhan@gmail.com

³**Үсенов Е.А.,**

ORCID ID: 0000-0001-7903-8674,

e-mail: yerbolat@physics.kz

^{1,2}**Оразбаев С.А.,**

PhD, ORCID ID: 0000-0002-7286-9990,

e-mail: sagi.oralbayev@gmail.com

¹Институт прикладных наук и информационных технологий, г. Алматы, Казахстан

²Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

³Принстонская лаборатория физики плазмы, Нью-Джерси, Принстон, США

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ СИНТЕЗОМ НАНОЧАСТИЦ В ПЛАЗМЕННОЙ СРЕДЕ С ИСПОЛЬЗОВАНИЕМ LABVIEW

Аннотация

В работе представлена система автоматического контроля процесса синтеза наноматериалов методом плазменно-усиленного химического осаждения из газовой фазы, реализованная на основе программной среды LabVIEW. Основная цель исследования заключается в разработке единой программно-аппаратной платформы, обеспечивающей последовательное управление ключевыми стадиями PECVD-процесса: подготовкой вакуумной камеры, контролем давления, подачей рабочего газа, зажиганием ВЧ-плазмы, согла-

сованием ВЧ-мощности, циклическим ростом наноматериалов и оптическим мониторингом наночастиц в плазменной среде. Использование LabVIEW позволило объединить управление исполнительными устройствами, регистрацию экспериментальных параметров и визуализацию процесса в режиме реального времени. Автоматизированный цикл начинается с откачки реакционной камеры до заданного базового давления. Переход к следующей стадии осуществляется только после достижения установленного порогового значения, что обеспечивает воспроизводимые начальные условия эксперимента. Далее программа управляет подачей рабочего газа через газовый расходомер. В работе рассмотрены два режима управления газовым потоком: аналоговый режим с подачей напряжения 0–5 В и цифровой режим связи через RS-232. Показано, что аналоговый способ требует точного масштабирования управляющего напряжения, поскольку подача 5 В приводит к полному открытию контроллера на максимальный расход. В то же время RS-232-интерфейс позволяет задавать расход непосредственно в единицах sccm, что повышает точность и удобство управления газовой средой. После стабилизации давления LabVIEW инициирует запуск ВЧ-разряда и выполняет алгоритм RF matching, направленный на минимизацию отраженной мощности и повышение стабильности плазменного процесса. Отдельный модуль программы реализует циклический режим роста наноматериалов, где задаются время горения плазмы, время паузы и количество циклов синтеза. Такой подход позволяет контролировать накопление материала на подложке и связывать параметры процесса с морфологическими характеристиками получаемых наноструктур. Завершающий модуль системы предназначен для оптического мониторинга плотности облака наночастиц в пылевой плазме. Для этого используется регистрация изменения интенсивности лазерного излучения, прошедшего через область плазмы, с помощью фотодетектора и измерительного блока Keithley 2401, связанного с LabVIEW через RS-232. Разница между исходной и измененной интенсивностью сигнала используется как диагностический параметр формирования и эволюции наночастиц. Разработанная система демонстрирует, что LabVIEW может эффективно применяться не только для автоматизации отдельных приборов, но и для построения полного цифрового цикла управления PECVD-синтезом наноматериалов.

Ключевые слова: PECVD, автоматизированный плазменный синтез, контроль роста наночастиц, диагностика пылевой плазмы, плазменная нанотехнология.

UDC 621.355
IRSTI 29.19.31

<https://doi.org/10.55452/1998-6688-2026-23-2-381-391>

¹**Kanatov Zh.S.,**

PhD student, ORCID ID: 0009-0003-2975-0999,
kanatovzhantilek@gmail.com

^{1*}**Kalkozova Zh.K.,**

Cand.Phys.-Math.Sc., Professor, ORCID ID: 0000-0002-4826-1678,

*e-mail: zhanar.kalkozova@kaznu.edu.kz

²**Mukash Zh.O.,**

PhD, Assistant Professor, ORCID ID: 0009-0004-7713-3925,

e-mail: zhanar.mukash@sdu.edu.kz

³**Mirzaeian M.,**

PhD, ORCID ID: 0000-0002-1474-8330,

e-mail: mojtaba.mirzaeian@uws.ac.uk

¹**Abdullin Kh.A.**

Dr.Phys.-Math.Sc., Professor, ORCID ID: 0000-0002-2729-2272,

e-mail: kh.abdullin@physics.kz

¹National Nanotechnology Laboratory of Open Type of Al–Farabi Kazakh National University,
Almaty, Kazakhstan

²School of Education and Humanities, SDU University, Kaskelen, Kazakhstan

³School of Computing, Engineering and Physical Sciences, University
of the West of Scotland, Paisley, United Kingdom

SURFACE TRANSFORMATION AND ENHANCED ELECTROCHEMICAL PERFORMANCE OF Ni₃S₂/NI FOAM ELECTRODES FOR HYBRID SUPERCAPACITORS

Abstract

The process of sulfurization of nickel foam surfaces to obtain Ni₃S₂ layers with high electrochemical capacitance and stability during electrochemical cycling has been extensively studied. However, the role of nickel hydroxide layers, which are expected to form under the electrochemical operating conditions of the Ni₃S₂/NF electrode, has not been sufficiently investigated. In the present work, it is demonstrated that the hydroxide phase makes a significant contribution to both electrochemical capacitance and cyclic stability. The Ni₃S₂/NF electrode was fabricated via a single-step hydrothermal method in the presence of thiourea at 160 °C. The initial structure of Ni₃S₂ on the NF surface was subsequently modified through electrochemical cycling in a KOH electrolyte. The increase in electrochemical capacitance of the electrode was accompanied by the formation of multiple nickel hydroxide phases, as identified by X-ray diffraction (XRD) and Raman spectroscopy. The electrode exhibited high performance stability over 20,000 galvanostatic charge–discharge (GCD) cycles at a current density of 20 A g^{−1}, retaining 90% of its maximum capacitance. The specific capacitance of the Ni₃S₂ electrode was 758 F g^{−1} at a current density of 2.7 A g^{−1}. When the current density increased to 90 A g^{−1}, the specific capacitance decreased to 233 F g^{−1}, corresponding to 30% of the capacitance at 2.7 A g^{−1}.

Keywords: Ni₃S₂, nickel foam, hybrid supercapacitor, nickel hydroxide, electrochemical activation, cycling stability.

Received April 7, 2026; accepted May 27, 2026.

Introduction

Renewable energy relies on inherently unstable energy sources. The development of reliable energy systems therefore requires advanced energy storage technologies. Supercapacitors, characterized by high energy efficiency, represent an important component of such systems, providing smoothing of power consumption fluctuations and thereby enhancing the stability of energy systems [1–3]. The main types of electrochemical supercapacitors used for energy storage include electrical double-layer capacitors (EDLCs), pseudocapacitors, and hybrid supercapacitors. Hybrid electrochemical supercapacitors combine the advantages of high specific power, enhanced specific energy, and long service life, integrating the properties of electrical double-layer capacitors and pseudocapacitors to achieve efficient energy storage.

The most promising materials for hybrid electrochemical supercapacitors include carbon nanomaterials (graphene, carbon nanotubes), transition metal oxides (MnO_2 , NiO), conductive polymers (polyaniline, polypyrrole), and their composites [4–7]. Metal sulfides, such as Ni_3S_2 and NiS , are also considered promising materials for hybrid electrochemical supercapacitors due to their high specific capacitance, good electrical conductivity, and structural stability associated with their layered structure and the ability to undergo reversible redox reactions.

Compared with metal oxides, metal sulfides exhibit several advantages in hybrid electrochemical supercapacitors, including higher electrical conductivity, enhanced specific capacitance due to rich redox activity and layered structure, as well as improved mechanical stability during cycling, which contributes to increased durability and efficiency.

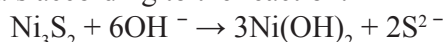
Many metal oxides are dielectric in nature. In contrast, sulfides generally exhibit higher electrical conductivity than oxides, which facilitates electron transfer, reduces internal resistance, and increases specific power [8]. Sulfides also demonstrate enhanced redox activity due to the presence of a greater number of redox-active centers and weaker metal–sulfur (M–S) bonds. As a result, they enable more extensive Faradaic redox reactions, associated with the multiple valence states of metals and the lower electronegativity of sulfur compared to oxygen, leading to higher specific capacitance.

The structural flexibility of sulfides contributes to improved mechanical stability during charge–discharge cycling, resulting in enhanced cyclic stability and improved energy storage performance. Furthermore, sulfides combined with oxides or carbon-based materials often outperform electrodes based solely on oxides, owing to synergistic effects that enhance ion transport and increase the active surface area.

Nickel sulfides (e.g., NiS , Ni_3S_2) are particularly promising for hybrid electrochemical supercapacitors due to their superior electrical conductivity compared to nickel oxides, which provides faster electron transfer and higher power density. In addition, their stable crystal structures, such as the layered morphology of NiS , provide excellent cyclic stability, maintaining performance over thousands of charge–discharge cycles. These characteristics, together with their cost-effectiveness and relative ease of synthesis, make nickel sulfides attractive electrode materials for efficient and durable hybrid supercapacitors.

Among nickel sulfides, Ni_3S_2 [9] is especially promising due to its high specific capacitance, good electrical conductivity, and structural stability. Ni_3S_2 exhibits a balanced combination of redox properties, while its developed morphology provides a high active surface area and facilitates ion diffusion, making it suitable for the fabrication of high-performance supercapacitors.

However, under electrochemical conditions in an alkaline environment, Ni_3S_2 can undergo oxidation at high positive potentials according to the reaction:



resulting in the formation of nickel hydroxide phases. Since these phases exhibit high electrochemical activity and possess high specific capacitance, their formation must be taken into account. However, to the best of our knowledge, this aspect of $\text{Ni}_3\text{S}_2/\text{NF}$ electrode modification has been investigated in only two studies [10,11]. In [10], the Ni_3S_2 layers were obtained by annealing, not by the hydrothermal method. In Ref. [11], the $\text{Ni}_3\text{S}_2/\text{NF}$ electrode was synthesized via a rapid hydrothermal

method in the presence of a nickel salt in the growth solution, which reduced the synthesis time to approximately 3 h.

In the present study, phase transformations in $\text{Ni}_3\text{S}_2/\text{NF}$ electrodes obtained using the conventional method [9, 12] of nickel foam sulfidation in thiourea are investigated.

Material and methods

Thiourea ($\text{CH}_4\text{N}_2\text{S}$) and nickel foam (NF) with a thickness of 1.5 mm (Shenzhen Tianchenghe Technology Co., Ltd.) and an areal density of 0.03 g cm^{-2} were used as starting materials. Milli-Q water ($18.2 \text{ M}\Omega \cdot \text{cm}$) was obtained using an AQUAMAX–Ultra 370 series water purification system (YL Instrument Co., Anyang, Korea).

Prior to hydrothermal synthesis, the NF substrates were purified by ultrasonic treatment in a 1 M HCl solution. The substrate was then placed in a Teflon beaker, and 30 mL of a 1 mM thiourea solution was added. The hydrothermal reactor with a Teflon liner was sealed and placed in an oven at 160°C for 24 h.

After synthesis and cooling to room temperature, the autoclave was removed from the furnace. The solution remained transparent after synthesis, and the resulting layers were dense and mechanically stable. The NF sample was rinsed with water, subjected to ultrasonic treatment for 30 min, and subsequently dried.

The surface morphology of the samples was examined using a Quanta 200i scanning electron microscope (FEI). X-ray diffraction (XRD) measurements were carried out using a MiniFlex diffractometer (Rigaku) with $\text{Cu K}\alpha$ radiation ($\lambda = 1.5418 \text{ \AA}$). Raman spectra were recorded using an NTEGRA spectrometer (NT–MDT, Zelenograd, Russia) with an excitation wavelength of 473 nm.

The electrochemical performance of the $\text{Ni}_3\text{S}_2/\text{NF}$ electrodes was investigated in a three-electrode configuration using a large-area carbon counter electrode and an Ag/AgCl reference electrode. Electrochemical measurements were performed using a CorrTest CS2350 bipotentiostat (Wuhan Corrtest Instruments Corp., Ltd., Wuhan, China), including cyclic voltammetry (CV), galvanostatic charge–discharge (GCD), and electrochemical impedance spectroscopy (EIS). The capacitance was calculated from CV and GCD measurements using Eqs. (1) and (2), respectively:

$$C = \frac{1}{2v(U_{\max} - U_{\min})} \oint I(V) dV, \quad (1)$$

$$C = \frac{I\Delta t}{\Delta U}, \quad (2)$$

where v (V s^{-1}) is the scan rate, and $U_{\max} - U_{\min}$ (V) is the potential window in the CV measurements. The current $I(V)$ is integrated over the full CV cycle. For GCD measurements, the capacitance was determined from the discharge current I (A), discharge time Δt (s), and voltage drop ΔU (V). To determine the specific capacitance, the mass of the NF samples was measured before and after synthesis. The areal mass of the initial nickel foam was approximately 0.0314 g cm^{-2} . Under the selected synthesis conditions, the mass increase for a sample area of 1 cm^2 was 1.2 mg cm^{-2} . Since the Ni_3S_2 layer is formed as a result of nickel sulfidation, the mass of the synthesized Ni_3S_2 on the NF substrate was estimated to be 4.5 mg cm^{-2} .

Results

Scanning electron microscopy (SEM) micrographs reveal the evolution of the $\text{Ni}_3\text{S}_2/\text{NF}$ electrode surface before and after cyclic voltammetry (CV). The initial morphology is characterized by thin, interwoven nanosheets forming an open porous structure, which provides a high specific surface area and facilitates electrolyte access to active sites.

After intensive electrochemical cycling in an alkaline medium within the potential range from -0.5 to 1 V , noticeable morphological changes are observed. The structural features become less distinct, appearing blurred, which may be attributed to the formation of new phases with dimensions

below the resolution limit of the microscope. During electrochemical oxidation and hydration, a nanocrystalline layer of nickel oxides/hydroxides forms on the surface of the sulfide, as confirmed by X-ray diffraction (XRD) data.

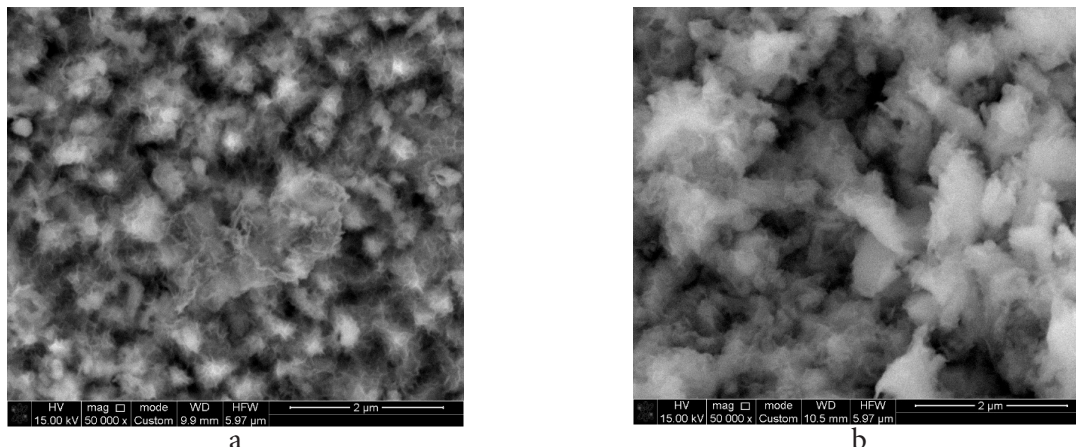


Figure 1 – SEM images of Ni_3S_2 layers on nickel foam grown by the hydrothermal method: a – after synthesis, b – after CV measurements in 3M KOH in the range of -0.5 V to $+1$ V vs Ag/AgCl

Figure 2 presents XRD patterns of the sample after synthesis (curve 1), after approximately 100 CV cycles and stabilization of the CV curves in a 3 M KOH electrolyte (curve 2), and after 20,000 GCD cycles (curve 3). The initial sample exhibits well-defined diffraction peaks corresponding to the crystalline Ni_3S_2 phase (JCPDS No. 01–073–0698), confirming the formation of nickel sulfide on the substrate. Intense peaks at 44.5° , 51.8° , and 76.3° are attributed to metallic nickel (Ni) from the NF substrate. After CV cycling in the potential range of -0.5 to $+1$ V (curve 2), the intensity of the Ni_3S_2 peaks decreases, and new diffraction maxima characteristic of hydrated nickel hydroxide appear. This indicates effective surface oxidation of the sulfide during redox processes over a wide potential window. After 20,000 GCD cycles (curve 3), the contribution of Ni_3S_2 decreases further, indicating an increase in the thickness of the nickel hydroxide layer. Additionally, significant broadening of the nickel hydroxide peaks is observed, suggesting a reduction in the crystallite size as a result of repeated Faradaic reactions.

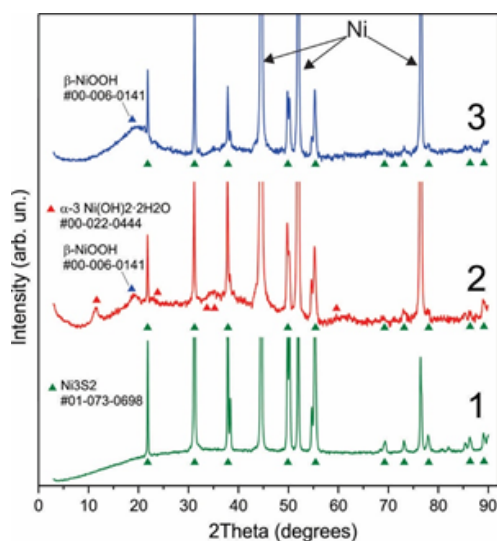


Figure 2 – XRD patterns of the electrode: 1 – initial (pristine) electrode, 2 – after cyclic voltammetry (CV) measurements in 3 M KOH in the potential range from -0.5 V to $+1$ V vs. Ag/AgCl, and 3 – after 20,000 galvanostatic charge–discharge (GCD) cycles in 3 M KOH in the potential range from 0 V to 0.5 V vs. Ag/AgCl

Raman spectroscopy was employed to identify the composition of the electrode surface before and after electrochemical measurements, which is particularly important for detecting thin layers of amorphous and nanocrystalline phases that are not observable by XRD. Figure 3 (curve 1) shows the Raman spectrum of the as-synthesized sample. The spectrum is characterized by a set of peaks in the region corresponding to the characteristic vibrational modes of the Ni_3S_2 crystal lattice. The main peaks are located at 145, 200, 245, 305, and 348 cm^{-1} , which is in good agreement with literature data for the Ni_3S_2 phase [13]. After short-term CV measurements (40–50 cycles) in 3 M KOH within the potential range from -0.5 to $+1$ V vs Ag/AgCl, the Raman spectrum undergoes significant changes (curve 2). The peaks corresponding to nickel sulfide almost completely disappear, while two broad bands centered at approximately 480 cm^{-1} and 560 cm^{-1} become dominant. These bands are attributed to vibrational modes associated with nickel oxyhydroxide (NiOOH) [14–17].

The obtained results confirm that, during deep anodic polarization (up to 1 V vs Ag/AgCl), the Ni_3S_2 surface undergoes irreversible electrochemical transformation. The formation of $\text{Ni}(\text{OH})_2/\text{NiOOH}$ phases indicates the development of a reaction layer on the sulfide surface, which provides Faradaic charge storage processes in an alkaline electrolyte.

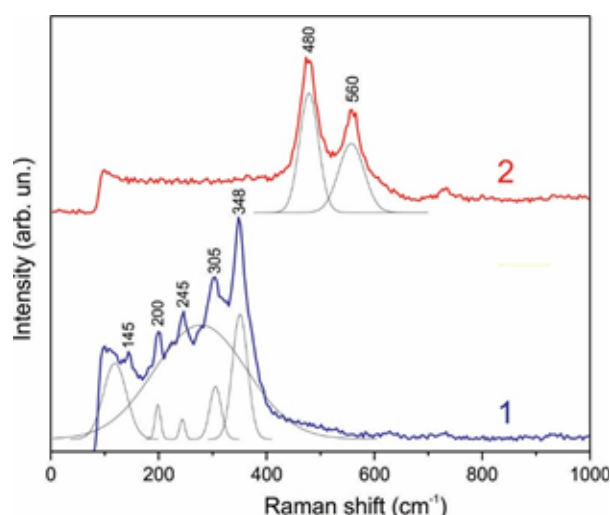


Figure 3 – Raman spectra of Ni_3S_2 layers on nickel foam: curve 1 – after synthesis, and curve 2 – after cyclic voltammetry (CV) measurements in 3 M KOH in the potential range from -0.5 V to $+1$ V vs. Ag/AgCl

Prior to the investigation of the long-term stability of the $\text{Ni}_3\text{S}_2/\text{NF}$ electrode, a preliminary activation procedure was carried out. It should be noted that the initial capacitance of the electrode immediately after synthesis was low. During the initial CV cycles, the current response gradually increased, and after approximately 50–80 cycles, the CV curves reached a stable state. This behavior may be attributed to the gradual penetration of the electrolyte into the porous structure, the formation of electrochemically active sites, and partial surface oxidation.

The long-term stability of the $\text{Ni}_3\text{S}_2/\text{NF}$ electrode was evaluated by galvanostatic charge–discharge (GCD) cycling over 20,000 cycles in a 3 M KOH electrolyte within the potential range of 0 – 0.5 V vs Ag/AgCl. Figure 4a shows the dependence of capacitance and Coulombic efficiency on the number of cycles, while Figure 4b presents a comparison of GCD charge–discharge curves at a current of 100 mA at the initial stage and after 20,000 cycles.

The cycling data reveal a characteristic process of continued, but slower, electrode activation during the first 2000 cycles, during which the areal capacitance increases from approximately 3.5 to 5.2 F cm^{-2} .

Considering that Raman spectroscopy results indicate rapid modification of the Ni_3S_2 surface layer, this increase in capacitance can be attributed to the formation of an electrochemically active

Ni(OH)₂/NiOOH layer. After approximately 2000 cycles, the electrode exhibits stabilization of its electrochemical properties and demonstrates exceptional cyclic stability. Even after 20,000 cycles, no decrease in capacitance or increase in series resistance is observed. The Coulombic efficiency remains stable at a value close to 100%, confirming the high reversibility of the Faradaic reactions.

The GCD curves recorded at the initial stage and after 20,000 cycles exhibit nearly identical shapes with pronounced voltage plateaus (Figure 4b), which is characteristic of battery-type behavior. A significant increase in charge and discharge times for the 20,000th cycle compared to the initial cycle confirms the increase in capacitance resulting from surface activation. At the same time, the preservation of curve symmetry indicates that the internal resistance (IR drop) of the electrode does not increase significantly over the entire cycling period.

Figures 5 and 6 present the GCD and CV characteristics of the electrode after 20,000 cycles, while Figure 7 shows the electrochemical impedance spectroscopy (EIS) results. At a low current density of 0.22 A g⁻¹, the capacitance of the electrode with an area of 1 cm² reaches 3.8 F, which, considering the mass loading of 4.5 mg cm⁻², corresponds to a specific capacitance of 853 F g⁻¹.

At an operating current density of 2.7 A g⁻¹, the specific capacitance is 758 F g⁻¹, and when the current density increases to 90 A g⁻¹, the specific capacitance decreases to 233 F g⁻¹, corresponding to 30% of the capacitance at 2.7 A g⁻¹. This indicates that the electrode retains a high specific capacitance even at high current densities.

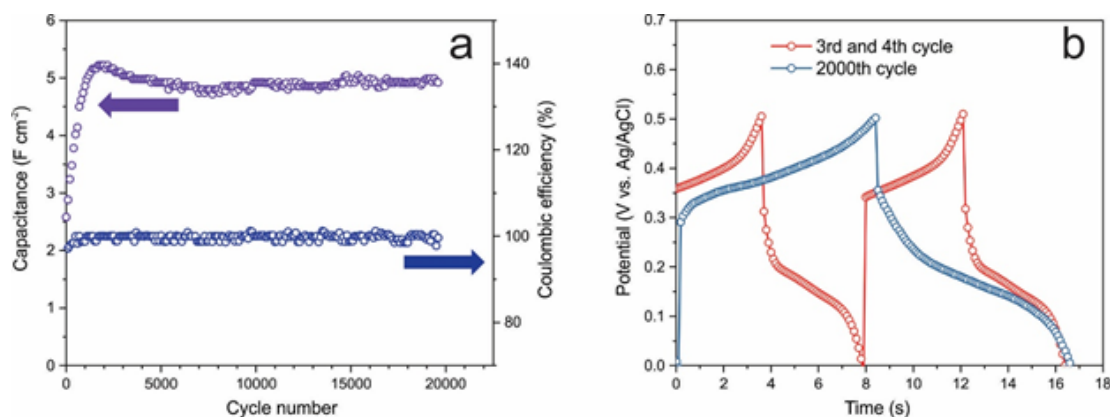


Figure 4 – (a) Capacitance and Coulombic efficiency of the Ni₃S₂/NF electrode as a function of cycle number; (b) comparison of galvanostatic charge–discharge (GCD) curves at a current of 100 mA (22 A g⁻¹) at the beginning of cycling and after 20,000 cycles in 3 M KOH in the potential range of 0 V to 0.5 V vs. Ag/AgCl

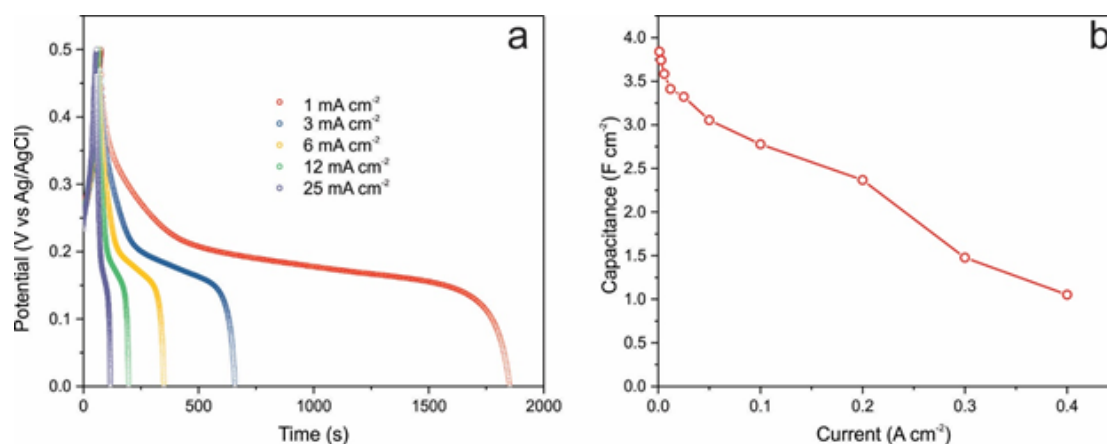


Figure 5 – (a) Galvanostatic charge–discharge (GCD) curves at different discharge current densities (charge current density fixed at 25 mA cm⁻² for all curves); (b) specific capacitance as a function of discharge current density.

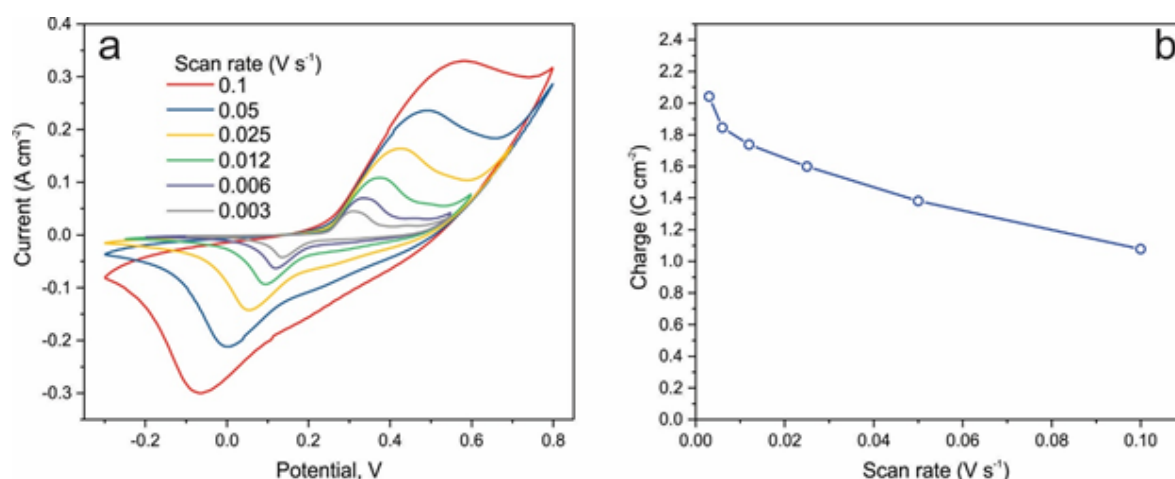


Figure 6 – (a) Cyclic voltammetry (CV) curves at different scan rates;
(b) specific capacitance as a function of scan rate.

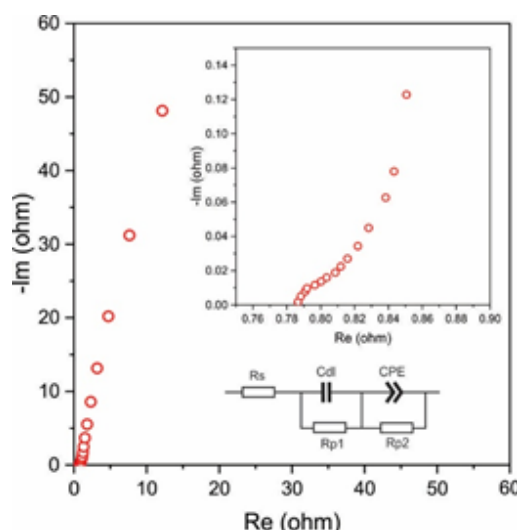


Figure 7 – Nyquist plot of the Ni₃S₂/NF electrode after 20,000 cycles.
Electrochemical impedance spectroscopy (EIS) measurements were performed
in the frequency range from 50 kHz to 0.0016 Hz at a DC offset of 300 mV vs. Ag/AgCl.
The inset shows the equivalent circuit model of the electrode.

Discussion

The results of the study demonstrate that the high electrochemical performance and stability of the Ni₃S₂/NF electrode are due to the phase and structural transformations of its surface during operation.

The phase transformation of the surface is confirmed by Raman spectroscopy. The spectra (Figure 3) show the disappearance of the characteristic vibrational modes of the Ni₃S₂ crystal lattice after electrochemical activation. Instead, bands at approximately 480 cm⁻¹ and 560 cm⁻¹, corresponding to stable nickel oxyhydroxide (NiOOH) phases, become dominant, indicating the formation of an electrochemically active reaction layer.

The structural transformation is supported by SEM and XRD data. Even relatively short-term CV cycling leads to a reduction in the contribution of the Ni₃S₂ phase and the appearance of new diffraction maxima characteristic of hydrated nickel hydroxide.

Prolonged cycling results in the formation of β -Ni(OH)₂ nanoparticles. Significant broadening of the diffraction peaks after 20,000 cycles indicates a decrease in crystallite size, which increases the active surface area and promotes Faradaic reactions.

As a consequence of these phase and structural transformations, the electrode demonstrates exceptional durability, retaining more than 90% of its maximum capacitance after 20,000 GCD cycles. The activation process during the first 2000 cycles, accompanied by an increase in capacitance from 3.5 to 5.2 F cm⁻², is directly associated with the formation of a developed Ni(OH)₂/NiOOH layer. The stability of the Coulombic efficiency at approximately 100% further confirms the high reversibility of the redox processes in the formed nanostructure.

Thus, the high capacitance and stability of the electrode are ensured by synergistic interaction between the conductive Ni₃S₂/NF framework and the nanocrystalline nickel hydroxide/oxyhydroxide layer formed under electrochemical conditions, which remains electrochemically active and mechanically stable over tens of thousands of cycles.

Conclusion

In this work, an electrode based on Ni₃S₂ nickel sulfide on nickel foam was synthesized using a simple one-step hydrothermal method. This approach is well established in the literature and has proven to be effective for the fabrication of electrodes for hybrid supercapacitors. However, despite numerous studies on the properties of Ni₃S₂ supercapacitor electrodes grown on NF via thiourea-assisted sulfidation, the role of nickel hydroxide layers formed under electrochemical conditions on the surface of Ni₃S₂ has not been sufficiently addressed.

For comparison, in Ref. [10], Ni₃S₂ electrodes were prepared by annealing an NF substrate in an H₂S-containing atmosphere (5% H₂S–95% N₂) at 400 °C for 60 min. In Ref. [11], nickel acetate was introduced into the growth solution to accelerate synthesis, reducing the processing time to approximately 3 h. However, the results obtained in the present study differ significantly from those reported in Ref. [11], particularly in terms of the superior stability of electrochemical performance during cycling.

It has been demonstrated that robust Ni₃S₂ layers obtained by sulfidation of NF in thiourea undergo surface modification during electrochemical activation in an alkaline medium, resulting in the formation of an active layer composed of nanocrystalline Ni(OH)₂ phases and stable NiOOH species. This composite structure provides high specific capacitance and excellent cycling stability.

The obtained results highlight the critical role of the synergistic interaction between the conductive sulfide framework and the nanocrystalline hydroxide/oxyhydroxide layer. This synergy makes the Ni(OH)₂/Ni₃S₂/NF system a promising electrode material for long-life hybrid supercapacitors.

Acknowledgements

The study was carried out with the financial support of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No AP23488976).

REFERENCES

- 1 Lakshmi, K.C.S., Vedhanarayanan, B. High-Performance Supercapacitors: A Comprehensive Review on Paradigm Shift of Conventional Energy Storage Devices. *Batteries*, 9(4), 202 (2023). <https://doi.org/10.3390/batteries9040202>.
- 2 Nargish Parvin, Dhananjaya Merum, Misook Kang, Sang Woo Joo, Jae Hak Jung and Tapas Kumar Mandal. Recent advances in hybrid supercapacitors: a review of high performance materials and scalable fabrication techniques. *Journal of Materials Chemistry A*, 13(30), 24320–24386 (2025). <https://doi.org/10.1039/D5TA02887F>.

3 Neeraj Singh, Virendra Singh, Neeraj Bisht, Puneet Negi, Archana Dhyani, Rajat Kumar Sharma, Tewari B.S. A comprehensive review on supercapacitors: Basics to recent advancements. *Journal of Energy Storage*. 121, 116498 (2025). <https://doi.org/10.1016/j.est.2025.116498>.

4 Reenu, Sonia, Lakshita Phor, Ashok Kumar, Surjeet Chahal, Electrode materials for supercapacitors: A comprehensive review of advancements and performance. *Journal of Energy Storage*. 84, Part B, 110698 (2024). <https://doi.org/10.1016/j.est.2024.110698>.

5 Mojtaba Mirzaeian, Qaisar Abbas, Abraham Ogwu, Peter Hall, Mark Goldin, Marjan Mirzaeian, Hassan Fathinejad Jirandehi. Electrode and electrolyte materials for electrochemical capacitors. *International Journal of Hydrogen Energy*. 42, 25565–25587 (2017). <https://doi.org/10.1016/j.ijhydene.2017.04.241>.

6 Muhammad Faisal Iqbal, Farooq Nasir, Fiza Shabbir, Zaheer Ud Din Babar, Muhammad Farooq Saleem, Kaleem Ullah, Nana Sun, Faizan Ali. Supercapacitors: An Emerging Energy Storage System. *Advanced Energy and Sustainability Research*. 6, 2400412 (2025). <https://doi.org/10.1002/aesr.202400412>.

7 Ander González, Eider Goikolea, Jon Andoni Barrena, Roman Mysyk. Review on supercapacitors: Technologies and materials. *Renewable and Sustainable Energy Reviews*. 58, 1189–1206 (2016). <https://doi.org/10.1016/j.rser.2015.12.249>.

8 Mengkang Zhu, Dan Wang, Zongyu Ge, Lin Pan, Yanli Chen, Wenchang Wang, Naotoshi Mitsuzaki, Shuyong Jiad Zhidong Chen. Recent advances in transition metal sulfide-based electrode materials for supercapacitors. *Chemical Communications*. 61, 8314–8326 (2025). <https://doi.org/10.1039/D5CC01411E>.

9 Dhakal, G., Sahoo, S., Sharma, K.P., Zhao, G.-L. A Review on the Recent Advancements of Ni-Based Sulfides and Mixed Sulfides for Supercapacitors and Electrocatalysis (Oxygen Evolution Reaction). *Molecules*. 30, 2877 (2025). <https://doi.org/10.3390/molecules30132877>.

10 Yan-Qiang Cao, Xu Qian, Wei Zhang, Min Li, Shan-Shan Wang, Di Wu, Ai-Dong Li. Self-formed porous Ni(OH)₂ on NiS₂/Ni foam during electrochemical cycling for high performance supercapacitor with ultrahigh areal capacitance. *Electrochimica Acta*. 303, 148–156 (2019). <https://doi.org/10.1016/j.electacta.2019.02.075>.

11 Abdullin K.A., Gabdullin M.T., Gritsenko L.V., Kalkozova Zh.K., Kanatov Zh.S., Markhabayeva A.A., Nemkayeva R.R., Zhapargali D., Mirzaeian M. In situ formation of nanocrystalline Ni(OH)₂ in alkaline electrolyte explains superior capacitance and cycling stability of Ni₃S₂/NF electrodes. *Scientific Reports* (2026). <https://doi.org/10.1038/s41598-026-42576-y>

12 Xuerui Yi, Caroline Kirk, and Neil Robertson. Achieving Complete Conversion from Nickel Foam to Nickel Sulfide Foam for a Freestanding Hybrid-Supercapacitor Electrode. *ChemElectroChem*. 11, e202400383, (2024). <https://doi.org/10.1002/celec.202400383>.

13 Jeng-Han Wang, Zhe Cheng, Jean-Luc Brédas, Meilin Liu. Electronic and vibrational properties of nickel sulfides from first principles. *Journal of Chemical Physics*. 127, 214705 (2007). <https://doi.org/10.1063/1.2801985>.

14 Julia Gallenberger, Harol Moreno Fernández, Achim Alkemper, Mohan Li, Chuanmu Tian, Bernhard Kaisera, Jan Philipp Hofmann. Stability and decomposition pathways of the NiOOH OER active phase of NiOx electrocatalysts at open circuit potential traced by ex situ and in situ spectroscopies. *Catalysis Science & Technology*. 13, 4693–4700 (2023). <https://doi.org/10.1039/d3cy00674c>.

15 Robert Kostecki, Frank McLarnon. Electrochemical and In Situ Raman Spectroscopic Characterization of Nickel Hydroxide Electrodes: I. Pure Nickel Hydroxide. *Journal of the Electrochemical Society*. 144, 485 (1997). <https://doi.org/10.1149/1.1837437>.

16 Hall D.S., Lockwood D.J., Bock C., MacDougall B.R. Nickel hydroxides and related materials: a review of their structures, synthesis and properties. *Proceedings of the Royal Society A*. 471, 20140792 (2015). <http://dx.doi.org/10.1098/rspa.2014.0792>.

17 Minjeong Lee, Yeongeun Jang, Gayoung Yoon, Seonghwa Lee, Gyeong Hee Ryu. Synthesis and electrochemical evaluation of nickel hydroxide nanosheets with phase transition to nickel oxide. *RSC Advances*. 14, 10172 (2024). <https://doi.org/10.1039/d4ra01120a>.

¹Қанатов Ж.С.,

докторант, ORCID ID: 0009-0003-2975-0999,

kanatovzhantilek@gmail.com

^{1*}Калкозова Ж.К.,

ф.-м.ғ.к., профессор, ORCID ID: 0000-0002-4826-1678,

*e-mail: zhanar.kalkozova@kaznu.edu.kz

²Мұқаш Ж.О.,

PhD, ассистент профессор, ORCID ID: 0009-0004-7713-3925,

e-mail: zhanar.mukash@sdu.edu.kz

³Мирзоян М.,

PhD, ORCID ID: 0000-0002-1474-8330,

e-mail: mojtaba.mirzaeian@uws.ac.uk

¹Абдуллин Х.А.,

ф.-м.ғ.д., профессор, ORCID ID: 0000-0002-2729-2272,

e-mail: kh.abdullin@physics.kz

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан²Білім беру және гуманитарлық ғылымдар мектебі, СДУ Университеті,

Қаскелең қ., Қазақстан

³Есептеу техникасы, инженерия және физикалық ғылымдар мектебі,

Батыс Шотландия университеті, Пейсли қ., Ұлыбритания

ГИБРИДТІ СУПЕРКОНДЕНСАТОРЛАРҒА АРНАЛҒАН $\text{Ni}_3\text{S}_2/\text{Ni}$ КӨБІК ЭЛЕКТРОДТАРЫНЫҢ БЕТКІ ТҮРЛЕНУІ ЖӘНЕ ЖАҚСАРТЫЛҒАН ЭЛЕКТРОХИМИЯЛЫҚ СИПАТТАМАЛАРЫ

Аңдатпа

Никель көбігінің бетін сульфидтеу процесі жоғары электрохимиялық сыйымдылығы және электрохимиялық циклдеу кезінде жоғары тұрақтылығы бар Ni_3S_2 қабаттарын алу мақсатында жан-жақты зерттелді. Алайда электрохимиялық жұмыс жағдайларында $\text{Ni}_3\text{S}_2/\text{NF}$ (никель көбігі) электродында түзіледі деп болжанатын никель гидроксиді қабаттарының рөлі әлі күнге дейін жеткілікті деңгейде зерттелмеген. Осы жұмыста гидроксидтік фазаның электрохимиялық сыйымдылыққа да, циклдік тұрақтылыққа да елеулі үлес қосатыны көрсетілді. $\text{Ni}_3\text{S}_2/\text{NF}$ электроды тиомочевина қатысында 160 °C температурада бірсатылы гидротермалдық әдіспен дайындалды. NF бетінде бастапқыда түзілген Ni_3S_2 құрылымы кейіннен КОН электролитінде электрохимиялық циклдеу нәтижесінде модификацияланды. Электродтың электрохимиялық сыйымдылығының артуы бірнеше никель гидроксиді фазаларының түзілуімен қатар жүрді, бұл рентгендік дифракция (XRD) және Раман спектроскопиясы әдістерімен расталды. Электрод 20 $\text{A} \cdot \text{г}^{-1}$ ток тығыздығында жүргізілген 20 000 гальваностатикалық заряд-разряд (GCD) циклі барысында максималды сыйымдылығының 90%-ын сақтап, жоғары циклдік тұрақтылық көрсетті. Ni_3S_2 электродының меншікті сыйымдылығы 2,7 $\text{A} \cdot \text{г}^{-1}$ ток тығыздығында 758 $\text{Ф} \cdot \text{г}^{-1}$ құрады. Ток тығыздығы 90 $\text{A} \cdot \text{г}^{-1}$ -ге дейін арттырылған кезде меншікті сыйымдылық 233 $\text{Ф} \cdot \text{г}^{-1}$ -ге дейін төмендеді, бұл 2,7 $\text{A} \cdot \text{г}^{-1}$ ток тығыздығындағы мәннің шамамен 30%-ын құрайды.

Түйін сөздер: Ni_3S_2 , никель көбігі, гибриді суперконденсатор, никель гидроксиді, электрохимиялық белсендіру, циклдік тұрақтылық.

¹Канатов Ж.С.,

PhD студент, ORCID ID: 0009-0003-2975-0999,

kanatovzhantilek@gmail.com

^{1*}Калкозова Ж.К.,

к.ф.-м.н., профессор, ORCID ID: 0000-0002-4826-1678,

*e-mail: zhanar.kalkozova@kaznu.edu.kz

²Мұқаш Ж.О.,

PhD, Ассистент профессор, ORCID ID: 0009-0004-7713-3925,

e-mail: zhanar.mukash@sdu.edu.kz

³Мирзоян М.

PhD, ORCID ID: 0000-0002-1474-8330,

e-mail: mojtaba.mirzaeian@uws.ac.uk

¹Абдуллин Х.А.,

д.ф.-м.н., профессор, ORCID ID: 0000-0002-2729-2272,

e-mail: kh.abdullin@physics.kz

¹Национальная нанотехнологическая лаборатория открытого типа,

Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

²Школа образования и гуманитарных наук, Университет СДУ, г. Каскелен, Казахстан

³Школа вычислительной техники, инженерии и физических наук,

Университет Западной Шотландии, г. Пейсли, Великобритания

ПРЕОБРАЗОВАНИЕ ПОВЕРХНОСТИ И УЛУЧШЕННЫЕ ЭЛЕКТРОХИМИЧЕСКИЕ ХАРАКТЕРИСТИКИ ЭЛЕКТРОДОВ $\text{Ni}_3\text{S}_2/\text{NI}$ ПЕНЫ ДЛЯ ГИБРИДНЫХ СУПЕРКОНДЕНСАТОРОВ

Аннотация

Процесс сульфидирования поверхности никелевой пены для получения слоёв Ni_3S_2 с высокой электрохимической емкостью и стабильностью при электрохимическом циклировании, был подробно изучен. Однако роль слоев гидроксида никеля, которые, как предполагается, образуются в условиях электрохимической работы электрода $\text{Ni}_3\text{S}_2/\text{NF}$ (никелевая пена), до сих пор не была достаточно изучена. В настоящей работе показано, что гидроксидная фаза вносит значительный вклад как в электрохимическую емкость, так и в циклическую стабильность. Электрод $\text{Ni}_3\text{S}_2/\text{NF}$ был изготовлен с помощью одноэтапного гидротермального метода в присутствии тиомочевины при температуре 160 °С. Первоначальная структура Ni_3S_2 на поверхности NF впоследствии была модифицирована в результате электрохимических циклов в электролите КОН. Увеличение электрохимической емкости электрода сопровождалось образованием нескольких фаз гидроксида никеля, что было подтверждено с помощью рентгеновской дифракции (XRD) и рамановской спектроскопии. Электрод продемонстрировал высокую стабильность характеристик в течение 20 000 циклов гальваностатического заряда-разряда (GCD) при плотности тока 20 $\text{A} \cdot \text{г}^{-1}$, сохранив 90% своей максимальной емкости. Удельная ёмкость электрода Ni_3S_2 составила 758 $\text{Ф} \cdot \text{г}^{-1}$ при плотности тока 2,7 $\text{A} \cdot \text{г}^{-1}$. При увеличении плотности тока до 90 $\text{A} \cdot \text{г}^{-1}$ удельная емкость снизилась до 233 $\text{Ф} \cdot \text{г}^{-1}$, что составляет 30% от значения при плотности тока 2,7 $\text{A} \cdot \text{г}^{-1}$.

Ключевые слова: Ni_3S_2 , никелевая пена, гибридный суперконденсатор, гидроксид никеля, электрохимическая активация, циклическая стабильность.

UDC 31.15.33
IRSTI 31.15.27<https://doi.org/10.55452/1998-6688-2026-23-2-392-400>¹**Rakhymbay L.,**PhD, Researcher, ORCID ID: 0000-0002-7924-837X,
e-mail: lunara.rakhymbay@nu.edu.kz^{2,3}**Myrzakhmetov B.,**PhD, Senior Researcher, ORCID ID: 0000-0002-6828-0649,
e-mail: bauyrzhan.myrzakhmetov@nu.edu.kz^{2,3}**Dauletbay A.,**PhD, Senior Researcher, ORCID ID: 0000-0003-0157-1449,
e-mail: akbar.dauletbay@nu.edu.kz^{1,2,3*}**Konarov A.,**PhD, Associate Professor, ORCID ID: 0000-0001-9352-8602,
*e-mail: aishuak.konarov@nu.edu.kz¹Department of Chemical and Materials Engineering, School of Engineering
and Digital Sciences, Nazarbayev University, Astana, Kazakhstan²National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan³Institute of New Materials and Energy Technologies, Nazarbayev University, Astana, Kazakhstan**OPTIMIZING ZnSO_4 - Li_2SO_4 ELECTROLYTES
FOR STABLE AQUEOUS ZINC-ION BATTERIES****Abstract**

Aqueous zinc-ion batteries (AZIBs) are considered as one of the most attractive candidates for safe, low-cost, and environmentally benign energy storage systems. However, the widespread implementation of these systems is still limited due to major problems with the zinc metal anode, including uncontrolled dendrite formation, hydrogen evolution, and the low reversibility of zinc plating and stripping. These issues lead to rapid capacity fading and shortened cycle life, highlighting the urgent need for electrolyte optimization as a simple and effective strategy to overcome anode instability. The purpose of this work was systematically investigated the electrochemical behavior of zinc anodes in a series of electrolytes with different ZnSO_4 - Li_2SO_4 compositions, namely 2M ZnSO_4 , 1.5M ZnSO_4 + 0.5M Li_2SO_4 , 0.5M ZnSO_4 + 1.5M Li_2SO_4 , 1M ZnSO_4 + 1M Li_2SO_4 , and 2M Li_2SO_4 . The electrochemical performance was evaluated using cyclic voltammetry (CV) and galvanostatic charge-discharge tests, and then post-cycle morphological characterization was performed using scanning electron microscopy (SEM). The results show that adding Li_2SO_4 to the ZnSO_4 electrolyte significantly changes the structure of the Zn^{2+} solution, thereby increasing the reversibility of zinc plating/stripping and suppressing dendrite formation. In particular, mixed electrolytes exhibit sharper voltage profiles and reduced polarization compared to single-salt systems. Among the tested formulations, the equimolar mixture of 1M ZnSO_4 + 1M Li_2SO_4 achieved the most balanced performance, delivering stable cycling and a uniform zinc dendrite morphology. This study highlights electrolyte engineering as a practical and scalable approach to stabilize zinc anodes, providing new insights into the design of high-performance aqueous zinc batteries for future large-scale energy storage applications.

Keywords: aqueous zinc-ion batteries, Zn anode, electrolyte, cyclic voltammetry, dendrite.*Received September 23, 2025; revised May 12, 2026; accepted June 6, 2026.***Introduction**

Electrochemical energy storage plays a pivotal role in enabling the large-scale integration of renewable energy and supporting the transition toward carbon neutrality. Among the current energy

storage technologies, lithium-ion batteries (LIBs) dominate the markets for portable electronics and electric vehicles, owing to their high energy density and dependable performance.

However, their broader deployment is hindered by critical challenges, including the high cost of lithium resources, uneven geographic distribution, and persistent safety concerns associated with flammable organic electrolytes. Moreover, the large-scale production of LIBs imposes significant environmental burdens, driving the pursuit of safer, more cost-effective, and sustainable alternatives [1–2].

In this context, aqueous zinc-ion batteries (AZIBs) have attracted considerable attention as next-generation energy storage devices. Zinc is abundant, inexpensive, and environmentally benign, with a global annual production exceeding 13 million tons. Its favorable electrochemical properties include a relatively low redox potential (-0.76 V vs. SHE) and a high theoretical capacity of 820 mAh g^{-1} , making it promising for high-performance energy storage. In addition, aqueous electrolytes are intrinsically safe and nonflammable, in sharp contrast to the volatile organic solvents used in LIBs, which positions AZIBs as strong contenders for large-scale stationary and grid-level applications [3–7].

Despite these advantages, the practical application of AZIBs faces several serious challenges, the most important of which is the uncontrolled formation of zinc dendrites during repeated coating/stripping cycles. These dendritic structures can pierce the separators, causing internal short circuits, rapid power degradation, and low coulombic efficiency [8]. In parallel, side reactions such as hydrogen evolution, corrosion, and passivation of the zinc surface further impair long-term cycling stability [8, 10]. Overcoming these interfacial instabilities requires carefully engineered strategies to ensure uniform Zn^{2+} deposition and stable electrode-electrolyte interactions.

To address these problems, researchers have explored multiple approaches. Anode surface modification through protective coatings or three-dimensional host structures reduces local current density and improves deposition homogeneity [9]. The design of functional separators with tailored porosity and ion selectivity has also been reported to regulate Zn^{2+} flux. On the cathode side, structural optimization aims to suppress dissolution and improve cycling durability. However, among all strategies, electrolyte engineering stands out as the most practical and scalable method. Previous studies have shown that tailoring electrolyte composition and solvation structures can regulate Zn^{2+} nucleation, suppress dendrite formation, and mitigate side reactions, thereby extending cycle life and advancing AZIBs toward practical deployment [10].

Electrolyte modification strategies have recently gained significant momentum. Conventional ZnSO_4 -based electrolytes provide high ionic conductivity but are prone to dendrite formation and hydrogen evolution. To mitigate these drawbacks, foreign cations such as Li^+ , Na^+ , and Mg^{2+} have been introduced into ZnSO_4 solutions. These ions influence the structure of the electric double layer and modify Zn^{2+} nucleation kinetics, thereby promoting uniform deposition [11–12]. Among these additives, Li^+ has received particular attention. It can homogenize current distribution across the Zn surface, facilitate compact deposits, and stabilize interfacial chemistry. For instance, Dai et al. demonstrated that Li-containing electrolytes form a protective solid-electrolyte interphase (SEI) composed of $\text{Li}_2\text{S}_2\text{O}_7$ and Li_2CO_3 , which mitigates dendrite propagation and suppresses parasitic hydrogen evolution [13]. Other studies have similarly reported that the incorporation of Li_2SO_4 improves Coulombic efficiency and extends cycling stability [14–15].

Beyond dendrite suppression, electrolyte engineering also modulates Zn^{2+} solvation chemistry and hydration structure. Recent reviews highlight that introducing Li^+ ions reshapes Zn^{2+} solvation shells, alters ion transport, and leads to favorable SEI compositions that stabilize electrode surfaces [16]. These findings underscore the importance of carefully designing the $\text{Zn}^{2+}/\text{Li}^+$ ratio to optimize both deposition morphology and interfacial chemistry.

Nevertheless, most previous reports have examined only narrow concentration ranges, leaving a lack of systematic comparisons across different ZnSO_4 - Li_2SO_4 ratios. Furthermore, the precise mechanisms by which Li^+ alters Zn deposition—whether through competitive adsorption at the electrode interface, modification of Zn^{2+} solvation, or stabilization of SEI layers—remain under

debate [17]. Addressing these questions requires carefully designed studies that isolate the role of $\text{Zn}^{2+}/\text{Li}^+$ ratio on electrochemical behavior and surface morphology.

Another critical but underexplored factor is the balance between dendrite suppression and available Zn^{2+} concentration. While dilution of Zn^{2+} by Li_2SO_4 may stabilize morphology and reduce side reactions, it can also lower the effective Zn^{2+} availability, limiting charge-storage capacity. Recent electrolyte design studies also emphasize the need to balance reversibility and capacity; in line with this, our results show that the equimolar $\text{Zn}^{2+}/\text{Li}^+$ ratio offers the most effective compromise between dendrite suppression and charge storage [18].

The purpose of this work is to fill this gap by systematically investigating five electrolyte formulations: 2M ZnSO_4 , 1.5M ZnSO_4 + 0.5M Li_2SO_4 , 0.5M ZnSO_4 + 1.5M Li_2SO_4 , 1M ZnSO_4 + 1M Li_2SO_4 , and 2M Li_2SO_4 . Using cyclic voltammetry, galvanostatic cycling, and post-cycling scanning electron microscopy, we examine how the $\text{Zn}^{2+}/\text{Li}^+$ ratio influences reversibility, dendrite formation, and cycling stability.

Materials and methods

Electrolytes. Five types of aqueous electrolytes were prepared in order to evaluate the effect of $\text{Zn}^{2+}/\text{Li}^+$ ratio on the electrochemical behavior of the cells. The electrolytes included: (i) 2M ZnSO_4 , (ii) 1.5M ZnSO_4 + 0.5M Li_2SO_4 , (iii) 0.5M ZnSO_4 + 1.5M Li_2SO_4 , (iv) 1M ZnSO_4 + 1M Li_2SO_4 and 2M Li_2SO_4 . All salts were of analytical grade and dissolved in deionized water under constant stirring until complete dissolution was achieved.

Electrodes. Zinc foil was used as the anode and cathode was lithium iron phosphate (LFP), which was prepared in the form of a slurry composed of active material (LFP), conductive carbon (AB), and polyvinylidene fluoride (PVDF) binder in a weight ratio of 8:1:1. The slurry then uniformly coated onto graphite foil using a doctor blade technique. The electrodes were dried under vacuum at 80 °C for 12 h.

Cell configuration. Electrochemical cells were assembled in a CR2032 coin-cell configuration. A glass fiber separator was used to physically isolate the Zn anode from the LFP cathode.

Electrochemical measurements. Cyclic voltammetry (CV) was conducted in the voltage range of 0.9-2.0 V at a scan rate of 0.2 mV s⁻¹ to evaluate the redox processes of the LFP cathode and the reversibility of Zn plating/stripping. Galvanostatic charge-discharge tests were performed at a constant current density of 20 mA g⁻¹ to assess specific capacity and cycling stability.

Characterization. To examine the influence of electrolyte composition on Zn deposition, the anode surface morphology was examined using scanning electron microscopy (SEM).

Results and discussion

The CV curves of Zn//LFP cells were recorded within a voltage window of 0.9-2.0 V at a scan rate of 0.2 mV s⁻¹ (Figure 1).

In Figure 1a 2M ZnSO_4 electrolyte, the CV profile shows broad and poorly defined redox features. A cathodic peak appears at ≈ 1.07 V, while the anodic process is not fully captured.

Upon partial substitution with Li_2SO_4 (1.5M ZnSO_4 + 0.5M Li_2SO_4), the electrochemical response improves significantly. The reduction peak lies at ≈ 1.10 -1.15 V, while the oxidation exhibits a shoulder at ≈ 1.42 V. With repeated cycling, the anodic peak stabilizes upon cycling, indicating reduced polarization and enhanced reversibility.

The equimolar mixture (1M ZnSO_4 + 1M Li_2SO_4) displays the sharpest and most symmetric peaks. The reduction peak is centered at ≈ 1.15 -1.19 V, while the oxidation peak occurs at ≈ 1.34 -1.48 V, yielding the smallest ΔE among all electrolytes. The strong current responses and nearly overlapping cycles demonstrate highly reversible Zn plating/stripping [19]. This composition thus provides the most favorable balance between Zn^{2+} concentration and Li^+ .

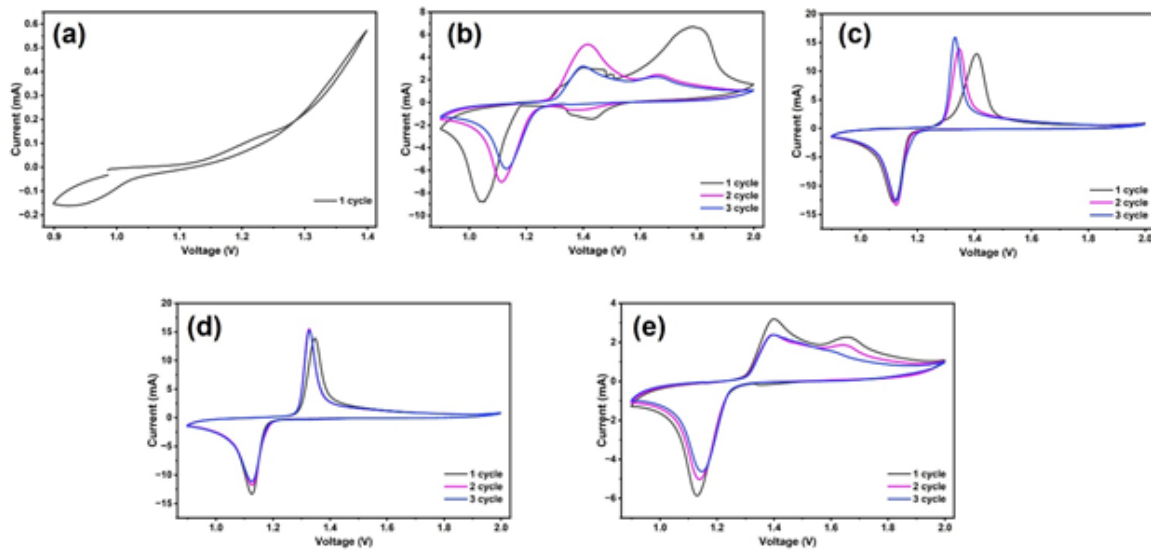


Figure 1 – Cyclic voltammograms of Zn//LFP cells: a) 2M ZnSO_4 ,
b) 1.5M ZnSO_4 + 0.5M Li_2SO_4 ; c) 1M ZnSO_4 + 1M Li_2SO_4 ,
d) 0.5M ZnSO_4 + 1.5M Li_2SO_4 and e) 2M Li_2SO_4

Further increasing the Li_2SO_4 fraction (0.5M ZnSO_4 + 1.5M Li_2SO_4) maintains sharp and reversible peaks at 1.12-1.16 V and $\approx$ 1.34-1.40 V. However, the overall current densities are lower than in the equimolar electrolyte, reflecting reduced Zn^{2+} availability. This suggests a trade-off between dendrite suppression and charge storage capability when Zn^{2+} concentration is diluted.

In the absence of ZnSO_4 (2M Li_2SO_4), the CV curves no longer display clear Zn^{2+}/Zn redox peaks. Instead, broad anodic responses are observed at $\approx$ 1.40-1.50 V with an additional weak shoulder. This confirms that Li_2SO_4 alone cannot sustain Zn plating/stripping and mainly drives parasitic side reactions.

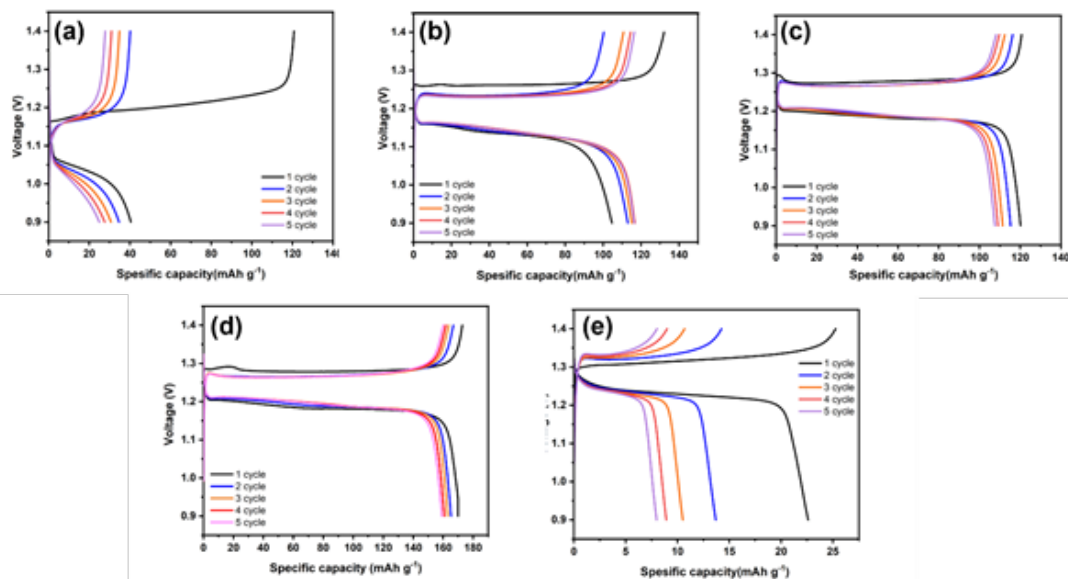


Figure 2 – Charge-discharge profiles of Zn//LFP cells in different electrolytes between 0.9-1.4 V

To validate the CV results, the cycling performance of Zn//LFP cells was further examined under repeated galvanostatic charge-discharge (Figure 2 and 3). The charge-discharge curves show clear differences across electrolytes. Pure ZnSO_4 exhibits large polarization and low capacity, while partial Li_2SO_4 substitution improves reversibility and sharpens voltage plateaus. The equimolar $1\text{M ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$ delivers the highest and most stable capacity with minimal hysteresis. In contrast, Li_2SO_4 alone provides steep sloping curves with very low storage, confirming the absence of reversible Zn^{2+} electrochemistry and indicating that the observed capacity mainly originates from side reactions rather than effective charge storage. In terms of cycling performance, the 2M ZnSO_4 electrolyte exhibits, the capacity rapidly declines from $\sim 40\text{ mAh g}^{-1}$ to around 10 mAh g^{-1} within 15 cycles. A similar rapid fading is observed in the $2\text{M Li}_2\text{SO}_4$ electrolyte, which delivers $<10\text{ mAh g}^{-1}$ after only a few cycles. By contrast, the mixed ZnSO_4 - Li_2SO_4 electrolytes exhibit markedly better cycling stability. Both the $1.5\text{M ZnSO}_4 + 0.5\text{M Li}_2\text{SO}_4$ and $0.5\text{M ZnSO}_4 + 1.5\text{M Li}_2\text{SO}_4$ compositions stabilize around $85\text{--}95\text{ mAh g}^{-1}$ after initial activation, retaining most of their capacity over 30 cycles.

The best performance is achieved with the equimolar $1\text{M ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$ electrolyte, which maintains $\sim 165\text{--}169\text{ mAh g}^{-1}$ [20] over 20 cycles with minimal fading, confirming highly reversible Zn plating/stripping and excellent interfacial stability. These results strongly corroborate the CV findings: while single-salt electrolytes suffer from poor reversibility and rapid capacity decay, mixed Zn-Li electrolytes, especially the equimolar formulation, ensure stable and reversible cycling behavior.

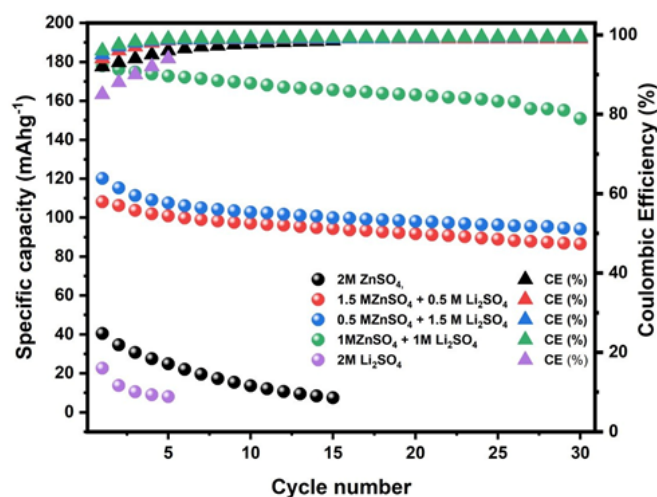


Figure 3 – Cycling performance of Zn//LFP cells in different electrolytes

In addition to capacity retention, the coulombic efficiency (CE) of the cells was evaluated to assess the reversibility of the electrochemical processes (Figure 3). The mixed $\text{ZnSO}_4 - \text{Li}_2\text{SO}_4$ electrolytes exhibit consistently high CE values close to $\sim 98\text{--}100\%$ throughout cycling, indicating highly reversible Zn plating/stripping behavior.

In particular, the equimolar $1\text{M ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$ electrolyte demonstrates the most stable CE ($\sim 99 - 100\%$), confirming minimal parasitic reactions and enhanced interfacial stability. In contrast, the 2M ZnSO_4 electrolyte shows slightly lower CE values, which can be attributed to side reactions such as hydrogen evolution and dendrite formation.

These results further confirm that electrolyte optimization with Li_2SO_4 effectively improves both reversibility and cycling stability of aqueous Zn-ion batteries.

Further, the surface morphologies of Zn anodes after 30 cycles in different electrolytes were examined by SEM (Figure 4). In the 2M ZnSO_4 electrolyte (Figure 4a), the surface is covered with large, irregular dendritic structures with lengths exceeding $1\text{ }\mu\text{m}$ ($1.3\text{--}1.6\text{ }\mu\text{m}$), indicating severe Zn dendrite growth and uncontrolled deposition. Such morphology explains the rapid capacity decay observed during cycling [16].

In the 1.5M $\text{ZnSO}_4 + 0.5\text{M Li}_2\text{SO}_4$ electrolyte (Figure 4b), the dendritic features are partially suppressed, but the surface still exhibits scattered plate-like deposits and roughness, reflecting incomplete stabilization. A more uniform morphology is obtained in the 1M $\text{ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$ electrolyte (Figure 4c), where the Zn surface is smoother and dendritic protrusions are largely absent. This confirms that the equimolar mixture promotes homogeneous Zn nucleation and growth. For the 0.5M $\text{ZnSO}_4 + 1.5\text{M Li}_2\text{SO}_4$ electrolyte (Figure 4d), the Zn deposits are finer and less aggregated compared to pure ZnSO_4 , though scattered needle-like features are still visible. The reduced Zn^{2+} concentration leads to thinner deposits, which may limit capacity despite improved stability. Finally, in the 2M Li_2SO_4 electrolyte (Figure 3e), the surface is dominated by irregular flakes and byproducts, lacking well-defined Zn deposits. This correlates with the poor electrochemical performance observed.

Taken together, the SEM results confirm that Li_2SO_4 -modified electrolytes suppress dendrite formation and promote smoother Zn deposition, with the 1:1 ZnSO_4 - Li_2SO_4 mixture producing the most uniform surface morphology, in line with electrochemical data.

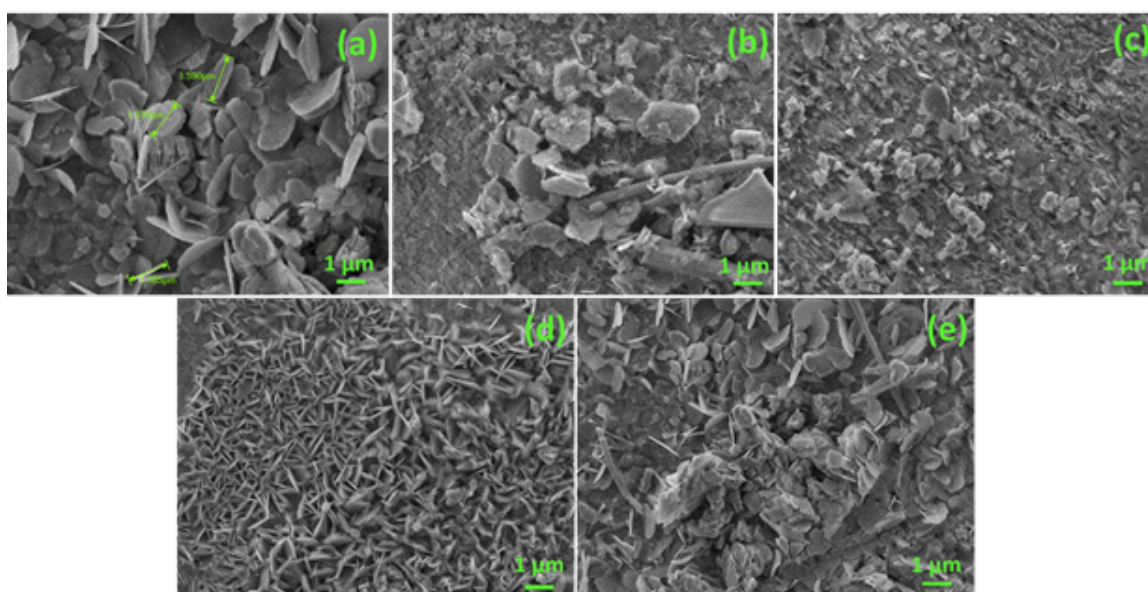


Figure 4 – SEM images of Zn anodes after cycling: a) 2M ZnSO_4 ,
b) 1.5M $\text{ZnSO}_4 + 0.5\text{M Li}_2\text{SO}_4$, c) 1M $\text{ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$,
d) 0.5M $\text{ZnSO}_4 + 1.5\text{M Li}_2\text{SO}_4$ and e) 2M Li_2SO_4

To further understand the origin of the improved electrochemical performance, the possible mechanisms are discussed below. The improved electrochemical performance in Li_2SO_4 -containing electrolytes can be attributed to two possible mechanisms. First, Li^+ ions modify the Zn^{2+} solvation structure, reducing the number of free water molecules in the Zn^{2+} solvation shell. This suppresses parasitic hydrogen evolution and promotes more uniform Zn deposition. Second, Li^+ may contribute to the formation of a more stable solid-electrolyte interphase (SEI), composed of inorganic species such as Li_2CO_3 and Li_2SO_4 -derived compounds, which protects the Zn surface and inhibits dendrite growth. Although direct evidence of SEI formation was not obtained in this study, the improved cycling stability and smoother morphology observed by SEM indirectly support this hypothesis.

The obtained results are consistent with recent advances in aqueous zinc-ion batteries, which highlight the critical role of electrolyte engineering in improving Zn anode stability and electrochemical performance. Previous studies have shown that tailoring the Zn^{2+} solvation structure, cycling stability [21] and introducing Li^+ -based additives can effectively suppress parasitic reactions, enhance Coulombic efficiency, and mitigate dendrite growth [22]. In most reported systems, modified electrolytes typically achieve capacities in the range of 90–150 mAh g^{-1} depending on electrolyte composition [23]. In contrast, the equimolar ZnSO_4 - Li_2SO_4 electrolyte (1M + 1M) investigated in

this work delivers a higher specific capacity of $\sim 165\text{--}169\text{ mAh g}^{-1}$, along with improved cycling stability and more uniform Zn deposition morphology.

These findings indicate that precise control of the $\text{Zn}^{2+}/\text{Li}^{+}$ ratio plays a decisive role in optimizing ion solvation structure and interfacial processes, leading to superior electrochemical performance compared to previously reported electrolyte systems.

Conclusion

In this research, the electrochemical behavior of Zn//LFP cells was investigated in different aqueous electrolytes. The results show that mixed $\text{ZnSO}_4\text{--Li}_2\text{SO}_4$ electrolytes significantly enhance Zn^{2+} plating/stripping reversibility and cycling stability compared to single-salt systems. Among the tested formulations, the $1\text{M ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$ electrolyte exhibited the best overall performance, delivering the highest capacity, excellent retention, and uniform Zn morphology. Overall, our results demonstrate that electrolyte engineering with Zn-Li co-salts is a viable route to achieving stable aqueous Zn-ion batteries, as evidenced by the $1\text{M ZnSO}_4 + 1\text{M Li}_2\text{SO}_4$ system delivering $\sim 165\text{--}169\text{ mAh g}^{-1}$ with excellent retention.

Acknowledgements: This work was supported by Research Targeted Program “Development of new technologies of materials and energy storage systems for a green economy” BR21882402 from the Ministry of Science and Higher Education of the Republic of Kazakhstan.

Conflict of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- 1 Tarascon J.-M., Armand M. Issues and challenges facing rechargeable lithium batteries. *Nature*, 2001, 414, 359–367. <https://doi.org/10.1038/35104644>
- 2 Nitta N., Wu F., Lee J.T., Yushin G. Li-ion battery materials: present and future. *Mater. Today*, 2015, 18, 252–264. <https://doi.org/10.1016/j.mattod.2014.10.040>
- 3 Parker J.F., et al. Rechargeable nickel–3D zinc batteries: An energy-dense, safer alternative to lithium-ion. *Science*, 2017, 356, 415–419. <https://doi.org/10.1126/science.aak9991>
- 4 Rakhman D., et al. Polyacrylamide-based hydrogel electrolyte for modulating water activity in aqueous hybrid batteries. *RSC Advances*, 2024, 14, 40222–40233. <https://doi.org/10.1039/D4RA07551J>
- 5 Konarov A., et al. Innovative hydrogel electrolytes for hybrid zinc-ion batteries. *Mechanics and Technologies*, 2025, 88, 156–164. <http://dx.doi.org/10.55956/PXTX1347>
- 6 Pan H., et al. Reversible aqueous zinc/manganese oxide energy storage from conversion reactions. *Nat. Energy*, 2016, 1, 16039. <https://doi.org/10.1038/nenergy.2016.39>
- 7 Zhang N., et al. Rechargeable aqueous zinc–manganese dioxide batteries with high energy and power densities. *Nat. Commun.*, 2017, 8, 405. <https://doi.org/10.1038/s41467-017-00467-x>
- 8 Yuan L., et al. Regulation methods for the Zn/electrolyte interphase and the effectiveness evaluation in aqueous Zn-ion batteries. *Energy Environ. Sci.*, 2021, 14, 5669–5689. <https://doi.org/10.1039/D1EE02021H>
- 9 Tang B., Shan L., Liang S., Zhou J. Issues and opportunities facing aqueous zinc-ion batteries. *Energy Environ. Sci.*, 2019, 12, 3288–3304. <https://doi.org/10.1039/C9EE02526J>
- 10 Zhang Q., et al. Interfacial design of dendrite-free zinc anodes for aqueous zinc-ion batteries. *Angew. Chem. Int. Ed.*, 2020, 59(32), 13180–13191. <https://doi.org/10.1002/anie.202000162>
- 11 Zhang P., et al. Zinc dendrite growth and inhibition strategies, *Mat. Today Energy*, 2021, 20, 100692. <https://doi.org/10.1016/j.mtener.2021.100692>
- 12 Guo X., et al. Alleviation of dendrite formation on zinc anodes via electrolyte additives. *ACS Energy Lett.*, 2021, 6, 395–403. <https://doi.org/10.1021/acsenergylett.0c02371>
- 13 Dai H., et al. Unraveling chemical origins of dendrite formation in zinc-ion batteries. *Nat. Commun.*, 2024, 15, 5123. <https://doi.org/10.1038/s41467-024-52651-5>
- 14 Zhang N., Chen X., Yu M., Niu Z. Materials chemistry for rechargeable zinc-ion batteries. *Chem. Soc. Rev.*, 2020, 49(13), 4203–4219. <https://doi.org/10.1039/C9CS00349E>

- 15 Ju Z., et al. Interfacial chemistry in multivalent aqueous batteries: fundamentals, challenges, and advances. *Chem. Soc. Rev.*, 2024, 53(18), 6290–6335. <https://doi.org/10.1039/D4CS00474D>
- 16 Zeng X., Zhang X., Qin H., et al. Dendrite-Free Zinc Deposition Induced by a Tridimensional Host for Stable Zinc-Ion Batteries. *Adv. Mater.*, 2021, 33, 2007416. <https://doi.org/10.1002/adma.202007416>
- 17 Zhangxing H., et al. Electrolyte additive engineering for aqueous Zn ion batteries, *Energy Storage Materials*, 2022, 51, 733 – 755. <https://doi.org/10.1016/j.ensm.2022.07.017>
- 18 Yang J., Yin B., Sun Y., et al. Zinc Anode for Mild Aqueous Zinc-Ion Batteries: Challenges, Strategies, and Perspectives. *Nano-Micro Lett.*, 2022, 14, 1029 (Article 78). <https://doi.org/10.1007/s40820-021-00782-5>
- 19 Kurmanbayeva I., et al. Tetrapropylammonium Hydroxide as a Zinc Dendrite Growth Suppressor for Rechargeable Aqueous Battery. *Front. Energy Res.*, 2020, 8, 599009. <https://doi.org/10.3389/fenrg.2020.599009>
- 20 Yesibolati N., et al. High performance Zn/LiFePO₄ aqueous rechargeable battery for large scale applications. *Electrochim. Acta*, 2015, 152, 505–511. <https://doi.org/10.1016/j.electacta.2014.11.168>
- 21 Shi, X., Xie, J., Wang, J. et al. A weakly solvating electrolyte towards practical rechargeable aqueous zinc-ion batteries. *Nat Commun* 2024, 15, 302. <https://doi.org/10.1038/s41467-023-44615-y>
- 22 Shen, Q.; Wang, Y.; Han, G.; Li, X.; Yuan, T.; Sun, H.; Gong, Y.; Chen, T. Recent Progress in Electrolyte Additives for Highly Reversible Zinc Anodes in Aqueous Zinc Batteries. *Batteries* 2023, 9, 284. <https://doi.org/10.3390/batteries9050284>
- 23 Kong J. et al. Highly improved aqueous Zn|| LiMn₂O₄ hybrid-ion batteries using poly (ethylene glycol) and manganese sulfate as electrolyte additives // *Sustainable Energy & Fuels*, 8(4), 826-836 (2024). <https://doi.org/10.1039/D3SE01295F>

¹Рахымбай Л.,

PhD, ғылыми қызметкер, ORCID: 0000-0002-7924-837X,
e-mail: lunara.rakhymbay@nu.edu.kz

^{2,3}Мырзахметов Б.,

PhD, аға ғылыми қызметкер, ORCID: 0000-0002-6828-0649,
e-mail: bauyrzhan.myrzakhmetov@nu.edu.kz

^{2,3}Дәулетбай А.,

PhD, аға ғылыми қызметкер, ORCID: 0000-0003-0157-1449,
e-mail: akbar.dauletbay@nu.edu.kz

^{1,2,3*}Қонаров А.,

PhD, қауымдастырылған профессор, ORCID: 0000-0001-9352-8602,
*e-mail: aishuak.konarov@nu.edu.kz

¹Назарбаев Университеті, Инженерия және цифрлық ғылымдар мектебі,
Химиялық және материалдар инженериясы департаменті, Астана қ., Қазақстан

²National Laboratory Astana, Назарбаев Университеті, Астана қ., Қазақстан

³Жаңа материалдар және энергетикалық технологиялар институты,
Назарбаев Университеті, Астана қ., Қазақстан

ТҰРАҚТЫ СУЛЫ МЫРЫШ-ИОНДЫҚ АККУМУЛЯТОРЛАР ҮШІН $\text{ZnSO}_4 - \text{Li}_2\text{SO}_4$ ЭЛЕКТРОЛИТТЕРІН ОҢТАЙЛАНДЫРУ

Аңдатпа

Сұлы мырыш-ионды батареялар (AZIBs) қауіпсіз, арзан және экологиялық таза энергия сақтау жүйелерінің болашағы зор түрлерінің бірі болып саналады. Алайда, мырыш анодының тұрақсыздығы, соның ішінде бақылаусыз дендриттік өсу, сутегінің бөлінуі және мырыштың қайтымсыз қапталу/шығарылу процестері олардың ауқымды қолданылуына кедергі келтіреді. Бұл мәселелер сыйымдылықтың тез төмендеуіне және қызмет ету мерзімінің қысқаруына әкеледі. Осыған байланысты электролитті оңтайландыру анодты тұрақтандырудың қарапайым әрі тиімді тәсілі ретінде ерекше мәнге ие. Жұмыстың мақсаты – әртүрлі құрамдағы $\text{ZnSO}_4\text{--Li}_2\text{SO}_4$ электролиттерінің (2 M ZnSO_4 , 1.5 M ZnSO_4 + 0.5 M Li_2SO_4 , 0.5 M ZnSO_4 + 1.5 M Li_2SO_4 , 1 M ZnSO_4 + 1 M Li_2SO_4 және 2 M Li_2SO_4) мырыш анодына әсерін жүйелі түрде зерттеу. Электрохимиялық сипаттамалар циклдік вольтамметрия (CV), гальваностатикалық заряд-разряд сынақтары арқылы зерттелді, ал циклдеуден кейін анод бетінің морфологиясы сканерлеуші электрондық микроскопия (СЭМ) әдісімен

талданды. Нәтижелер ZnSO_4 электролитіне Li_2SO_4 тұзын енгізу Zn^{2+} иондарының сольватациялық ортасын өзгертетінін көрсетті, соның арқасында мырыштың қайтымдылығы артып, дендриттердің түзілуі бәсеңдеді. Әсіресе, аралас электролиттер бір компонентті жүйелермен салыстырғанда төмен поляризацияны және айқынырақ кернеу қисықтарын көрсетті. Олардың ішінде $1 \text{ M ZnSO}_4 + 1 \text{ M Li}_2\text{SO}_4$ теңмолярлы құрамы ең оңтайлы нәтижелерге қол жеткізді: тұрақты циклдеу, төмен интерфейстік кедергі және біртекті, дендритсіз морфология. Осылайша, электролиттік инженерия мырыш анодтарын тұрақтандырудың тиімді әдісі болып табылады және болашақта сулы мырыш-ионды батареяларды ірі көлемді энергия сақтау мақсатында дамытуға жаңа мүмкіндіктер ашады.

Түйін сөздер: сулы мырыш-ионды аккумуляторы, Zn анод, электролит, циклдік вольтамметрия, дендрит.

¹Рахымбай Л.,

PhD, научный сотрудник, ORCID ID: 0000-0002-7924-837X,

e-mail: lunara.rakhymbay@nu.edu.kz

^{2,3}Мырзахметов Б.,

PhD, старший научный сотрудник, ORCID ID: 0000-0002-6828-0649,

e-mail: bauyrzhan.myrzakhetov@nu.edu.kz

^{2,3}Даулетбай А.,

PhD, старший научный сотрудник, ORCID ID: 0000-0003-0157-1449,

e-mail: akbar.dauletbay@nu.edu.kz

^{1,2,3*}Конаров А.,

PhD, ассоциированный профессор, ORCID ID: 0000-0001-9352-8602,

*e-mail: aishuak.konarov@nu.edu.kz

¹Департамент химической и материаловедческой инженерии,

Школа инженерии и цифровых наук, Назарбаев Университет, г. Астана, Казахстан

²National Laboratory Astana, Назарбаев Университет, г. Астана, Казахстан

³Институт новых материалов и энергетических технологий, Назарбаев Университет,
г. Астана, Казахстан

ОПТИМИЗАЦИЯ ЭЛЕКТРОЛИТОВ $\text{ZnSO}_4 - \text{Li}_2\text{SO}_4$ ДЛЯ СТАБИЛЬНЫХ ВОДНЫХ ЦИНК-ИОННЫХ БАТАРЕЙ

Аннотация

Водные цинковые аккумуляторы (AZIBs) рассматриваются как одни из наиболее перспективных кандидатов для безопасного, недорогого и экологически чистого накопления энергии. Тем не менее их практическое применение ограничено нестабильностью цинкового анода, вызванной неконтролируемым ростом дендритов, выделением водорода и низкой обратимостью процессов осаждения/растворения цинка. Эти факторы приводят к быстрой деградации емкости и сокращению срока службы батареи. В связи с этим оптимизация электролита является эффективным и доступным подходом к стабилизации анода. Целью работы является систематическая оценка поведения цинкового анода в электролитах различного состава 2 M ZnSO_4 , $1.5 \text{ M ZnSO}_4 + 0.5 \text{ M Li}_2\text{SO}_4$, $0.5 \text{ M ZnSO}_4 + 1.5 \text{ M Li}_2\text{SO}_4$, $1 \text{ M ZnSO}_4 + 1 \text{ M Li}_2\text{SO}_4$ и $2 \text{ M Li}_2\text{SO}_4$. Электрохимические характеристики исследованы методами циклической вольтамперометрии (CV), гальваностатического циклирования и после испытаний морфология поверхности анализировалась с помощью сканирующей электронной микроскопии (СЭМ). Данные показали, что добавление Li_2SO_4 в раствор ZnSO_4 существенно изменяет сольватационную оболочку Zn^{2+} , улучшая обратимость процессов осаждения/растворения и подавляя рост дендритов. Смешанные электролиты продемонстрировали более четкие вольт-амперные профили и меньшую поляризацию по сравнению с одноионными системами. Наиболее сбалансированные характеристики достигнуты в электролите $1 \text{ M ZnSO}_4 + 1 \text{ M Li}_2\text{SO}_4$: обеспечены стабильное циклирование, низкое межфазное сопротивление и однородная бездендритная морфология. Таким образом, инженерия электролита является практическим и масштабируемым решением для стабилизации цинковых анодов и открывает путь к созданию высокоэффективных водных аккумуляторов для крупномасштабных систем хранения энергии.

Ключевые слова: водные цинк-ионные аккумуляторы, Zn-анод, электролит, циклическая вольтамперометрия, дендрит.

UDC 621.039.51
IRSTI 58.33.29

<https://doi.org/10.55452/1998-6688-2026-23-2-401-417>

¹Aryngazin A.,

MSc, ORCID ID: 0009-0007-7861-4917,
e-mail: ansar.aryngazin@nu.edu.kz

¹Kurbanova B.,

MSc, ORCID ID: 0000-0003-0928-521X,
e-mail: bayan.kurbanova@nu.edu.kz

²Sizyuk Y.,

PhD, e-mail: yuriy.sizyuk@nu.edu.kz

³Temirbayeva A.

M.Sc., ORCID ID: 0000-0002-9526-3617,
e-mail: assem.zkgmu@gmail.com

¹Alsar Zh.,

PhD, ORCID ID: 0000-0001-7287-7555,
e-mail: zhanna.alsar@nu.edu.kz

^{1*}Insepov Z.,

D.Sc. PhD, Adj. Professor, ORCID ID: 0000-0002-8079-6293,
*e-mail: zinsepov@nu.edu.kz

¹Nazarbayev University Research Administration, Astana, Kazakhstan

²Depart. Mech. Nucl. Engin, Virginia Commonwealth University, VA USA

³Eurasian National University named after Gumilev, Astana, Kazakhstan

⁴School of Nuclear Engineering, Purdue University IN, USA

NEUTRONICS AND THERMAL-HYDRAULICS OF PRESSURIZED WATER REACTOR WITH THORIUM-BASED FUEL

Abstract

In the present study, the coupled behavior of thermal-hydraulic processes and solid mechanics phenomena was investigated using the finite element method implemented within the COMSOL Multiphysics environment. Neutronic analyses were performed using the OpenMC code under the steady-state neutron transport approximation. To optimize the core configuration of a thorium-fueled pressurized water reactor (PWR), high-pressure light water was selected as the coolant and moderator. The reactor core was composed of 49 fuel rods containing thorium-based fuel compounds arranged in a regular lattice configuration. Monte Carlo neutron transport simulations conducted with OpenMC enabled the evaluation of the effective neutron multiplication factor (k_{eff}) and the spatial power distribution within the reactor core. Thermal-hydraulic calculations yielded detailed temperature and coolant velocity distributions, while solid mechanics analyses provided the corresponding spatial distributions of stresses and deformations arising under operational conditions. The obtained results demonstrate the applicability and effectiveness of a coupled multiphysics approach integrating steady-state neutronics with thermal-hydraulic and structural analyses for the assessment and optimization of thorium-fueled pressurized water reactor systems.

Keywords: nuclear reactor physics, thorium fuel, neutronics, thermal-hydraulics, solid mechanics, COMSOL, OpenMC.

Received February 25, 2026; revised May 28, 2026; accepted June 9, 2026.

1. Introduction

The growing demand for sustainable and proliferation-resistant nuclear energy systems has stimulated renewed interest in alternative fuel cycles capable of improving resource utilization and reducing long-lived radioactive waste. Among such alternatives, thorium-based fuels have attracted considerable attention due to their favorable neutronic characteristics, high natural abundance, and potential for reduced production of minor actinides compared with conventional uranium fuel cycles [1]. The thorium fuel cycle has been studied for several decades, beginning with early experimental demonstrations such as the Molten Salt Reactor Experiment conducted at Oak Ridge National Laboratory in the 1960s [2]. Since then, numerous investigations have explored the feasibility of incorporating thorium into both existing and advanced reactor systems.

In light-water reactors, and particularly in pressurized water reactors (PWRs), partial or full replacement of uranium fuel with thorium-based compositions has been proposed as a promising pathway toward improved fuel utilization and waste reduction. Thorium dioxide exhibits several advantageous thermophysical properties, including higher thermal conductivity and lower thermal expansion compared with uranium dioxide [1]. These characteristics can contribute to reduced temperature gradients, improved thermal margins, and enhanced structural stability of fuel elements during operation. At the same time, incorporation of fissile isotopes such as U-233 or low-enriched uranium enables thorium to sustain nuclear fission through neutron capture and subsequent transmutation processes.

Previous studies have investigated neutronic performance, burnup characteristics, and safety parameters of thorium-based fuels in thermal and fast reactor systems. Monte Carlo transport simulations using codes such as MCNP [3, 4] and OpenMC [5], supported by evaluated nuclear data libraries including ENDF [6] and JEFF [7], have demonstrated the feasibility of thorium utilization in PWR assemblies with acceptable reactivity coefficients and fuel cycle performance [8–17]. In parallel, thermal-hydraulic and thermo-mechanical analyses have shown that the superior thermal conductivity of ThO₂-based fuels can reduce peak fuel temperatures and mitigate fuel-cladding mechanical interaction [18–25]. However, most existing investigations treat neutronics, thermal-hydraulics, and structural behavior either independently or within simplified coupling frameworks, which may limit the accuracy of safety and performance assessments.

Accurate prediction of reactor behavior requires consistent coupling between neutronic power generation and the resulting thermal-hydraulic and mechanical responses of the fuel system. While time-dependent neutronic simulations provide detailed transient behavior, they are computationally expensive when combined with high-fidelity multiphysics models [26–28]. An alternative approach involves coupling steady-state neutronics with transient thermal-hydraulic and structural calculations [29, 30]. This approach exploits the difference in characteristic timescales between neutron transport and thermal-hydraulic processes: neutron populations typically reach equilibrium much faster than temperature and flow fields evolve. As a result, steady-state neutronic solutions can be used to provide spatial power distributions that drive time-dependent thermal and mechanical simulations with significantly reduced computational cost.

Such multiphysics coupling has been applied in recent years to both conventional light-water reactors and advanced reactor concepts [26, 29, 31]. Studies indicate that steady-state neutronics combined with transient thermal-hydraulics can provide reliable predictions of temperature evolution, stress development, and flow stability while maintaining acceptable computational efficiency. Nevertheless, further investigation is required to assess the applicability of this approach to thorium-based fuel assemblies in PWR configurations and to evaluate the resulting thermo-mechanical behavior under transient heating conditions.

By combining steady-state neutronics with time-dependent thermal-hydraulic and structural modeling, this study aims to provide a consistent multiphysics assessment of thorium-based fuel

performance in a PWR assembly and to evaluate the suitability of this coupling strategy for future high-fidelity reactor safety analyses.

In this paper a Design-Basis Core of a conventional PWR with Thorium/Uranium Oxide fuel, Zircaloy-4 cladding and light water coolant was considered. The neutronics study was performed using OpenMC v.0.15.3 with ENDF.B.VIII-0 nuclear cross-section library data. The thermal-hydraulics, solid mechanics and fuel safety study were performed in COMSOL v.6.3. The paper is organized as follows: Section II describes main methods and materials used for the study. In Section III the results are obtained and discussed. In Section IV conclusions and future research goals are presented.

2. Methods and models

2.1. OpenMC

OpenMC is an open-source community-developed Monte-Carlo neutron and photon transport simulation software package. It uses constructive solid geometry or CAD representation to perform various neutronics calculations with capabilities of computation parallelism via MPI, OpenMP or hybrid programming model [5].

The model in consideration of neutronics calculations is an assembly of 7x7 pins arranged in rectangular lattice with $(\text{Th}/^{233}\text{U})\text{O}_2$ fuel domain, enveloped by Zircaloy-4 alloy as a cladding domain. These pillars are surrounded by light water coolant domain. Featured important specifications of a reactor core are listed in Table 1. The top view slice at half-height of the assembly can be seen in Figure 1.

Table 1 – Parameters of a PWR Design Core

Parameter	Value
Core lattice geometry	Rectangular
Number of pins	49 (7x7)
Fuel material	$(\text{Th}_{0.985}, ^{233}\text{U}_{0.015} \text{ at\%})\text{O}_2$
Fuel density	10.0 g/cm ³
Fuel temperature	900 K
Fuel thickness	1 cm
Fuel element height	100 cm
Cladding material	Zircaloy-4 ($\text{Zr}_{0.985}, \text{Sn}_{0.014}, ^{5}\text{Fe}_{0.0021}, \text{Cr}_{0.0009} \text{ at\%}$)
Cladding density	6.55 g/cm ³
Cladding thickness	0.2 cm
Coolant material	H ₂ O
Coolant density	1.0 g/cm ³
Boundaries	X: -0.75 – +21.75 cm Y: -0.75 – +21.75 cm Z: -50 – +50 cm
Active core volume	50 625 cm ³
Mesh dimensions	500x500x500

The calculations parameters have been chosen to optimize between the complexity and accuracy and time of calculations that depends on computing power of processing units. According to execution settings for the specific OpenMC mode called “k-eigenvalue”, a single run simulates a selected by the user number of particles, that combine into fission generations that store resulting fission neutrons,

the spatial coordinates of the fission site, the sampled outgoing energy and direction of the fission neutron, and the weight of the neutron of the current generation. Then, the next successive generation uses random sampling of stored generation to ensure that the neutron population won't grow in a rapid manner. Successive generations combine into batches for statistical significance. Some batches are needed to be discarded in order for the source – not known a priori – to evolve closer to true distribution at each iteration. After that the tallies that contain execution information are starting to accumulate. In the simulations presented here, 96000 particles with 25 generations per batch, 240 batches and 100 inactive batches were used.

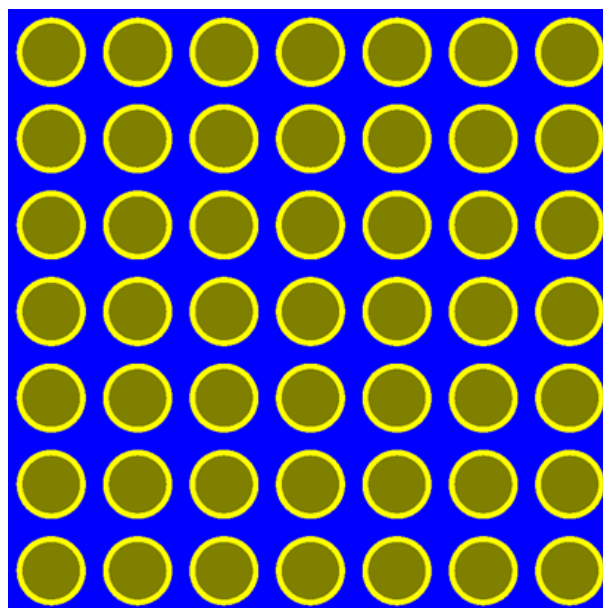


Figure 1 – 2D slice of a PWR Design Core at half-height of the design core.
Designations: blue – coolant, yellow – cladding, olive – fuel

For the accurate calculations of neutron flux and heat distribution throughout the assembly without heavy overstress of processing power the assembly space was divided into 3D mesh with moderate resolution of 60 blocks per dimension. This value ensures enough accuracy for calculations while providing easiness on post-processing of obtained raw results.

Additionally in the execution settings, a source of neutrons constrained inside the fuel only domain with uniform spatial distribution, isotropic angular distribution, Watt energy distribution and strength of source at 98% was included. To get the analogous perspective of neutron energy spectrum from similar reactor cores, the closest to PWR cross-section data library available, called SHEM-361, was used.

2.2. Thermal-hydraulics and solid mechanics with COMSOL

The radial power distribution from the neutronic analysis is used to perform thermal-hydraulic and solid-mechanics simulations. The objective is to capture and analyze these conditions through numerical simulations.

Based on the power distribution, the hottest fuel rod can be identified, representing the most limiting and safety-critical operating conditions within the reactor core. The solution procedure begins with discretizing the geometry into smaller elements, a process referred to as meshing. Subsequently, the governing partial differential equations (PDEs) are applied to the generated mesh. The heat transfer equation is:

$$\rho C_p \left(\frac{\partial T}{\partial t} + \mathbf{u}_{\text{trans}} \cdot \nabla T \right) + \nabla \cdot \mathbf{q} = Q + Q_{\text{ted}} \quad (1)$$

where equation parameters are listed in Table 2.

Table 2 – Heat transfer equation parameters

Parameter	Unit	Definition
ρ	kg/m^3	Density
C_p	$\text{J}/(\text{kg} \cdot \text{K})$	specific heat at constant pressure
T	K	Temperature
u_{trans}	m/s	velocity vector of translational motion
$q = -k\nabla T$	W/m^2	conduction heat flux
k	$\text{W}/(\text{m} \cdot \text{K})$	Thermal conductivity
Q	W/m^3	additional heat sources
$Q_{\text{ted}} = \alpha T : \frac{dS}{dt}$	W/m^3	thermoelastic damping term
α	$1/\text{K}$	thermal expansion coefficient
S	Pa	Piola–Kirchhoff stress tensor
$d/dt = \partial/\partial t + u \cdot \nabla$	$1/s$	Material derivative

and $:$ stands for double-dot tensor product (double contraction).

A free triangular fine mesh was employed in the simulations and is illustrated in Figure 2.

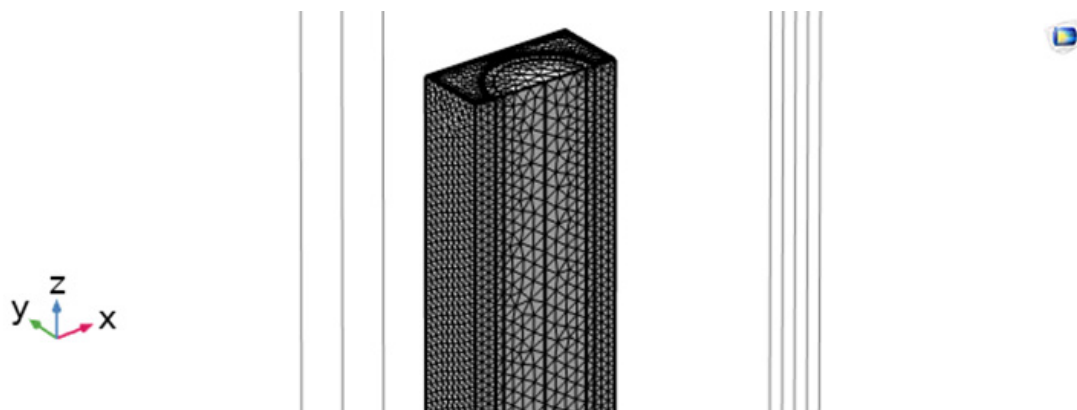


Figure 2 – Mesh of the single pint, from inward:
1 – fuel rod, 2 – cladding and 3 – coolant

With the thermal power applied as a heat source over the cross-sectional surface of the core, the equation (1) is solved using the finite element methods to determine the spatial temperature distribution within the fuel, cladding, and coolant regions. Heat generation was applied only in the fuel domain, while heat transfer through the fuel–cladding and cladding–coolant interfaces was solved by the coupled heat-transfer and fluid-flow physics. All boundaries are assumed to be thermally insulated, except for the top, bottom, and cross-sectional surfaces and are explicitly formulated as follows:

$$-\mathbf{n} \cdot \mathbf{q} = 0, \quad (2)$$

where $\mathbf{n}$ is a vector normal to the surface.

The adiabatic boundary condition was applied only to the external lateral boundaries of the computational domain to avoid artificial heat losses from the selected fuel-rod/coolant subdomain. This assumption represents a simplified local-domain approximation and does not suppress heat transfer between the fuel, cladding, and coolant. Convective heat removal was modeled explicitly through the coupled heat-transfer and turbulent-flow interfaces in the coolant region, with inlet and outlet boundaries imposed at the bottom and top surfaces, respectively

Thermal properties of fuel [32] and clad materials [33] as temperature-dependent expressions are listed in Table 3.

Table 3 – Thermodynamical properties of fuel and cladding as temperature-dependent polynomials

	$(\text{Th}, {}^{233}\text{U})\text{O}_2$	Zircaloy-4
C_p	$239.9 + 8.512 \times 10^{-5} \cdot T$ $-5.754 \times 10^{-8} \cdot T^2$ $+1.528 \times 10^{-11} T^3$ [J/(mol · K)]	$7.1 \cdot 10^{-2}$ $+1.7 \cdot 10^{-5} \cdot T$ $-0.89 \cdot 10^{-3} \cdot T^2$ [cal/g · K]
k	14.242 $-0.022 \cdot T$ $+1.764 \times 10^{-5} \cdot T^2$ $-6.380 \times 10^{-9} \cdot T^3$ $+8.038 \times 10^{-13} \cdot T^4$ [W/(m · K)]	0.11804 $+0.00012 \cdot T$ $-4.33441 \cdot 10^{-8} \cdot T^2$ $+5.90277 \times 10^{-11} \cdot T^3$ [W/(cm · °C)]
ρ	9326 $-0.0004 \cdot T$ $+2.54 \times 10^{-7} \cdot T^2$ $-7.897 \times 10^{-11} \cdot T^3$ [kg/m ³]	10950 $-0.35 \cdot T$ $-1.2 \cdot 10^{-4} \cdot T^2$ [kg/m ³]

The Navier–Stokes PDE governing momentum and mass conservation are solved to obtain the coolant pressure along the axial (z) direction as well as the corresponding velocity field – see Equation 3:

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \nabla \cdot \boldsymbol{\sigma} + \mathbf{F}_v, \quad (3)$$

where equation parameters are listed in Table 4.

Table 4 – Navier-Stokes equation parameters

Parameter	Unit	Definition
ρ	kg/m ³	Mass density
$\mathbf{u}$	J/(kg · K)	Displacement field
$\boldsymbol{\sigma}$	Pa	Cauchy stress tensor
$\mathbf{F}_v$	N	Body force

When the turbulent flow (k– ω) model is coupled with the heat-transfer physics, two key quantities are determined: (i) the convective heat-transfer coefficient h , which depends on the coolant velocity, channel geometry, and temperature-dependent coolant properties, and (ii) the axial pressure

drop resulting from the Navier–Stokes solution. The coolant flow was prescribed through the axial direction of the fuel-rod channel. The bottom surface of the coolant domain was defined as the inlet, where the inlet velocity and coolant temperature were specified. The top surface was defined as the outlet, where pressure was fixed to the reference outlet pressure. No-slip wall conditions were applied at the solid–fluid interfaces.

Coupled equation determines influence of thermal expansion on material’s strain:

$$\varepsilon_{th} = \alpha(T)(T - T_{ref}) \quad (4)$$

Within the solid mechanics interface, the governing elasticity equations are solved to evaluate thermally induced deformation, displacement, and strain in the fuel and cladding, fully coupled with the heat-transfer physics. The fuel and cladding domains were coupled to the temperature field obtained from the heat-transfer calculation. Thermally induced expansion was included through the temperature-dependent thermal expansion coefficient. The top and bottom boundaries of the solid domain were mechanically constrained to represent axial support of the fuel rod, while the remaining surfaces were allowed to deform freely unless otherwise specified. This setup allows evaluation of thermally induced displacement, volumetric strain, and von Mises stress under constrained axial conditions

For mechanical properties, the following thermal expansion expressions for fuel [34] and clad [32], presented in Table 5, were used:

Table 5 – Mechanical properties of fuel and cladding as temperature-dependent polynomials

	(Th, ^{233}U)O ₂	Zircaloy-4
α	$5.416 \cdot 10^{-6} + 3.47 \cdot 10^{-9} \cdot T + 3.0107 \cdot 10^{-15} \cdot T^2 - 1.2212 \cdot 10^{-15} \cdot T^3 + 4.181 \cdot 10^{-19} \cdot T^4$ [1/K]	$(5.22 + 1.82 \cdot 10^{-3} \cdot T) \cdot 10^{-6}$ [1/K]
E	$259.87 \cdot 10^9 \cdot (1 - P^2)$ [Pa]	$9.21 \cdot 10^{10} - 4.05 \cdot 10^7 \cdot T$ [Pa]

where P is the porosity of material.

3. Results and Discussion

3.1. Effective neutron multiplication coefficient, heat power distribution and peak power density

The maximum normalized heating density value has been obtained:

$$\omega = 4109.2892 \text{ W/cm}^3.$$

In this work, the hottest fuel rod was selected based on the OpenMC volumetric power distribution, and the maximum normalized heating volumetric density value was used in COMSOL to represent the most limiting thermal-hydraulic and thermo-mechanical condition. The maximum value was selected because it corresponds to the hottest fuel rod and it represents the most safety-critical location in the core, where the highest temperature, thermal stress, strain, and displacement are expected.

Effective multiplication coefficient $k_{eff} = 1.01777 \pm 0.00005$. Normalized neutron spectrum of the model in comparison with SHEM-361 is presented in Figure 3. The values of the flux are significantly higher for the model than for the reference group. This is explained by a multitude of factors, despite SHEM-361 being suitable for comparison of various LWR reactors: geometry or tally

constraints – as tallies for the calculations were constrained only in the fuel domain, and energy-group mismatch – since for the purposes of this paper the energy was divided into smaller bins, getting 500-group energy spectrum.

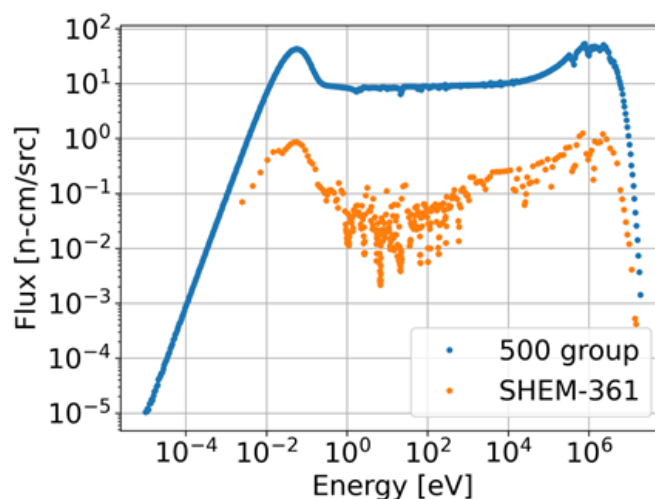


Figure 3 – Neutron spectrum of a Design-Basis Core (blue), compared with SHEM-361 energy group (orange)

Neutron flux and heat distribution, normalized on a 900 kW reactor output, are presented in Figure 4. The temperature distribution in Figure 4 (b) is concentrated on the fuel pins, with the coldest ones in the center of the assembly, confirming that heat is generated only in the fuel zones. The relatively colder pins are located in the center of the core and the heat increase is shown to have a radially outward direction, where at the edges the hottest pins can be seen. This trend can be observed in [35, 36], while other studies show exactly the inverse behavior [12, 16].

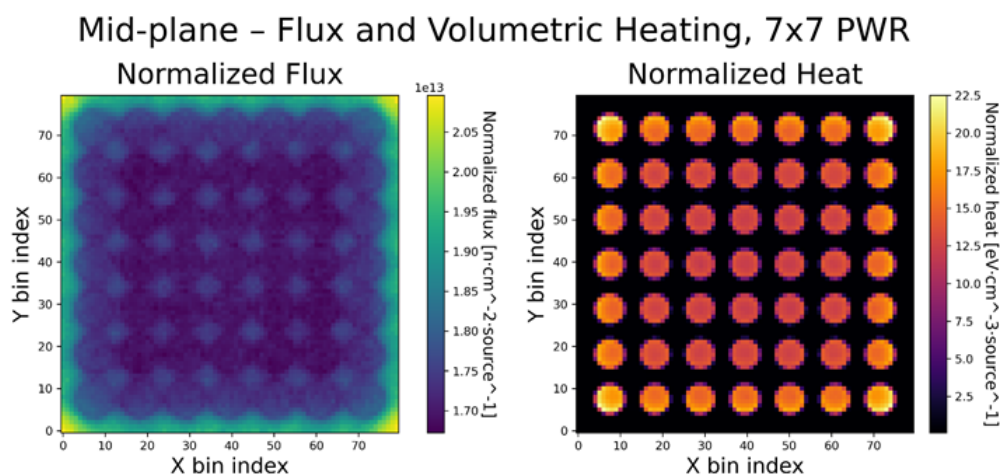


Figure 4 – Normalized flux (a) and heat (b) distribution of the core assembly.
Scale is as follows: colder colors – lower flux/temperature

3.2. Thermal-hydraulics, solid mechanics and fuel safety

3.2.1. Temperature

Figures 5 (a) and (b) present the temperature contours in the yx plane at 0.2 s and 0.8 s, respectively, following the application of the maximum volumetric heat source in the fuel. At 0.2 s, the highest temperatures are confined to the fuel region, while a clear radial temperature gradient develops across

the fuel-cladding interface and into the surrounding coolant. As time increases to 0.8 s, the heated zone expands radially, and the temperature increases throughout the fuel and cladding and reaching a maximum, indicating enhanced heat diffusion and thermal penetration into the adjacent materials. Figure 4 (c) shows the temperature profiles along the fuel rod for different time periods (0–0.8 s). The profiles reveal a relatively uniform temperature within the fuel at early times, followed by a gradual decrease across the cladding and a sharp temperature drop within the coolant region. The normalized neutron flux and volumetric heat distributions for the 7×7 PWR assembly are presented in Figure 4. The thermal field is primarily concentrated within the fuel regions, confirming that volumetric heat generation occurs exclusively inside the fuel pins. The radial heat distribution reveals comparatively lower temperatures in the central region of the assembly and progressively increasing temperatures toward the peripheral fuel pins, where the maximum thermal intensity is observed. This behavior is associated with the spatial neutron-flux distribution and geometric characteristics of the assembly. Similar thermal trends have been reported in previous studies of thorium-based and conventional fuel systems, although several investigations have demonstrated inverse radial distributions depending on fuel composition and reactor configuration [35].

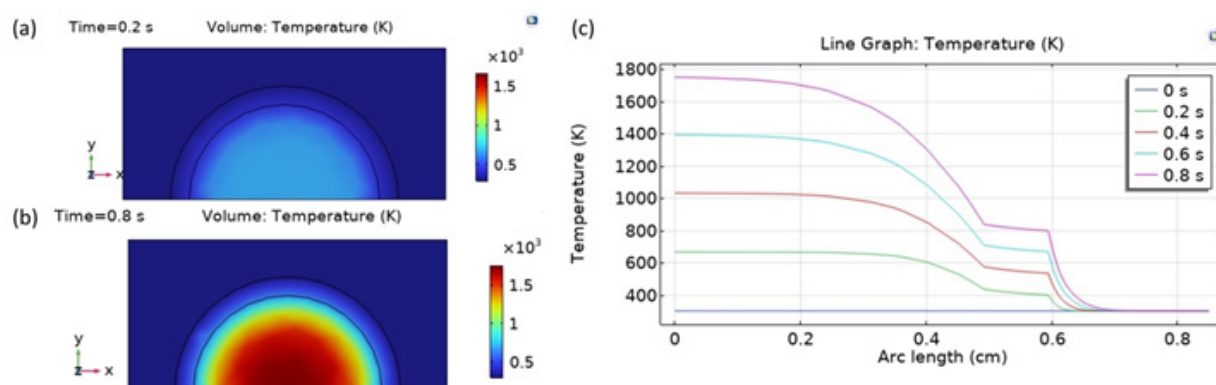


Figure 5 – The temperature profile along the y-x direction at (a) 0.2s and (b) 0.8s; (c) the temperature distribution through different materials of the fuel rod after application of the maximum heat source for different times (from 0 to 0.8 seconds)

3.2.2. Coolant velocity

Figure 6 (a) presents a three-dimensional slice of the velocity magnitude at $t=0.8$ s. The color map indicates a predominantly uniform flow along the axial direction, with higher velocities concentrated toward the central region of the channel and lower velocities near the walls. This distribution reflects the combined effects of channel geometry and viscous boundary layers, resulting in a developed velocity profile along the flow direction. Figure 6 (b) shows the corresponding velocity magnitude along the rod length at two different times, 0.2 s and 0.8 s. At 0.2 s, the velocity profile remains relatively flat, indicating that the flow is still developing. As time increases to 0.8 s, the velocity increases along the channel length, with a noticeable rise near the outlet region up to 1.1 m/s. The higher velocities observed at later times suggest flow acceleration due to thermal effects and coupling with heat transfer, while the sharp increase near the end of the channel highlights the outlet and boundary influences. However, no variation in pressure is observed over the 0.8 s interval, despite the increase in heat sources. Heat removal from the cladding occurs through the surrounding coolant domain. In the coolant, the heat transfer is governed by the coupled heat-transfer and fluid-flow physics. Therefore, the coolant removes the heat mainly by forced convection along the axial direction. The calculated velocity field reflects the imposed inlet/outlet flow conditions and the development of the coolant flow inside the channel. Thus, the temperature distribution results from the combined effect of internal fuel heat generation, conductive heat transfer through the solid fuel and cladding, and convective heat removal by the coolant. A nearly unchanged pressure indicates that, under the present modeling assumptions, the coolant flow remains stable during the short heating transient.

However, a more realistic assessment of pressure evolution would require a full-assembly or system-level thermal-hydraulic model with realistic inlet/outlet pressure losses, pump characteristics, and feedback between coolant heating, density variation, and flow resistance.

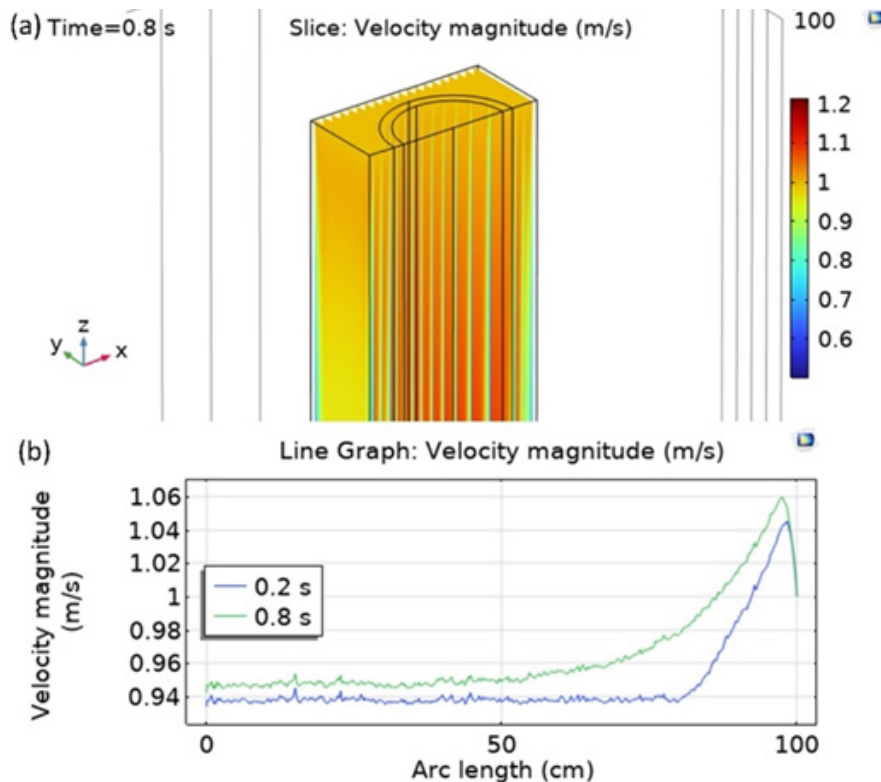


Figure 6 – (a) 3D slice of the velocity magnitude at $t=0.8$ s;
(b) corresponding velocity magnitude along the rod at $t=0.2$ s and $t=0.8$ s.

3.2.3. Mechanical stresses

Figures 7 (a) and (b) show the spatial distribution of von Mises stress at 0.2 s and 0.8 s, respectively. As time increases to 0.8 s, the magnitude of von Mises stress increases significantly throughout the active zone, reflecting the development of thermally induced stresses due to temperature gradients and constrained thermal expansion. Higher stress concentrations are observed along the axial direction and near geometrical or material transition regions, where mechanical constraints and thermal gradients are more pronounced. Figure 7 (c) shows the corresponding von Mises stress profiles along the rod length for both time instants. The stress remains nearly constant along most of the active zone, indicating a relatively uniform mechanical loading, while a noticeable increase in stress magnitude is observed at later times. The higher stress levels confirm the cumulative effect of thermal loading over time.

Figures 8 (a) and (b) present the spatial distribution of volumetric strain at 0.2 s and 0.8 s, respectively. At 0.2 s, the strain field is already dominated by positive volumetric strain, indicating thermal expansion of the material as heat is deposited in the active zone. The highest strain values are concentrated in the central region, while lower strain magnitudes appear near the outer region, reflecting the presence of mechanical constraints and cooler surrounding regions. At 0.8 s the magnitude of volumetric strain increases noticeably throughout the active zone, demonstrating the cumulative effect of continued heating. The expansion becomes more pronounced and spatially extended, while the overall strain pattern remains similar, suggesting a stable deformation mode governed by the temperature distribution rather than abrupt mechanical instabilities. Figure 8 (c) shows the volumetric strain profiles along the rod length for both time instants. The strain remains relatively uniform along most of the active zone, with higher values observed at 0.8 s compared

to 0.2 s. Minor fluctuations along the arc length may be attributed to geometric features, material interfaces, or numerical discretization effects. A small reduction near the boundaries indicates the influence of mechanical constraints imposed at the top and bottom surfaces. An increased volumetric strain in the fuel compared to the cladding material was also observed in the literature [36].

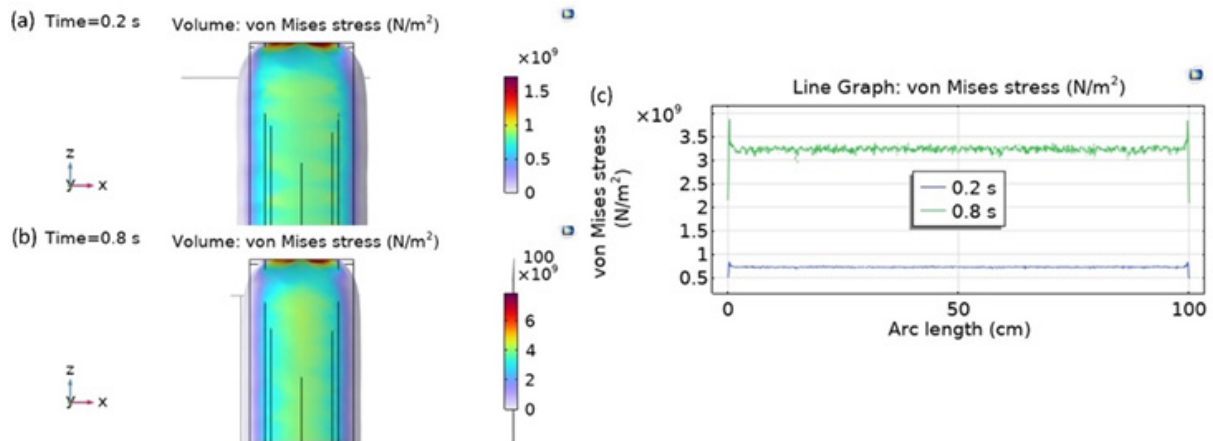


Figure 7 – Spatial distribution of von Mises stress at (a) 0.2 s and (b) 0.8 s, respectively; (c) the corresponding von Mises stress profiles along the rod length for both time instants

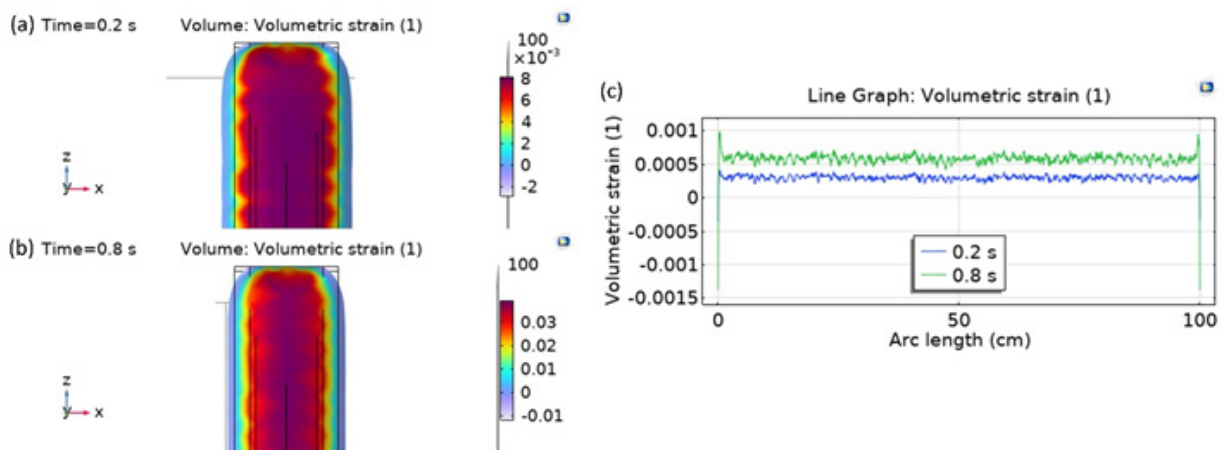


Figure 8 – Spatial distribution of volumetric strain at (a) 0.2 s and (b) 0.8 s, respectively; (c) volumetric strain profiles along the rod length for both time instants

Figures 9 (a) and (b) show the spatial distribution of the displacement magnitude at 0.2 s and 0.8 s, respectively. At the early time of 0.2 s, the displacement is relatively small throughout the structure, indicating the initial stage of thermal expansion following heat deposition with. The deformation is distributed smoothly along the axial direction, with minimal distortion near the boundaries due to mechanical constraints. At 0.8 s, the displacement magnitude increases significantly, reflecting the cumulative effect of sustained heating. The deformation becomes more pronounced in the central region of the active zone, while lower displacement values are observed near the top and bottom boundaries, consistent with the imposed structural constraints. The overall displacement pattern remains symmetric, suggesting a stable and uniform thermo-mechanical response without localized buckling or instability. Figure 9 (c) presents the displacement magnitude along the rod length for both time instants. The displacement increases gradually from the boundaries toward the mid-span, reaching a maximum near the center of the core, and then decreases toward the opposite boundary. The substantially higher displacement at 0.8 s compared to 0.2 s highlights the strong time dependence

of thermal expansion. The displacement profile suggests non-uniform deformation along the radial length, which is influenced by local thermal and mechanical loading distributions [37].

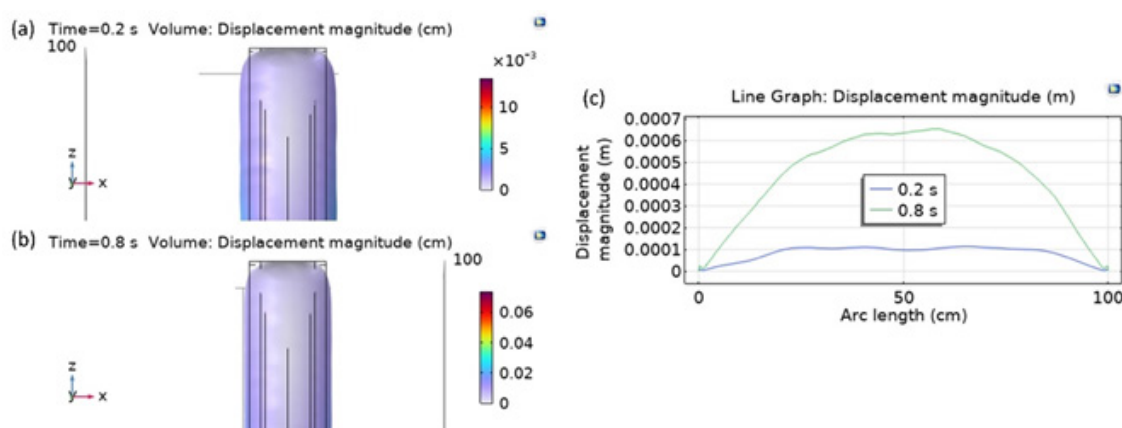


Figure 9 – Spatial distribution of the displacement magnitude at (a) 0.2 s and (b) 0.8 s, respectively; (c) displacement magnitude along the rod length for both time instants

Figures 10 (a), (b) and (c) show the von Mises stress, strain and displacement magnitude in the cladding. At 0.2 s, the stress level and displacement remain relatively low and fairly uniform along the rod length, indicating the initial response to thermal loading. As time increases to 0.8 s, the von Mises stress rises noticeably across the entire cladding, reflecting the accumulation of thermally induced stresses due to temperature gradients and constrained expansion. Slight variations near the boundaries may be attributed to geometric effects and boundary conditions, but overall, the stress distribution remains smooth, suggesting the absence of localized stress concentrations. At 0.8 s, the displacement magnitude increases significantly, reaching a maximum near the center of the rod length and decreasing toward the constrained boundaries. This symmetric displacement profile reflects thermally driven expansion constrained at the ends, leading to maximum deformation in the central region. Positive volumetric strain is observed at both time instants, indicating thermal expansion of the cladding material. The strain magnitude is consistently higher at 0.8 s than at 0.2 s, demonstrating progressive expansion as heating continues. The relatively uniform strain profile along the arc length implies that the cladding undergoes largely homogeneous volumetric deformation, with only minor fluctuations associated with numerical discretization or local thermal variations.

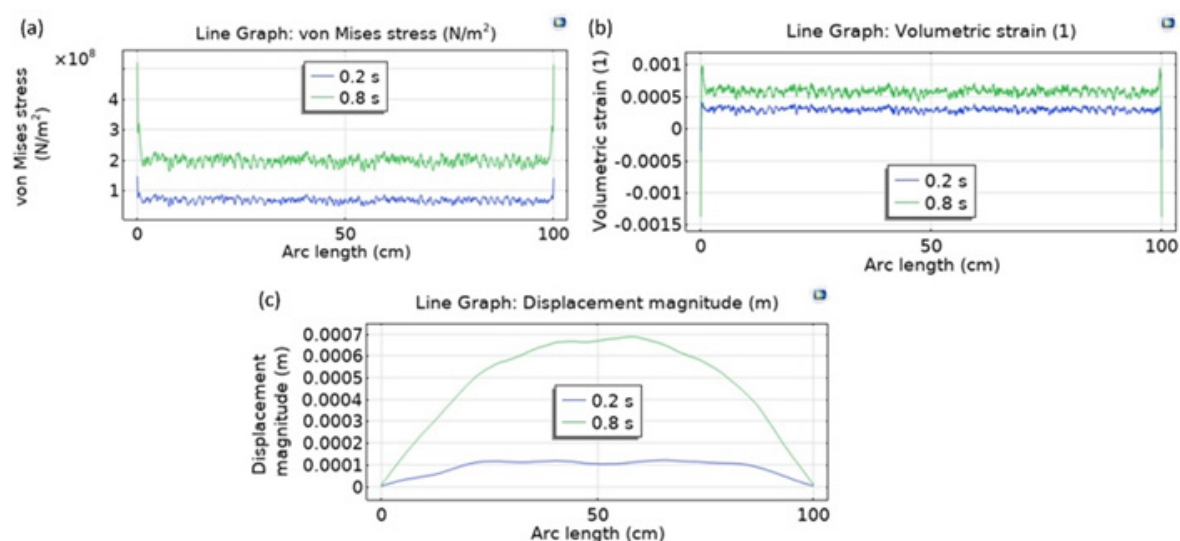


Figure 10 – (a) von Mises stress, (b) strain and (c) displacement magnitude in the cladding at 0.2 s and 0.8 s.

Unusually high values observed in temperature and stresses evaluation, while conducting a safety analysis, indicate a very high risk in designing the PWR with this material composition and power distribution. Unless comprehensive analysis has been conducted, it is early to judge the possibility of replacing uranium fuel with thorium in PWR cores. The analysis, according to the authors' opinions, should include realistic power output of the design, more detailed material composition and, possibly, feedback loop between the coupled physics processes.

4. Conclusions

A coupled neutronic, thermal-hydraulic, and thermo-mechanical analysis of a thorium-based PWR fuel assembly was successfully performed using OpenMC and COMSOL Multiphysics. The neutronic calculations provided the effective neutron multiplication factor and spatial power distribution required for transient thermal-hydraulic simulations. The obtained results demonstrated progressive temperature accumulation within the fuel region, accompanied by the development of thermal gradients, thermally induced stresses, volumetric strain, and structural displacement during transient operation.

The study confirms that coupling steady-state neutronics with transient thermal-hydraulic analysis represents a computationally efficient approach for multiphysics reactor simulations while preserving the dominant physical interactions governing reactor behavior. At the same time, the elevated temperature and stress values obtained during short transient periods indicate the necessity for further refinement of the current model, including more realistic power distributions, improved thermal-hydraulic boundary conditions, and incorporation of additional physics feedback mechanisms.

Future work will focus on extending the model to full-core reactor configurations and implementing more advanced coupled multiphysics approaches to improve the realism and predictive capability of thorium-based PWR safety analyses.

Acknowledgment. The article includes the results of research carried out within the framework of project-targeted funding under the project IRN BR24993225 funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

REFERENCES

- 1 World Nuclear Association. Thorium. URL: <https://world-nuclear.org/information-library/current-and-future-generation/thorium>
- 2 Oak Ridge National Laboratory. Molten-Salt Reactor Experiment 1965–1972 (Oak Ridge, TN: ORNL, 2015). URL: https://web.archive.org/web/20160303211133if_/https://dl.dropboxusercontent.com/u/15726934/Historic_Molten_Salt_Reactor_Experiment_Brochure_ORNL_1965-1972.pdf
- 3 Carter, L.L. Monte Carlo Code Development in Los Alamos. Los Alamos Scientific Laboratory Report (March 1975).
- 4 James, M.R. MCNPX 2.7.x – New Features Being Developed. Los Alamos National Laboratory Technical Report.
- 5 Romano, P.K., Horelik, N.E., Herman, B.R., Nelson, A.G., Forget, B., and Smith, K. OpenMC: A State-of-the-Art Monte Carlo Code for Research and Development. *Annals of Nuclear Energy*, 82, 90–97 (2015). <https://doi.org/10.1016/j.anucene.2014.07.048>
- 6 Brown, D.A., Chadwick, M.B., Capote, R., Kahler, A.C., Trkov, A., Herman, M.W., Sonzogni, A.A., Danon, Y., Carlson, A.D., Dunn, M., et al. ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-Project Cross Sections, New Standards and Thermal Scattering Data. *Nuclear Data Sheets*, 148, 1–142 (2018). <https://doi.org/10.1016/j.nds.2018.02.001>

7 Plompen, A.J.M., Cabellos, O., De Saint Jean, C., Fleming, M., Algora, A., Angelone, M., Archier, P., et al. The Joint Evaluated Fission and Fusion Nuclear Data Library, JEFF-3.3. *European Physical Journal A*, 56 (7), 181 (2020). <https://doi.org/10.1140/epja/s10050-020-00141-9>

8 Tsige-Tamirat, H. Neutronics Assessment of the Use of Thorium Fuels in Current Pressurized Water Reactors. *Progress in Nuclear Energy*, 53 (6), 717–721 (2011). <https://doi.org/10.1016/j.pnucene.2010.12.001>

9 Trellue, H.R., Bathke, C.G., and Sadasivan, P. Neutronics and Material Attractiveness for PWR Thorium Systems Using Monte Carlo Techniques. *Progress in Nuclear Energy*, 53 (6), 698–707 (2011). <https://doi.org/10.1016/j.pnucene.2011.02.002>

10 Chaudri, K.S., Tian, W., Su, G., and Qiu, S. Coupled Neutronics/Thermal Hydraulics Evaluation for Thorium Based Fuels in Thermal Spectrum SCWR. *Progress in Nuclear Energy*, 68, 55–64 (2013). <https://doi.org/10.1016/j.pnucene.2013.04.005>

11 Arzhannikov, A.V., Shmakov, V.M., Modestov, D.G., Bedenko, S.V., Prikhodko, V.V., Lutsik, I.O., and Shamanin, I.V. Facility to Study Neutronic Properties of a Hybrid Thorium Reactor with a Source of Thermonuclear Neutrons Based on a Magnetic Trap. *Nuclear Engineering and Technology*, 52 (11), 2460–2470 (2020). <https://doi.org/10.1016/j.net.2020.04.017>

12 Kumar, E.R.S.A., Pancholi, M.K., Darnowski, P., and Dzido, A. Neutronic Performance of a Thorium Based Mixed Oxide Fuel in a Burner Sodium-Cooled Fast Reactor. *Energy*, 212, 118744 (2020). <https://doi.org/10.1016/j.energy.2020.118744>

13 Galahom, A.A., Mohsen, M.Y., and Amrani, N. Explore the Possible Advantages of Using Thorium-Based Fuel in a Pressurized Water Reactor (PWR). Part 1: Neutronic Analysis. *Nuclear Engineering and Technology*, 54 (1), 1–10 (2022). <https://doi.org/10.1016/j.net.2021.07.012>

14 Li, J., Li, X., and Cai, J. Neutronic Characteristics and Feasibility Analysis of Micro-Heterogeneous Duplex ThO₂-UO₂ Fuel Pin in PWR. *Nuclear Engineering and Design*, 382, 111382 (2021). <https://doi.org/10.1016/j.nucengdes.2021.111382>

15 Kabach, O., Chetaine, A., Benchrif, A., Amsil, H., and El Banni, F. A Comparative Analysis of the Neutronic Performance of Thorium Mixed with Uranium or Plutonium in a High-Temperature Pebble-Bed Reactor. *International Journal of Energy Research*, 45 (11), 16824–16841 (2021). <https://doi.org/10.1002/er.6863>

16 Mirvakili, S.M., Kavafshary, M.A., and Vaziri, A.J. Comparison of Neutronic Behavior of UO₂, (Th-233U)O₂ and (Th-235U)O₂ Fuels in a Typical Heavy Water Reactor. *Nuclear Engineering and Technology*, 47 (3), 315–322 (2015). <https://doi.org/10.1016/j.net.2014.12.005>

17 Tucker, L.P., Alajo, A., and Usman, S. Thorium-Based Mixed Oxide Fuel in a Pressurized Water Reactor: A Beginning-of-Life Feasibility Analysis with MCNP. *Annals of Nuclear Energy*, 76, 323–334 (2015). <https://doi.org/10.1016/j.anucene.2014.10.020>

18 Pillai, C.G.S., and Raj, P. Thermal Conductivity of ThO₂ and Th_{0.98}U_{0.02}O₂. *Journal of Nuclear Materials*, 277 (1), 116–119 (2000). [https://doi.org/10.1016/S0022-3115\(99\)00237-8](https://doi.org/10.1016/S0022-3115(99)00237-8)

19 Kutty, T.R.G., Kulkarni, R.V., Sengupta, P., Khan, K.B., Bhanumurthy, K., Sengupta, A.K., Panakkal, J.P., Kumar, A., and Kamath, H.S. Development of CAP Process for Fabrication of ThO₂-UO₂ Fuels. Part II: Characterization and Property Evaluation. *Journal of Nuclear Materials*, 373 (1–3), 309–318 (2008). <https://doi.org/10.1016/j.jnucmat.2007.06.011>

20 Bakker, K., Cordfunke, E.H.P., Konings, R.J.M., and Schram, R.P.C. Critical Evaluation of the Thermal Properties of ThO₂ and Th_{1-y}U_yO₂ and a Survey of the Literature Data on Th_{1-y}Pu_yO₂. *Journal of Nuclear Materials*, 250 (1), 1–12 (1997). [https://doi.org/10.1016/S0022-3115\(97\)00241-9](https://doi.org/10.1016/S0022-3115(97)00241-9)

21 Tyagi, A.K., Mathews, M.D., Ambekar, B.R., and Ramachandran, R. Thermal Expansion of ThO₂-2, 4 and 6 wt.% UO₂ by HT-XRD. *Thermochimica Acta*, 421 (1–2), 69–71 (2004). <https://doi.org/10.1016/j.tca.2004.04.025>

22 Tyagi, A.K., and Mathews, M.D. Thermal Expansion of ThO₂-2 wt.% UO₂ by HT-XRD. *Journal of Nuclear Materials*, 278 (1), 123–125 (2000). [https://doi.org/10.1016/S0022-3115\(99\)00277-9](https://doi.org/10.1016/S0022-3115(99)00277-9)

23 Anthonysamy, S., Panneerselvam, G., Bera, S., Narasimhan, S.V., and Vasudeva Rao, P.R. Studies on Thermal Expansion and XPS of Urania-Thoria Solid Solutions. *Journal of Nuclear Materials*, 281 (2–3), 308–316 (2000). [https://doi.org/10.1016/S0022-3115\(00\)00185-9](https://doi.org/10.1016/S0022-3115(00)00185-9)

- 24 Ghosh, P.S., Somayajulu, P.S., Krishnan, K., Pathak, N., Arya, A., and Dey, G.K. Thermal Expansion and Thermal Conductivity of (Th,U)O₂ Mixed Oxides: A Molecular Dynamics and Experimental Study. *Journal of Alloys and Compounds*, 650, 165–177 (2015). <https://doi.org/10.1016/j.jallcom.2015.07.260>
- 25 Xiao, H., Long, C., Tian, X., Chen, H., and Duan, X. Effect of Thorium Addition on the Thermophysical Properties of Uranium Dioxide: Atomistic Simulations. *Materials & Design*, 96, 335–340 (2016). <https://doi.org/10.1016/j.matdes.2016.02.019>
- 26 Griggs, D.P., Henry, A.F., and Kazimi, M.S. Development of a Three-Dimensional Two-Fluid Code with Transient Neutronic Feedback for LWR Applications. Massachusetts Institute of Technology Technical Report (1981).
- 27 Mahadevan, V.S. High Resolution Numerical Methods for Coupled Non-Linear Multi-Physics Simulations with Applications in Reactor Analysis. PhD Dissertation, Texas A&M University, College Station, TX, USA (2010).
- 28 Pijls, T. Developing a GPU-Accelerated Filter-Matrix Lattice Boltzmann Multiphysics Tool for the Transient Thermal-Hydraulics and Neutronics of a Molten Salt Fast Reactor Core. MSc Thesis, Delft University of Technology, Delft, The Netherlands (2025). URL: https://martinrohde.nl/theses/msc/MSc_Thomas_Pijls.pdf
- 29 Hanumantharao, G., Vijay, S., and Venkateswara Rao, M. Transient Heat Transfer Analysis for Optimum Temperature Distribution to Reduce Thermal Stresses. *International Journal of Engineering Research and Technology*, 1, 1–7 (2012).
- 30 Fuertes-Miquel, V.S. Transient Flows: Mathematical Models, Laboratory Tests, Protection Elements and Systems. *Water*, 17, 1454 (2025). <https://doi.org/10.3390/w17111454>
- 31 Hu, R., Fang, J., Nunez, D., Tano, M., Giudicelli, G., and Salko, R. Development of Integrated Thermal Fluids Modeling Capability for MSRs. Argonne National Laboratory Report ANL/NSE-22/56 (2022).
- 32 Mohsen, M.Y.M., Abdel-Rahman, M.A.E., and Galahom, A.A. Ensuring the Possibility of Using Thorium as a Fuel in a Pressurized Water Reactor (PWR). *Nuclear Science and Techniques*, 32, 137 (2021). <https://doi.org/10.1007/s41365-021-00981-0>
- 33 Hammer, R.R. Zircaloy-4, Uranium Dioxide, and Materials Formed by Their Interaction: A Literature Review with Extrapolation of Physical Properties to High Temperatures. Idaho Nuclear Corporation Report IN-1093 (Idaho Falls, ID, USA, 1967).
- 34 Shields, A.E., Hernandez, S.E.R., and de Leeuw, N.H. Theoretical Analysis of Uranium-Doped Thorium Dioxide: Introduction of a Thoria Force Field with Explicit Polarization. *AIP Advances*, 5, 087118 (2015). <https://doi.org/10.1063/1.4928438>
- 35 Jiao, G., Xia, G., Zhu, H., Zhou, T., and Peng, M. Thermal-Mechanical Coupling Characteristics and Heat Pipe Failure Analysis of Heat Pipe Cooled Space Reactor. *Annals of Nuclear Energy*, 192, 110025 (2023). <https://doi.org/10.1016/j.anucene.2023.110025>
- 36 Mohsen, M.Y.M., Soliman, A.Y., and Abdel-Rahman, M.A.E. Thermal-Hydraulic and Solid Mechanics Safety Analysis for VVER-1000 Reactor Using Analytical and CFD Approaches. *Progress in Nuclear Energy*, 130, 103568 (2020). <https://doi.org/10.1016/j.pnucene.2020.103568>
- 37 Mohsen, M.Y.M., Abdel-Rahman, M.A.E., Hassan, M.S., and Galahom, A.A.G. Searching for the Most Optimum Burnable Absorbers (BAs) for AP-1000 from the Neutronic, Thermal-Hydraulic, and Solid Mechanics Points of View. *Nuclear Engineering and Design*, 391, 111728 (2022). <https://doi.org/10.1016/j.nucengdes.2022.111728>

¹**Арынгазин А.,**

магистр, ORCID ID: 0009-0007-7861-4917,

e-mail: ansar.aryngazin@nu.edu.kz

¹**Курбанова Б.,**

магистр, ORCID ID: 0000-0003-0928-521X,

e-mail: bayan.kurbanova@nu.edu.kz

²**Сизюк Ю.,**

PhD, e-mail: yuriy.sizyuk@nu.edu.kz

³**Темирбаева А.,**

магистр, ORCID ID: 0000-0002-9526-3617,

e-mail: assem.zkgmu@gmail.com

¹**Алсар Ж.,**

PhD, ORCID ID: 0000-0001-7287-7555,

e-mail: zhanna.alsar@nu.edu.kz

^{1*}**Инсепов З.,**

PhD, жетекші зерттеуші, адъюнкт-профессор, ORCID ID: 0000-0002-8079-6293,

*e-mail: zinsepov@nu.edu.kz

¹Назарбаев университетінің зерттеу басқармасы, Астана қ., Қазақстан;

²Механикалық ядролық инженерия кафедрасы, вирджиния достастық университеті,
Вирджиния, АҚШ;

³Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан;

⁴Ядролық инженерия мектебі, Пердью университеті, Индиана, АҚШ.

ТОРИЙ ОТЫНЫ ҚОЛДАНЫЛАТЫН ҚЫСЫМДАҒЫ СУ РЕАКТОРЫНЫҢ НЕЙТРОНДЫҚ-ФИЗИКАЛЫҚ ЖӘНЕ ЖЫЛУ-ГИДРАВЛИКАЛЫҚ СИПАТТАМАЛАРЫН ЗЕРТТЕУ

Аңдатпа

Жұмыста жылу-гидравликалық үдерістердің және қатты дене механикасы құбылыстарының динамикасы COMSOL Multiphysics бағдарламалық кешенінде соңғы элементтер әдісі негізінде модельденді. Нейтрондық-физикалық сипаттамаларды зерттеу үшін OpenMC бағдарламалық пакеті стационарлық нейтроника жуықтауында қолданылды. Торий негізіндегі реактордың белсенді аймағын оңтайландыру мақсатында салқындатқыш ретінде жоғары қысымды жеңіл су пайдаланылды. Белсенді аймақта торий қосылыстарымен толтырылған 49 отын штангасы орналастырылды. OpenMC пакетінде жүзеге асырылған Монте-Карло әдісі арқылы нейтрондардың тасымалдануын модельдеу нәтижесінде қарастырылған реактор үлгісінің тиімді көбейту коэффициенті (k_{eff}) мен қуаттың таралуы анықталды. Жылу-гидравликалық есептеулер нәтижесінде температуралық профильдер мен ағын жылдамдығының мәндері алынды, ал қатты дене механикасы бойынша жүргізілген талдау нәтижесінде кернеулер мен деформациялардың кеңістіктік таралуы анықталды. Алынған нәтижелер стационарлық нейтроника мен өтпелі жылу-гидравликаны байланыстыра есептеу әдісін қолданудың тиімділігін растады.

Түйін сөздер: ядролық реактор физикасы, торий отыны, нейтроника, термиялық гидравлика, қатты дене механикасы, COMSOL, OpenMC.

¹Арынгазин А.,

MSc, ORCID ID: 0009-0007-7861-4917,

e-mail: ansar.aryngazin@nu.edu.kz,

¹Курбанова Б.

PhD, ORCID ID: 0000-0003-0928-521X,

e-mail: bayan.kurbanova@nu.edu.kz,

²Сизюк Ю.,

PhD, e-mail: yuriy.sizyuk@nu.edu.kz,

³Темирбаева А.,

M.Sc. ORCID ID: 0000-0002-9526-3617,

e-mail: assem.zkgmu@gmail.com

¹Алсар Ж.,

PhD, ORCID ID: 0000-0001-7287-7555,

e-mail: zhanna.alsar@nu.edu.kz

^{1*}Инсепов З.,

доктор наук, PhD, адъюнкт-профессор, ORCID ID: 0000-0002-8079-6293,

*e-mail: zinsepov@nu.edu.kz

¹Исследовательская администрация Назарбаевского университета, г. Астана, Казахстан

²Кафедра машиностроения и ядерной инженерии, Университет Содружества Вирджинии,
Вирджиния, США

³Евразийский национальный университет им. Л.Н. Гумилева, г. Астана, Казахстан

⁴Школа ядерной инженерии, Университет Пердью, Индиана, США

ИССЛЕДОВАНИЕ НЕЙТРОНИКИ И ТЕПЛОГИДРАВЛИКИ РЕАКТОРА С ВОДОЙ ПОД ДАВЛЕНИЕМ И ТОПЛИВОМ НА ОСНОВЕ ТОРИЯ

Аннотация

В данной работе динамика тепло-гидравлических явлений, а также явлений механики твердого тела была смоделирована методом конечных элементов с использованием Comsol Multiphysics. Для изучения нейтроники был применен OpenMC в приближении стационарной нейтроники. Для оптимизации активной зоны ториевого реактора в качестве охладителя использовалась легкая вода под высоким давлением. В активной зоне располагались 49 топливных стержней, заполненные соединениями тория, организованные в виде регулярной – решеточной геометрии. Моделирование переноса нейтронов методом Монте-Карло в пакете OpenMC позволило получить эффективное значение коэффициента размножения k_{eff} и пространственное распределение мощности рассматриваемой модели. Были получены детальные тепло- и гидравлические расчеты – температурный и скоростные профили, в то время как механика твердого тела определила пространственное распределение напряжений и деформаций. Результаты подтвердили применимость и эффективность использования метода совмещения стационарной нейтроники с термогидравликой и структурного анализа для оценки и оптимизации ториевых реакторов с водным охлаждением.

Ключевые слова: физика ядерных реакторов, ториевое топливо, нейтронная физика, теплогидравлика, механика твердого тела, COMSOL, OpenMC.

МУНАЙ ГАЗ ИНЖЕНЕРИЯСЫ ЖӘНЕ ГЕОЛОГИЯ
OIL AND GAS ENGINEERING, GEOLOGY
НЕФТЕГАЗОВАЯ ИНЖЕНЕРИЯ И ГЕОЛОГИЯ

UDC 622.271
IRSTI 52.13.15

<https://doi.org/10.55452/1998-6688-2026-23-2-418-434>

^{1*}Abiken B.I.,

Master's Student, ORCID ID: 0009-0003-1343-4748,

*e-mail: be_abiken@kbtu.kz

²Stokes O.,

Bachelor, Geology, ORCID ID: 0009-0000-4110-1543,

e-mail: oliver.stokes@hexagon.com

³Gambal M.,

PhD, Mining Engineer, ORCID ID: 0009-0007-0357-2994,

e-mail: gambalmaxim@gmail.com

¹Tulemissova Zh.S.,

PhD, Associate Professor, ORCID ID: 0000-0003-1803-4535,

e-mail: z.tulemissova@kbtu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

²Hexagon Mining, Perth, Australia

³Leica Geosystems Kazakhstan LLC, Almaty, Kazakhstan

**BLAST MOVEMENT MONITORING IN GRADE CONTROL
OF ZHAIREM MINESITE CONSIDERING ORE LOSSES AND DILUTION**

Abstract

This paper examines the economic efficiency of Blast Movement Monitors (BMMs) in a grade control environment within open-pit mining operations. Blast-induced displacement of ore-waste contacts remain a major constraint on effective grade control in open-pit mining. Unaccounted movement during blasting frequently leads to ore dilution, loss of valuable material, and misclassification, ultimately reducing processing efficiency and economic returns. The study is based on the premise that blast-induced rock mass movement is inherently stochastic and cannot be reliably reproduced using deterministic modelling alone; therefore, high-precision grade control requires direct in-situ measurement. By capturing three-dimensional displacement vectors from multiple monitoring points within a blast block, BMM enables more accurate reconstruction of post-blast ore boundaries. This multi-point approach accounts for spatial variability and provides a robust basis for updating polygons prior to excavation. In 2024, the pilot implementation of the BMM system was started at the Zhairem Mining and Concentrating Complex (MCC), and the system is being implemented at several Kazakhstan mining companies. Using field data from the Zhairem site, the study quantifies the operational consequences of uncontrolled blast movement, including ore loss, dilution, and misclassification. The results indicate that improved boundary delineation enhances resource utilisation and contributes directly to increased profitability. Based on these findings, the paper proposes a financial

value assessment tool designed to support investment decisions related to blast movement monitoring systems, enabling operations to systematically evaluate the cost-benefit ratio of BMM implementation and optimise grade control strategies within a broader economic framework.

Keywords: Blast Movement Monitors, grade control, ore loss, dilution, misclassification, ROI, Zhairam, Kazzinc, Glencore, ore, open-pit, zinc, barite, mineralisation type, excavation.

Received March 13, 2026; revised May 7, 2026; accepted June 2, 2026.

Introduction

Minesite profits are significantly impacted by throughput and cost efficiency of the comminution process, starting with ‘Drill and Blast’. Fragmenting rock with explosives is substantially cheaper per tonne than size reduction in the plant, so optimising drill-and-blast is one of the largest levers for improving economic performance [1–2]. Blasting outcomes also affect downstream unit operations, affecting excavation, haulage, and mill performance [1–3]. It is therefore essential not only to design effective blasts, but also to measure and understand their impacts on both rock fragmentation and ore-waste spatial relationships.

Grade control aims to maximise ore recovery and minimise waste processing in plant. In open pit operations, ore-waste boundaries are typically defined using resource models based on diamond core and RC drilling, then refined with blast-hole sampling [4–5]. However, the initial step of the comminution chain, production blasting – induces complex, highly variable rock movement that displaces these ore-waste contacts [6]. If the post-blast movement is not accounted for, ore is sent to the waste dump (ore loss), waste is sent to the plant (dilution), or material is routed to suboptimal destinations (misclassification). Case studies from large gold operations report ore misclassification of 3–8% and theoretical ore losses of 9–24%, with multi-million-dollar impacts on project value depending on orebody geometry and blast conditions [6–7].

In parallel with a robust grade control system, BMM data enables the derivation of corrected post-blast dig polygons, reducing ore loss and dilution and improving reconciliation and project profitability [8–9]. Despite demonstrated technical benefits and reported returns on investment, the adoption of BMMs is often constrained by perceptions of cost and operational complexity. This study presents a comprehensive analysis of the economic efficiency of BMM implementation in grade control, quantifying costs and benefits under different geological and operational scenarios and identifying conditions under which BMMs deliver the greatest economic value.

Scientific contribution

This study provides several contributions to the existing body of knowledge in grade control and blast movement analysis:

1. It establishes a direct quantitative link between blast-induced displacement and economic performance, enabling the evaluation of ore loss, dilution, and misclassification in monetary terms.
2. It introduces misclassification as an independent economic category, particularly relevant for polymetallic deposits with multiple processing streams, where conventional approaches tend to aggregate this effect within dilution.
3. It explains a practical value-based framework for decision-making, allowing mining operations to assess the return on investment (ROI) of BMM systems under varying geological and economic conditions.

Materials and methods

The study demonstrates first use of BMM technology in a geologically complex polymetallic deposit in Kazakhstan, where multiple ore types with different economic values and processing routes coexist. This enables the explicit consideration of misclassification as a separate economic category,

which is often overlooked in conventional grade control assessments. Although based on a case study, the proposed workflow provides a practical framework for linking blast-induced displacement with value-based decision-making in open-pit mining.

Understanding the Zhairem Deposit

One of the key technical challenges in contemporary open-pit mining, especially at structurally complex deposits such as the Zhairem Mining and Processing Complex (MCC), is achieving alignment between planned ore grades and the material that ultimately reaches the processing plant. A major source of this mismatch is the uncontrolled movement of ore-waste boundaries and contacts between different ore types during blasting operations. The goal of the Blast Movement Monitoring (BMM) system is to provide a direct means of tracking this displacement.

The core issue lies in the fact that displacement within a blast block is neither uniform nor easily predictable. Field experience indicates that the movement vector varies significantly and is governed by numerous interacting factors, including blast design parameters, geomechanical characteristics of the rock mass, and surrounding rock conditions. This simplified approach does not reflect the spatial heterogeneity of blast-induced displacement, which published studies [10–13] have shown may deviate from average horizontal movement values by as much as $\pm 50\%$. Moreover, displacement direction can vary unpredictably, further limiting the reliability of extrapolation from limited measurements.

The precision requirements of modern grade control, often targeting boundary definition within ± 0.2 m cannot be consistently achieved through predictive simulation alone. Many mine sites have High Precision Systems on excavators to accurately dig to plan. Despite continuous improvements in both hardware and computational tools, blast modelling remains insufficiently representative for detailed contact delineation. This limitation stems primarily from uncertainties in input parameters (such as in-situ rock mass properties) and from the simplifying assumptions embedded in numerical algorithms, which cannot fully reproduce the complex, dynamic processes occurring during full-scale blasting. Practical observations repeatedly demonstrate a marked discrepancy between modelled forecasts and displacement data obtained from BMM systems.

Hexagon's Blast Movement Monitoring (BMM) system comprises:

- ◆ Blast Movement Monitors, radio-frequency sensors installed in dedicated holes prior to firing
- ◆ Handheld detector to locate the post-blast BMM positions
- ◆ Software that computes movement vectors and translates ore/waste polygons to their true post-blast locations.

The system is used to reduce ore loss, dilution and misclassification by redefining dig boundaries based on measured movement, thereby increasing ore recovery and economic performance.

Workflow:

- ◆ BMMs are activated and installed in the blast pattern, typically at ore-waste boundaries or other key locations.
- ◆ After blasting, the BMMs are detected on the muckpile surface using a handheld detector.
- ◆ The pre and post-blast positions are compared to generate a movement vector.
- ◆ These vectors are input into software (BMM Explorer), which translates the ore and waste boundaries to their true post-blast locations.
- ◆ Updated dig polygons are provided to the mining team, ensuring that ore is sent to the mill and waste to the dump, as intended.

Value Calculator

The Value Calculator transforms physical movement into economic metrics. For each category, the net value is computed from recoverable metal revenue less mining and processing costs. The calculator uses block-by-block data (tonnage, grades, costs, recovery, prices) to estimate the net value impact for each category.

Definitions:

- ◆ Ore Loss – The rock mass that has moved outside the original ore boundary because of blasting (Figure 1).

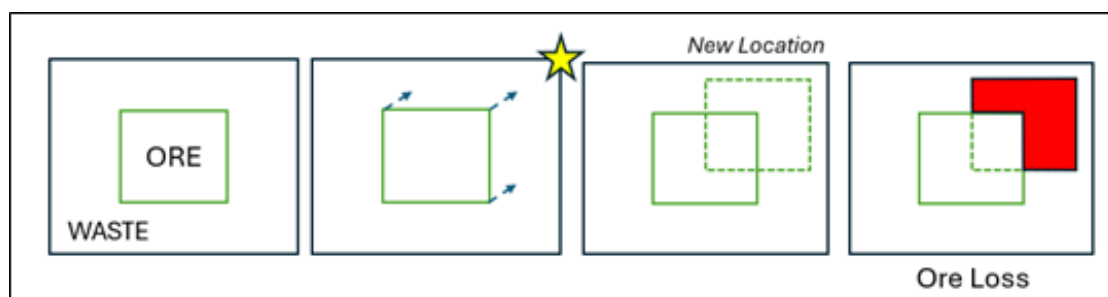


Figure 1 – A visual representaion of 'Ore Loss' (marked red)

♦ Dilution – Waste rock mass that has moved inside the original ore boundary because of blasting. (Figure 2)

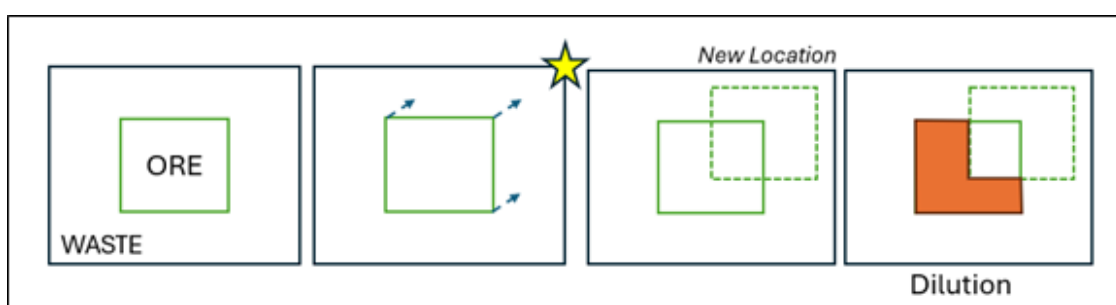


Figure 2 – A visual representation of 'Dilution' (marked brown)

♦ Misclassification – Ore or waste that is incorrectly classified, leading to suboptimal processing or disposal (Figure 3).

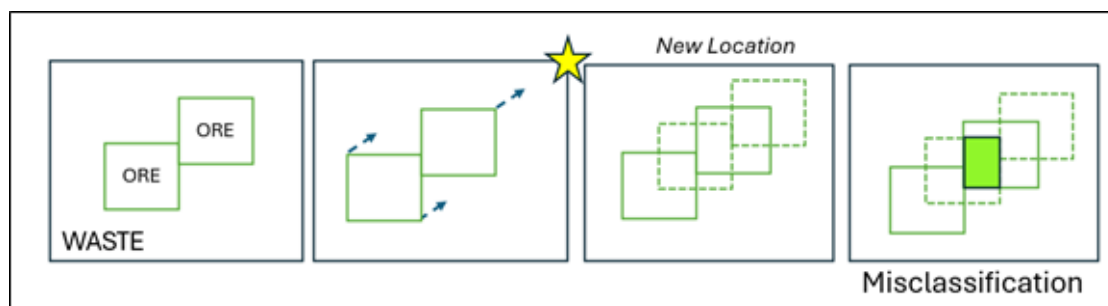


Figure 3 – A visual representation of 'Misclassification' (marked green)

- ♦ Grade (%) – Percentage of valuable material (ore) within the rock mass.
- ♦ Example: 100t of ore @ with a grade of 2 grams/tonne would contain 200 grams of ore.
- ♦ Recovery (%) – Percentage of valuable material recovered during the metallurgical processing stage.
- ♦ Example: A processing plant with 90% efficiency was fed 100t of ore @ with a grade of 2 grams/tonne. Only 90% of the material (180 grams) would be extracted from the rock mass. The remaining 20 grams would become tailings.
- ♦ Processing Cost (\$) – The milling unit cost (USD \$/ton) represents the total expense to process one tonne of ore through the plant. Only ore tonnes are considered.
- ♦ Mining Cost (\$) – The total expense incurred to move one tonne of material (both ore and waste) in the open pit.

Calculating Ore Loss

As mentioned above, Ore Loss is the (high value) ore blocks that have moved outside the original ore boundary during blasting. This material is sent to the waste dump. Ore loss is represented in several ways.

- ♦ Ore Loss (t) – The weight of material that has moved outside of the original ore block.
- ♦ Ore Loss (%) – Tonnage of material outside the original ore block / the total tonnage of the ore block * 100.
- ♦ Ore Loss (\$) – The value of the ore lost is calculated, then the mining and processing cost is subtracted from the total. Even though this material is treated as waste, the processing cost is still attributed to the value of the ore.

$$[(\text{Ore Loss} \times \text{Grade} \times \text{Recovery} \times \text{Price})] - [(\text{Ore Loss} \times \text{Mining Cost}) + (\text{Ore Loss} \times \text{Processing Cost})]$$

Calculating Dilution

When the blast occurs, waste material can move inside the original ore boundary. If site was not to account for blast movement, and mined insitu, the waste material would be processed as ore. Despite having no economic value, it is treated as ore and incurs additional economic harm.

- ♦ Dilution (t) – Weight of waste material (tonnes) that has moved inside the original ore boundary.
- ♦ Dilution (%) – Tonnage of material that has moved inside the original ore boundary / the total tonnage of waste in the blast * 100.
- ♦ Dilution (\$) – The cost to mine and process material that has no economic value.

$$[\text{Dilution (t)} \times (\text{Mining Cost})] + [\text{Dilution (t)} \times (\text{Processing Cost})]$$

Calculating Misclassification

When the blast occurs, ore may move into another ore block. In some cases, this may be okay (i.e. a low value material moving into a high value ore block. However, in other situations this can have a significant financial impact, particularly when these materials are processed different ways. As an example, in iron ore mines, sites can have ore bodies with high sulphur or phosphorus can have significant downstream effects. This can affect the quality of the steel product and may involve penalties for the company for providing an out of spec raw material.

Zhairem Mine Site (Kazzinc/Glencore)

Zhairem is one of the largest polymetallic deposits in Kazakhstan, hosting substantial zinc-lead, barite-polymetallic and iron-manganese ore reserves in the Ulytau Region near the settlement of Zhairem. The mining and processing complex is operated by Kazzinc, in which Glencore International AG is a shareholder. Zhairem forms part of the Atasu ore district and is a major iron-manganese ore field situated within the Zhailma volcano-tectonic depression, a Devonian graben-syncline whose structure and magmatism strongly control ore formation and mining conditions [14]. The operating area includes the Zapadny, Dalnezapadny and Eastern Zhairem deposits, characterised by complex geology and hydrogeology that are critical for exploration strategies, slope stability and mine drainage design [14–15]. Currently, open-pit mining is conducted only at the Dalnezapadny pit [15]. The reserves of the Zhairem deposit recorded on the state balance sheet are classified by mineralisation types:

- ♦ In the Dalnezapadny pit:
- ♦ Oxidised polymetallic type;
- ♦ Sulfide polymetallic type;
- ♦ Oxidised barite-polymetallic type;
- ♦ Sulfide barite-polymetallic type;
- ♦ Barite type.

Table 1 – Mineralisation Types of Dalnezapadny pit

No.	WEATH/FRESH	BaSO ₄	ZnEq	Mineralisation Types	Mineralisation Type Code
1	Fresh	< 40	≥ 1	Sulfide polymetallic	LBS
2	Weathered	≥ 40	≥ 1	Oxidised barite-polymetallic	HBO
3	Weathered	< 40	≥ 1	Oxidised polymetallic	LBO
4	Fresh	≥ 40	≥ 1	Sulfide barite-polymetallic	HBS
5	–	≥ 40	0.01 < ZnEq < 1	Barite	MONOBA
6	otherwise			Waste rock	WASTE

- ♦ Formula of Zinc Equivalent (ZnEq) : $ZnEq = (11.45 * Zn) + (7.32 * Pb) + (0.42 * Ag)$.
- ♦ Grade of Barite : BaSO₄
- ♦ WEATH/FRESH – spatial position of waste/ore bodies relative to the boundary of the weathering (oxidation) zone.

Blast 220-6

Blast 220-6 was a V-pattern shot fired toward a free face with five BMMs installed. Blast Parameters are listed in Table 2. A plan view of the blast is shown in Figure 4, displaying the blast holes, centreline, initiation point (IP) and timing contours. 253 Blast holes were drilled, along with 5 BMMs in dedicated holes. BMMs were installed towards the free face of the blast, on ore/waste boundaries where there is a high chance of ore loss occurring.

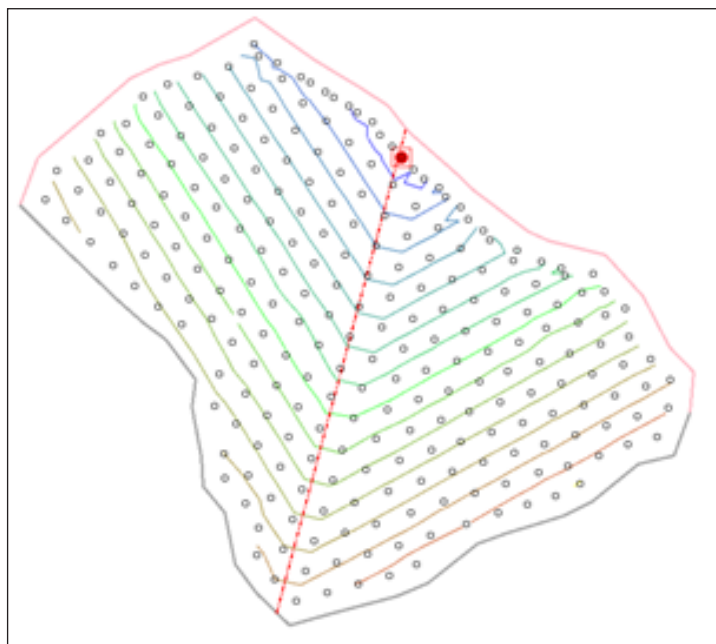


Figure 4 – Blast 220-6 Design

Table 2 – Blast Parameters

Bench Height	8m
Hole Depth	8.6m
Hole Diameter	171mm
Burden	6m

Continuation of table 2

Spacing	7m
Stemming Length	2.9m
Powder Factor	1.11kg/m ³
Explosive Type	Water gel explosive
Blast Holes	253
Volume	33,109m ³
Tonnage	107,546t

Results and discussion

All BMMs were recovered by onsite personnel. Post-blast detection recorded significant horizontal movement to the North-Northwest. As per the data in Table 3, BMM horizontal movement varied between 2.55m and 14.32m, and inclination varied by 92.13°. Average horizontal movement was 11.16m and the standard deviation was 4.87 m. BMMs generally moved slightly upwards as expected (creating heave) except for BMM 5 (Figure 5).

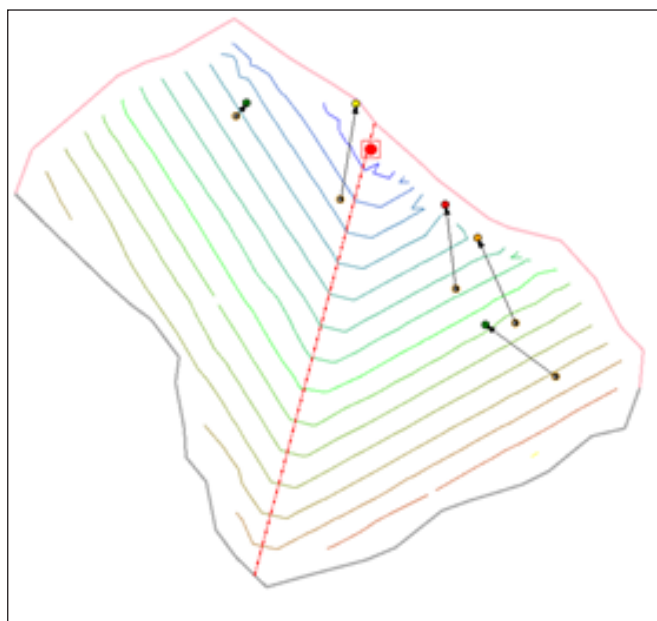


Figure 5 – Blast 220-6 BMM Vectors

Table 3 – Vector results

BMM#	Horizontal Distance (m)	Direction (°)
1	2.55	39.15
2	14.32	9.11
3	12.52	353.47
4	13.71	336.39
5	12.71	307.02

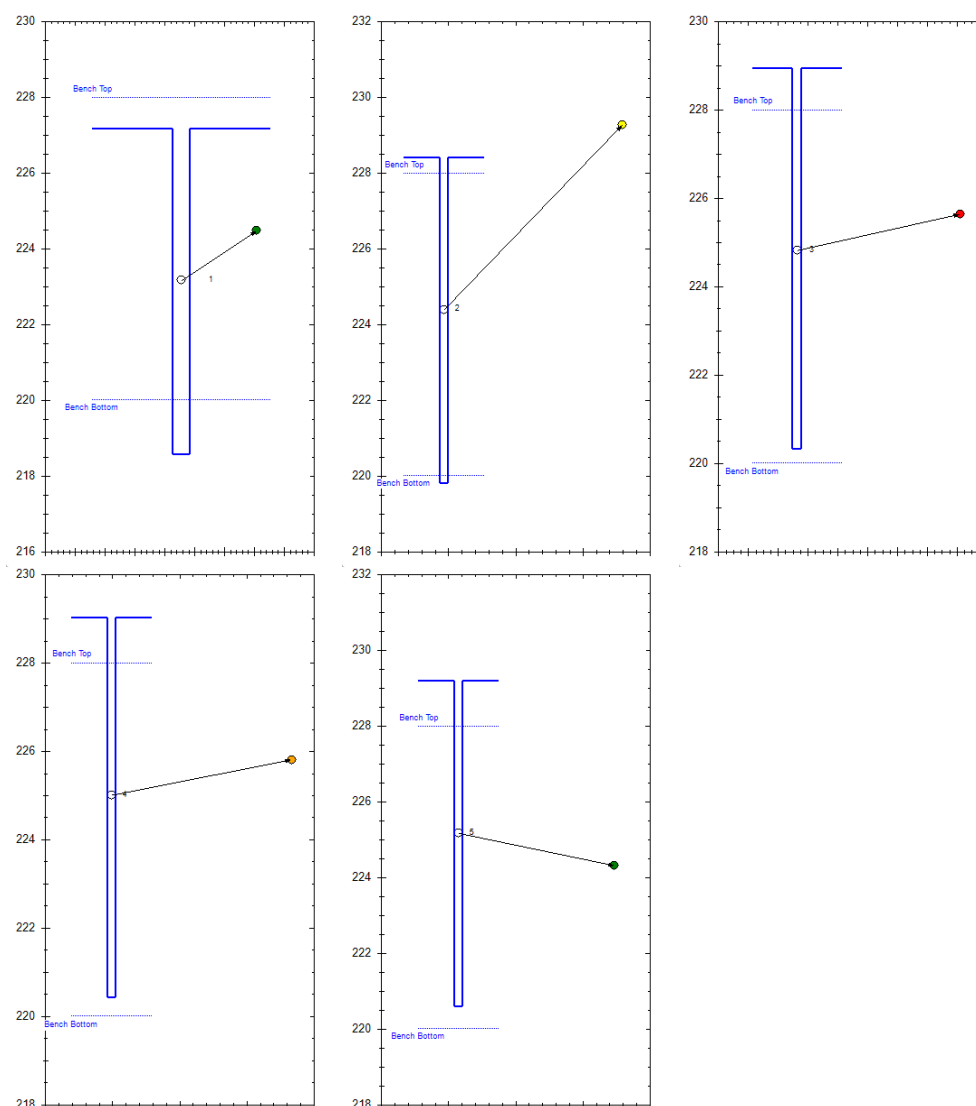


Figure 6 – Side View of BMM Vectors

The negative vertical displacement values recorded by several BMM sensors can mainly be attributed to their close proximity of the bench crest (break line). Vertical displacement data are particularly important in areas where high-grade ore zones are concentrated. Long term BMM data collected indicates that horizontal rock mass displacement is typically minimal at the bench surface, increases with depth and reaches values close to the maximum in the vicinity of the upper section of the explosive charge column, and then decreases again toward the bench toe – this is referred to as the ‘D curve’. For this reason, surface scanning alone is insufficient for accurately determining the displacement of the entire blasted rock mass.

At the same time, specialists from the Chief Geologist’s Department of Zhairem MCC emphasize that rock mass displacement within individual zones of a blast block remains highly variable and difficult to predict. The movement of blasted material strongly depends on the geo-mechanical properties of the rock mass, including lithology, structural features, and rock strength.

Economic Outcomes (Value Calculator)

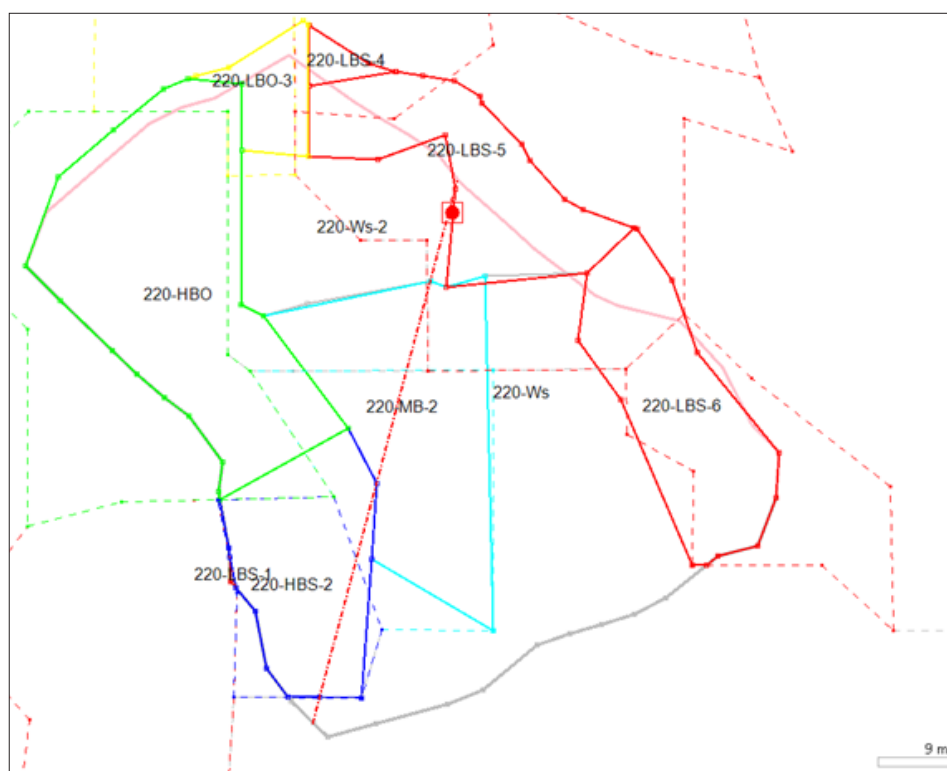


Figure 7 – Pre and Post Blast Blocks

The BMM Explorer software translates the ore blocks using the BMM vectors. Figure 7 shows the pre-blast ore-blocks (dashed lines) and translated post-blast ore-blocks (solid lines).

Table 4 – Ore Block Inputs

	Zinc			Barite			
Material	Min	Max	Average	Min	Max	Average	Density
HBO	2.19	2.19	2.19	57.13	57.13	57.13	3.29
HBS	1.96	6.16	3.57	39.15	67.12	56.12	3.36
LBO	3.63	6.41	4.95	0.00	11.07	2.10	2.46
LBS	2.28	8.02	5.45	0.00	27.11	9.27	3.27
MB	0.22	0.44	0.36	57.34	63.61	60.73	3.43
OOL	1.44	2.05	1.61	0.00	0.15	0.11	2.76
Waste	0.07	0.70	0.20	0.00	15.15	4.49	3.07

Table 5 – Economic Inputs

Ore Price – Barite	6.14 USD/tonne
Ore Price – Zinc	962.2439 USD/tonne
Mining Cost	5.6 USD/tonne
Processing Cost	10.43 USD/tonne
Processing Efficiency	75%

Table 6 – Economic Results

Ore Loss			Dilution			Misclassification	
Tonne	%	\$	Tonne	%	\$	T	%
6,689	8.6	68,858	8,645	11.2	138,582	8,686	11.20

The results above in Table 6 are results if blast movement had not been accounted for. These values are represented as red (ore-loss), dilution (brown) and misclassification (green) on Figure 8 below. It is noted that these values are an unreconciled potential value.

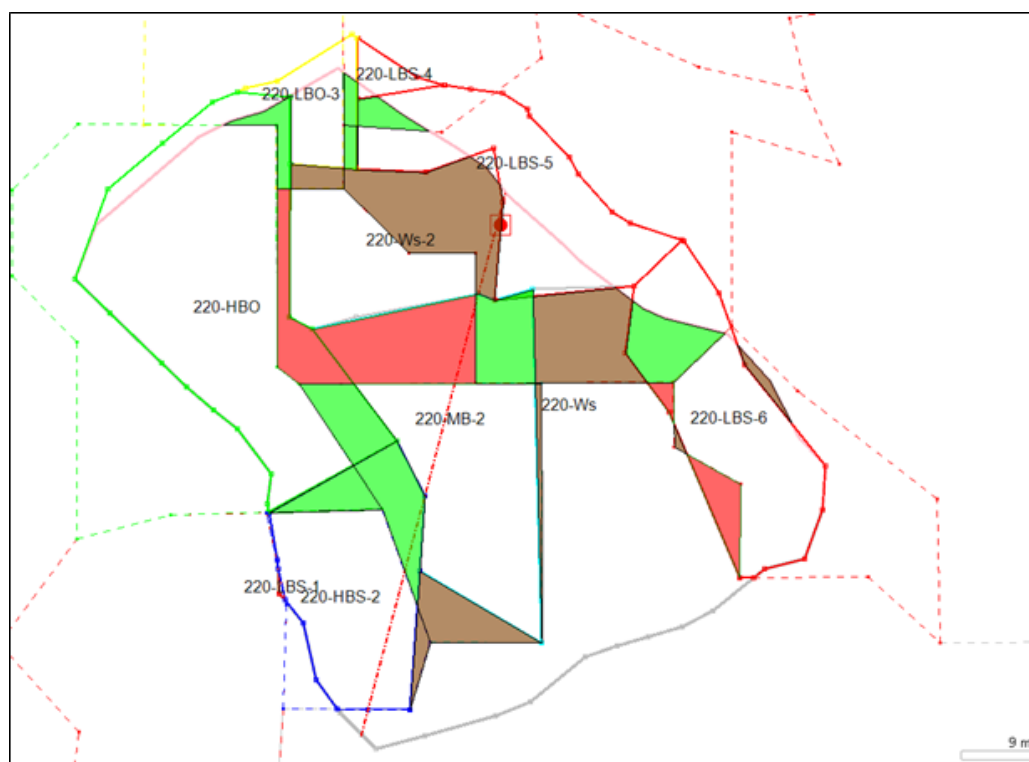


Figure 8 – Ore Loss, Dilution and Misclassification of Blast 220-6

Further Analysis

Using the BMM results, we can generate secondary results based on the movement averages and compare these to the original BMM vectors. This is akin to assumptions made in other alternative software. For this calculation we have generated vectors with an average horizontal distance of 11.16m and direction of -6.97°. In this case the discrepancy compared with the Average movement is over \$55,000 in opportunity loss, highlighting the inconsistency of blasting.

Table 7 – Discrepancy between BMM System and Average Vector

	Ore Loss			Dilution			Misclassification	
	T	%	\$	T	%	\$	T	%
BMM System	6,689	9%	68,858	8,645	11%	138,583	8,686	11%
Average Vector	5,867	3%	46,139	10,713	10%	171,732	2,537	2%
Discrepancy	822	5%	22,719	2,068	1%	33,149	6,149	9%

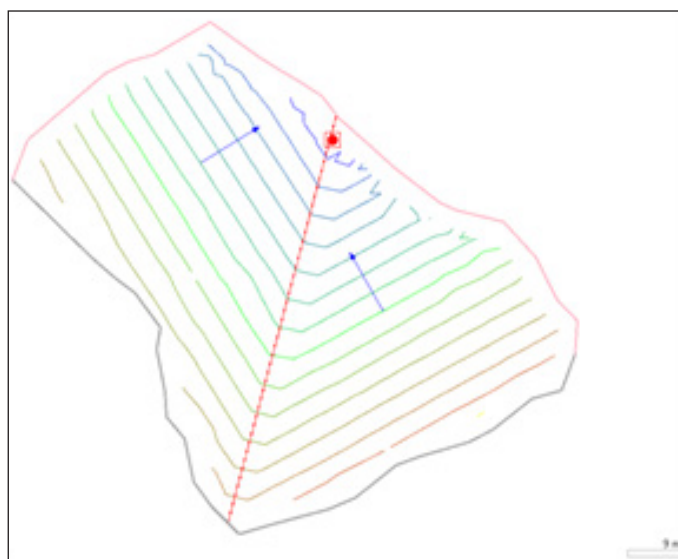


Figure 9 – Textbook Vectors

Other blasting software have traditionally relied on the assumption that ore moves perpendicular to the timing contours. This is consistent with textbook blast design. Using the software, we can replicate this assumption and compare financial results. An example of the blast direction has been displayed in Figure 9, and the horizontal distance has been entered as the average of BMM movements (11.16m). Results indicate a \$60,000 discrepancy between relying on the textbook direction of movement and using real measured vectors.

Table 8 – Discrepancy between BMM system and Textbook Vector

	Ore Loss			Dilution			Misclassification	
	Tonne	%	\$	Tonne	%	\$	T	%
BMM System	6,689	9%	68,858	8,645	11%	138,583	8,686	11%
Textbook Vector	7,225	7%	96,827	10,643	10%	170,611	9,329	12%
Discrepancy	536	2%	27,969	1,998	1%	32,029	643	1%

Table 9 – Variability across Blast Blocks of Zhairam

No.	Bench	Blast	Min (m)	Max (m)	Block Tonnage (t)	Ore Loss (t)	Dilution (t)	Misclassification (t)
1	208	208-1-1	2.78	2.78	18,698	276	0	0
2	212	212-10 (214)	0.63	2.12	111,192	1	1,630	1,955
3	212	212-12	0.31	0.73	70,918	41	1,412	702
4	212	212-16	2.96	6.70	77,835	2,047	328	362
5	212	212-27	10.21	10.85	41,438	4,417	0	0
6	212	212-29	5.12	14.62	68,254	4,662	2,807	0
7	212	212-30	12.36	17.73	121,969	18,224	594	15,343
8	212	212-31	8.09	11.95	68,454	7,739	3,024	1,231
9	220	220-1	1.87	5.25	58,731	1,081	42	2,188

Continuation of table 9

10	220	220-18	2.64	2.64	18,631	226	115	545
11	220	220-2	1.54	4.59	90,286	659	0	4,348
12	220	220-2-1	1.54	4.59	96,058	2,995	2,271	873
13	220	220-22	1.74	4.06	121,723	131	2,308	866
14	220	220-4	3.31	13.62	69,859	791	1,334	1,676
15	220	220-5	8.33	8.33	86,782	1,012	3,675	4,562
16	220	220-6	2.55	14.32	107,546	6,689	8,645	8,686
17	224	224-3	3.47	3.47	45,671	230	1,080	0
18	224	224-6	2.36	2.60	55,599	981	0	1,343
19	224	224-8	3.05	3.05	34,403	169	9	0
20	228	228-10-1	7.84	12.88	104,693	6,155	3,529	7,395
21	228	228-15	2.99	5.43	142,884	3,725	633	6,099
22	228	228-16	5.38	6.00	65,514	989	669	3,800
23	228	228-9	3.56	6.31	47,321	4	789	5,705
24	232	232-1	2.86	2.86	69,653	447	0	0
25	232	232-4	3.62	4.68	96,724	201	1,990	2,251
26	236	236-10a	2.43	2.64	111,070	2,381	0	0
27	236	236-13s	2.55	2.55	27,093	327	0	502
28	236	236-2	8.99	12.10	83,046	3,410	1,069	3,426
29	236	236-2-1	8.99	12.10	71,000	11,080	10,971	6,401
30	236	236-8	4.63	4.63	80,729	428	2,674	0
31	240	240-1	1.50	3.63	85,204	976	365	193
32	244	244-13	2.46	7.43	34,877	2,270	818	0
33	244	244-5	3.31	3.31	77,919	431	2,026	817
34	244	244-6	4.69	8.08	82,408	1,913	0	1,802
35	248	248-1	5.79	5.79	71,388	79	830	0
36	248	248-2	1.49	2.56	64,957	578	368	448
37	308	308-13	7.44	15.97	70,268	2,225	6,163	4,631
38	308	308-22	0.78	7.07	85,679	2,918	1,444	4,763
39	316	316-16	7.23	13.12	112,426	11,237	5,193	1,820

The data presented in Table 9 indicate high spatial variability in blast-induced displacement, with no consistent relationship to bench level, block geometry, or tonnage, reflecting the complex interaction between geological conditions and blast design. Several blocks (e.g., 5, 7, 8, 29) exhibit disproportionately high ore loss relative to their tonnage, suggesting localized zones of intensified displacement or poor ore-waste boundary preservation.

Dilution values are frequently comparable to or exceed ore loss (notably in blocks 16, 20, 23, 37), which supports the conclusion that dilution represents a major economic risk, as previously identified in the economic analysis section. This reflects inefficient separation of ore and waste during blasting and highlights the sensitivity of grade control to blast-induced mixing.

Misclassification values are also significant and, in several cases (e.g., 7, 10, 12, 20), exceed both ore loss and dilution. This is particularly important in the context of polymetallic deposits such as Zhairem, where incorrect routing of material between processing streams leads to additional economic penalties beyond simple dilution effects.

Overall, Table 9 provides empirical evidence that blast-induced displacement is a dominant and highly variable driver of ore loss, dilution, and misclassification, further justifying the need for direct measurement approaches such as BMM for effective grade control.

The results from Blast 220-6 at the Zhairem Mining and Concentrating Complex clearly demonstrate that blast-induced movement is both significant in magnitude and highly spatially variable, with direct consequences for grade control performance and economic outcomes. Measured displacement ranges (2.55–14.32 m) and directional variability exceeding 90° demonstrate that blast movement cannot be reliably represented using a single average vector or deterministic approach.

At Zhairem, as in many contemporary open-pit operations, excavation accuracy of ± 0.2 m is achievable due to high-precision guidance systems on loading equipment. However, the results show that pre-blast ore polygons can be displaced by more than an order of magnitude greater than excavation tolerances. Without correction, this discrepancy leads directly to systematic ore loss, dilution, and misclassification, regardless of how accurately the dig design is followed. In this sense, blast movement represents a dominant source of grade control error, overriding downstream improvements in equipment accuracy if left unaccounted for.

Mining based on in-situ polygons without accounting for blast movement would result in approximately 6,689 t of ore lost and 8,645 t of dilution, with a combined economic impact exceeding USD 200,000 for a single blast. This highlights the strong economic sensitivity of grade control to spatial uncertainty at ore boundaries.

The BMM system is typically supplied as part of an integrated solution; however, for the purposes of this analysis, a simplified cost assumption is adopted. It is assumed that drilling and labour costs amount to approximately USD 500 per BMM unit installed. Under this assumption, the total cost associated with deploying five BMMs to track blast-induced movement would be approximately USD 2,500. On this basis, the resulting return on investment (ROI) can be evaluated by comparing this implementation cost against the quantified value preserved through reductions in ore loss, dilution, and misclassification.

Dilution represents a major economic penalty due to additional mining and processing costs, while misclassification further increases losses by directing material to inappropriate processing streams.

A key methodological insight from this study is the demonstrated inadequacy of using a limited number of BMMs to represent entire ore-waste contacts. Although only five BMMs were installed in Blast 220-6, the observed variability suggests that even this number may be insufficient for complex geological settings with multiple ore types and irregular boundaries. This finding supports the argument that BMM deployment strategies must be tailored to geological complexity and economic risk, rather than applied as a uniform, minimum-cost solution. Increasing sensor density at high-value or high-risk contacts is likely to yield disproportionate economic returns.

While the results from Zhairem demonstrate clear benefits, several limitations should be acknowledged. While detailed economic evaluation is demonstrated using Blast 220-6, the analysis is supported by a dataset of 20 blasts from the Zhairem operation, enabling assessment of variability in ore loss, dilution, and misclassification across different conditions. A simplified sensitivity analysis indicates that the economic impact remains significant under a wide range of input assumptions. For example, a $\pm 20\%$ variation in metal prices or processing recovery results in proportional changes in calculated value; however, the relative magnitude of losses due to blast movement remains consistently high. Future work should focus on multi-blast datasets to assess variability over time, optimise BMM placement strategies, and further integrate blast movement data with comminution and plant performance models.

Correlation analysis

A correlation analysis was performed to evaluate the relationship between blast-induced displacement and grade control performance indicators. Displacement values were approximated using the midpoint of the reported range for each blast. The results indicate a moderate to strong positive correlation between displacement and ore loss, suggesting that increased blast movement leads to higher loss of valuable material. Similar relationships were observed for dilution and misclassification, although with slightly lower correlation strength.

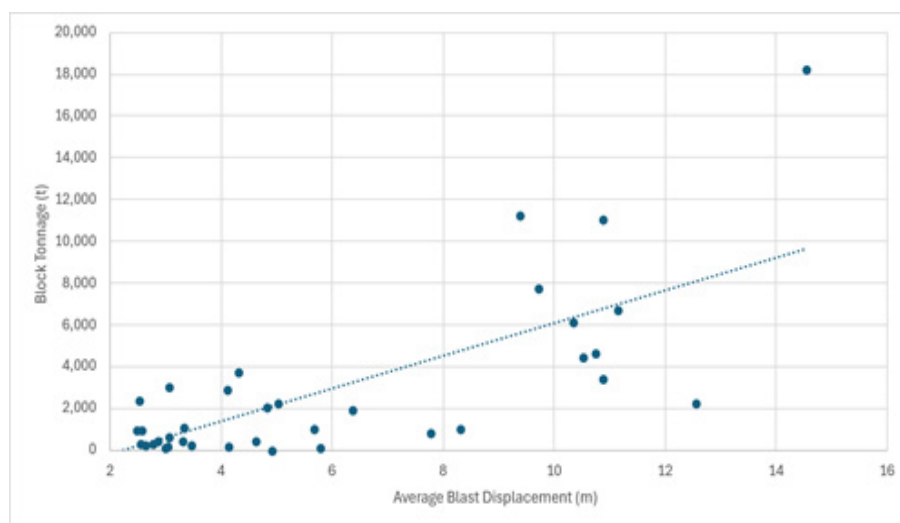


Figure 10 – Correlation between Displacement and Ore Loss

The relationship between blast-induced displacement and ore loss is illustrated in Figure 10 ($R^2 = 0.553$). A positive trend can be observed, indicating that larger displacement distances generally correspond to increased ore loss. Although some scatter is present due to geological variability, the overall pattern confirms that displacement is a key driver of ore loss in open-pit operations.

Conclusions

This study demonstrates that blast-induced rock movement represents a critical and quantifiable source of error in grade control for open-pit mining. The results show that displacement is highly variable and cannot be reliably approximated using deterministic or averaged approaches.

The integration of BMM data into grade control workflows enables the correction of ore boundaries prior to excavation, significantly reducing ore loss, dilution, and misclassification. The economic analysis confirms that even a single blast can generate value losses exceeding USD 200,000 when movement is not accounted for.

From a methodological perspective, the study highlights the limitations of predictive modelling and supports the use of direct measurement techniques in complex geological environments. The findings also indicate that the effectiveness of BMM depends on appropriate sensor placement and should be aligned with geological variability and economic risk.

Although a detailed analysis is presented for Blast 220-6, the study is supported by a broader dataset of multiple blasts from the Zhairem minesite, enabling validation of observed trends across varying conditions. This integration of detailed case analysis with multi-blast variability enhances the robustness of the results and supports the wider applicability of the proposed methodology for linking blast-induced displacement with economic performance.

REFERENCES

- 1 Carrasco, C., Keeney, L., Napier-Munn, T., Bode, P. Unlocking additional value by optimizing communication strategies to process Grade Engineering streams. *Minerals Engineering*, 103, 2–10 (2017). <https://doi.org/10.1016/j.mineng.2016.07.020>.
- 2 Navarro, J., Seidl, T., Hartlieb, P., Sanchidrián, J., Segarra, P., Couceiro, P., Schimek, P., Godoy, C. Blastability and ore grade assessment from drill monitoring for open pit applications. *Rock Mechanics and Rock Engineering*, 54, 3209–3228 (2021). <https://doi.org/10.1007/s00603-020-02354-2>.
- 3 Akbar, S., Abdolmaleki, M., Ghadernejad, S., Esmaili, K. Applying knowledge-based and data-driven methods to improve ore grade control of blast hole drill cuttings using hyperspectral imaging. *Remote Sensing*, 16, 2823 (2024). <https://doi.org/10.3390/rs16152823>.
- 4 Díaz, A., Fernández, C., Álvarez, I. Grade control in one of the biggest open pit mines in Europe: Corta Atalaya, Riotinto. *Minerals*, 15 (2025). <https://doi.org/10.3390/min15010044>.

- 5 Dimitrakopoulos, R., Godoy, M. Grade control based on economic ore/waste classification functions and stochastic simulations: examples, comparisons and applications. *Mining Technology*, 123, 90–106 (2014). <https://doi.org/10.1179/1743286314Y.0000000062>.
- 6 Thornton, D. The implications of blast-induced movement to grade control (2009).
- 7 Rogers, W. Understanding blast movement to optimise grade control practices at the Ahafo Gold Mine in Ghana (2014). <https://doi.org/10.14264/uql.2014.322>.
- 8 Eshun, P., Dzigbordi, K. Control of ore loss and dilution at AngloGold Ashanti Iduapriem Mine using blast movement monitoring system. *Ghana Mining Journal*, 16, 49–59 (2016). <https://doi.org/10.4314/gm.v16i1.6>.
- 9 Hmoud, S., Kumral, M. Risk-based optimization of post-blast dig-limits incorporating blast movement and grade uncertainties with multiple destinations in open-pit mines. *Natural Resources Research*, 34, 193–214 (2024). <https://doi.org/10.1007/s11053-024-10428-z>.
- 10 Tordoir, A., Weatherley, D., Onederra, I., Bye, A. A new 3D simulation framework to model blast induced rock mass movement using physics engines. In: *Proceedings of the Ninth International Symposium on Rock Fragmentation by Blasting (Fragblast)*, Granada, Spain, pp. 381–388 (2009).
- 11 Fitzgerald, M., York, S., Cooke, D., Thornton, D. Blast monitoring and blast translation – case study of a grade improvement project at the Fimiston Pit, Kalgoorlie, Western Australia. In: *Proceedings of the Eighth International Mining Geology Conference*, Queenstown, New Zealand, pp. 285–297 (2011).
- 12 Thornton, D. The application of electronic monitors to understand blast movement dynamics and improve blast designs. In: *Proceedings of the Ninth International Symposium on Rock Fragmentation by Blasting (Fragblast)*, Granada, Spain, pp. 287–300 (2009).
- 13 Yang, R.L., Kavetsky, A.P. A three dimensional model of muckpile formation and grade boundary movement in open pit blasting. *International Journal of Mining and Geological Engineering*, 8, 13–34 (1990).
- 14 Malchenko, E., Roman, A., Portnov, V., Askarova, N. Peculiarities of the formation of the Zhailma volcano-tectonic deep. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu* (2023). <https://doi.org/10.33271/nvngu/2023-1/026>.
- 15 Bilimova, I.S., Ponomareva, E.V., Ponomareva, M.V. Studying hydrogeological conditions for the purpose of developing methods of drying the walls of the Zhairam deposit Dalnezapadny site. *Trudy Universiteta*, 1(94), 170–175 (2024). https://doi.org/10.52209/1609-1825_2024_1_170.

^{1*}Әбікен Б.І.,

магистрант, ORCID ID: 0009-0003-1343-4748,

*e-mail: be_abiken@kbtu.kz

²Стоукс О.,

бакалавр, ORCID ID: 0009-0000-4110-1543,

e-mail: oliver.stokes@hexagon.com

³Гамбаль М.,

PhD, тау-кен инженері, ORCID ID: 0009-0007-0357-2994,

e-mail: gambalmaxim@gmail.com

¹Түлемисова Ж.С.,

PhD, қауымдастырылған профессор, ORCID ID: 0000-0003-1803-4535,

e-mail: z.tulemissova@kbtu.kz

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

²Hexagon Mining, Перт қ., Аустралия

³Leica Geosystems Kazakhstan LLC, Алматы қ., Қазақстан

ЖАЙРЕМ КЕНОРНЫНДА ЖАРЫЛЫС ЖҰМЫСТАРЫ КЕЗІНДЕ КЕННІҢ ЖОҒАЛЫМДАРЫ МЕН ҚҰНАРСЫЗДАНУЫН ЕСКЕРЕ ОТЫРЫП, ТАУЖЫНЫСТАРДЫҢ ЖЫЛЖУЫ ЖӘНЕ КЕН САПАСЫН БАҚЫЛАУ

Аңдатпа

Мақала карьерлерде кен сапасын бақылау тұрғысынан жарылыс жұмыстары кезінде таужыныстардың жылжуын бақылаудың (ВММ) экономикалық тиімділігін талқылайды. Жарылыс жұмыстары кезінде таужыныстардың жылжуы карьерлерде кен сапасын тиімді бақылауға кедергі келтіретін негізгі факторлардың

бірі. Жарылыс кезіндегі есепке алынбаған ығысулар жиі кеннің құнарсыздануына, алынатын кеннің жоғалуына және қате классификацияға әкеліп соғып, нәтижесінде өңдеу мен экономикалық тиімділікті төмендетеді. Зерттеу жарылыс нәтижесінде пайда болатын таужыныс жылжуының табиғаты стохастикалық екендігіне және оны тек детерминистік модельдеу арқылы сенімді түрде қайта жаңғырту мүмкін еместігіне негізделген. Сондықтан жоғары дәлдіктегі сапаны бақылау үшін кен орнында тікелей өлшеулер жүргізу қажет. Жарылыс блогындағы бірнеше бақылау нүктесінен үшөлшемді ығысу векторларын тіркеу арқылы ВММ жүйесі жарылыстан кейін кен шекараларын дәлірек қалпына келтіруге мүмкіндік береді. Бұл көпнүктелі тәсіл кеңістіктік өзгергіштікті ескереді және қазу жұмыстарына дейінгі құрам жоспарларын түзету үшін сенімді негіз ұсынады. 2024 ж. Жайрем кен байыту комбинатында ВММ жүйесінің пилоттық енгізілуі жүзеге асырылды, ал қазіргі уақытта бұл жүйе Қазақстандағы бірнеше тау-кен компаниясында қолданылуда. Жайрем кен орнынан алынған деректерді пайдалана отырып, зерттеу бақылаусыз таужыныс ығысуының операциялық салдарын, соның ішінде кеннің жоғалуын, құнарсыздануын және қате классификацияны сандық түрде бағалайды. Нәтижелер бос жыныстар мен кен арасындағы шекараларды дәлірек анықтау ресурс тиімділігін арттырып, пайдалылықтың тікелей өсуіне ықпал ететінін көрсетеді. Осы нәтижелер негізінде мақалада жарылысты бақылау жүйелеріне инвестициялық шешімдерді қолдауға арналған қаржылық шығындарды бағалау құралы ұсынылады. Бұл құрал операторларға ВММ енгізудің шығын-пайда арақатынасын жүйелі түрде бағалауға және сапаны бақылау стратегияларын кең экономикалық шеңберде оңтайландыруға мүмкіндік береді.

Түйін сөздер: жарылыс кезінде кеннің ығысуын бақылау, кеннің сапасын бақылау, кен жоғалымдары, кенді құнарсыздандыру, қате классификация, инвестициялардың өтелімділігі, Жайрем, Казцинк, Glencore, кен, карьер, цинк, барит, минерализация түрі, экскавация.

^{1*}Әбікен Б.І.,

магистрант, ORCID ID: 0009-0003-1343-4748,

*e-mail: be_abiken@kbtu.kz

²Стоукс О.,

бакалавр, ORCID ID: 0009-0000-4110-1543,

e-mail: oliver.stokes@hexagon.com

³Гамбаль М.,

PhD, горный инженер, ORCID ID: 0009-0007-0357-2994,

e-mail: gambalmaxim@gmail.com

¹Тулемисова Ж.С.,

PhD, ассоциированный профессор, ORCID ID: 0000-0003-1803-4535,

e-mail: z.tulemissova@kbtu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

²Hexagon Mining, г. Перт, Австралия

³Leica Geosystems Kazakhstan LLC, г. Алматы, Казахстан

МОНИТОРИНГ СМЕЩЕНИЯ ГОРНЫХ ПОРОД ВО ВРЕМЯ ВЗРЫВНЫХ РАБОТ ПРИ КОНТРОЛЕ КАЧЕСТВА РУДЫ НА МЕСТОРОЖДЕНИИ ЖАЙРЕМ С УЧЕТОМ ПОТЕРЬ И РАЗУБОЖИВАНИЯ РУДЫ

Аннотация

В данной статье рассматривается экономическая эффективность мониторинга сдвигов породы при взрывах (ВММ) с учетом контроля качества руды в карьерах. Перемещения горной массы при взрывах остаются одним из основных факторов, ограничивающих эффективный контроль качества руды в карьерах. Неучтенные смещения при взрывах часто приводят к разубоживанию руды, потере полезного ископаемого и неправильной классификации, что в итоге снижает эффективность переработки и экономическую рентабельность. Исследование основано на предпосылке, что смещение горных пород, вызванное взрывом, по своей природе является стохастическим и не может быть надежно воспроизведено с помощью одного только детерминированного моделирования; поэтому для высокоточного контроля качества требуются прямые измерения на месторождении. За счет фиксации трехмерных векторов смещения из нескольких точек мониторинга в пределах взрывного блока ВММ позволяет более точно восстанавливать границы руды после взрыва. Этот многоточечный подход учитывает пространственную изменчивость и обеспечивает надежную

основу для корректировки сортовых планов перед началом экскавации. В 2024 г. пилотная реализация системы ВММ была начата на Жайремском горно-обогатительном комбинате (ГОК), и система применяется в нескольких горнодобывающих компаниях Казахстана. Используя полевые данные с Жайремского месторождения, в исследовании количественно оцениваются эксплуатационные последствия неконтролируемого сдвига породы, включая потери руды, разубоживание и неправильную классификацию. Результаты показывают, что улучшенное определение границ между пустой породой и рудой повышает эффективность использования ресурсов и напрямую способствует увеличению прибыльности. На основе этих выводов в статье предлагается инструмент оценки финансовой стоимости, предназначенный для поддержки инвестиционных решений, связанных с системами мониторинга движения взрыва, что позволяет операторам систематически оценивать соотношение затрат и выгод внедрения ВММ и оптимизировать стратегии контроля качества в более широких экономических рамках.

Ключевые слова: мониторинг сдвига пород при взрыве, контроль за качеством руды, потеря руды, разубоживание, ошибочная классификация, окупаемость инвестиций, Жайрем, Казцинк, Glencore, руда, карьер, цинк, барит, тип минерализации, экскавация.

UDC 553.43:550.83:004.8
IRSTI 38.35.19

<https://doi.org/10.55452/1998-6688-2026-23-2-435-450>

¹*Saduov A.,

Senior Lecturer, ORCID ID: 0000-0003-1501-7772,

*e-mail: a.saduov@satbayev.university

¹Satbayev University, Almaty, Kazakhstan

UNSUPERVISED DELINEATION OF PROSPECTIVITY ZONES FOR STRATIFORM CU-CO IN THE SOUTHERN COPPERBELT MARGIN (ZAMBIA)

Abstract

This study presents an unsupervised workflow for delineating prospectivity zones for stratiform copper-cobalt mineralization in the southern margin of the Central African Copperbelt, Zambia, using only airborne geophysical data (gravity, magnetics, radiometry, terrain). Three clustering algorithms-K-Means, Fuzzy C-Means (FCM), and Self-Organizing Maps (SOM)-were applied, followed by consensus clustering to isolate robust target zones. Key geophysical filters (tilt derivative, total horizontal derivative, analytic signal) and radiometric ratios (U/Th, U/K) were computed and compared across clusters. The most prospective cluster was validated against literature-calibrated geophysical thresholds and was found to match known exploration criteria and regional structural trends. A final probabilistic prospectivity map was generated from FCM membership, classifying targets into three confidence levels. While lack of drill-hole data prevents quantitative accuracy assessment, this approach demonstrates that fully unlabeled machine learning on airborne data can effectively guide early-stage exploration in data-sparse regions. The proposed workflow offers a reproducible framework for AI-driven prospectivity mapping in frontier terrains.

Keywords: copper-cobalt, unsupervised learning, clustering, airborne geophysics, Zambia.

Received March 2, 2026; accepted May 28, 2026.

Introduction

The Central African Copperbelt (Zambia and the Democratic Republic of the Congo) has been recognized as the world's largest sediment-hosted copper province and a major cobalt source. The Neoproterozoic Katangan basin has been reported to contain ~8.7 billion tonnes of copper ore at ~2.6 wt% Cu, and mineralization along the southern Zambian margin has been interpreted to extend beneath sedimentary cover into greenschist- to amphibolite-facies rocks affected by the Lufilian orogeny [1]. In contrast to the well-drilled northern districts, the Southern Copperbelt Margin has remained weakly constrained by geochemical sampling and drilling, and stratiform Cu–Co potential has therefore been expected to be largely concealed beneath cover [1].

Traditional mineral prospectivity mapping (MPM) approaches, including knowledge-driven overlays and supervised machine-learning models, have been shown to depend on reliable positive (deposit) and negative (barren) labels [2, 3]. In greenfield settings, occurrences are commonly absent and true negatives cannot be verified, which limits the applicability of supervised calibration and can render weakly supervised strategies inconclusive [3]. Consequently, fully unsupervised approaches have been required to delineate exploration targets without deposit labels.

Unsupervised learning has been increasingly used to detect multivariate domains and anomalies directly from geodata. Clustering methods have been applied to delineate coherent geoscientific signature zones, yet many studies have relied on a single algorithm and/or on expert-guided cluster selection. To reduce algorithm-specific bias, a consensus framework has been adopted by integrating

K-Means, Fuzzy C-Means (FCM), and Self-Organizing Maps (SOM), and only areas consistently identified across methods have been retained as high-priority zones [4]. The complementary behaviour of hard partitions (K-Means), gradual memberships (FCM), and topology-preserving mapping (SOM) has been leveraged to improve robustness under label-scarce conditions [5].

Because drill-hole and systematic geochemical data were unavailable, prospectivity was inferred solely from airborne geophysical proxy variables. Magnetic derivatives, including tilt derivative and total horizontal derivative, were used to emphasize structural edges and shallow sources, while analytic signal amplitude was used to highlight source-centred responses. Radiometric ratios (U/Th and U/K) were further used as proxies for alteration and redox-related processes, and the applicability of radiometric criteria in exploration has been supported by regional case studies [6]. These six co-registered layers were used as the core input to the clustering ensemble.

Cluster prioritization was performed by comparing cluster summary statistics to literature-derived thresholds representing geophysical responses expected in favourable structural and alteration settings. This external benchmarking was intended to constrain cluster selection objectively and reduce interpretive subjectivity in the absence of local training data. The study was therefore designed to deliver an objective, data-driven prospectivity map for stratiform Cu–Co mineralization in a true greenfield setting and to demonstrate the utility of ensemble unsupervised analytics for frontier exploration targeting.

Geological and Geophysical Context

Geological Setting

The study area has been located near Chililabombwe and Chingola (northern Zambia) along the southern margin of the Central African Copperbelt. The regional basement has been overlain by the Neoproterozoic Katanga Supergroup (~880–520 Ma), which hosts stratiform Cu–Co mineralization predominantly within the Lower Roan Subgroup. In particular, the carbonaceous Ore Shale Formation at the transition from red-bed siliciclastics to overlying reduced sediments has been reported to contain disseminated chalcopyrite, bornite, and chalcocite hosted by bedding-parallel fabrics and carbonate bands [7].

Late Neoproterozoic to Early Paleozoic Pan-African deformation has been expressed by the Lufilian Arc, a NW-SE-trending fold-and-thrust belt, where northeast-vergent thrusting has emplaced Roan sequences onto pre-Katangan crystalline basement. Basement highs, composed mainly of Paleoproterozoic to Mesoproterozoic granites and gneisses, have been interpreted to control sub-basin segmentation during deposition and to influence the preservation and localization of Cu–Co mineralization; for example, basement ridges have been associated with Roan thinning/absence and ore concentration along ridge flanks.

Along the southern Copperbelt margin, Roan stratigraphy has been largely concealed beneath thrust sheets and younger Kundelungu cover. The Chililabombwe-Chingola district has exemplified this configuration, where high-grade Ore Shale orebodies have been exploited at major operations (e.g., Konkola and Nchanga) and have been interpreted to continue downdip beneath thrusts to depths exceeding 1 km, remaining partly blind at surface. Structural duplication and folding of the Ore Shale horizon have been expected to generate synclines and faulted blocks capable of hosting concealed repeats or extensions of mineralization [7].

Airborne survey and preprocessing

A high-resolution airborne survey (KoBold Metals NRG2155) has been acquired over the study area (2022) and the publicly released dataset has been accessed via Kaggle [8]. Gravity, magnetics, radiometrics, and terrain data have been collected along NE-SW flight lines spaced 200 m apart with orthogonal tie-lines at ~2 km, at a nominal terrain clearance of ~80 m. Instrumentation has been reported to include a cesium-vapor magnetometer, an airborne gravimeter stabilized by GPS/inertial measurements, and a 32 L NaI gamma-ray spectrometer; dual-frequency GPS positioning has been used for georeferencing.

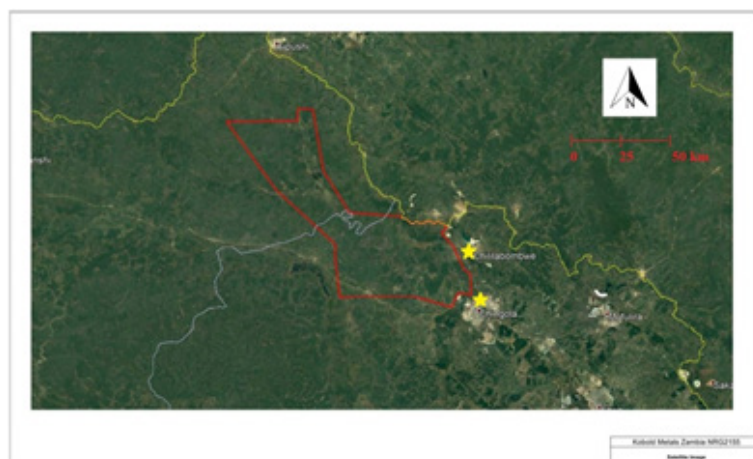


Figure 1 – Geographical location of the airborne survey polygon (red outline) near Chililabombwe-Chingola on the southern margin of the Central African Copperbelt. The polygon from the provided KML overlay is draped on Google Earth imagery, showing the proximity to the DRC-Zambia border and known mining areas (yellow outline)



Figure 2 – Simplified geological map of the southern margin of the Central African Copperbelt in northern Zambia (adapted from Porter GeoConsultancy 2024). The Roan Group (hosting stratiform Cu-Co ore) is exposed in the cores of Lufilian fold structures and otherwise concealed beneath Nguba and Kundelungu Group formations. Purple stars indicate major Cu-Co deposits (e.g., Konkola, Nchanga). The survey area (yellow outline) spans a zone of complex thrusting and folding where Roan units are largely hidden, highlighting the potential for concealed mineralization.

Standard corrections and derivative transformations have been applied to prepare analysis-ready layers. Gravity observations have been corrected for latitude, free-air, terrain ($\rho = 2.67 \text{ g cm}^{-3}$), and Bouguer effects to derive Bouguer anomalies and gravity disturbance grids, followed by computation of first vertical derivative and total horizontal derivative to enhance shallow and lateral density contrasts [10]. Magnetic total-field data have been leveled, diurnally corrected, and reduced to the pole to center anomalies over their sources, and derivative products (first vertical derivative and total horizontal derivative) have been generated for edge enhancement [11]. Tilt derivative and analytic signal amplitude have been computed to emphasize subtle sources and source-centered responses, respectively. Radiometric K, eTh, and eU channels have been noise-filtered using NASVD, and ratio maps (U/Th and U/K) have been produced as proxies for alteration and redox-related processes. A digital terrain model has been derived to provide elevation context.

All layers listed in Table 1 have been resampled to a common 50 m grid (WGS84 UTM Zone 35S).

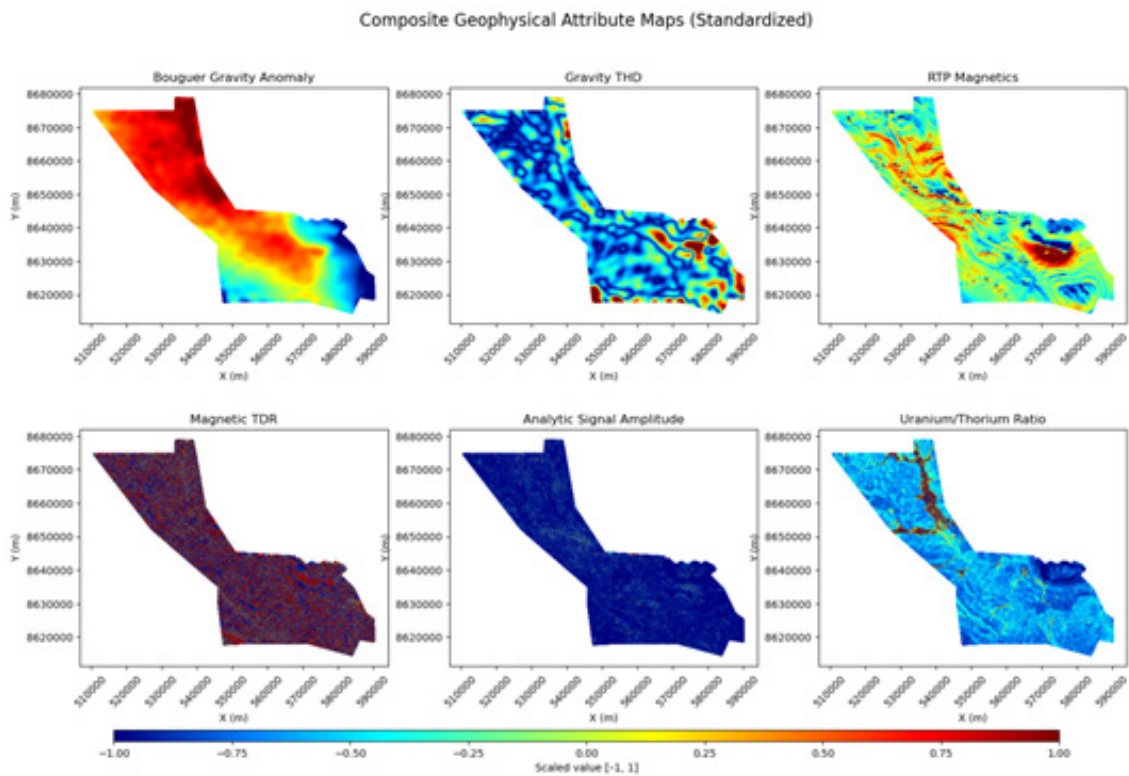


Figure 3 – Composite geophysical attribute maps (standardized) for the airborne survey area. Top row (left to right): Bouguer gravity anomaly; gravity total horizontal derivative (THD); reduced-to-pole (RTP) magnetic intensity. Bottom row (left to right): magnetic tilt derivative (TDR); analytic signal amplitude; uranium/thorium (U/Th) ratio

Table 1 – Summary of input geophysical and radiometric data layers from the airborne survey (NRG2155). All grids are 50 m resolution

Data Layer	Description (Measurement)
X, Y	Coordinates (projection: WGS84 UTM Zone 35S, in meters)
Bouguer267_Processed	Bouguer gravity anomaly (terrain-corrected, density 2.67g/cc) in mGal
Gravity_disturbance	Gravity disturbance (residual anomaly after regional removal) in mGal
Grav_FVD	Gravity first vertical derivative (enhances shallow density contrasts)
Grav_THD	Gravity total horizontal derivative (highlights lateral density gradients)

Continuation of table 1

Grav_Gaussian_HP	Gravity high-pass filter (Gaussian filter isolating local high-frequency gravity signal)
Processed_magnetics_RTP	Magnetic intensity, reduced to pole (nT; highlights total field anomalies centered over sources)
Gradient_levelled	Leveled magnetic vertical gradient (nT/m; measured or computed vertical magnetic gradient, leveled)
Mag_FVD	Magnetic first vertical derivative (sharpens shallow magnetic sources)
Mag_THD	Magnetic total horizontal derivative (highlights edges of magnetic bodies)
Mag_Tilt_Derivative	Magnetic tilt derivative (degrees; angle of total field, useful for subtle source detection)
Mag_Analytical_Signal	Magnetic analytic signal amplitude (nT/m; combined 3D gradient magnitude, peaks over magnetic source centers)
Potassium_NASVD	Potassium (% K) from radiometric survey (NASVD noise-filtered)
Uranium_NASVD	Uranium (ppm eU) from radiometric survey (NASVD filtered)
Thorium_NASVD	Thorium (ppm eTh) from radiometric survey (NASVD filtered)
U_Th_ratio	Uranium/Thorium ratio (dimensionless; highlights U enrichment or Th depletion)
U_K_ratio	Uranium/Potassium ratio (dimensionless; can indicate uranium mobility vs K)
Digital_terrain	Digital Terrain Model (m; elevation above sea level)

Methodology

Data Standardization and Feature Selection

All eighteen geophysical and radiometric variables prepared in Section 2.2 were considered during exploratory analysis. For clustering and subsequent quantitative validation, six diagnostic features were retained: Grav_THD, Mag_THD, Mag_Tilt_Derivative, Mag_Analytical_Signal, U/Th, and U/K. This selection was justified by the availability of published threshold intervals and methodological constraints enabling external benchmarking of cluster signatures [10–14]. Variables lacking transferable validation thresholds (e.g., Bouguer anomaly, DTM, and individual K-Th-U channels) were not prioritized for validation due to limited comparability across terrains and potential sensitivity to surficial processes; in particular, K and Th responses may be strongly influenced by regolith and weathering and may not reflect subsurface mineralization [9].

No features were removed or down-weighted within the adopted six-feature workflow. Prior to clustering, each retained feature was standardized to zero mean and unit variance to ensure comparable scaling between gravity derivatives, magnetic derivatives, and radiometric ratios and to prevent dominance of any variable within distance-based clustering.

3.2. Unsupervised clustering algorithms

K-Means clustering was used to obtain hard spatial partitions by minimizing within-cluster variance [10]. Solutions with $k = 3$ and $k = 4$ were evaluated to test partition stability and to assess whether additional clusters produced meaningful spatial structure.

Fuzzy C-Means (FCM) clustering was used to represent gradational boundaries via membership degrees in $[0, 1]$. The fuzzifier was set to $m = 2$ as a conventional compromise between membership smoothness and numerical stability, and $c = 3$ and $c = 4$ were examined. Membership fields were interpreted as continuous measures of similarity to each cluster.

Self-Organizing Maps (SOM) were trained using a 6×6 lattice to preserve topological relations in the reduced feature space. The MiniSom implementation was used for training and mapping. After convergence, SOM prototypes were aggregated into three groups to obtain a spatial label map consistent with the $k = 3$ partitioning used across methods.

These algorithms were selected to exploit complementary behaviour: compact partitions (K-Means), fuzzy transitions (FCM), and topology-preserving representation (SOM). Supervised

and semi-supervised schemes were not adopted because reliable positive and negative labels were unavailable in this greenfield setting [10].

3.3. Consensus Clustering

To obtain a robust ensemble partition, K-Means, FCM, and SOM were applied independently to the same six standardized features with k (or c) = 3. For FCM, hard labels were derived by selecting the maximum-membership class per sample. For SOM, neuron prototypes were grouped into three clusters and labels were propagated to all samples.

The three label vectors were assembled into an $N \times 3$ label matrix, where each row encoded the triplet of method-specific assignments for one sample. A secondary K-Means run with $k = 3$ was then performed on this label matrix to derive the final consensus partition (consensus_k3). This strategy was intended to down-weight method-specific artefacts and to emphasize spatial patterns persistent across algorithms [11].

Validation using literature-based feature thresholds

Consensus clusters were validated by comparing their feature-wise mean values to literature-derived anomaly intervals for the six retained features. Threshold ranges reported for potential-field edge detection, tilt derivative and analytic signal applications, as well as radiometric alteration indicators, were used to define expected anomaly magnitudes over structurally complex or mineralization-favourable zones. Because anomaly amplitudes may be attenuated by depth and cover, gravity and magnetic derivative thresholds were adjusted downward by 15–30% to account for amplitude decay. Radiometric thresholds were similarly reduced to account for weathering-related depletion effects reported for uranium and potassium responses in near-surface materials [12].

Threshold Calibration

For cross-feature comparability, both cluster means and threshold bounds were transformed into Z-scores using the global mean (μ) and standard deviation (σ) of each feature:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

where μ and σ were computed from the full dataset. A cluster was considered consistent with the benchmark if its standardized mean exceeded the standardized threshold criterion.

To address skewness and to emphasize proportional anomalies, a log10 transformation was additionally applied to selected features (notably ASA and U/K) and to the corresponding thresholds:

$$L = \log_{10}(X) \quad (2)$$

and agreement with benchmark ranges was evaluated in log space. This supplementary check was used to ensure that conclusions were not driven by extreme-value asymmetry and to support interpretation under multiplicative anomaly behaviour [13–15].

Results

Clustering Results

The optimal cluster number was evaluated using the elbow criterion and silhouette scores. A clear inflection and the highest silhouette values were observed at $k = 3$, whereas improvements beyond $k = 3$ were marginal and $k = 4$ produced additional fragmentation, particularly in the eastern sector.

K-Means partitioning with $k = 3$ produced three spatially coherent domains. The $k = 4$ solution subdivided the northeastern domain without a clear gain in spatial interpretability. Stable convergence was achieved using multiple random initializations under Lloyd's optimization scheme.

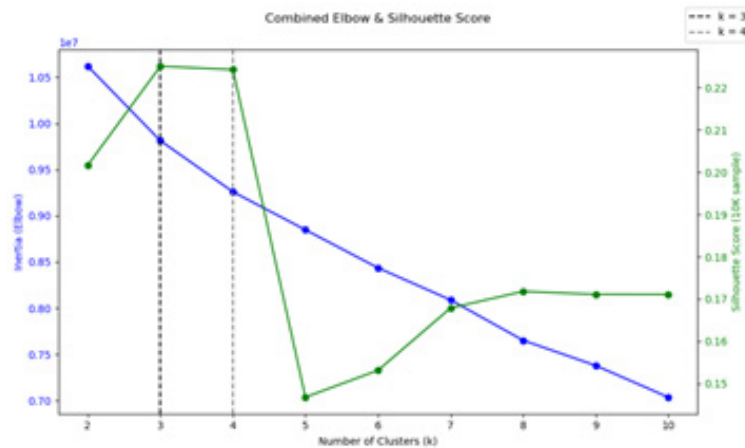


Figure 4 – Elbow method (sum of squared errors versus k) and silhouette score plot for $k = 2-6$. The inflection point in the elbow curve and peak in the silhouette coefficient at $k = 3$ support the selection of three clusters.

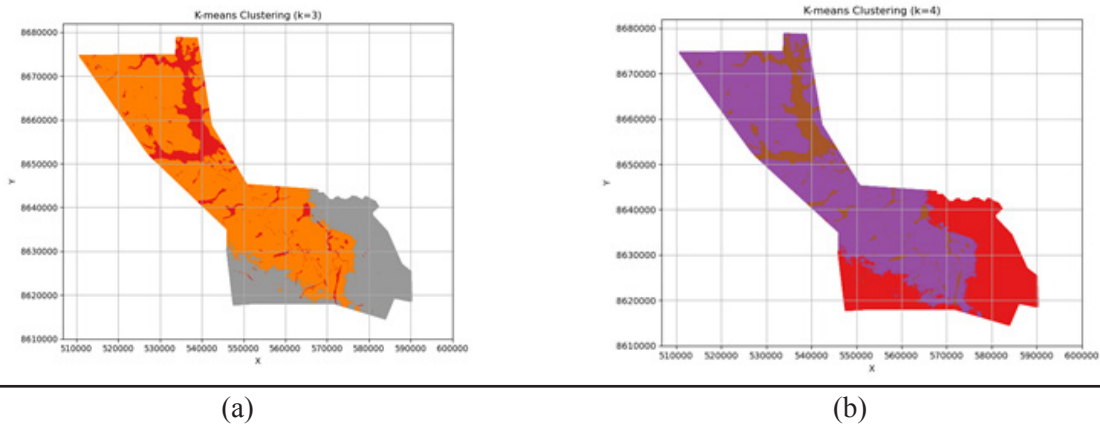


Figure 5 – K-Means clustering: (a) $k = 3$ showing three compact zones; (b) $k = 4$ illustrating over-fragmentation in the northeast

SOM training on a 6×6 lattice yielded 36 prototype neurons summarizing the six-feature input space [16–17]. The resulting neuron map captured local continuity and discontinuity in multivariate geophysical signatures.

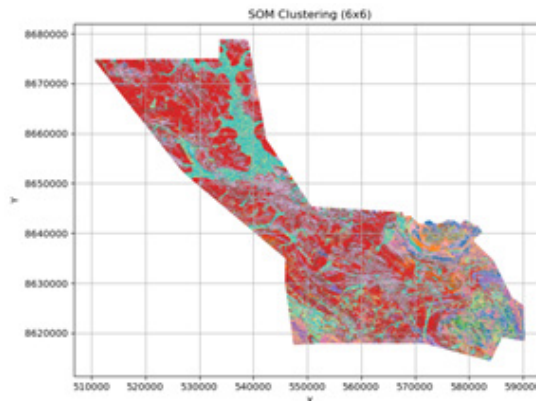


Figure 6 – Self-Organizing Map (6×6) projection encoding the six geophysical features onto 36 neurons. Adjacent neurons on the map correspond to similar input signatures, highlighting local geophysical variability and structural lineaments

Post-clustering of SOM codebook vectors using K-Means ($k = 3$) generated smoother, more spatially continuous domains relative to direct K-Means. As with K-Means, the $k = 4$ variant introduced additional fragmentation.

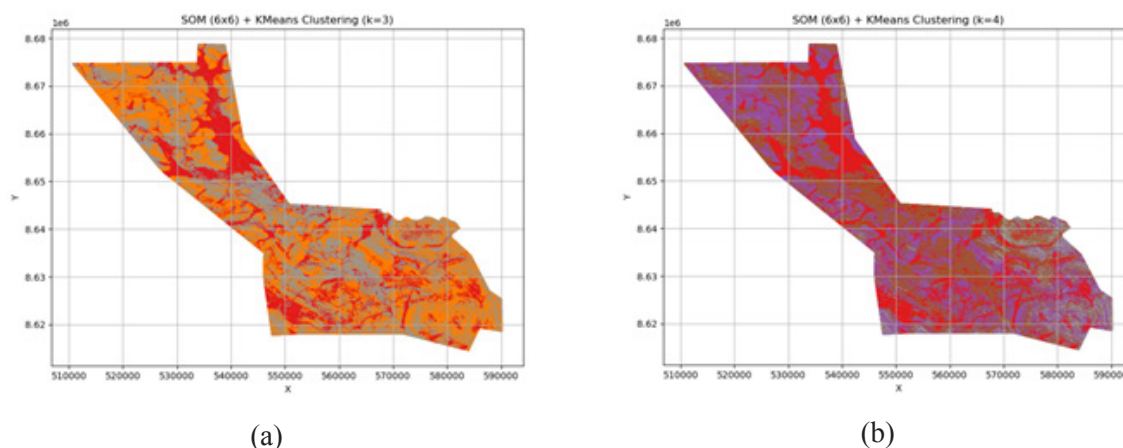


Figure 7 – Clustering of SOM prototypes with K-Means: (a) $k = 3$ yields spatially coherent domains; (b) $k = 4$ introduces unnecessary subdivisions

FCM with $m = 2$ and $c = 3$ produced spatial membership gradients, emphasizing transitional zones relative to hard partitioning. The $c = 4$ variant increased partition complexity with limited additional structure.

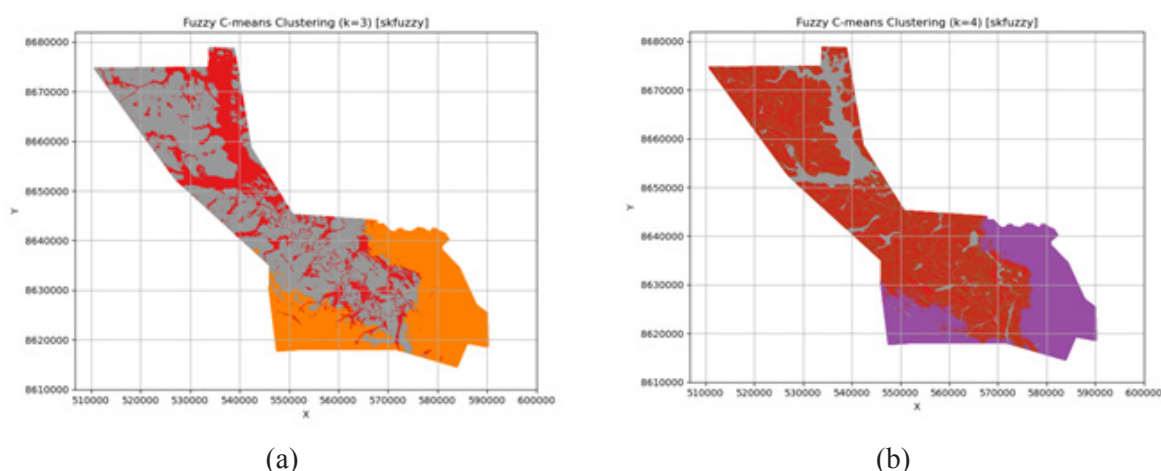


Figure 8 – Fuzzy C-Means clustering: (a) $c = 3$ showing membership gradations; (b) $c = 4$ illustrating increased complexity with limited interpretive gain

Feature discrimination was quantified using ANOVA F-scores computed from cluster partitions. In K-Means, the largest F-scores were associated with gravity and radiometric variables, indicating their dominant role in partition separation. Comparable patterns were observed for FCM and SOM-derived partitions.

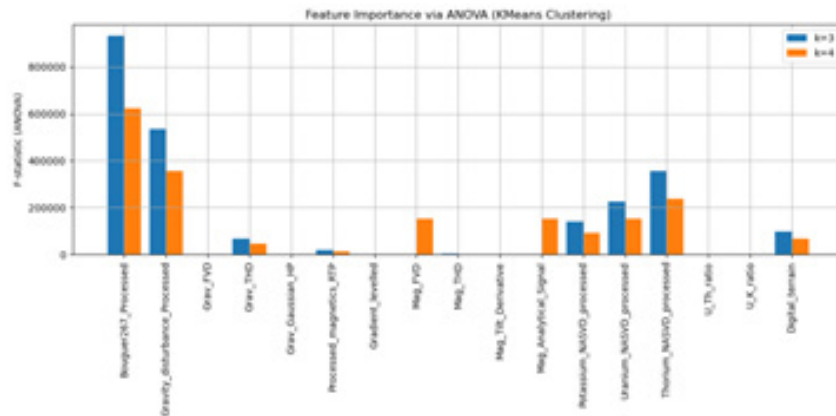


Figure 9 – ANOVA F-values for each geophysical feature in KMeans clustering (k=3 and k=4). Highest discrimination observed for gravity and radiometric features

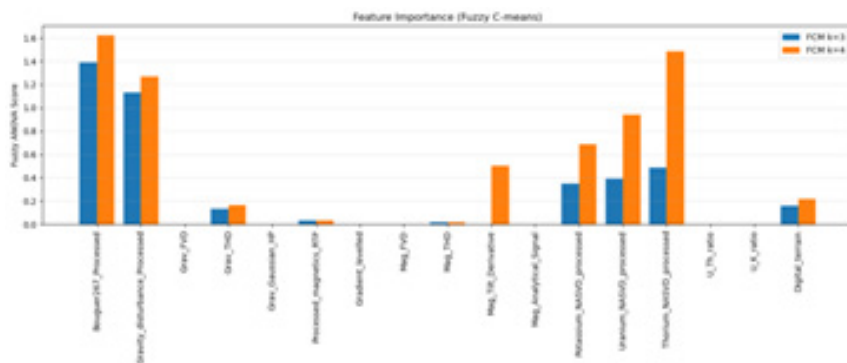


Figure 10 – Feature importance for Fuzzy C-Means clustering based on ANOVA F-scores. The top features align with those found in KMeans

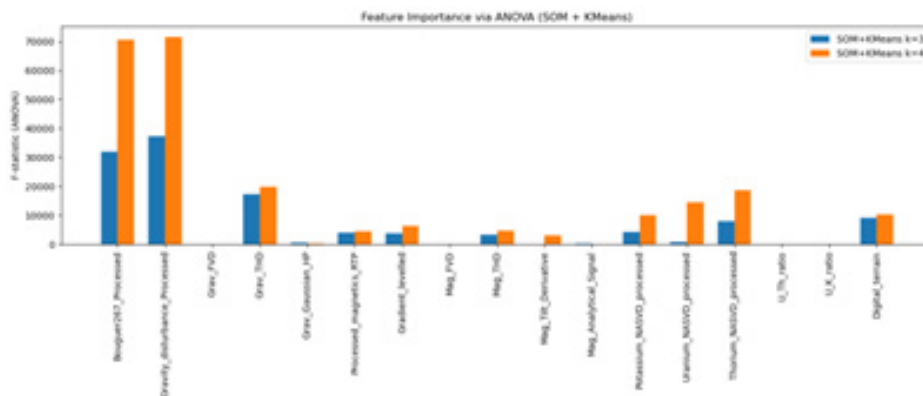


Figure 11 – ANOVA F-values for SOM + KMeans clustering across all features. The results confirm the dominance of gravity and radiometric attributes in cluster differentiation

Consensus clustering was then generated by meta-clustering the label vectors from K-Means, FCM, and SOM using K-Means (k = 3), producing the final consensus_k3 partition. The resulting consensus map highlighted zones of multi-method agreement.

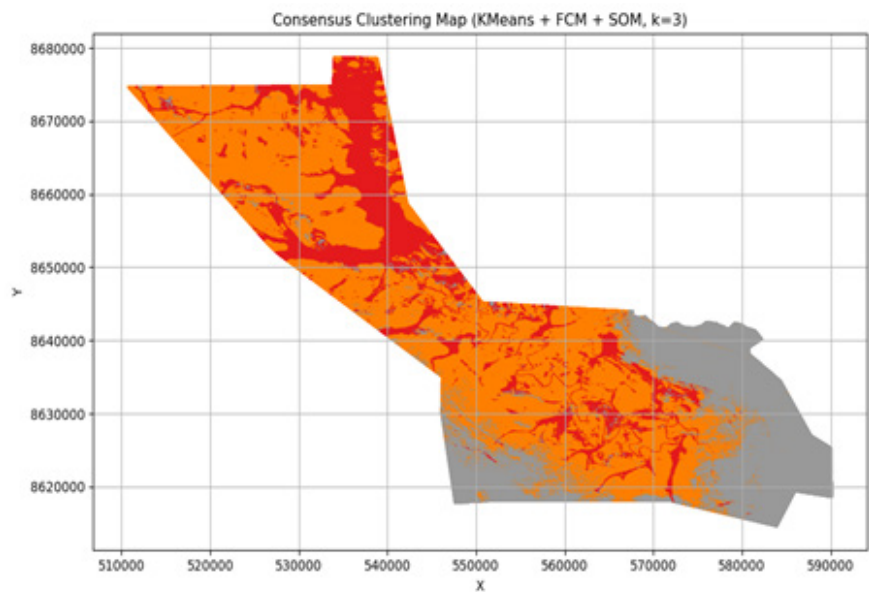


Figure 12 – Consensus clusters (consensus_k3) derived from meta-K-Means on method label vectors. Regions consistently identified by all three methods are emphasized as priority targets

Cluster Validation Against Geophysical Thresholds

Validation of consensus_k3 was performed by comparing feature-wise cluster means with literature-based intervals for six diagnostic variables (Grav_THD, Mag_THD, Mag_Tilt_Derivative, Mag_Analytical_Signal, U/Th, and U/K). The adopted ranges and sources are summarized in Table 2.

Table 2 – Calibrated geophysical intervals used for cluster validation

Feature	Literature Interval	Justification / Source
Grav_THD	0.0025-0.0041	Zhu et al. (2022); Gupta et al. (2018)
Mag_THD	1.66-3.13	Radhakrishna et al. (2014)
Mag_Tilt_Derivative	1.52-1.55	Stewart and Miller (2018)
Mag_Analytical_Signal	150-750	Onyedim and Awoyemi (2005)
U/Th ratio	0.5-1.5	Ciputra et al. (2024)
U/K ratio	5-12	IAEA (2003)

Standardized (Z-score) cluster means were compared against the corresponding standardized threshold bounds, and Cluster 1 was identified as the cluster most consistently meeting or exceeding the benchmark intervals.

A complementary comparison was performed using log10-transformed cluster means and log10-scaled thresholds, confirming the same priority ranking.

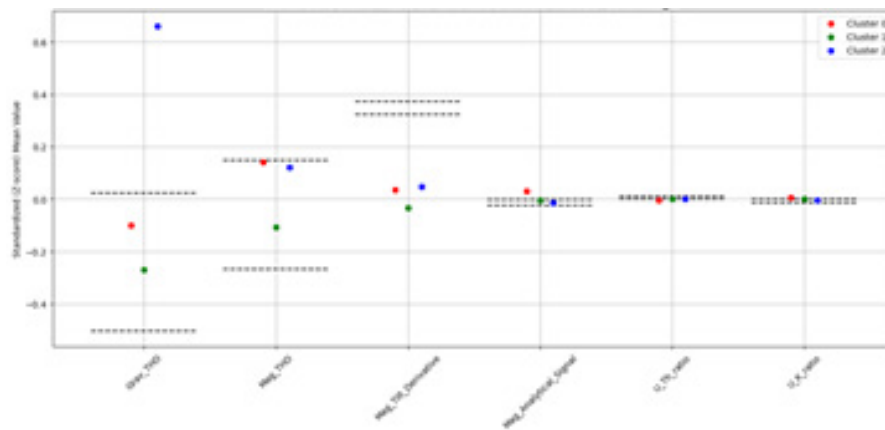


Figure 13 – Standardized (Z-score) cluster means for consensus_k3 compared to literature-based geophysical intervals. Horizontal dashed lines indicate the lower and upper bounds of each threshold interval derived from published studies. Cluster 1 exhibits consistent alignment with or elevation above these bounds across multiple features, reinforcing its geophysical plausibility

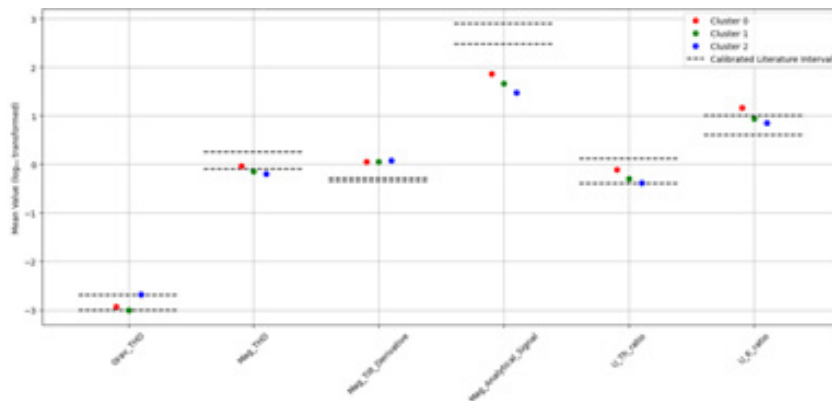


Figure 14 – Log₁₀-transformed cluster means for consensus_k3 compared to log₁₀-scaled intervals. Horizontal dashed lines denote the transformed interval limits

Final Cu-Co Prospectivity Map

A continuous prospectivity map was generated from the FCM membership to the target cluster (Cluster 1, k = 3). Membership values were interpreted as continuous similarity scores to the selected prospective domain.

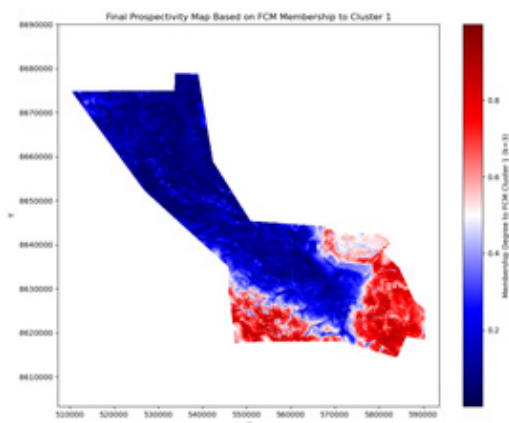


Figure 15 – Continuous prospectivity map based on FCM membership to Cluster 1 (k = 3). Color bar indicates membership values from 0.0 (blue) to 1.0 (red)

For decision-oriented targeting, membership values were discretized into three classes: Low (0.0-0.5), Medium (0.5-0.8), and High (0.8-1.0), based on the empirical membership distribution.

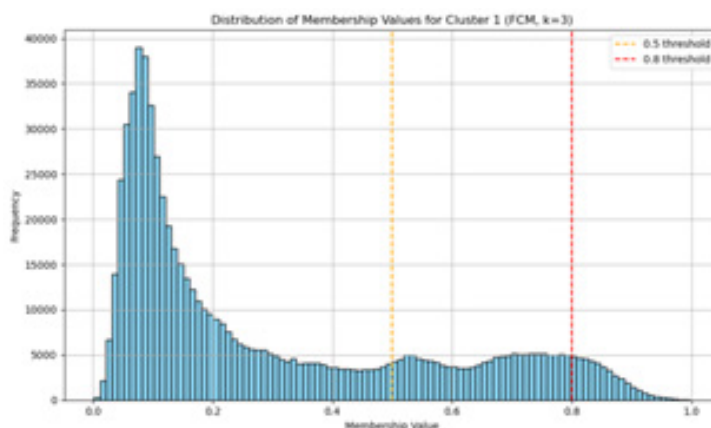


Figure 16 – Distribution of FCM membership values for Cluster 1 ($k = 3$) with 50 bins. Dashed lines at 0.5 and 0.8 mark the thresholds for medium and high prospectivity classes

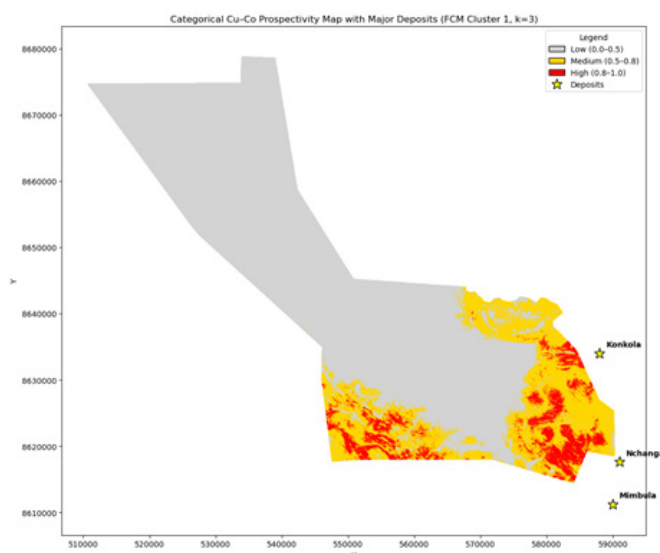


Figure 17 – Categorical Cu-Co prospectivity map (Low/Medium/High) based on FCM membership to Cluster 1 ($k = 3$)

Thus, the High-prospectivity zones (red) demarcate prime targets for Cu-Co exploration, particularly on the southeastern and southern flanks of the survey area.

Discussion

The results demonstrate that ensemble unsupervised clustering of airborne gravity, magnetic, and radiometric data can delineate structurally coherent Cu-Co target zones in the data-scarce southern Copperbelt margin without the use of labeled training data. By integrating K-Means, Fuzzy C-Means, and Self-Organizing Maps within a consensus framework and validating cluster signatures against literature-derived geophysical intervals, an objective and reproducible targeting scheme was obtained. The high-prospectivity domain forms a spatially coherent belt concentrated in the southern and southeastern sectors of the survey area and occurs in proximity to major stratiform

Cu-Co districts, suggesting consistency with the fold-and-thrust architecture of the Lufilian Arc [18–20]. Elevated magnetic edge indicators and radiometric ratios within the priority cluster are compatible with structurally prepared zones and alteration-related contrasts reported in analogous Cu-Co systems [21–22]. Methodologically, the integration of consensus clustering with quantitative threshold benchmarking reduces subjective cluster selection and mitigates algorithm-specific bias, as recommended in ensemble clustering theory [23], while ensuring consistency with established geophysical anomaly behaviour [24]. Nevertheless, several limitations remain. The absence of drilling or direct geological validation prevents quantitative accuracy assessment; literature-based threshold intervals represent generalized analogues and may not fully capture local lithological or cover variability; radiometric responses may be attenuated by weathering and regolith processes [25]; and multivariate similarity does not guarantee geological uniqueness of causative sources. Therefore, the presented results should be interpreted as a preliminary, data-driven targeting framework requiring subsequent geological verification and potential integration with geochemical or semi-supervised extensions.

Conclusion

A fully unsupervised workflow for delineating prospective stratiform Cu-Co zones in the southern Copperbelt margin (Zambia) has been developed using multi-parameter airborne geophysical data. By combining K-Means, Fuzzy C-Means, and Self-Organizing Maps within a consensus clustering scheme and validating cluster signatures against literature-calibrated geophysical thresholds, a priority cluster was objectively identified and translated into a continuous probabilistic prospectivity map. The selected cluster occupies approximately 12% of the survey area and concentrates exploration focus on the most geophysically anomalous terrain. Spatial concordance with the regional Lufilian structural framework and proximity to established Cu-Co districts support the geological plausibility of the delineated targets [18,19]; however, quantitative validation remains contingent upon ground truth data. The proposed ensemble unsupervised framework provides a reproducible and transferable early-stage targeting approach that can be extended through geochemical integration, weak labeling strategies, and application across other metallogenic provinces.

REFERENCES

- 1 Ndonfack, K.I.A., Yang, Z., Zhang, J., Whattam, S.A., Xie, Y. Geology, geochemistry, and exploration of the Central African Copperbelt: a review. *International Geology Review* (2024). <http://doi.org/10.1080/00206814.2024.2426200>
- 2 Yousefi, M., Carranza, E.J.M. Data-driven index overlay and Boolean logic mineral prospectivity modeling in greenfields exploration. *Natural Resources Research*, 25(1), 3–18 (2016). <https://doi.org/10.1007/s11053-014-9261-9>
- 3 Saremi, M., Hezarkhani, A., Seyyed Mirzabozorg, S.A.A., DehghanNiri, R., Shirazy, A., Shirazi, A. Unsupervised anomaly detection for mineral prospectivity mapping using Isolation Forest and Extended Isolation Forest algorithms. *Minerals*, 15(4), 411 (2025). <https://doi.org/10.3390/min15040411>
- 4 Xiong, Y., Zuo, R. A positive and unlabeled learning algorithm for mineral prospectivity mapping. *Computers & Geosciences*, 147, 104667 (2021). <https://doi.org/10.1016/j.cageo.2020.104667>
- 5 Wang, C., Chen, J., Ouyang, Y. Determination of predictive variables in mineral prospectivity mapping using supervised and unsupervised methods. *Natural Resources Research*, 31, 2081–2102 (2022). <https://doi.org/10.1007/s11053-021-09930-5>
- 6 Sharapatov, A., Saduov, A., Assirbek, N., Abdyrov, M., Zhumabayev, B. Prediction of rare and anomalous minerals using anomaly detection and machine learning techniques. *Applied Computing and Geosciences*, 26, 100250 (2025). <https://doi.org/10.1016/j.acags.2025.100250>
- 7 Rezapour, M.J., Abedi, M., Bahroudi, A., Rahimi, H. A clustering approach for mineral potential mapping: a deposit-scale porphyry copper exploration targeting. *Geopersia*, 10(1), 149–163 (2020). <http://doi.org/10.22059/geope.2019.284414.648488>

- 8 Martin, R., Boisvert, J. Ensemble spatial and multivariate clustering with geomodeling-specific clustering metrics. *Computers & Geosciences*, 120, 82–96 (2018). <https://doi.org/10.1016/j.cageo.2018.08.005>
- 9 Strehl, A., Ghosh, J. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583–617 (2002). <https://doi.org/10.1162/153244303321897735>
- 10 Miller, H.G., Singh, V. Potential field tilt – A new concept for location of potential field sources. *Journal of Applied Geophysics*, 32(2-3), 213–217 (1994). [https://doi.org/10.1016/0926-9851\(94\)90022-1](https://doi.org/10.1016/0926-9851(94)90022-1)
- 11 Cooper, G.R.J. Balancing images of potential-field data. *Geophysics*, 74(3), L17–L20 (2009). <https://doi.org/10.1190/1.3096615>
- 12 Yan, T.-J., Wu, Y.-G., Yuan, Y., Chen, L.-N. Edge detection of potential field data using an enhanced analytic signal tilt angle. *Chinese Journal of Geophysics*, 59(4), 341–349 (2016).
- 13 Kianoush, P., Faraj Khah, N.K., Hosseini, S.A., Jamshidi, E., Afzal, P., Ebrahimabadi, A. Geobody estimation by Bhattacharyya method utilizing nonlinear inverse modeling of magnetic data in Baba-Ali iron deposit, NW Iran (2023). <https://doi.org/10.21203/rs.3.rs-2433584/v6>
- 14 Abdel-Hakeem, M., El-Tahir, M., Abu Zeid, E., et al. Genetic implications of Th/U, Th/K, and U/K ratios for U mineralizations: a case study from El-Missikat and El-Erediya shear zones, Eastern Desert, Egypt. *Geochemical Transactions*, 24(1), 3 (2023). <https://doi.org/10.1186/s12932-023-00083-3>
- 15 Sharapatov, A., Assirbek, N., Saduov, A., Abdyrov, M., Zhumabayev, B. Consolidated geological and geophysical characteristics of uranium deposit rocks and prospects for their utilization (Shu-Sarysu Province, Kazakhstan). *News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences*, 2024(6), 210–229 (2024). <https://doi.org/10.32014/2024.2518-170X.471>
- 16 Torremans, K., Gauquier, J., Boyce, A.J., Barrie, C.D., Muchez, Ph., Visser, W., Cnudde, V. Remobilisation features and structural control on ore grade distribution at the Konkola stratiform Cu-Co ore deposit, Zambia. *Journal of African Earth Sciences*, 79, 10–23 (2013). <https://doi.org/10.1016/j.jafrearsci.2012.10.005>
- 17 Cailteux, J.L.H., Kampunzu, A.B., Lerouge, C., Kaputo, A.K., Milesi, J.P. Genesis of sediment-hosted stratiform copper-cobalt deposits, Central African Copperbelt. *Journal of African Earth Sciences*, 42(1–5), 134–158 (2005). <https://doi.org/10.1016/j.jafrearsci.2005.08.001>
- 18 Jackson, M.P.A., Warin, O.N. Neoproterozoic allochthonous salt tectonics during the Lufilian orogeny in the Katangan Copperbelt, central Africa. *Geological Society of America Bulletin*, 115(3), 314–330 (2003). [https://doi.org/10.1130/0016-7606\(2003\)115<0314:NASTDT>2.0.CO;2](https://doi.org/10.1130/0016-7606(2003)115<0314:NASTDT>2.0.CO;2)
- 19 Hitzman, M.W., et al. The Central African Copperbelt: diverse stratigraphic, structural, and temporal settings in the world's largest sedimentary copper district. *University of Tasmania, Journal contribution* (2012). <https://hdl.handle.net/102.100.100/572953>
- 20 Sharapatov, A., Saduov, A., Assirbek, N. Sravnitel'nyi analiz vozmozhnostei algoritmov mashinnogo i glubokogo obucheniia v geologii [Comparative analysis of machine and deep learning capabilities in geology]. *Nauchno-tehnicheskii i proizvodstvennyi Gornyi zhurnal Kazakhstana* [Scientific-Technical and Industrial Mining Journal of Kazakhstan], 11 (223), 14–21 (2023). <https://doi.org/10.48498/minmag.2023.223.11.002> (in Russian).
- 21 Wilford, J. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma*, 183–184, 124–142 (2012). <https://doi.org/10.1016/j.geoderma.2010.12.022>
- 22 Carranza, E.J.M., Hale, M. Where are regional geochemical anomalies found? *Geochemistry: Exploration, Environment, Analysis*, 1(3), 173–184 (2001). <https://doi.org/10.1144/geochem.1.3.173>
- 23 Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666 (2010). <https://doi.org/10.1016/j.patrec.2009.09.011>
- 24 Bezdek, J.C., Ehrlich, R., Full, W. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2–3), 191–203 (1984). [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
- 25 Sharapatov, A., Zhumabayev, B.T., Saduov, A.B., Assirbek, N. Geomagnitnye dannye i ikh ispol'zovanie pri reshenii zadach geonauk [Geomagnetic data and their use in solving geoscience problems]. *Nauchno-tehnicheskii i proizvodstvennyi Gornyi zhurnal Kazakhstana* [Scientific-Technical and Industrial Mining Journal of Kazakhstan], 12 (224), 30–36 (2023). <https://doi.org/10.48498/minmag.2023.224.12.004> (in Russian).

¹*Садуов А.,

аға оқытушы, ORCID ID: 0000-0003-1501-7772

e-mail: a.saduov@satbayev.university

¹Satbayev University, Алматы қ., Қазақстан

ЗАМБИЯДАҒЫ МЫС БЕЛДЕУІНІҢ ОҢТҮСТІК ЖИЕГІНДЕ СТРАТИФОРМДЫ МЫС-КОБАЛЬТ МИНЕРАЛДАНУЫНЫҢ ПЕРСПЕКТИВАЛЫҚ АЙМАҚТАРЫН МҰҒАЛІМСІЗ ОҚЫТУ ӘДІСТЕРІ НЕГІЗІНДЕ АЙҚЫНДАУ

Аңдатпа

Мақалада Орталық Африка Мыс белдеуінің оңтүстік жиегінде (Замбия) орналасқан стратиформды мыс-кобальт минералдануының перспективалы аймақтарын тек аэрогеофизикалық деректер негізінде айқындауға арналған тәсіл ұсынылған. Пайдаланылған деректер гравиметриялық, магниттік, радиометриялық және бедерлік параметрлерді қамтиды. Талдау барысында кластерлеудің үш алгоритмі – K-Means, Fuzzy C-Means (FCM) және Кохоненнің өзін-өзі ұйымдастыратын карталары (Self-Organizing Maps, SOM) қолданылып, кейін олардың нәтижелері тұрақты мақсатты аймақтарды бөліп көрсету мақсатында консенсустық кластерлеу аясында біріктірілді. Әрбір кластер үшін негізгі геофизикалық туынды көрсеткіштер, оның ішінде еңіс туындысы, толық көлденең туынды және аналитикалық сигнал, сондай-ақ U/Th және U/K радиометриялық қатынастары есептеліп, өзара салыстырылды. Ең перспективалы кластер әдебиеттерге негізделген геофизикалық шекті мәндер арқылы валидацияланып, оның белгілі іздеу критерийлері мен өңірлік құрылымдық заңдылықтарға сәйкес келетіні көрсетілді. Қорытынды перспективалылықтың ықтималдық картасы FCM алгоритмінің мүшелік мәндері негізінде құрастырылып, кейін сенімділіктің үш деңгейіне жіктелді. Бұрғылау деректерінің болмауы нәтижелердің дәлдігін тікелей сандық бағалауға мүмкіндік бермегенімен, алынған қорытындылар көппараметрлі аэрогеофизикалық деректерге қолданылған мұғалімсіз оқыту әдістерінің геологиялық ақпарат тапшы жағдайларда барлау жұмыстарының бастапқы кезеңдерінде тиімді пайдаланылуы мүмкін екенін көрсетті. Ұсынылған тәсілді аз зерттелген аумақтарда жасанды интеллектке негізделген перспективалылықты бағалаудың қайта жаңғыртылатын әдістемелік негізі ретінде қарастыруға болады.

Түйін сөздер: мыс-кобальт, мұғалімсіз оқыту, кластерлеу, аэрогеофизика, перспективалықты болжау, Замбия.

¹*Садуов А.,

ст. преподаватель, ORCID ID: 0000-0003-1501-7772,

e-mail: a.saduov@satbayev.university

¹Satbayev University, г. Алматы, Казахстан

ВЫДЕЛЕНИЕ ЗОН ПЕРСПЕКТИВНОСТИ СТРАТИФОРМНОЙ МЕДНО-КОБАЛЬТОВОЙ МИНЕРАЛИЗАЦИИ НА ОСНОВЕ МЕТОДОВ ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ НА ЮЖНОЙ ОКРАИНЕ МЕДНОГО ПОЯСА (ЗАМБИЯ)

Аннотация

В статье представлен подход к выделению перспективных зон стратиформной медно-кобальтовой минерализации на южной окраине Центрально-Африканского Медного пояса (Замбия) на основе исключительно аэрогеофизических данных, включающих гравиметрические, магнитные, радиометрические и топографические параметры. Для анализа были использованы три алгоритма кластеризации – K-Means, Fuzzy C-Means (FCM) и самоорганизующиеся карты Кохонена (Self-Organizing Maps, SOM), после чего результаты были объединены в рамках консенсусной кластеризации для выделения устойчивых целевых зон. По каждому кластеру были рассчитаны и сопоставлены ключевые геофизические производные, включая производную наклона, полную горизонтальную производную и аналитический сигнал, а также радиометрические

отношения U/Th и U/K . Наиболее перспективный кластер был валидирован по литературно обоснованным геофизическим пороговым значениям и показал соответствие известным поисковым критериям и региональным структурным закономерностям. Итоговая вероятностная карта перспективности была построена по значениям принадлежности FCM и дополнительно разделена на три уровня уверенности. Несмотря на отсутствие буровых данных, не позволяющее выполнить прямую количественную оценку точности, полученные результаты показывают, что методы обучения без учителя, примененные к многопараметрическим аэрогеофизическим данным, могут эффективно использоваться для ранних стадий поисково-разведочных работ в условиях дефицита геологической информации. Предложенный подход может рассматриваться как воспроизводимая основа для ИИ-ориентированного прогнозирования перспективности в слабо изученных территориях.

Ключевые слова: медь-кобальт, обучение без учителя, кластеризация, аэрогеофизика, прогнозирование перспективности, Замбия.

ӘОЖ 622.276:532.5
FTAMP 52.47.15

<https://doi.org/10.55452/1998-6688-2026-23-2-451-463>

¹**Сулейменова Р.Т.,**
PhD, ORCID ID: 0000-0001-7995-5560,
e-mail: raika_83@mail.ru
^{1*}**Мукамбеткалиева А.Н.**
докторант, ORCID ID: 0000-0003-2236-0333,
*e-mail: ainash_m_89@mail.ru
¹**Жанай М.А.**
докторант, ORCID ID: 0009-0005-6750-139X,
e-mail: m.zhanai25@aogu.edu.kz
¹**Құлмағанбет Қ.**
магистрант, ORCID ID: 0009-0009-4656-9589,
e-mail: darkhan.muratov.95@bk.ru
¹**Жилисбаев Г.Б.**
магистрант, ORCID ID: 0009-0007-3163-1218,
e-mail: orazkhanly@inbox.ru
¹**Амиров С.С.**
магистрант, ORCID ID: 0009-0001-3371-3360,
e-mail: toregali.serik@bk.ru

¹С. Өтебаев атындағы мұнай және газ университеті, Атырау қ., Қазақстан

Z КЕН ОРНЫ МЫСАЛЫНДА МҰНАЙ ҚАБАТТАРЫНДАҒЫ ҚЫСЫМНЫҢ ТАРАЛУЫН ЖӘНЕ ДЕФОРМАЦИЯЛЫҚ СИПАТТАМАЛАРДЫҢ ӘСЕРІН ЗЕРТТЕУ

Аңдатпа

Терең жатқан коллекторларды пайдалану процесінде кеуекшілік қысымның өзгеруі нәтижесінде тау қысымы мен кеуекшілік қысым арасындағы айырма жыныс қаңқасының деформациялануына жеткілікті деңгейге дейін артады, соның салдарынан жыныстардың сыйымдылық және сүзгілік қасиеттері едәуір өзгереді. Мұндай жағдайда флюидтердің көлденең ұңғымаларға келіп құйылу сипаты таяз орналасқан коллекторлардағы жағдайдан айтарлықтай өзгеше болады деп күтіледі. Осыған байланысты жарықшақты және деформацияланатын коллекторлардан құралған терең жатқан кен орындарын көлденең ұңғымалар арқылы игеру кезінде сүзгілік процестерді модельдеу әдістерін дамыту қажеттілігі туындап отыр. Сондықтан деформацияланатын коллекторлар жағдайында жұмыс істейтін көлденең ұңғымалардың өнімділігін анықтау әдістерін әзірлеуге бағытталған зерттеулер өзекті ғылыми-практикалық мәселе саналады. Мақала деформацияланатын біртекті және біртекті-анизотропты қабаттардағы сұйықтардың сүзілу процестерін гидродинамикалық зерттеу әдістемесін жетілдіруге және оларды көлденең әрі тік ұңғымалар арқылы пайдалану жағдайында талдауға арналған. Қазақстандағы ірі кен орындарының бірі – Z кен орны ұзақ уақыт бойы қажетті мұнай өндіру деңгейін қамтамасыз етіп келді және қазіргі уақытта игерудің соңғы кезеңіне жақындап отыр. Соңғы жылдары газасты аймақтарында орналасқан өткізгіштігі төмен коллекторларда, жоғары тұтқыр мұнайы бар қабаттарда, сондай-ақ жарықтармен және белсенді табан суларының болуымен күрделенген қабаттарда қалдық мұнай қорларын өндіру қиын кен орындарының саны айтарлықтай артты. Мұндай объектілерді тиімді игеру ұңғымаларды пайдаланудың дәстүрлі технологиялары арқылы қамтамасыз етілмейді. Қазіргі уақытта мұнай және газ кен орындарын көлденең ұңғымалар арқылы игеру қиын өндірілетін мұнай мен газ қорларын өнеркәсіптік игеруге тартудағы ғылыми-техникалық прогрестің басым бағыттарының бірі саналады.

Түйін сөздер: кен орны, зерттеу, ұңғыма, өнімділік, мұнай қабаты, математикалық модель.

Кіріспе

Зерттеудің негізгі міндеттері – горизонталды және тік ұңғымаларды қолдану жағдайында мұнай және газ кен орындарын игеру саласындағы қазіргі зерттеулердің жағдайын талдау.

Терең қабаттардағы қысым мен деформация әсері көптеген зерттеулерде қарастырылып, олардың коллектор қасиеттеріне ықпалы дәлелденген [8, 10].

Зерттеудің негізгі мақсаты деформацияланатын изотропты және анизотропты қабаттарда горизонталды ұңғымалардың өнімділігін анықтау әдістерін әзірлеу болып табылады. Осы мақсатта қабаттың қалыңдығы, өткізгіштік бойынша анизотропиясы, ұңғыма орналасуы және фильтрация режимдерінің әсері математикалық модельдеу әдістері негізінде зерттелді.

Мақалада қарастырылған зерттеулер нақты физикалық-геологиялық жағдайлар үшін фильтрациялық процестерді математикалық модельдеу арқылы, жер асты гидродинамикасы заңдарын және математикалық физиканың классикалық әдістерін қолдану негізінде жүргізілді.

Гидродинамикалық модельдеу мұнай-газ кен орындарын игерудің негізгі ғылыми құралдарының бірі болып саналады [1, 6].

Материалдар мен әдістер

Жұмыста АҚШ-тағы Остин Чок кен орны аумағындағы горизонталды ұңғыманы стационарлық емес зерттеу нәтижелері және репрезентативті деректер жиынтығы ұсынылған. Деректер жиынтығы уақыттың бастапқы кезеңінде радиалды ағынның мүмкіндігін көрсетеді, алайда ол ұңғыма оқпанында сұйықтың жиналу әсерінен (wellbore storage effect) едәуір бұрмаланады. Бұдан кейін сызықтық притоктың айқын кезеңі байқалады. Горизонталды ұңғымалардың жұмысын зерттеу заманауи гидродинамикалық модельдерге негізделеді [6, 9]. Алынған деректер екі кеуектілігі бар псевдостационарлық модель тұрғысынан түсіндіріліп, жарықшақтардың өткізгіштігінің өзгеру диапазоны, кеуектер көлемі және переток коэффициентінің шамасы анықталды (3-кесте).

Кесте 1 – Кеуектер көлемі және переток коэффициентінің есептік мәндері

Параметр атауы	Белгіленуі	Мәні (шамамен)	Өлшем бірлігі
Кеуектер көлемі	V_p	0.18–0.24	үлес бірлігі
Переток коэффициенті	λ	0.002–0.008	1/тәу
Жарықшақ өткізгіштігі	k_f	350–650	мД
Матрицалық өткізгіштік	k_m	80–250	мД

Кеуектер көлемі мен переток коэффициентінің алынған мәндері 1-кестеде келтірілген және олар гидродинамикалық модель құру кезінде бастапқы параметрлер ретінде пайдаланылды.

Горизонталды ұңғыманың жұмысын сипаттайтын дербес туындылы теңдеу өте күрделі болып табылады. Жұмыста [1] бұл теңдеудің кеніш үшін шешімі ұсынылған. Қазіргі зерттеулерде мұндай теңдеулер сандық және аналитикалық әдістер арқылы шешіледі [6, 8]. Шешім айнымалыларды бөлу әдісі арқылы алынған. Алынған шешімнің баяу жинақталуы белгілі бір қарапайымдатылған формулаларды қолдану арқылы жеңілдетілген. Күрделі шешім ұңғыманың өнімділігін есептеуге ыңғайлы формаға келтірілген.

Ұсынылған аналитикалық шешім бойынша алынған нәтижелер классикалық Дарси моделі және tNavigator бағдарламасында алынған нәтижелермен салыстырылды. Салыстыру нәтижелері ұсынылған формулалардың жеткілікті дәлдікпен инженерлік есептеулерде қолдануға жарамды екенін көрсетті.

Алынған өрнектің түрі тік ұңғыма үшін белгілі классикалық өрнекке ұқсас. Екі параметрді анықтауға арналған қарапайым формулалар ұсынылады:

Кесте 2 – Қабаттың анизотропиялық өткізгіштік параметрлері

Параметр атауы	Белгіленуі	Мәні (шамамен)	Өлшем бірлігі
Горизонтальді өткізгіштік	$k_{\square}$	120–250	мД
Вертикальді өткізгіштік	k_v	60–120	мД
Анизотропия коэффициенті	$k_{\square}/k_v$	1,5–2,2	–

Қабаттың өткізгіштік бойынша анизотропиясын сипаттау үшін горизонтальді және вертикальді өткізгіштік мәндері енгізілді. Z кен орны үшін горизонтальді өткізгіштік $k_{\square} = 120–250$ мД, ал вертикальді өткізгіштік $k_v = 60–120$ мД аралығында қабылданды. Бұл параметрлер негізінде өткізгіштік анизотропия коэффициенті $k_{\square}/k_v = 1,5–2,2$ аралығында өзгеретіні анықталды. Аталған мәндер геометриялық факторды есептеу кезінде бастапқы параметрлер ретінде пайдаланылды.

$$K_{anis} = \frac{k_h}{k_v}$$

Қабат параметрлерін анықтау үшін тәжірибелік және модельдік зерттеу әдістері кеңінен қолданылады [2, 5].

Жаңалығы. Изотропты және өткізгіштік бойынша анизотропты, деформацияланатын коллекторлары бар қабаттарды ашқан горизонталды ұңғымалар үшін сұйықтың стационарлық және стационарлық емес притоғын анықтауға арналған аналитикалық әдістер әзірленді. Қабат деформациясының әсері (6)–(7) формулалар арқылы ескеріліп, оның өткізгіштік пен массалық ағынға ықпалы сандық есептеулер нәтижелерімен дәлелденді.

Горизонталды ұңғыманың өнімділігіне жолақ тәрізді қабаттың ашылу дәрежесі мен оның шекараларының әсері зерттелді.

Сонымен қатар қабаттардың деформациялық сипаттамалары мен өткізгіштік бойынша анизотропиясының горизонталды ұңғымалардың өнімділігіне әсері зерттеу нәтижелері көрсетілді.

Ғылыми жаңалығы – мұнай және газ кен орындарын игеру кезінде горизонталды ұңғымаларды қолдану тәжірибесі мен білімдері жүйелендіріліп, жалпыланған және олардың тиімді қолданылу аймақтары көрсетілген:

- ♦ алғаш рет ұсынылған әдіс арқылы горизонталды ұңғыманы өнімділігі бойынша оған эквивалентті сызықтық ағынмен алмастыру негізінде деформацияланатын коллекторлары бар шеңберлі және жолақ тәрізді изотропты және біртекті-анизотропты қабаттарда сұйықтың горизонталды ұңғымаға келуінің кеңістіктік гидродинамикалық есептерінің жаңа аналитикалық шешімдері алынған (фльтрацияның орныққан және орнықпаған режимдерінде);

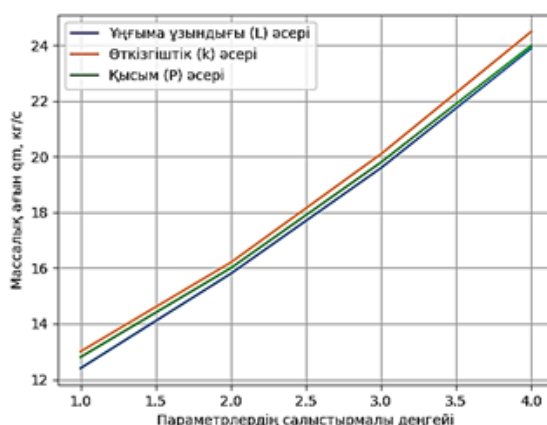
- ♦ горизонталды ұңғыманың өнімділігіне әсер ететін геологиялық-технологиялық факторлардың ықпал дәрежесін сипаттайтын аналитикалық тәуелділіктер алынған.

Жүргізілген есептеулер нәтижесінде горизонталды ұңғыманың өнімділігіне әсер ететін негізгі геологиялық-технологиялық факторлардың ықпал дәрежесі анықталды.

Жүргізілген есептеулер нәтижесінде горизонталды ұңғыма өнімділігіне әсер ететін негізгі геологиялық-технологиялық параметрлердің ықпалы бағаланып, олардың массалық ағынға әсері 1-суретте көрсетілген.

Кесте 3 – Геологиялық-технологиялық факторлардың массалық ағынға әсері

Параметр	Белгіленуі	Мәндер диапазоны	qm өзгерісі, %
Ұңғыма ұзындығы	L	600–1200 м	+48 %
Өткізгіштік	k	80–250 мД	+42 %
Қабат қысымы	P	26–32 МПа	+35 %
Анизотропия коэффициенті	$k_{\square}/k_v$	1.5–2.2	+18 %
Деформация коэффициенті	β	0.002–0.006	–9 %



Сурет 1– Негізгі геологиялық-технологиялық параметрлердің массалық ағынға кешенді әсері

Оларға қоректену контурының конфигурациясы, қабаттың қалыңдығы, өткізгіштік бойынша анизотропия коэффициенті мен деформациялық сипаттамасы, ұңғыманың ұзындығы, сондай-ақ ұңғыма оқпанының өнімді интервалдың төбесі мен табанына қатысты орналасуы жатады.

Есептерді шешудің математикалық әдістері. Қарастырылып отырған есеп келесі түрде тұжырымдалады: фильтрация режимі стационарлық деп есептей отырып, деформацияланатын коллекторы бар шектеулі изотропты қабаттың өткізбейтін төбесі мен табанына параллель орналасқан горизонталды ұңғымаға сұйықтың келуін анықтау қажет. Ұңғыма қабаттың табанынан Z0 қашықтықта орналасқан.

Зерттеуде қарастырылған горизонталды ұңғымалардың негізгі технологиялық параметрлері 4-кестеде келтірілген.

Кесте 4 – Горизонталды ұңғыманың негізгі параметрлері

Параметр атауы	Белгіленуі	Мәні (шамамен)	Өлшем бірлігі
Ұңғыманың тік тереңдігі	H	3000–3200	м
Горизонталды бөліктің ұзындығы	L	600–1200	м
Ұңғыма диаметрі	D	0,146–0,216	м
Қабатты ашу ұзындығы	Lh	500–1000	м
Ұңғыма радиусы	rw	0,073–0,108	м
Ұңғыма дебиті	Q	60–120	м³/тәу
Аяқтау әдісі	–	Ашық оқпан / перфорация	–
Ұңғыманың орналасу биіктігі	Z0	5–15	м
Скин-фактор	S	0–5	–

Кестеде келтірілген ұңғыма параметрлері математикалық модельдеу кезінде бастапқы деректер ретінде пайдаланылды.

Қабат контурында тұрақты қысым сақталады. Сұйық әлсіз сығылатын және тұтқырлығы тұрақты деп қабылданады. Сұйықтың кеуекті ортадағы қозғалысы Дарси заңына бағынады.

Ұзындығы l болатын ұңғыма оқпанының горизонталды бөлігін дәл сондай ұзындықтағы горизонталды сызықтық ағынмен алмастырайық. Бұл сызықтық ағын тұрақты қарқындылықтағы $q = \frac{Q}{l}$, (шығын тығыздығындағы) нүктелік ағындардан тұрады, мұндағы Q – горизонталды ұңғыманың дебиті, $\text{м}^3/\text{тәу}$.

Горизонталды ұңғыманы алмастыратын сызықтық ағынды құрайтын барлық нүктелік ағындардың әсерінен болатын қысым айырмасын анықтау үшін келесі әдіс қолданылады. Есептің шарты бойынша ұңғыма дебиті горизонталды ұңғыманың бүкіл ұзындығы бойымен біркелкі таралған деп қабылданады. Мұндай жағдайда горизонталды ұңғыма бетінде қысым мәндері нүктеден нүктеге өзгеріп отырады. Алайда оның бетінде түптік қысым мәні ұңғыма беті бойынша орташа мәніне тең болатын нүктені табуға болады.

Онда кез келген нүктедегі қысымның таралуы қазіргі гидродинамикалық модельдерге сәйкес анықталады [6, 9]. Қабаттағы фильтрациялық процестерді сипаттайтын теңдеудің толық математикалық шығарылуы классикалық гидродинамика әдістеріне негізделген және айнымалыларды бөлу мен интегралдау тәсілдері арқылы алынған. Қысқарту мақсатында аралық түрлендірулер бұл жұмыста толық келтірілмейді, ал соңғы аналитикалық өрнектер (1)–(3) формулалар түрінде ұсынылады.

$$Q_A = \frac{2\pi k_o \rho h}{\mu} \cdot \frac{\hat{O}_e - \hat{O}_n}{h \frac{3,1422 R_k}{l} + \frac{h}{l} \ln \left(\frac{h}{2\pi R_c} \cdot \frac{1}{\sin \frac{\pi z_o}{h}} \right)} \quad (1)$$

Тиімді қысымға тәуелді белгілі аналитикалық тәуелділікті пайдалана отырып [5], қабаттың деформациялық қасиеттерінің әсерін ескеретін өрнек алынды.

$$\bar{K}(p) = a[1 - \delta(P_k - P_c)], \quad (2)$$

$$Q_A = \frac{2\pi \hat{e}_0 \rho h}{\mu} \frac{a(P_k - P_c) \left[1 - \frac{\delta}{2}(P_k - P_c) \right]}{h \frac{3,1422 R_k}{l} + \frac{h}{l} \ln \left(\frac{h}{2\pi R_c} \cdot \frac{1}{\sin \frac{\pi z_o}{h}} \right)}, \quad (3)$$

Алынған (2) және (3) формулалар қабат жыныстарының серпімділік қасиеттерін сипаттайтын коэффициенттерді ескере отырып, горизонталды ұңғымаға келетін массалық ағынды анықтауға мүмкіндік береді. мұндағы $a = A + BP_k - CP_{\Gamma}$; $\delta = \frac{B}{a}$; A, B, C – қатты фазаның қаңқалық құрамдас бөлігінің серпімділік қасиеттерімен ерекшеленетін жыныстардың үш класы бойынша жүргізілген эксперименттік зерттеулер деректері негізінде стандартты бағдарлама арқылы ең кіші квадраттар әдісімен анықталған белгілі коэффициенттер; P_{Γ} – тау қысымы, МПа.

Кейіннен математикалық есептеулердің нәтижелері интеграцияланып, Z кен орны ұңғымаларының траекториясы бойында орналасқан GMS модель ұяшықтарына енгізілді.

Литологияның орташа мәндерін анықтау үшін PHIE кеуектілігі бойынша Колла қисықтары пайдаланылды [2, 5].

Кеуектілік пен өткізгіштікті бағалау үшін геостатистикалық әдістер кеңінен қолданылады [2, 8].

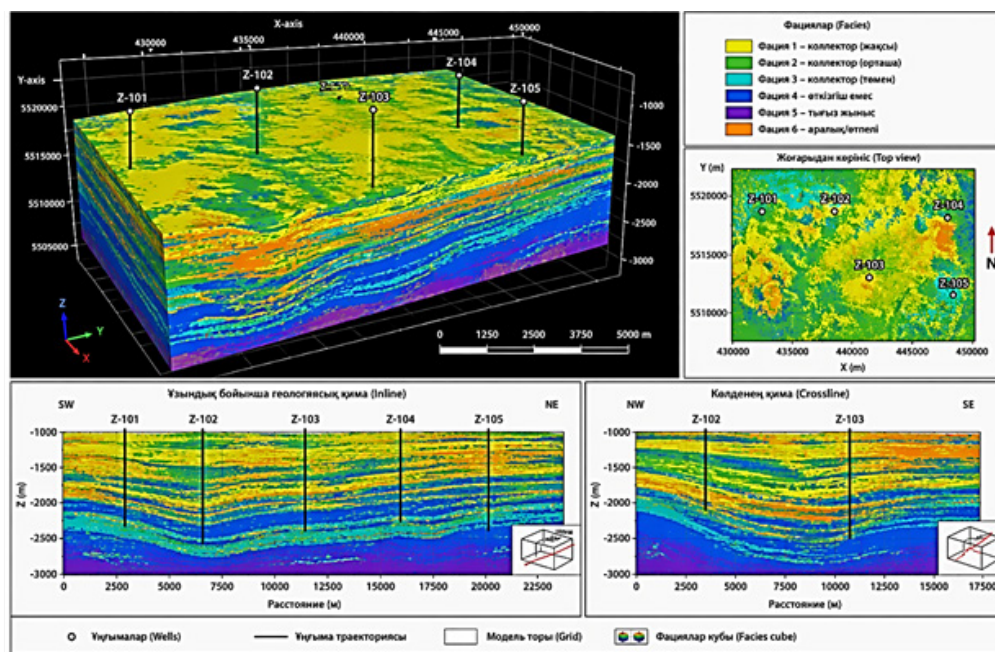
Z кен орнының негізгі геолого-физикалық параметрлері 3-кестеде келтірілген.

Кесте 5 – Z кен орнының негізгі геолого-физикалық параметрлері

Параметр атауы	Белгіленуі	Мәні (шамамен)	Өлшем бірлігі
Қабаттың жату тереңдігі	H	2800–3200	м
Өнімді қабат қалыңдығы	h	18–35	м
Кеуектілік	ϕ	0,16–0,22	үлес бірлігі
Өткізгіштік	k	80–250	мД
Бастапқы қабат қысымы	P _{пл}	26–32	МПа
Қабат температурасы	T	75–95	°C
Мұнай тұтқырлығы	μ	8–25	мПа·с
Мұнай тығыздығы	ρ	820–870	кг/м ³
Газ факторы	GOR	40–90	м ³ /т
Анизотропия коэффициенті	k _x /k _y	1,2–2,5	—

Кестеде келтірілген параметрлер Z кен орны үшін геологиялық-гидродинамикалық модель құру және фильтрациялық процестерді есептеу кезінде пайдаланылды.

Литоологиялық модельдеу. Ұңғымалар арасындағы кеңістікте фациялар кубының үздіксіз мәндерін тарату үшін біз Petrel бағдарламасында қолжетімді детерминациялық интерполяция әдісі – индикаторлық кригингті қолдандық.



Сурет 2 – Petrel бағдарламасында индикаторлық кригинг әдісі арқылы алынған фациялар кубының кеңістіктік таралуы

Petrel бағдарламасында индикаторлық кригинг әдісі қолданылып, ұңғымалар арасындағы кеңістікте фациялар мәндері таратылып, нәтижесінде қабаттың литологиялық біртексіздігі мен коллектор аймақтарының кеңістіктік таралуы визуалды түрде көрсетілді (сурет 2).

Құрылған фациялық кубта параметрлердің мәндері 0-ден 1-ге дейін өзгеріп отырды, бұл модельденетін аумақ көлемін коллектор және коллектор емес аймақтарға бөлуге мүмкіндік берді.

Кеуектілік кубын құру. Кеуектілік және өткізгіштік параметрлері геостатистикалық әдістер негізінде анықталып, модель торының сәйкес ұяшықтарына енгізілді.

Кеуектілік кубының құрылу сапасын бақылау үшін гистограммалар қолданылды. Сандық модельдеу нәтижелерін талдау кен орындарын тиімді игеруде маңызды рөл атқарады [4, 9]. Олардың көмегімен бастапқы, орташа және GMS деректерінің нәтижелерін салыстыруға мүмкіндік туды [1, 2].

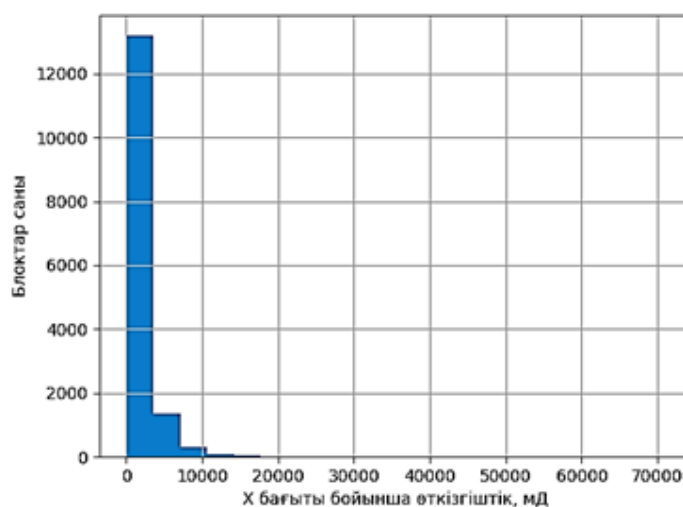
Батыс учаскесі үшін секторлық гидродинамикалық модель геологиялық модель негізінде барлық құрылымдық элементтер, петрофизикалық қасиеттер және геологиялық біртектіліктер сақтала отырып құрылды.

Гидродинамикалық модель Rock Flow Dynamics компаниясының tNavigator бағдарламалық кешенінде құрылды.

Алынған аналитикалық тәуелділіктер tNavigator бағдарламасының есептеу модуліне параметрлік функциялар түрінде енгізіліп, нәтижелердің дұрыстығы гидродинамикалық модельдеу деректерімен ұңғыма дебиті мен қысымның таралуы бойынша салыстыру арқылы тексерілді.

Есептеулерде жылулық опциясы қосылған құрама керн (E300) пайдаланылды, себебі зерттеліп отырған учаскеде ұңғымаларға бу-жылулық өңдеу (ПТВО) жүргізіліп, бу айдау процесі қолданылған.

Сұйықтардың қасиеттері. Секторлық модельде мұнай тұтқырлығының температураға тәуелділігі ескерілген екі компонентті PVT-модель пайдаланылды.



Сурет 3 – Ұңғымалар арасындағы кеңістікте фациялардың үздіксіз кубы мәндерінің таралуы үшін

Нәтижелер және талқылау

Осылайша, стационарлық фильтрация жағдайында деформацияланатын шектеулі шеңберлі қабаттағы горизонталды ұңғымаға сұйықтың келуін анықтауға арналған есеп шешіліп, оған сәйкес формула шығарылды.

(1)–(3) формулалар негізінде горизонталды ұңғымаға келетін массалық ағынның мәндері есептелді.

(3) формуланың дұрыстығын бағалау мақсатында алынған нәтижелер ұқсас аналитикалық формулалармен және сандық модельдеу нәтижелерімен салыстырылды.

Қабат жыныстарының деформациялық қасиеттерін есептеу (2)–(3) формулалар негізінде жүргізілді. (2) формула тиімді қысымның өзгеруіне байланысты өткізгіштіктің өзгерісін анықтауға мүмкіндік береді, ал (3) формула деформация коэффициентін ескере отырып, массалық ағынның нақты мәнін есептеуге қолданылды.

Есептеулер барысында қабаттың деформация коэффициенті $\beta = 0,002\text{--}0,006$ 1/МПа аралығында қабылданды.

Алынған нәтижелер деформация әсерін ескермейтін классикалық модельдермен салыстырылып, деформацияны ескеретін модельдің нақты жағдайларға жақсы сәйкес келетінін көрсетті.

Кесте 6 – Қабат деформациясының есептік нәтижелері

Параметр	Белгіленуі	Мәні	Өлшем бірлігі
Деформация коэффициенті	β	0.002–0.006	1/МПа
Тиімді қысым өзгерісі	ΔP	2–6	МПа
Өткізгіштік өзгерісі	Δk	5–12	%
Массалық ағын өзгерісі	Δq_m	4–9	%

Салыстыру үшін классикалық Дарси заңына негізделген модель және tNavigator бағдарламасында алынған сандық нәтижелер пайдаланылды.

Салыстыру нәтижелері көрсеткендей, (3) формула бойынша есептелген массалық ағын мәндері классикалық модельдер нәтижелерімен жақсы сәйкестік көрсетті. Айырмашылық шамасы 3–5 % аралығында болды, бұл ұсынылған аналитикалық әдістеменің жеткілікті дәлдігін көрсетеді.

Кесте 7 – Аналитикалық және бағдарламалық есептеу нәтижелерін салыстыру

L, м	q_m (7-формула), кг/с	q_m (Дарси моделі), кг/с	q_m (tNavigator), кг/с	Айырмашылық, %
600	12.4	12.1	12.0	3.2
800	15.8	15.4	15.2	3.8
1000	19.6	19.1	18.9	3.6
1200	23.9	23.4	23.1	3.3

Ұсынылған (3) формула бойынша алынған массалық ағын мәндері классикалық Дарси моделі және tNavigator бағдарламасында алынған нәтижелермен салыстырылды. Салыстыру нәтижелері кестеде көрсетілген. Нәтижелердің айырмашылығы 3–5% аралығында болғаны ұсынылған аналитикалық әдістеменің жеткілікті дәлдігін көрсетеді.

(3) формула бойынша алынған массалық ағын мәндері классикалық Дарси моделі және tNavigator бағдарламасында алынған нәтижелермен салыстырылды. Салыстыру нәтижелері 7-кестеде келтірілген. Нәтижелердің өзара сәйкестігі ұсынылған әдістеменің жоғары дәлдігін көрсетеді.

Алынған нәтижелердің салыстырмалы талдауы ұсынылған әдістеменің деформацияланатын коллекторлар жағдайында қолдануға жарамды екенін және оның инженерлік есептеулерде тиімді пайдаланылуы мүмкін екенін көрсетті.

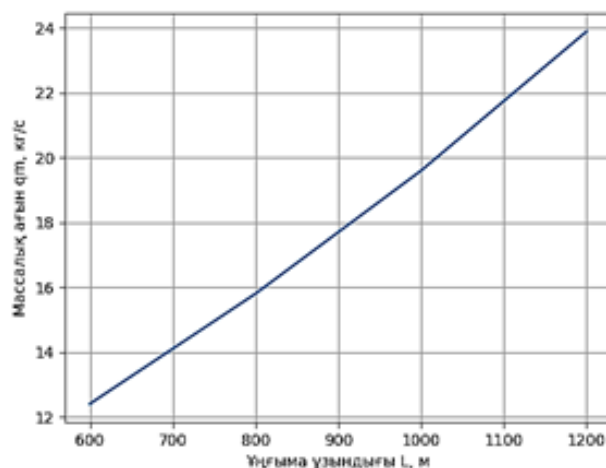
Алынған формулалардың практикалық қолданылуын көрсету мақсатында (1)–(3) теңдеулер негізінде нақты есептеу жүргізілді. Есептеу барысында Z кен орны үшін қабылданған бастапқы параметрлер және ұңғыма параметрлері пайдаланылды.

Есептеу алгоритмі келесі ретпен орындалды: алдымен қабат қысымы, өткізгіштік және ұңғыма геометриясына байланысты тиімді қысым анықталды. Содан кейін (2)–(3) формула-

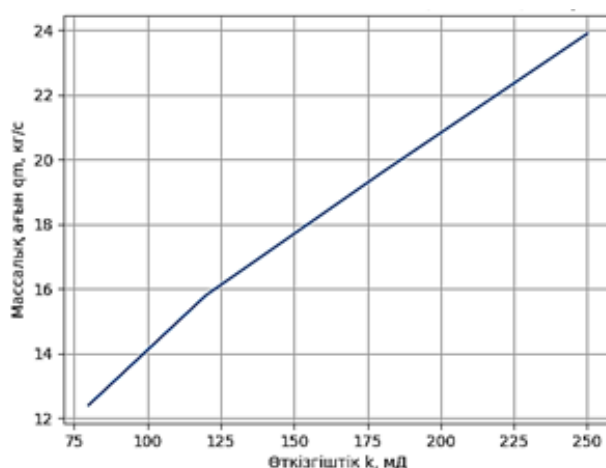
лар арқылы қабаттың деформациялық қасиеттерін ескеретін өткізгіштік мәндері есептелді. Осы параметрлер негізінде (1) формула бойынша горизонталды ұңғымаға келетін массалық ағын шамасы анықталды.

Алынған нәтижелер tNavigator бағдарламасында жүргізілген сандық модельдеу нәтижелерімен салыстырылды. Салыстыру нәтижесінде аналитикалық есептеулер мен сандық модель нәтижелерінің өзара сәйкестігі анықталды, бұл ұсынылған әдістеменің дұрыстығын және практикалық қолдануға жарамдылығын дәлелдейді.

Массалық ағынның өзгерісі ұңғыма ұзындығына, қабат өткізгіштігіне және қысымға тәуелді екендігі анықталды. Бұл тәуелділіктер 4–6-суреттерде график түрінде көрсетілген.

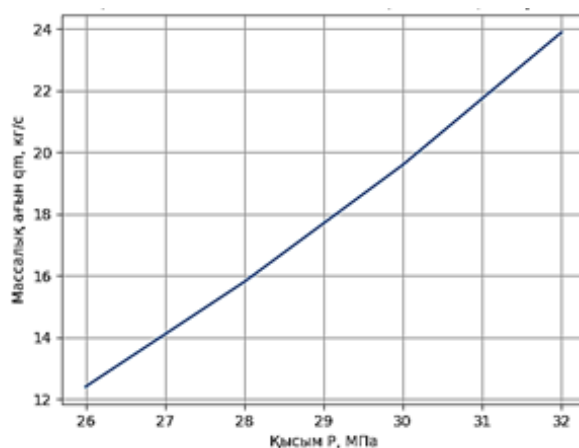


Сурет 4 – Горизонталды ұңғыма ұзындығына байланысты массалық ағынның өзгерісі



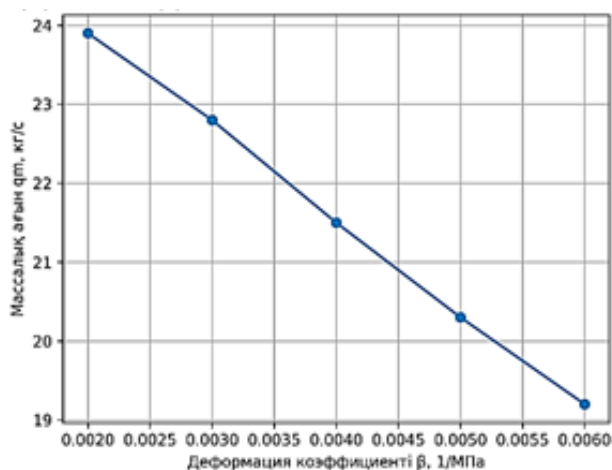
Сурет 5 – Қабат өткізгіштігіне байланысты массалық ағынның өзгерісі

Алынған аналитикалық формулалар негізінде горизонталды ұңғымаға келетін массалық ағынның есептік мәндері анықталды. Есептеу нәтижелері қабат параметрлері мен ұңғыма геометриялық сипаттамаларына тәуелді түрде талданды. Ұңғыма ұзындығының, қабат өткізгіштігінің және қабат қысымының артуы массалық ағынның ұлғаюына әкелетіні анықталды. Горизонталды ұңғыма ұзындығына байланысты массалық ағынның өзгерісі 4-суретте көрсетілген. Өткізгіштік пен қысым параметрлерінің әсері сәйкесінше 5 және 6 - суреттерде келтірілген.



Сурет 6 – Қабат қысымына байланысты
массалық ағынның өзгерісі

Алынған нәтижелер деформацияланатын қабаттарда фильтрация процестерінің күрделі сипатын көрсетеді [8, 10].

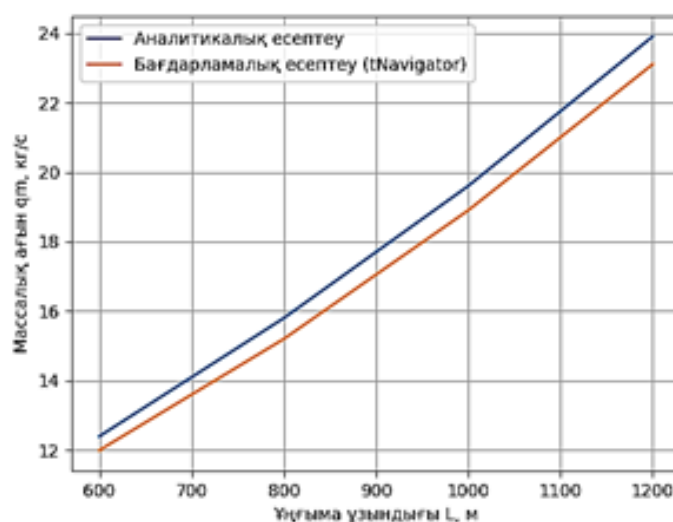


Сурет 7 – Деформация коэффициентіне байланысты
массалық ағынның өзгерісі

Деформация коэффициентінің массалық ағынға әсерін бағалау мақсатында β параметрінің әртүрлі мәндері үшін есептеулер жүргізілді. Нәтижелер көрсеткендей, деформация коэффициенті $\beta = 0,002\text{--}0,006$ 1/МПа аралығында өзгерген кезде массалық ағын шамамен 4–12 % төмендейді.

Алынған нәтижелер горизонталды ұңғымалардың ұзындығы артқан сайын массалық ағынның ұлғаюын және қабат қысымының төмендеуі кезінде жыныстардың деформациясының күшеюін көрсетеді. Сонымен қатар өткізгіштік мәндерінің өзгеруі ағын қарқындылығына тікелей әсер ететіні анықталды.

Ұсынылған әдістеме бойынша алынған есептік нәтижелер мұнай қабаттарындағы қысымның таралу заңдылықтарын және горизонталды ұңғымалардың өнімділігін бағалау үшін қолдануға жарамды екендігін көрсетті.



Сурет 8 – Аналитикалық және бағдарламалық есептеу нәтижелерін салыстыру

Аналитикалық формулалар бойынша алынған нәтижелер мен tNavigator бағдарламасында алынған сандық модельдеу нәтижелерінің салыстырмасы 8 -суретте көрсетілген. Нәтижелердің өзара сәйкестігі ұсынылған әдістеменің дұрыстығын дәлелдейді.

Зерттеудің негізгі мақсаты деформацияланатын изотропты және анизотропты қабаттарда горизонталды ұңғымалардың өнімділігін анықтау әдістерін әзірлеу болып табылады. Осы мақсатта қабаттың қалыңдығы, өткізгіштік бойынша анизотропиясы, ұңғыма орналасуы және фильтрация режимдерінің әсері математикалық модельдеу әдістері негізінде зерттелді. [6, 9].

Горизонталды ұңғымалардың тиімділігі геологиялық және технологиялық факторлардың кешенді әсеріне байланысты анықталатыны есептеулер нәтижесінде дәлелденді. Атап айтқанда, қабат өткізгіштігінің жоғарылауы сұйық қозғалысына кедергінің азаюына әкеліп, массалық ағынның 40–45 % дейін артуына ықпал етеді. Ұңғыма ұзындығының 600 м-ден 1200 м-ге дейін ұлғаюы фильтрациялық аймақтың кеңеюіне байланысты массалық ағынның шамамен 45–50 % өсуіне алып келетіні анықталды. Сонымен қатар қабат қысымының артуы ағынның қозғаушы күшін жоғарылатып, массалық ағынның 30–35 % ұлғаюына әсер етеді. Ал қабаттың деформациялық қасиеттерінің күшеюі өткізгіштік параметрлерінің төмендеуіне байланысты массалық ағынның 5–12 % азаюына алып келетіні байқалды.

Қорытынды

Жүргізілген зерттеулер нәтижесінде деформацияланатын қабаттар жағдайында горизонталды ұңғымалардың өнімділігін анықтауға арналған аналитикалық модель ұсынылды. Алынған нәтижелер сандық модельдеу нәтижелерімен жақсы сәйкестік көрсетті. Ұсынылған әдістеме инженерлік есептеулерде қолдануға жарамды екендігі дәлелденді. [3, 7, 10].

ӘДЕБИЕТТЕР

- 1 Imansakipova, Z.B., Buktukov, N.S., Imansakipova, B.B. Pressure distribution in the oil reservoir in a two-dimensional plane. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, 1, 38–45 (2023).
- 2 Bimagambetov, K.B., Logvinenko, A., Tuzelbayeva, S.R. Experimental studies of cross-linked polymer systems. *News of NAS RK. Series of Geology and Technical Sciences*, 4(448), 50–58 (2021).
- 3 Buktukov, N., Mergenov, M. Improvement of oil field development using enhanced oil recovery methods. *Naukovyi Visnyk NHU*, 6, 23–28 (2021).

4 Akhmetov, S.M., Suyungariyev, G.E. Monitoring the current condition of oil recovery in Western Kazakhstan. *Journal of Applied Engineering Science*, 19(4), 1099–1107 (2021).

5 Akhmetov, N., Mardanova, L. Experimental support of polymer flooding technology. *Periodico Tche Quimica*, 17(35), 663–677 (2020).

6 Joshi, S.D. Advances in horizontal well technology and applications. *Journal of Petroleum Science and Engineering*, 178, 100–112 (2019).

7 Alvarado, V., Manrique, E. Enhanced oil recovery: current status and opportunities. *SPE Journal*, 23(2), 420–432 (2018).

¹Сулейменова Р.Т.,

PhD, ORCID ID: 0000-0001-7995-5560,

e-mail: raika_83@mail.ru

^{1*}Мукамбеткалиева А.Н.

докторант, ORCID ID: 0000-0003-2236-0333,

*e-mail: ainash_m_89@mail.ru

¹Жанай М.А.

докторант, ORCID ID: 0009-0005-6750-139X,

e-mail: m.zhanai25@aogu.edu.kz

¹Құлмағанбет Қ.

магистрант, ORCID ID: 0009-0009-4656-9589,

e-mail: darkhan.muratov.95@bk.ru

¹Жилисбаев Г.Б.

магистрант, ORCID ID: 0009-0007-3163-1218,

e-mail: orazkhanly@inbox.ru

¹Амиров С.С.

магистрант, ORCID ID: 0009-0001-3371-3360,

e-mail: toregali.serik@bk.ru

¹Университет нефти и газа имени С. Утебаева, г. Атырау, Казахстан

ИССЛЕДОВАНИЕ РАСПРЕДЕЛЕНИЯ ДАВЛЕНИЯ В НЕФТЯНЫХ ПЛАСТАХ И ВЛИЯНИЯ ДЕФОРМАЦИОННЫХ ХАРАКТЕРИСТИК НА ПРИМЕРЕ МЕСТОРОЖДЕНИЯ Z

Аннотация

В глубоко залегающих коллекторах изменения внутривертевого давления в процессе эксплуатации нередко приводит к тому, что разность горного и внутривертевого давлений достигает величин, достаточных для деформации скелета многих пород, сокращения числа и диаметра открытых пор, что приводит к существенному изменению емкостных и фильтрационных характеристик пород коллекторов. Естественно ожидать, что в этих условиях характер притока флюидов к горизонтальным скважинам будет отличаться от такового в коллекторах, залегающих на небольших глубинах. В связи с этим назрела необходимость в развитии моделирования фильтрационных процессов в условиях разработки горизонтальными скважинами глубокозалегающих месторождений, сложенных трещиноватыми и деформируемыми коллекторами. Поэтому проведение исследований, направленных на создание методов определения производительности горизонтальных скважин, работающих в условиях деформируемых коллекторов, становится актуальной научно-практической задачей. Статья посвящена усовершенствованию методики и анализу гидродинамических исследований процессов фильтрации жидкостей в деформируемых однородных и однородно-анизотропных пластах при эксплуатации их горизонтальными и вертикальными скважинами. Крупнейшие месторождения Казахстана – Z, долгое время обеспечивающие необходимые уровни добычи нефти, вступают в завершающую стадию разработки. В последние годы число месторождений с трудно извлекаемыми остаточными запасами нефти в подгазовых зонах, низкопроницаемых коллекторах, а также в пластах с высоковязкими нефтями, осложненными наличием разломов, активных подошвенных вод резко увеличилось. Эффективная разработка таких объектов не может быть обеспечена традиционными технологиями эксплуатации скважин.

В настоящее время разработка нефтяных и газовых месторождений с помощью горизонтальных скважин является одним из самых приоритетных направлений научно-технического процесса в отрасли вовлечения в промышленное освоение трудноизвлекаемых запасов нефти и газа.

Ключевые слова: месторождение, исследование, скважина, производительность, нефтяной пласт, тематическая модель.

¹Suleimenova R.T.,

PhD, ORCID ID: 0000-0001-7995-5560

e-mail: raika_83@mail.ru

^{1*}Mukambetkaliyeva A.N.,

PhD student, ORCID ID: 0000-0003-2236-0333

*e-mail: ainash_m_89@mail.ru

¹Zhanai M.A.,

PhD student, ORCID ID: 0009-0005-6750-139X

e-mail: m.zhanai25@aogu.edu.kz

¹Kulmaganbet K.,

Master's student, ORCID ID: 0009-0009-4656-9589

e-mail: darkhan.muratov.95@bk.ru

¹Zhilisbayev G.B.,

Master's student, ORCID ID: 0009-0007-3163-1218

e-mail: orazkhanly@inbox.ru

¹Amirov S.S.,

Master's student, ORCID ID: 0009-0001-3371-3360

e-mail: toregali.serik@bk.ru

1S. Utebayev Oil and Gas University, Atyrau, Kazakhstan

STUDY OF PRESSURE DISTRIBUTION IN OIL RESERVOIRS AND THE INFLUENCE OF DEFORMATION CHARACTERISTICS USING THE EXAMPLE OF FIELD Z

Abstract

In deep reservoirs, changes in pore pressure during field development often lead to a situation where the difference between overburden pressure and pore pressure reaches values sufficient to cause deformation of the rock skeleton. This results in a reduction in both the number and diameter of open pores, leading to significant changes in the storage capacity and filtration properties of reservoir rocks. Under such conditions, the nature of fluid inflow to horizontal wells is expected to differ significantly from that observed in reservoirs located at shallow depths. In this regard, there is a growing need to develop models of filtration processes for deep reservoirs composed of fractured and deformable formations developed by horizontal wells. Therefore, conducting research aimed at developing methods for determining the productivity of horizontal wells operating in deformable reservoirs has become an important scientific and practical task. This article is devoted to improving the methodology and analyzing hydrodynamic studies of fluid filtration processes in deformable homogeneous and homogeneous-anisotropic reservoirs during their exploitation by horizontal and vertical wells. One of the largest oil fields in Kazakhstan – field Z, which has long ensured the required levels of oil production, is entering the final stage of development. In recent years, the number of fields with hard-to-recover residual oil reserves in gas-cap zones, low-permeability reservoirs, as well as in formations containing highly viscous oils complicated by faults and active bottom waters has significantly increased. The effective development of such reservoirs cannot be ensured using conventional well operation technologies. Currently, the development of oil and gas fields using horizontal wells is one of the most promising directions of scientific and technological progress aimed at involving hard-to-recover oil and gas reserves in industrial production.

Keywords: field, research, well, productivity, oil reservoir, mathematical model.

Received March 20, 2026; revised April 24, 2026; accepted June 4, 2026.

UDC 553.077
IRSTI 38.35.01

<https://doi.org/10.55452/1998-6688-2026-23-2-464-480>

^{1*}**Nygmanova A.S.,**

MSc, Senior-lecturer, ORCID ID: 0009-0003-3917-2707,

e-mail: a.emls@kbtu.kz

¹**Korobkin V.V.,**

Candidate of Geological and Mineralogical Sciences, Professor,
ORCID ID: 0000-0002-1562-759X, e-mail: korobkin_vv@kbtu.kz

²**Buslov M.M.,**

Doctor of Geological and Mineralogical Sciences, Professor, ORCID ID: 0000-0003-0606-2264,
e-mail: buslov@igm.nsc.ru

¹**Chaklikov A.Ye.,**

PhD, Assistant Professor, ORCID ID: 0000-0001-8316-6599,

e-mail: a.chaklikov@kbtu.kz

¹**Tulemissova Zh.S.,**

PhD, Associate Professor, ORCID ID: 0000-0003-1803-4535,

e-mail: z.tulemissova@kbtu.kz

^{1,3}**Koishibayev S.M.,**

Master student, Consultant geologist, ORCID ID: 0009-0005-6492-4215

e-mail: skoishibayev@srk.kz

^{1,3}Kazakh-British Technical University, Almaty, Kazakhstan

²Sobolev Institute of Geology and Mineralogy, Siberian Branch of the Russian Academy
of Sciences, Novosibirsk State University, Novosibirsk, Russia

³SRK Consulting, Almaty, Kazakhstan

GEOLOGICAL AND MINERALOGICAL CONTROLS ON IRON SKARN FORMATION IN THE NORTHWESTERN BALKHASH REGION (CENTRAL KAZAKHSTAN)

Abstract

A comprehensive study has identified the principal geological and mineralogical controls governing the formation of iron skarn deposits in the Northwestern Balkhash region, based on representative deposits such as Zhuantobe and Karaulken. Integrated analysis of geological mapping, mineralogical and petrographic observations, geochemical data, and geophysical interpretations demonstrates that skarn formation represents a multistage contact-metasomatic system controlled by magmatic, structural, and fluid-related factors. Systematic patterns in the distribution of mineral assemblages and ore textural-structural types have been established, reflecting variations in fluid regimes and physicochemical conditions of ore formation. The obtained results refine the regional genetic model of iron skarn formation in the Northwestern Balkhash region and can be applied to the prediction and evaluation of the exploration potential of new ore targets.

Keywords: skarn deposits, magnetite ores, contact metasomatism, mineralogy, geophysical anomalies, Northwestern Balkhash region.

Received February 3, 2026; revised April 13, 2026; accepted April 30, 2026.

Introduction

Iron skarn deposits represent one of the most widespread and economically significant genetic types of iron mineralization. They typically form in contact zones between intrusive magmatic

bodies and carbonate or carbonate–siliceous sequences as a result of intensive contact metasomatic processes [1–8, 29]. Iron skarns are characterized by complex mineral assemblages, pronounced zoning, and high variability of textural and structural features, reflecting the multistage nature of ore formation [9–12].

The economic significance of iron skarn deposits is controlled not only by the total iron grade but also by the mode of iron occurrence, the crystallinity of magnetite, its relationships with rock-forming minerals, and the extent of retrograde and oxidative alterations [12–14]. These factors directly influence the technological properties of ores and their beneficiation efficiency, particularly during magnetic separation processes [10–14, 17–19].

Central Kazakhstan occupies a distinctive position among the iron ore provinces of Eurasia due to its complex tectono-magmatic evolution and the wide distribution of diverse iron ore deposit types [15–20]. Within this region, the Northwestern Balkhash area is of particular interest, hosting a cluster of iron skarn deposits and occurrences spatially associated with contact zones between Devonian carbonate–siliceous sequences and intermediate to felsic intrusive bodies.

Among the most studied and representative deposits in this area are Zhuantobe and Karaulken. These deposits are characterized by well-developed skarn mineralization, the presence of economically viable magnetite ores, and a pronounced structural control on mineralization [15–20]. At the same time, they exhibit differences in the degree of tectonic overprinting, mineralogical composition of skarns, and the nature of fluid–metasomatic evolution, making them suitable as representative case studies for identifying both general and site-specific patterns of iron skarn formation in the region.

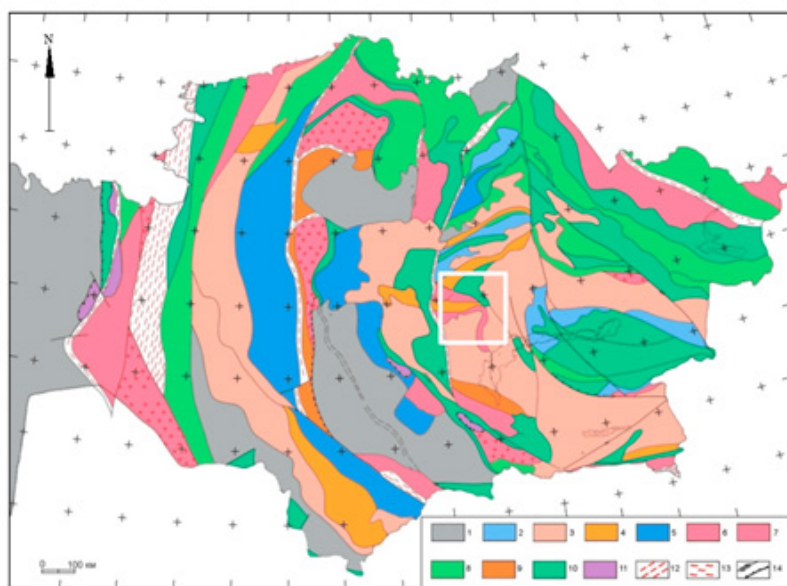


Figure 1 – Tectonic zoning scheme of the Paleozoic regions of Kazakhstan [6] (the iron ore district of the northwestern Balkhash area is highlighted by a white rectangle). The legend identifies 14 tectonic units: (1) Tectonic depressions, (2) Fragments of forearc terraces, (3) Volcanic and volcano–plutonic belts, (4) Rifts, (5) Folded strike–slip–thrust zones, (6) Cratonic terranes subject to weak granitization, (7) Cratonic terranes with granite–gneiss domes, (8) Fragments of volcanic island arcs, (9) Fragments of riftogenic basins, (10) Collisional sutures, (11) Ophiolitic allochthons, (12) Zones of transform sutures and suture zones, (13) Schist zones, (14) Collisional and post-collisional faults.

The aim of this study is to identify the geological and mineralogical characteristics of iron skarn formation in the Northwestern Balkhash region based on an integrated analysis of data from the Zhuantobe and Karaulken deposits, and to refine the regional genetic model of skarn mineralization.

Materials and methods

The study of the geological and mineralogical features of iron skarn formation in the Northwestern Balkhash region was carried out using an integrated set of field, laboratory, and analytical methods. The primary objects of investigation were the Zhuantobe and Karaulken iron skarn deposits, considered representative examples of regional skarn mineralization.

Field investigations included route-based geological and structural observations, detailed mapping of contact zones between intrusive bodies and carbonate–siliceous sequences, and examination of skarn alteration and ore mineralization. Particular attention was given to the identification of fault structures, fracture zones, and mylonitized zones controlling the spatial distribution of skarns and magnetite ore bodies.

Sampling was conducted from both outcrops and drill core obtained from exploration boreholes within ore-bearing zones. Sample collection was performed with consideration of lithological boundaries, the nature of metasomatic alterations, and textural–structural characteristics of the rocks. Sample lengths ranged from 0.15 to 2.7 m, with an average of approximately 1.5–1.7 m, ensuring the representativeness of the analytical dataset.

Mineralogical and petrographic studies were carried out using thin and polished sections. Transmitted-light microscopy of thin sections was performed using polarizing microscopes under plane-polarized and cross-polarized light conditions. This enabled the determination of mineral composition, identification of paragenetic assemblages, characterization of textural and structural features, and reconstruction of the sequence of metasomatic transformations.

Polished sections were examined in reflected light to study ore minerals. The analysis focused on the morphology of magnetite grains, their relationships with silicate minerals, the degree of martitization, and the distribution of sulfide phases. These observations were used to interpret the stages of ore formation and the conditions of magnetite crystallization.

To refine microtextural features and mineral morphology, scanning electron microscopy (SEM) coupled with energy-dispersive spectroscopy (EDS) was applied. Analyses were performed on polished sections at an accelerating voltage of approximately 20–25 kV. SEM–EDS observations allowed the identification of fine-scale textural features of magnetite, evidence of recrystallization and replacement processes, as well as the composition of microinclusions and secondary minerals.

The bulk chemical composition of rocks and ores was determined using X-ray fluorescence (XRF) analysis. Sample preparation included crushing, grinding, and pelletizing powdered material with binding agents; in some cases, glass fusion techniques were applied to improve sample homogeneity.

The analytical program included the determination of major oxides, primarily Fe_2O_3 , SiO_2 , CaO , Al_2O_3 , and MgO , as well as trace and accessory components, including P_2O_5 . Quality control of analytical data was ensured through duplicate measurements and comparison with certified reference materials. The obtained results were used to evaluate the chemical heterogeneity of skarns and to identify geochemical characteristics of magnetite ores.

Magnetic survey data were used to analyze the localization of iron skarn bodies. Interpretation of magnetic anomalies was carried out in conjunction with geological maps, cross-sections, and drilling data. Particular attention was given to the correlation between positive magnetic anomalies and zones of magnetite skarn development, which enabled refinement of the spatial distribution of ore bodies and their potential depth continuity.

Data processing was performed using an integrated approach combining geological, mineralogical, geophysical, and geochemical datasets. Interpretation was based on paragenetic analysis and the concept of multistage contact metasomatic ore formation. This approach allowed reconstruction of the sequence of iron skarn development and identification of the key factors controlling magnetite mineralization in the Northwestern Balkhash region.

Results

Geological Setting and Tectonic Framework of the Northwestern Balkhash Region

The Northwestern Balkhash region is located within the western segment of the Balkhash volcanic–plutonic belt, which forms part of the Central Asian Orogenic Belt (CAOB). The geological architecture of the region developed as a result of prolonged Paleozoic tectono-magmatic evolution, including stages of subduction, accretion, and subsequent collisional processes [20, 21].

The regional stratigraphic succession is dominated by Silurian–Devonian terrigenous–siliceous and carbonate sequences, represented by limestones, dolomites, siliceous shales, sandstones, and siltstones [10–17]. Carbonate units play a key role in the formation of iron skarns, acting as chemically reactive host rocks for interaction with magmatic and hydrothermal fluids.

Intrusive complexes in the region are represented by diorites, quartz diorites, granodiorites, and granites emplaced along deep-seated fault zones. These intrusions served as sources of heat and metal-bearing fluids and played a fundamental role in driving contact metamorphism and metasomatic processes.

The tectonic framework of the Northwestern Balkhash region is characterized by the development of strike-slip, thrust, and composite fault systems with northwestern, sublatitudinal, and meridional orientations [20]. These structures exerted a primary control on both the emplacement of intrusive bodies and the localization of skarn formation and ore accumulation.

Intersections of faults of different orders created zones of enhanced permeability, which facilitated the concentration of fluid flow and promoted the formation of magnetite ore bodies [18–20].

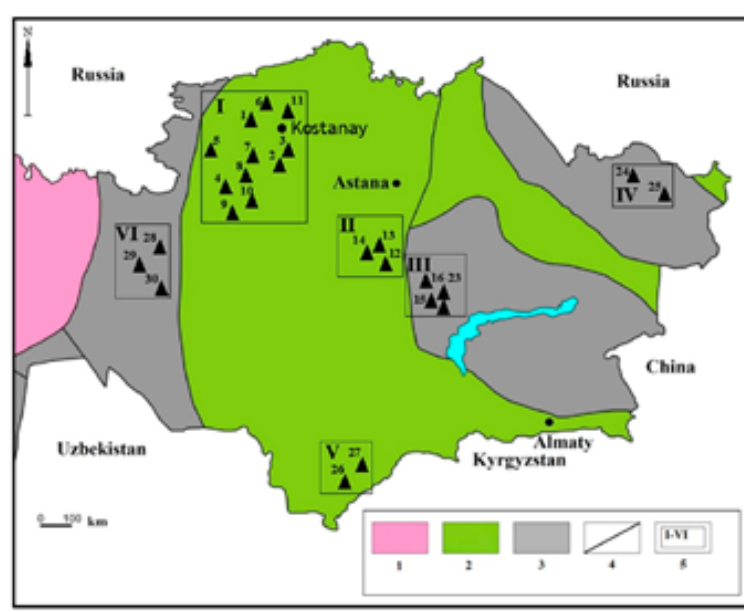


Figure 2 – Scheme of the main tectonic units and distribution of major industrial types of iron ore deposits in Kazakhstan: (1) East European Craton; (2, 3) Kazakh segment of the Central Asian Fold Belt: (2) Caledonian fold systems, (3) Hercynian fold systems; (4) geological boundaries; (5) outlines of iron ore districts.

The deposits are grouped into six regions: I – Northern Torgay: 1. Kachar, 2. Sarbai, 3. Sokolov, 4. Lisakov, 5. Ayat, 6. Aleshino, 7. Lomonosov, 8. South Sarbai, 9. Sor, 10. Shagyrkol, 11. Korzhynkol; II – Central Kazakhstan: 12. Karazhal, 13. Ushkatyn, 14. Bolshoi Ktai; III – Northwestern Balkhash Region: 15. Bapy, 16. Zhuantobe, 17. Karaulken, 18. Akchagyl, 19. Ushtobe, 20. Kiyik, 21. Taitobe, 22. Tomashev, 23. Kyzyl-Sayak; IV – Eastern Kazakhstan: 24. Kholzun, 25. Rodionov Log; V – Southern Kazakhstan: 26. Irissu, 27. Abail; VI – Western Kazakhstan: 28. Velikhovskoye, 29. Kokbulak, 30. Taldy-Espe. III marks the position of the North Balkhash ore region, which hosts the Zhuantobe deposit

Geological Characteristics of Iron Skarn Deposits (Zhuantobe and Karaulken Case Studies)

The Zhuantobe iron skarn deposit is located in the central part of the Bapin ore field and is spatially associated with the contact zone between Silurian–Devonian carbonate and terrigenous–siliceous rocks and intrusive bodies of the granitoid complex. The geological structure of the deposit is complicated by a system of northwest-trending faults, which exert a primary control on the distribution of ore bodies [18–20].

Skarns at Zhuantobe are predominantly developed in exocontact zones and are represented by calcic–magnesian varieties dominated by pyroxene–garnet assemblages. Magnetite ore bodies occur as lenticular and stratiform bodies, characterized by significant thickness variability and frequently confined to zones of intense fracturing and brecciation. A distinctive feature of the Zhuantobe deposit is the widespread occurrence of massive and disseminated magnetite textures, reflecting intensive iron metasomatism. Near-surface zones are affected by oxidation processes, resulting in the formation of hematite and limonite.

The Karaulken deposit is located within the Akbastau fault zone and exhibits a more complex structural framework. The host rocks consist of Devonian carbonate–siliceous sequences intruded by diorite and diorite-porphyry bodies. Skarns are developed in both exocontact and endocontact zones. Karaulken is characterized by a more extensive development of retrograde skarns, represented by epidote–calcite and amphibole assemblages. Magnetite mineralization displays pronounced textural heterogeneity, including vein-type, brecciated, and crystallization-schistose ore varieties.

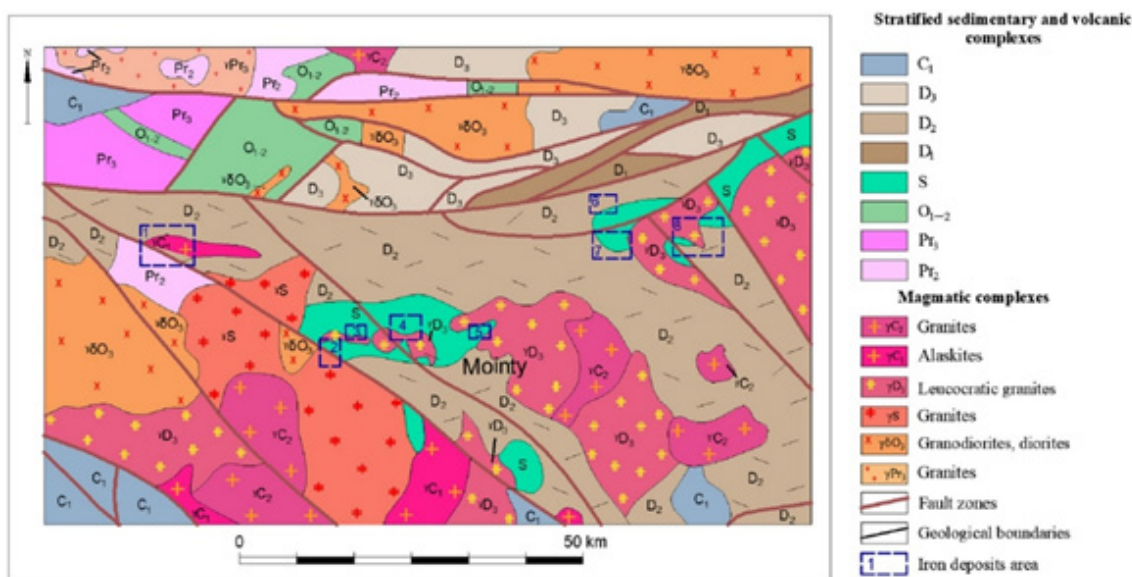


Figure 3 – The scheme of the geological structure of the iron ore region of the North-West Balkhash, compiled according to data [12, 14, 22]. The deposits are marked with blue contours: (1) Kiyik, (2) Western Bapy, (3) Bapy, (4) Eastern Bapy, (5) Bale, (6) Ushtobe, (7) Akchagyl, (8) Zhuantobe

Geophysical Indicators of Iron Skarn Localization

Magnetic surveying represents one of the most informative methods for detecting iron skarn ore bodies in the Northwestern Balkhash region. Magnetite-bearing ores generate positive magnetic anomalies, the intensity and geometry of which are directly related to the thickness, morphology, and spatial distribution of ore bodies.

At the Zhuantobe deposit, high-amplitude magnetic anomalies are recorded, indicating the presence of massive magnetite accumulations and enabling the prediction of their depth continuity. In contrast, the Karaulken deposit exhibits a more complex magnetic field pattern, reflecting its block-faulted structural framework and the heterogeneous distribution of magnetite mineralization.

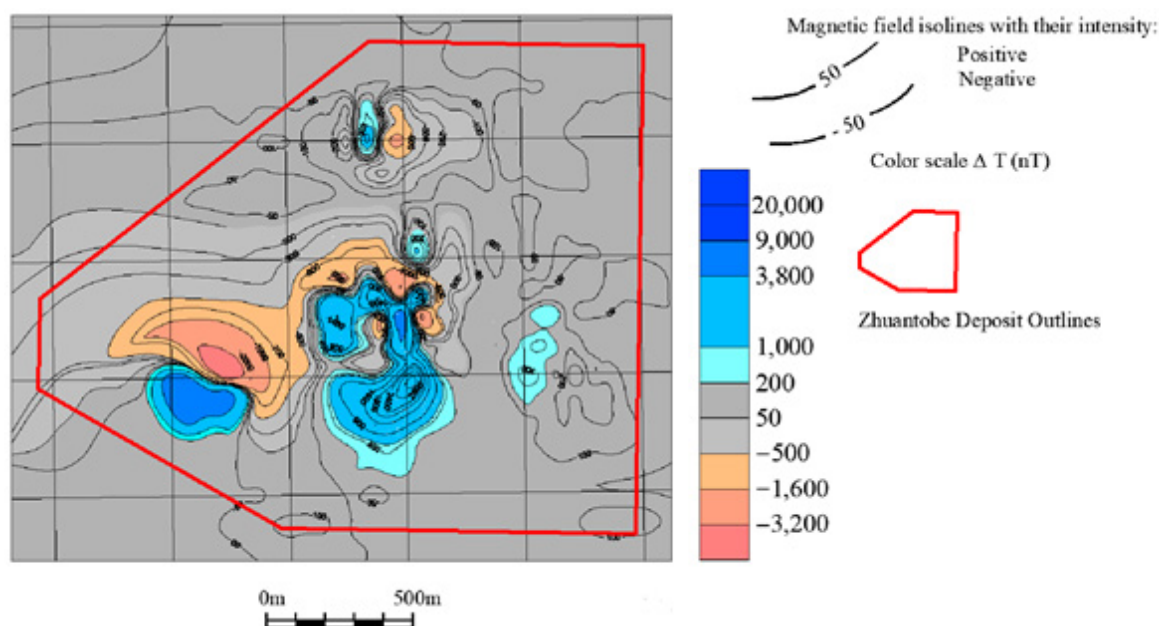


Figure 4 – Map of the magnetic field of the Zhuantobe deposit. Positive anomalies correspond to ore-bearing skarn zones. The map was compiled using field investigation results and supplemented with data from geophysical surveys [14, 22]

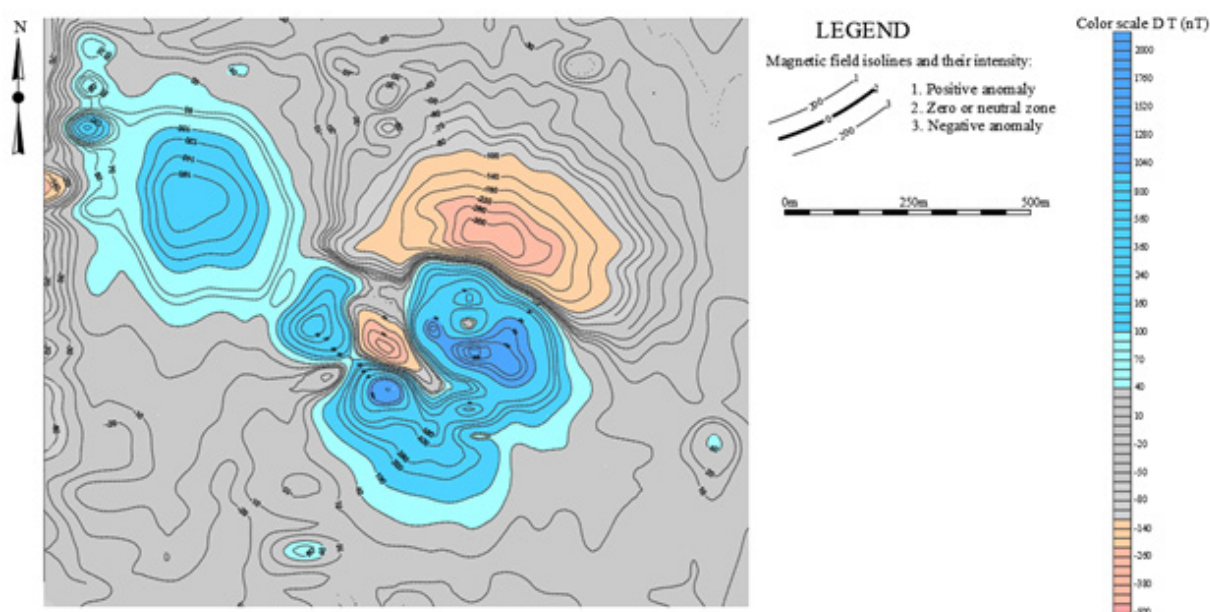


Figure 5 – Magnetic field map of the Karaulken deposit. Positive anomalies correspond to skarn ore-bearing zones. The map is based on field survey results and supplemented by geophysical data from [14, 22]

Mineralogical–Petrographic Characteristics and Composition of Skarns

Iron skarn formations of the Northwestern Balkhash region are characterized by considerable variability in mineral assemblages, textural–structural features, and degrees of metasomatic alteration of host rocks. Mineralogical and petrographic studies indicate that skarn assemblages at

the Zhuantobe and Karaulken deposits range from weakly altered marbles and hornfels to intensely metasomatized magnetite-bearing skarns and skarnoids.

The principal rock-forming minerals of the skarns include pyroxenes, garnets, epidote-group minerals, amphiboles, and calcite. Their relative abundance and morphological characteristics vary depending on the position within the contact zone, the intensity of fluid influence, and the stage of metasomatic evolution.

Pyroxenes are predominantly represented by diopside and, less commonly, hedenbergite. They occur as euhedral to subhedral crystals as well as aggregates of prismatic grains forming massive to weakly foliated textures. Pyroxenes are characteristic of the early anhydrous skarn stage and are widely developed in prograde skarns, particularly within exocontact zones.

Garnets are mainly represented by andradite and its solid-solution varieties, forming granular and porphyroblastic aggregates. They commonly contain inclusions of magnetite, epidote, and relic silicate minerals, indicating crystallization under conditions of intense iron metasomatism. Garnet-dominated skarns are especially characteristic of the Zhuantobe deposit, where they are closely associated with massive magnetite mineralization.

Epidote and epidote-group minerals are широко developed in retrograde skarns and skarnoids. They form fine-grained aggregates, veinlets, and replacement textures after pyroxenes and garnets. The occurrence of epidote reflects decreasing temperatures and an increasing role of aqueous fluids during the late stages of metasomatic evolution.

Amphiboles are represented mainly by hornblende and actinolite. They occur as prismatic crystals and fibrous aggregates replacing earlier pyroxenes. Amphiboles are particularly abundant in the Karaulken skarns, reflecting a more pronounced retrograde overprint.

Calcite is one of the dominant minerals of the late-stage assemblage and is widely distributed both within skarns and in vein and veinlet systems. It fills fractures, pores, and brecciated zones, forming mosaic and vein-type textures.

Magnetite is the principal ore mineral of the Northwestern Balkhash skarns. It occurs as massive accumulations, disseminated grains, veinlets, and brecciated aggregates. The morphology of magnetite ranges from equigranular and subhedral crystals to complex interstitial and vein-controlled forms, reflecting the dynamics of fluid flow and evolving physicochemical conditions. In near-surface zones, magnetite is partially replaced by hematite and limonite, forming martitized and limonitized ore varieties.

Sulfide minerals are represented predominantly by pyrite and, less commonly, chalcopyrite. They occur as fine disseminations, veinlets, and microinclusions within magnetite and the silicate matrix. Sulfide mineralization is mainly associated with the late hydrothermal stage and is subordinate relative to magnetite mineralization.

Mineralogical and petrographic data allow the identification of several successive stages of skarn formation:

1. Contact metamorphic stage, characterized by the formation of hornfels and marbles;
2. Prograde anhydrous skarn stage, marked by the development of pyroxene–garnet assemblages and primary magnetite;
3. Retrograde hydrothermal stage, accompanied by the formation of amphiboles, epidote, calcite, and sulfides;
4. Supergene oxidation stage, resulting in the replacement of magnetite by hematite and limonite.

Textural–Structural Features of Iron Skarns

Textural and structural characteristics of iron skarns in the Northwestern Balkhash region represent key indicators of their formation conditions and reflect the multistage evolution of contact metasomatic processes. Analysis of skarn textures and structures from the Zhuantobe and Karaulken deposits allowed the identification of several characteristic types, the formation of which is controlled by variations in thermodynamic conditions, fluid regimes, and tectonic activity.

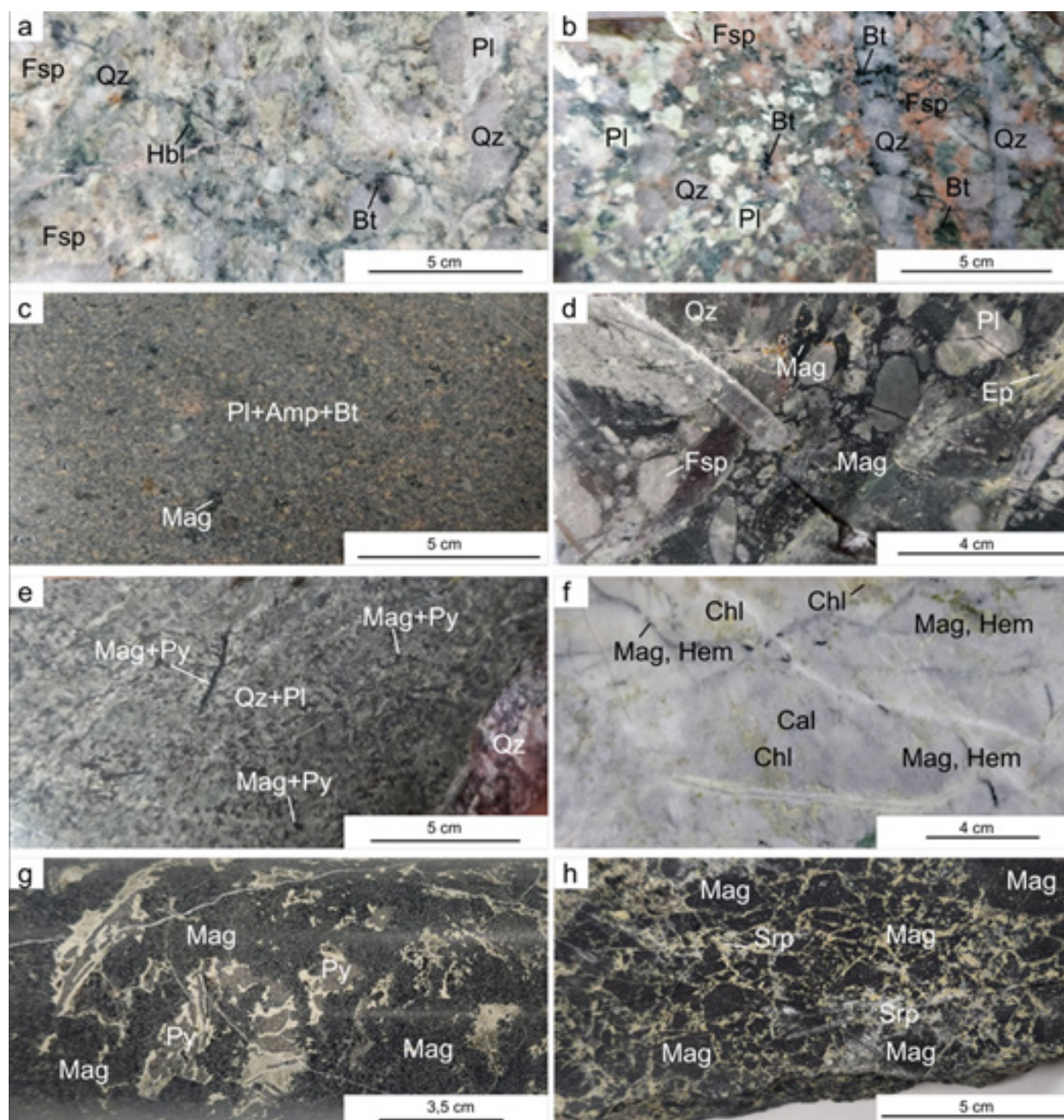


Figure 6 – Samples of the main rock types of the Zhuantobe deposit, details of Figure 5.

(a) Cataclastic coarse-grained granites, well Z01, from the depth 150.8 m; (b) silicified coarse-grained granites, well Z02, from the depth 142.0 m; (c) diorites, well Z01, from the depth 216.2 m; (d) cataclastic lithoclastic rhyolite tuffs, well Z02, from the depth 47.7; (e) skarns on sandstone, well Z04, from the depth 48.6 m; (f) marbles, well Z03, from the depth 121.7 m; (g) magnetite ores, well Z04, from the depth 40.1 m; (h) magnetite ores, well Z02, from the depth 150.2 m. Conventional indices of minerals [32]: feldspar (Fsp); quartz (Qz); hornblende (Hbl); biotite (Bt); plagioclase (Pl); amphibole (Amp); magnetite (Mag); epidote group minerals (Ep); pyrite (Py); chlorite (Chl); calcite (Cal); hematite (Hem); serpentine (Srp)

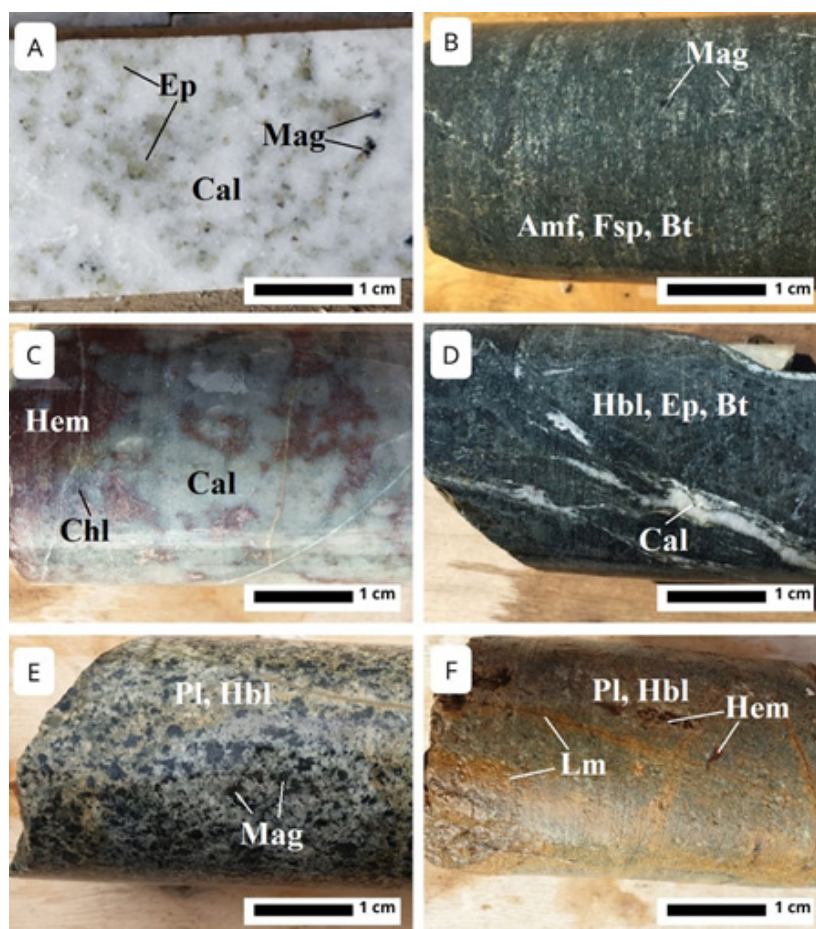


Figure 7 – Representative rock types of the Karaulken iron-skarne system (Central Kazakhstan). (A) Saccharoidal calcite marble with disseminated diopside–epidote grains from the contact zone; (B) Amphibolite from the prograde metamorphic zone; (C) Epidote–calcite skarn from the retrograde hydrothermal; (D) Actinolite–quartz–chlorite skarnoid from the transitional zone between prograde and retrograde stages; (E) Diorite from the intrusive complex; (F) Limonite–hematite skarnoid from the oxidation zone. Conventional indices of minerals (Whitney and Evans, 2010): epidote group minerals (Ep); magnetite (Mag); calcite (Cal); amphibole (Amp); feldspar (Fsp); biotite (Bt); hematite (Hem); chlorite (Chl); hornblende (Hbl); plagioclase (Pl); limonite (Lm)

Massive textures are typical of zones of intense iron metasomatism and are represented by dense accumulations of magnetite with a minor proportion of silicate matrix. Magnetite forms equigranular to subhedral crystals tightly intergrown, giving the ores a compact, monolithic appearance. These textures develop under conditions of high Fe-rich fluid activity and relatively stable thermal regimes.

Disseminated textures are characterized by an uneven distribution of magnetite within a silicate skarn matrix. Ore minerals occur as isolated grains or small aggregates interstitially distributed among pyroxene, garnet, epidote, and amphibole. Grain size varies from microscopic to millimeter scale. These textures reflect less intensive iron metasomatism and indicate conditions where ore formation occurred contemporaneously with the growth of silicate minerals. Disseminated textures are widely developed at both Zhuantobe and Karaulken and are typical of transitional zones between ore-bearing and barren skarns.

Vein and veinlet textures are represented by networks of magnetite, quartz–calcite, and epidote–calcite veinlets crosscutting the skarn matrix. Magnetite within these structures forms elongated

aggregates commonly aligned along fractures and zones of tectonic brecciation. Vein–disseminated textures reflect a pulsatory fluid flow regime and repeated reactivation of fault structures. These textures are particularly characteristic of the Karaulken deposit, where ore zones are closely associated with strike-slip and thrust fault systems. Genetically, they are interpreted as products of late-stage skarn formation and the onset of hydrothermal overprinting.

Brecciated textures develop within zones of intense tectonic deformation and consist of angular to subrounded fragments of skarns and host rocks cemented by magnetite, calcite, epidote, and secondary silicates. The variability in fragment size and morphology indicates multiple episodes of fragmentation followed by mineral precipitation. The presence of brecciated textures highlights the significant role of tectonics in ore formation and indicates repeated reopening of the system for fluid circulation. These textures are especially characteristic of Karaulken and reflect a prolonged and complex deformation history.

Crystallization-schistose textures form under conditions of directed stress and ductile deformation. They are expressed by the preferred orientation of prismatic and tabular amphibole and epidote crystals, as well as elongated magnetite aggregates. Such textures develop during late stages of skarn evolution under the combined influence of tectonic deformation and hydrothermal processes.

In terms of structure, skarns range from coarse- and medium-grained to fine-grained and cryptocrystalline varieties. Coarse-grained structures are typical of early prograde skarns, whereas fine-grained and recrystallized textures reflect retrograde alteration and repeated fluid activity.

Magnetite ores commonly exhibit replacement, interstitial, and corrosion textures, indicating complex relationships between ore and rock-forming minerals. These features confirm the multistage nature of iron skarn formation in the Northwestern Balkhash region.

The integrated analysis of textural–structural features allows the interpretation of iron skarn formation as a result of a multistage evolutionary system involving:

- ♦ Prograde skarn formation under high-temperature conditions and intensive iron metasomatism;
- ♦ Subsequent hydrothermal overprinting, leading to the development of vein and brecciated textures;
- ♦ Tectonic reorganization of ore zones, resulting in oriented and schistose structures;
- ♦ Near-surface oxidation processes, causing secondary alteration of magnetite into hematite and limonite.

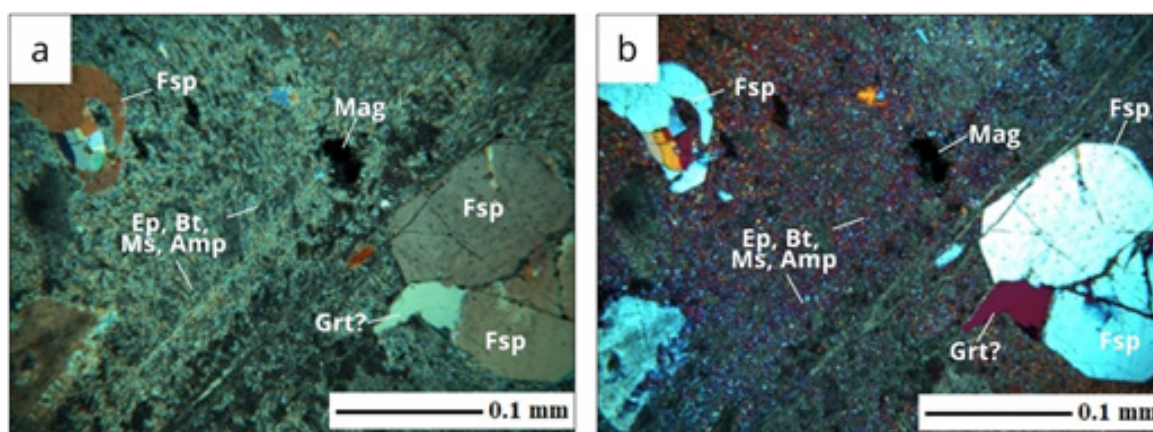


Figure 8 – Garnet–epidote skarns with magnetite from Zhuantobe deposit: (a) parallel nicols; (b) crossed nicols with compensator; magnification $\times 40$. Conventional indices of minerals [32]: feldspar (Fsp); magnetite (Mag); epidote group minerals (Ep); biotite (Bt); muscovite (Ms); amphibole (Amp); probable garnet (Grt?)

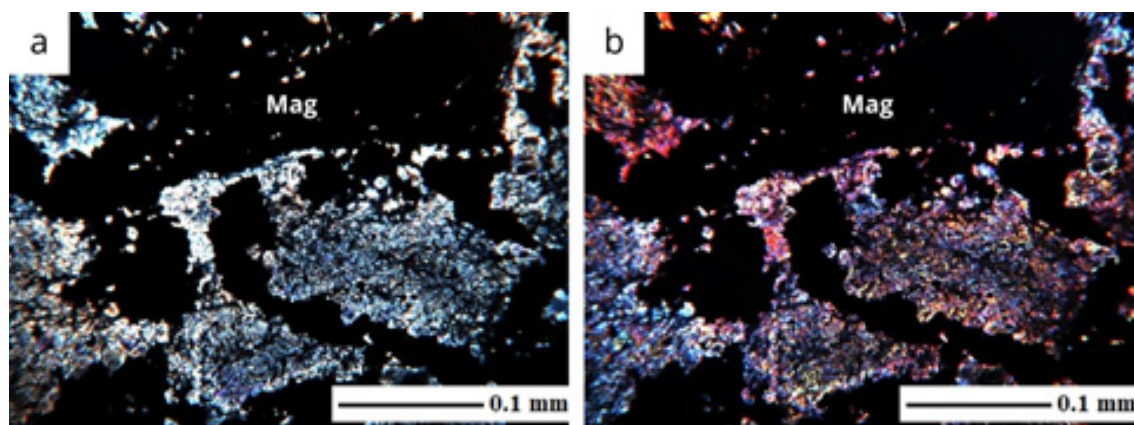


Figure 9 – Magnetite ores from Karaulken deposit: (a) parallel nicols; (b) crossed nicols with compensator; magnification $\times 40$. Conventional indices of minerals [32]: magnetite (Mag).

Bulk Composition and Geochemical Characteristics of Iron Skarns

Geochemical data obtained from the Zhuantobe and Karaulken deposits provide key insights into the distribution patterns of iron and associated elements within skarn systems of the Northwestern Balkhash region.

The chemical composition of skarns is controlled by the interplay between the composition of carbonate–siliceous host rocks and the nature of magmatic–hydrothermal processes. Most skarn varieties are characterized by elevated contents of Fe_2O_3 , CaO , and SiO_2 , with variable concentrations of Al_2O_3 , MgO , and MnO . This compositional pattern reflects intensive silicification and iron metasomatism associated with the formation of magnetite mineralization.

Bulk Composition and Geochemical Characteristics of Iron Skarns

Geochemical data obtained from the Zhuantobe and Karaulken deposits provide key insights into the distribution patterns of iron and associated elements within skarn systems of the Northwestern Balkhash region.

The chemical composition of skarns is controlled by the interplay between the composition of carbonate–siliceous host rocks and the nature of magmatic–hydrothermal processes. Most skarn varieties are characterized by elevated contents of Fe_2O_3 , CaO , and SiO_2 , with variable concentrations of Al_2O_3 , MgO , and MnO . This compositional pattern reflects intensive silicification and iron metasomatism associated with the formation of magnetite mineralization.

Fe_2O_3 contents in skarns and ores vary widely, ranging from moderate values in weakly altered rocks to high concentrations in massive magnetite bodies. Maximum values are typical of prograde skarns, where iron is predominantly concentrated in magnetite. In retrograde assemblages, iron is redistributed among magnetite, hematite, and iron hydroxides.

Calcium and silicon represent the principal components of the skarn matrix and are mainly incorporated into pyroxene, garnet, epidote, and calcite. The relative proportions of CaO and SiO_2 depend on the degree of replacement of carbonate rocks by silicate minerals and the intensity of subsequent hydrothermal alteration.

Ore bodies at Zhuantobe and Karaulken are dominated by magnetite as the principal iron-bearing phase. Geochemical data indicate that magnetite ores are characterized by relatively low silica contents and elevated iron concentrations, which positively influence their metallurgical properties.

At the Zhuantobe deposit, magnetite ores generally exhibit a more homogeneous chemical composition and relatively stable Fe_2O_3 contents. This is attributed to the predominance of massive and disseminated textures formed under conditions of sustained and intensive iron metasomatism. These ores are characterized by low concentrations of deleterious elements, making them favorable for beneficiation.

In contrast, Karaulken ores display pronounced geochemical heterogeneity, expressed by fluctuations in iron content, elevated SiO_2 concentrations, and local enrichment in P_2O_5 . These features are related to a more complex structural setting, multi-stage reactivation of fault zones, and intense retrograde alteration of skarns.

Trace and associated elements in iron skarns are mainly represented by Mn, Ti, Al, and P, occurring in both ore and rock-forming minerals. Manganese commonly substitutes for iron in magnetite and pyroxenes, whereas titanium is present as fine inclusions of ilmenite or titanomagnetite.

Phosphorus is one of the most critical technological components affecting ore quality. Elevated P_2O_5 contents in certain zones of Karaulken are associated with the presence of apatite and secondary phosphate minerals formed during late hydrothermal stages. The heterogeneous distribution of P_2O_5 necessitates a differentiated approach to resource evaluation and processing strategies.

Sulfur contents are generally low and are associated with the presence of pyrite and, less commonly, chalcopyrite. Sulfide mineralization is subordinate and does not significantly affect the overall composition of the ores; however, its presence serves as an important indicator of late hydrothermal processes.

The analysis of bulk composition demonstrates that geochemical characteristics are closely linked to the stages of skarn system evolution. The prograde stage is marked by maximum iron concentration and the formation of relatively simple magnetite-dominated assemblages. During the retrograde stage, redistribution of components occurs, accompanied by enrichment in calcium and silicon and localized accumulation of trace elements.

The bulk composition of iron skarns directly influences their technological properties. The predominance of magnetite and relatively low levels of harmful impurities in Zhuantobe ores result in high beneficiation efficiency and the production of high-grade concentrates. In contrast, the more complex composition of Karaulken ores requires detailed mineralogical and technological evaluation and the application of combined processing methods.

Overall, the geochemical characteristics of iron skarns in the Northwestern Balkhash region reflect the complex evolution of contact metasomatic systems and provide a robust basis for genetic interpretations and practical applications in exploration and resource development.

Table 1 – Generalized bulk composition of iron skarns in the Northwestern Balkhash region (Zhuantobe and Karaulken)

Component	Content range, wt. %	Average, wt. %	Mineral host	Genetic interpretation
Fe_2O_3	25.0 – 55.0	38.0 – 42.0	Magnetite, hematite	The main ore component, the prograde stage of skarn formation
SiO_2	15.0 – 35.0	22.0 – 27.0	Pyroxenes, garnets, quartz	Silicate matrix of skarns, intensity of silicification
CaO	10.0 – 30.0	18.0 – 22.0	Calcite, garnets, epidote	Metasomatic substitution carbonate breeds
Al_2O_3	3.0 – 10.0	5.0 – 7.0	Garnets, epidote, amphiboles	The introduction of aluminum during magmatic -hydrothermal action
MgO	2.0 – 8.0	3.5 – 5.0	Pyroxenes, amphiboles	Calcium-magnesium type skarnov
MnO	0.1 – 1.5	0.3 – 0.7	Magnetite, pyroxenes	Isomorphic impurities in ore minerals
TiO_2	0.1 – 1.0	0.2 – 0.5	Ilmenite, titanomagnetite	Magmatic component ore-bearing fluids
P_2O_5	0.05 – 3.0	0.3 – 1.2	Apatite	Late hydrothermal processes, a technologically important component
SO_3	0.01 – 1.0	0.1 – 0.3	Pyrite, sulfates	Subordinate sulfide mineralization

Discussion

A comparative analysis of the geology, mineralogy, and formation conditions of iron skarns in the Northwestern Balkhash region with global analogues reveals both fundamental similarities and region-specific features. International studies demonstrate that iron skarn deposits represent significant sources of high-grade iron ores, typically formed at the contact between magmatic intrusions and carbonate sequences under the influence of magmatic–hydrothermal fluids [23, 27, 30–33]. Established genetic models emphasize multistage skarn evolution, including prograde and retrograde stages involving magnetite, pyroxene, garnet, and epidote assemblages.

For example, studies of the Zanzan iron skarn deposits (Iran) [23] indicate the development of both magnesian and calcic skarns (endoskarn and exoskarn), formed at the contact between granitic intrusions and carbonate rocks and characterized by pyroxene–garnet–magnetite associations. Similarly, investigations of the high-grade Jinling iron skarn deposit (North China) reveal well-defined mineralogical zoning and clear stage evolution, where early iron enrichment is associated with Fe-rich hydrothermal fluids, followed by retrograde mineralization and secondary silicate formation [30–33].

The comparative analysis indicates that iron skarns of the Northwestern Balkhash region conform to the principal characteristics of global iron skarn systems:

- ♦ spatial association with contact zones between intermediate to felsic intrusions and carbonate sequences;
- ♦ multistage evolution involving prograde and retrograde phases;
- ♦ mineral assemblages dominated by magnetite, pyroxene, garnet, and epidote;
- ♦ strong structural control by fault systems facilitating fluid migration.

Despite this general genetic framework, several distinctive features of the Northwestern Balkhash skarns can be identified:

- ♦ more pronounced tectonic overprinting, expressed by the development of vein systems and brecciated structures, typical of the Central Asian Orogenic Belt;
- ♦ significant textural heterogeneity related to repeated reactivation of fault zones, comparable to skarn systems in the Tianshan–Beishan region;
- ♦ geochemical variability, including elevated concentrations of impurity elements (e.g., phosphorus), similar to patterns observed in North China skarns.

Global syntheses indicate that, although iron skarn deposits form under a wide range of geological conditions, they are unified by a common contact metasomatic mechanism driven by magmatic heat and fluid circulation. This framework is consistent with the genetic model proposed for the Northwestern Balkhash region, confirming its classification as a typical iron skarn system within an active continental margin–type geodynamic setting.

Conclusion

This study establishes the key geological and mineralogical patterns governing the formation of iron skarns in the Northwestern Balkhash region, based on detailed investigation of the Zhuantobe and Karaulken deposits. Integrated analysis of geological mapping, mineralogical–petrographic observations, geochemical data, and geophysical interpretation demonstrates that skarn formation represents a multistage contact metasomatic system controlled by magmatic, structural, and fluid-related factors.

Mineralogical and petrographic data confirm the multistage evolution of skarn assemblages. The prograde anhydrous stage is characterized by the formation of pyroxene–garnet assemblages and primary magnetite mineralization. During the retrograde stage, skarns undergo hydrothermal reworking accompanied by the development of amphiboles, epidote, calcite, and subordinate sulfide mineralization. Near-surface oxidation processes result in partial replacement of magnetite by hematite and limonite, leading to additional mineralogical and geochemical differentiation of ores.

Fault-block tectonics plays a fundamental role in controlling the emplacement of intrusions and the migration of ore-forming fluids, thereby governing the localization of mineralization. The more интенсив tectonic overprinting observed at Karaulken results in increased geochemical heterogeneity compared to the более stable conditions of ore formation at Zhuantobe.

Comparison with global iron skarn systems indicates that the Northwestern Balkhash deposits correspond to classical iron skarns associated with active continental margin environments and the Central Asian Orogenic Belt.

The identified geological, mineralogical, and geochemical criteria can be applied in the prediction and exploration of new iron ore targets, as well as in the assessment of technological properties of ores and the selection of optimal processing strategies for their industrial development.

REFERENCES

- 1 Satpayev, K.I. Main patterns of spatial distribution of endogenous mineralization zones in Central Kazakhstan. *Soviet Geology*, 58, 93–109 (1957). (in Russian).
- 2 Novokhatsky, I.P. Genetic types of iron ore deposits in Central Kazakhstan. In: *Iron Ore Deposits of Central Kazakhstan and Ways of Their Use* (Moscow: USSR Academy of Sciences, 1960). (in Russian).
- 3 Satpayev, K.I. *Selected Works. Problems of Metallogeny and Mineral Resources of Kazakhstan*, Vol. 3 (Alma-Ata: Nauka, 1968), 312 p. (in Russian).
- 4 Shcherba, G.N. Atasu type deposits in Kazakhstan. *Endogenous Ore Deposits*, pp. 185–196 (1968). (in Russian).
- 5 Bogdanov, A.A. Tectonic zoning of the Paleozooids of Central Kazakhstan and Tien Shan. *Bulletin of the Moscow Society of Naturalists*, 40, 8–42 (1965). (in Russian).
- 6 Bepalov, V.F. *Tectonic Map of the Kazakh SSR and Adjacent Territories of the Union Republics at a Scale of 1:1,500,000. Explanatory Note* (Alma-Ata, 1975), 160 p. (in Russian).
- 7 Dymkin, A.M. *Contact-Metasomatic Iron Deposits of the Southern Part of the Main Ore Belt of Turgay* (Novosibirsk: Publishing House of the Siberian Branch of the USSR Academy of Sciences, 1962), 239 p. (in Russian).
- 8 Esenov, Sh.E., Lyapichev, G.F., Sidorenko, A.V., and Shlygin, E.D. *Geology of the USSR. Vol. XX. Central Kazakhstan. Geological Description. Book 1* (Moscow: Nedra, 1972), 532 p. (in Russian).
- 9 Zaicev, Yu.A. *Geology of the USSR. Vol. XX. Central Kazakhstan. Minerals* (Moscow: Nedra, 1979), 541 p. (in Russian).
- 10 Abdulin, A.A. *Geology and Mineral Resources of Kazakhstan* (Almaty: Gylym, 1994), 394 p. (in Russian).
- 11 Zholtayev, G.Zh., Zhukov, N.M., and Antonenko, A.A. *Atlas of Solid Mineral Deposits of the Republic of Kazakhstan* (Almaty: K.I. Satpayev Institute of Geological Sciences, 2024), 263 p. (in Russian).
- 12 Bekzhanov, G.R., Koshkin, V.Y., and Nikitchenko, I.I. *Geological Structure of Kazakhstan* (Almaty: AMR of the Republic of Kazakhstan, 2000), 396 p. (in Russian).
- 13 Bekmukhametov, A.E. *Magmatogenic Iron Ore Formations* (Moscow: Nedra, 1987), 212 p. (in Russian).
- 14 Kubeeva, L.V., Zhabkin, V.F., Yugay, V.D., Baimuldina, N.N., and Muzgina, V.S. *Geocological studies during the development of the Bapy iron ore deposit in the Republic of Kazakhstan. Subsoil Use in the 21st Century*, 1, 81–83 (2010). (in Russian).
- 15 Bekmukhametov, A.E., Pankratova, N.L., and Kryukov, A.S. *Iron Ore Formations. Chu-Ili Belt. Minerals* (Alma-Ata: Nauka, 1980). (in Russian).
- 16 Zhukov, N.M., Akylbekov, S.A., Antonenko, A.A., and Amanbayev, R. *Iron Deposits of Kazakhstan: A Handbook* (Almaty, 2016), 258 p. (in Russian).
- 17 Daukeev, S.Zh., Uzhkenov, B.S., Abdulin, A.A., Bespaev, Kh.A., Votsalevsky, E.S., Lyubetsky, V.N., Mazurov, A.K., and Miroshnichenko, L.A. *Deep Structure and Geodynamics. In: Deep Structure and Mineral Resources of Kazakhstan, Vol. 1* (Almaty: National Academy of Sciences of the Republic of Kazakhstan, 2002), 224 p. (in Russian).
- 18 Daukeev, S.Zh., Uzhkenov, B.S., Abdulin, A.A., Miroshnichenko, L.A., Zhukov, N.M., Mazurov, A.K., Bespaev, Kh.A., Dolgoplov, B.M., Akylbekov, S.A., Zhautikov, T.M., and Gubaidulin, F.G. *Metallogeny. In: Deep Structure and Mineral Resources of Kazakhstan, Vol. 2* (Almaty: National Academy of Sciences of the Republic of Kazakhstan, 2002), 265 p. (in Russian).

- 19 Nikitchenko, I.I. Minerals of Kazakhstan: Explanatory Note to the Map of Minerals of Kazakhstan at a Scale of 1:1,000,000 (Kokshetau: Ministry of Energy and Mineral Resources of the Republic of Kazakhstan, 2002), 188 p. (in Russian).
- 20 Korobkin, V.V., and Buslov, M.M. Tectonics and geodynamics of the western Central Asian Fold Belt (Kazakhstan Paleozoides). *Russian Geology and Geophysics*, 52, 1600–1618 (2011). <https://doi.org/10.1016/j.rgg.2011.11.011>
- 21 Korobkin, V.V., Chaklikov, A.Y., and Tulemissova, Z.S. Tectonic zoning of Paleozoides of Kazakhstan and its oil and gas-bearing regions. *Kazakhstan Journal for Oil & Gas Industry*, 4(1), 39–49 (2022). <https://doi.org/10.54859/kjogi107854>
- 22 Boyko, O.N., Filatova, G.V., Zyabkin, V.F., and Lim, D.K. Report on the Results of Exploration of Iron-Containing Ores in the Bapy Area in the Karaganda Region for 2015–2018 (Karaganda, 2019).
- 23 Zhao, W.W., and Zhou, M.F. In-situ LA-ICP-MS trace elemental analyses of magnetite: The Mesozoic Tengtie skarn Fe deposit in the Nanling Range, South China. *Ore Geology Reviews*, 65(4), 867–883 (2015). <https://doi.org/10.1016/j.oregeorev.2014.09.019>
- 24 Korobkin, V., Chaklikov, A., Tulemissova, Z., Samatov, I., and Dobrovolskaya, Y. Results of the Study of Epigenetic Changes of Famennian–Tournaisian Carbonate Rocks of the Northern Marginal Shear Zone of the Caspian Syncline (Kazakhstan). *Minerals*, 13, 249 (2023). <https://doi.org/10.3390/min13020249>
- 25 Korobkin, V., Samatov, I., Chaklikov, A., and Tulemissova, Z. Peculiarities of Dynamics of Hypergenic Mineral Transformation of Nickel Weathering Crusts of Ultramafic Rocks of the Kempirsay Group of Deposits in Western Kazakhstan. *Minerals*, 12, 650 (2022). <https://doi.org/10.3390/min12050650>
- 26 Korobkin, V.V., Tulemissova, Z.S., Samatov, I.B., and Chaklikov, A.Y. Secondary calcite in carbonate reservoirs of oil fields: a method for its quantitative determination. *Kazakhstan Journal for Oil & Gas Industry*, 7(2), 84–95 (2025). (in Russian). <https://doi.org/10.54859/kjogi108822>
- 27 Vezzoni, S., Rocchi, S., and Dini, A. Campiglia Marittima Skarn (Tuscany): A Challenging Example for the Evolution of Skarn-Forming Models. *Minerals*, 13, 482 (2023). <https://doi.org/10.3390/min13040482>
- 28 Vom Rath, G. Die Berge von Campiglia in der Toskanischen Maremma. *Zeitschrift der Deutschen Geologischen Gesellschaft*, 20, 307–364 (1868).
- 29 Reverdatto, V.V. Facies of Contact Metamorphism (Moscow: Nedra, 1970), 272 p. (in Russian).
- 30 Chen, Y., Cao, Y., Liu, L., Wang, C., Yang, W., Gai, Y., Xie, T., Song, L., and Xie, F. Elucidating the Genetic Mechanism and the Ore-Forming Materials of the Kaladawan Iron Deposit in the North Altyn Tagh, Western China. *Minerals*, 14, 589 (2024). <https://doi.org/10.3390/min14060589>
- 31 Yi, J., Shi, X., Ji, G., Zhang, L., Wang, S., and Deng, H. The Geochemical Characteristics of Trace Elements in Magnetite and Fe Isotope Geochemistry of the Makeng Iron Deposit in Southwest Fujian and Their Significance in Ore Genesis. *Minerals*, 14, 217 (2024). <https://doi.org/10.3390/min14030217>
- 32 Feng, Q., Gao, M., Fu, C., Li, S., Li, Y., Gao, J., Ma, M., Wang, Z., Zhu, Y., Wu, B., Duan, Z., and Dang, Z. Phlogopite $^{40}\text{Ar}/^{39}\text{Ar}$ Geochronology for Guodian Skarn Fe Deposit in Qihe–Yucheng District, Luxi Block, North China Craton: A Link between Craton Destruction and Fe Mineralization. *Minerals*, 14, 690 (2024). <https://doi.org/10.3390/min14070690>
- 33 Li, Y., Jiang, Z., Wang, D., Zhang, Z., and Duan, S. Genesis of the Beizhan Iron Deposit in Western Tianshan, China: Insights from Trace Element and Fe–O Isotope Compositions of Magnetite. *Minerals*, 14, 304 (2024). <https://doi.org/10.3390/min14030304>
- 34 Matiole, E., Couto, H.J.B., Lima, N., Silva, K., and de Freitas, A.S. Improving recovery of iron using column flotation of iron ore slimes. *Minerals Engineering*, 158, 106608 (2020). <https://doi.org/10.1016/j.mineng.2020.106608>
- 35 Nakhaei, F., and Irannajad, M. Reagents types in flotation of iron oxide minerals: A review. *Mineral Processing and Extractive Metallurgy Review*, 39(2), 89–124 (2018). <https://doi.org/10.1080/08827508.2017.1391245>
- 36 Silva, R.C.F., Ardisson, J.D., Cotta, A.A.C., Araujo, M.H., and de Carvalho Teixeira, A.P. Use of iron mining tailings from dams for carbon nanotubes synthesis in fluidized bed for 17α -ethinylestradiol removal. *Environmental Pollution*, 260, 114099 (2020). <https://doi.org/10.1016/j.envpol.2020.114099>
- 37 Araujo, A.C., Amarante, S.C., Souza, C.C., and Silva, R.R.R. Ore mineralogy and its relevance for selection of concentration methods in processing of Brazilian iron ores. *Mineral Processing and Extractive Metallurgy*, 112(1), 54–64 (2003). <https://doi.org/10.1179/037195503225011439>
- 38 Lima, R.M.F., and Abreu, F.P.V.F. Characterization and concentration by selective flocculation/magnetic separation of iron ore slimes from a dam of Quadrilátero Ferrífero – Brazil. *Journal of Materials Research and Technology*, 9, 12356–12362 (2020). <https://doi.org/10.1016/j.jmrt.2019.12.034>

39 Muwanguzi, A.J.B., Karasev, A.V., Byaruhanga, J.K., and Jönsson, P.G. Characterization of the Physical and Metallurgical Properties of Natural Iron Ore for Iron Production. ISRN Materials Science, Article 147420 (2012). <https://doi.org/10.5402/2012/147420>

40 Yellishetty, M., Ranjith, P.G., and Tharumarajah, A. Iron ore and steel production trends and material flows in the world: Is this really sustainable? Resources, Conservation and Recycling, 54(12), 1084–1094 (2010). <https://doi.org/10.1016/j.resconrec.2010.03.003>

41 Zhang, X., Gu, X., Han, Y., Parra-Álvarez, N., Claremboux, V., and Kawatra, S.K. Flotation of Iron Ores: A Review. Mineral Processing and Extractive Metallurgy Review, 42(3), 184–212 (2021). <https://doi.org/10.1080/08827508.2019.1689494>

42 Ostadrahimi, M., Farrokhpay, S., Zare, T., Aghajanlou, M., Gharibi, K., and Salari Rad, M.M. An Integrated Separation Approach for Beneficiation of Low-Grade Iron Ore. Minerals, 15(9), 958 (2025). <https://doi.org/10.3390/min15090958>

43 Prasad, J., Venkatesh, A.S., Sahoo, P.R., Singh, S., and Sylvestre Kanouo, N. Geological Controls on High-Grade Iron Ores from Kiriburu-Meghahatuburu Iron Ore Deposit, Singhbhum-Orissa Craton, Eastern India. Minerals, 7(10), 197 (2017). <https://doi.org/10.3390/min7100197>

44 Mahanta, C.R., Sahoo, P.R., Mohanta, M.K., Rath, R.K., Dey, S., Tripathy, S.K., Prasad, J., and Venkatesh, A.S. Mineralogical Characteristics of Hematitic Iron Ore: A Geometallurgical Study on Ore from Eastern India. Minerals, 13(9), 1194 (2023). <https://doi.org/10.3390/min13091194>

45 Maniteja, M., Samanta, G., Gebretsadik, A., Tsae, N.B., Rai, S.S., Fissaha, Y., Okada, N., and Kawamura, Y. Advancing Iron Ore Grade Estimation: A Comparative Study of Machine Learning and Ordinary Kriging. Minerals, 15(2), 131 (2025). <https://doi.org/10.3390/min15020131>

^{1*}Ныгманова А.С.,

магистр, аға оқытушы, ORCID ID: 0009-0003-3917-2707,

e-mail: a.emls@kbtu.kz

¹Коробкин В.В.,

г.-м.ф.к., профессор, ORCID ID: 0000-0002-1562-759X,

e-mail: korobkin_vv@kbtu.kz

²Буслов М.М.,

г.-м.ф.д., профессор, ORCID ID: 0000-0003-0606-2264,

e-mail: buslov@igm.nsc.ru

¹Чакликов А.Е.,

PhD, ассистент-профессор, ORCID ID: 0000-0001-8316-6599,

e-mail: a.chaklikov@kbtu.kz

¹Тулемисова Ж.С.,

PhD, қауымдастырылған профессор, ORCID ID: 0000-0003-1803-4535,

e-mail: z.tulemissova@kbtu.kz

^{1,3}Койшибаев С.М.,

магистрант, кеңесші геолог, ORCID ID: 0009-0005-6492-4215

e-mail: skoishibayev@srk.kz

^{1,3}Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

²В.С. Соболев атындағы Геология және минералогия институты, Ресей ғылым академиясының Сібір бөлімшесі, Новосібір мемлекеттік университеті, Новосібір қ., Ресей

³SRK Consulting, Алматы қ., Қазақстан

ОРТАЛЫҚ ҚАЗАҚСТАННЫҢ СОЛТҮСТІК-БАТЫС БАЛҚАШ ӨңІРІНДЕГІ ТЕМІР СКАРНДАРЫНЫҢ ҚАЛЫПТАСУЫН БАҚЫЛАЙТЫН ГЕОЛОГИЯЛЫҚ ЖӘНЕ МИНЕРАЛОГИЯЛЫҚ ФАКТОРЛАР

Аңдатпа

Кешенді зерттеу Солтүстік-Батыс Балқаш өңіріндегі темір скарн кенорындарының қалыптасуын анықтайтын негізгі геологиялық және минералогиялық факторларды, Жуантөбе және Қарауылкен сияқты үлгілік кенорындар негізінде айқындады. Геологиялық карталау, минералогиялық және петрографиялық

бақылаулар, геохимиялық деректер мен геофизикалық интерпретациялардың интеграцияланған талдауы скарн түзілуінің магматикалық, құрылымдық және флюидтік факторлармен бақыланатын көпсатылы контактылы-метасоматикалық жүйе екенін көрсетеді. Минералдық қауымдастықтардың таралуы мен кеннің текстуралық-құрылымдық типтеріндегі жүйелі заңдылықтар анықталып, олар флюидтік режимдердің және кен түзілуінің физика-химиялық жағдайларының өзгерістерін бейнелейді. Алынған нәтижелер Солтүстік-Батыс Балқаш өңіріндегі темір скарндарының аймақтық генетикалық моделін нақтылап, жаңа кен нысандарының болжамын жасау және барлау әлеуетін бағалау үшін қолданылуы мүмкін.

Түйін сөздер: скарн кенорындары; магнетитті кендер; контактылы метасоматоз; минералогия; геофизикалық аномалиялар; Солтүстік-Батыс Балқаш өңірі.

^{1*}**Ныгманова А.С.,**

магистр, ст. преподаватель, ORCID ID: 0009-0003-3917-2707,

e-mail: a.emls@kbtu.kz

¹**Коробкин В.В.,**

кандидат геолого-минералогических наук, профессор, ORCID ID: 0000-0002-1562-759X,

e-mail: korobkin_vv@kbtu.kz

²**Буслов М.М.,**

доктор геолого-минералогических наук, профессор, ORCID ID: 0000-0003-0606-2264,

e-mail: buslov@igm.nsc.ru

¹**Чакликов А.Е.,**

PhD, ассистент-профессор, ORCID ID: 0000-0001-8316-6599,

e-mail: a.chaklikov@kbtu.kz

¹**Тулемисова Ж.С.,**

PhD, ассоциированный профессор, ORCID ID: 0000-0003-1803-4535,

e-mail: z.tulemissova@kbtu.kz

^{1,3}**Койшибаев С.М.,**

магистрант, консультант геолог, ORCID ID: 0009-0005-6492-4215

e-mail: skoishibayev@srk.kz

^{1,3}Казахстанско-Британский технический университет, г. Алматы, Казахстан

²Институт геологии и минералогии им. В.С. Соболева СО РАН, Новосибирский
государственный университет, г. Новосибирск, Россия

³SRK Consulting, г. Алматы, Казахстан

ГЕОЛОГИЧЕСКИЕ И МИНЕРАЛОГИЧЕСКИЕ ФАКТОРЫ КОНТРОЛЯ ФОРМИРОВАНИЯ ЖЕЛЕЗНЫХ СКАРНОВ В СЕВЕРО-ЗАПАДНОМ ПРИБАЛХАШЬЕ (ЦЕНТРАЛЬНЫЙ КАЗАХСТАН)

Аннотация

Комплексное исследование позволило выявить основные геологические и минералогические факторы, контролирующие формирование железных скарновых месторождений Северо-Западного Прибалхашья на примере репрезентативных объектов, таких как Жуантобе и Караулкен. Интегрированный анализ данных геологического картирования, минералогических и петрографических наблюдений, геохимических исследований и геофизических интерпретаций показывает, что формирование скарнов представляет собой многостадийную контактно-метасоматическую систему, контролируемую магматическими, структурными и флюидными факторами. Установлены закономерности распределения минеральных ассоциаций и текстурно-структурных типов руд, отражающие изменения флюидных режимов и физико-химических условий рудообразования. Полученные результаты уточняют региональную генетическую модель формирования железных скарнов Северо-Западного Прибалхашья и могут быть использованы для прогноза и оценки перспектив новых рудных объектов.

Ключевые слова: скарновые месторождения, магнетитовые руды, контактный метасоматоз, минералогия, геофизические аномалии, Северо-Западное Прибалхашье.

UDC 553.411.071
IRSTI 38.49.19

<https://doi.org/10.55452/1998-6688-2026-23-2-481-493>

¹**Umarbekova Z.T.,**

PhD, ORCID ID: 0000-0001-7890-1851,
e-mail: zama7777@mail.ru

¹**Gadeev R.,**

Researcher, ORCID ID: 0000-0002-9428-8142,
e-mail: gadeev_ravil@mail.ru

^{1,2}**Togizov K.**

PhD, ORCID ID: 0000-0002-4830-405X,
e-mail: k.togizov@satbayev.university

¹**Mashrapova M.A.,**

PhD, ORCID ID: 0000-0002-6009-9730,
e-mail: moldir_m_m@mail.ru

^{1,2*}**Miniskul Sh.D.,**

Researcher, ORCID ID: 0000-0001-8878-2008,
*e-mail: shat_230393@mail.ru

¹**Amanbaev R.A.,**

Junior Researcher, ORCID ID: 0000-0002-7798-1076,
e-mail: rustemamanbaev_85@mail.ru

¹K.I. Satpayev Institute of Geological Sciences, Almaty, Kazakhstan

²Satbayev University, Almaty, Kazakhstan

PHASE FORMS OF GOLD AND THEIR INFLUENCE ON RECOVERY IN THE OXIDATION ZONE OF THE ARKHARLY GOLD-SILVER DEPOSIT (KAZAKHSTAN)

Abstract

The aim of this study is to determine the phase forms of gold in the oxidation zone of the Arkharly deposit and to assess their influence on the efficiency of gold and silver recovery. The main research methods included mineralogical analysis, electron microprobe studies, and technological testing. The study identified and characterized silver halides, established their genetic relationship with gold, and substantiated the role of supergene processes in the redistribution of gold. The Arkharly gold-silver deposit (Kazakhstan) is characterized by a complex composition of ores and a variety of gold occurrence forms, which significantly affect processing efficiency. This paper examines the phase forms of gold in the oxidation zone and their impact on technological recovery indicators. Based on mineralogical, electron microprobe, and technological studies, it was found that gold occurs in various forms: native gold of different fineness, electrum, as well as finely dispersed particles associated with silver minerals. Particular attention is given to silver halides, widely developed in the oxidation zone and forming complex intergrowths with native silver and gold. Textural and structural features of mineral aggregates indicate the redistribution of gold under supergene conditions. The formation of gold-bearing phases is associated with chloride and, presumably, colloidal migration of matter, leading to the formation of zonal structures and inclusions of high-fineness gold. Technological test results showed low efficiency of gravity concentration due to the fine-dispersed nature of gold, and high efficiency of cyanidation and sorption leaching (gold recovery up to 93.5%). A direct relationship between the phase forms of gold and recovery indicators has been established. The obtained results expand the understanding of gold behavior in the oxidation zone and can be used in developing effective technologies for processing complex gold-silver ores.

Keywords: Arkharly deposit, gold, oxidation zone, silver halides, phase forms of gold, supergene processes, cyanidation, sorption leaching.

Received April 15, 2026; revised May 12, 2026; accepted May 28, 2026.

Introduction

Epithermal gold–silver deposits are characterized by a significant diversity of ore forms [1–3]. This diversity exerts a decisive influence on their technological properties and the efficiency of ore processing. One of the key challenges in the development of such deposits is the presence of finely dispersed and refractory gold associated with various minerals and products of supergene alteration.

A special role in the redistribution of gold is played by processes occurring within the oxidation zone [4, 5]. Under these conditions, primary mineral associations are transformed, secondary minerals are formed, and noble metals undergo migration and reprecipitation. In a number of cases, supergene processes lead to the formation of high-fineness gold, the development of secondary enrichment zones, and the emergence of complex phase forms of the metal that differ significantly from hypogene ones.

Despite the considerable number of studies devoted to the mineralogy of gold–silver deposits, issues related to the role of silver halide compounds [6] remain insufficiently investigated. This is particularly true for deposits where complex mineral associations involving silver chlorides, bromides, and iodides are formed within the oxidation zone and may serve as indicators of specific geochemical conditions.

The Arkharly deposit, located in Southern Dzhungaria, represents a typical example of gold–silver epithermal mineralization [7], characterized by widespread quartz veins, low sulfide content, and complex mineral zonation. Unlike many other deposits in Kazakhstan, it is distinguished by elevated contents of lead and zinc, an almost complete absence of arsenic, and the widespread occurrence of adularia, amethyst, and hematite within the ores.

In the oxidation zone of the deposit, diverse supergene mineral associations are formed, accompanied by the redistribution of gold and silver [8]. The morphology and composition of gold vary significantly, indicating complex mechanisms of its transport and deposition. In particular, chloride and colloidal forms of migration may play an important role, leading to the formation of zonal structures and finely dispersed gold inclusions within silver minerals.

At the same time, the influence of gold phase forms on technological processing indicators, including gravity concentration and cyanidation, remains insufficiently studied. Practical experience shows that finely dispersed and structurally bound gold significantly reduces the efficiency of conventional extraction methods and requires a deeper understanding of its mineralogical nature.

This paper examines the characteristics of gold phase forms in the oxidation zone of the Arkharly deposit, their genetic relationship with silver halides, and the influence of the identified mineralogical features on the efficiency of technological processes for gold and silver recovery.

Geological Setting

The Arkharly deposit is located in Southern Dzhungaria and is confined to the Saryozek synclinorium, within the southern limb of the Arkharly anticline. The geological structure of the area is defined by the development of volcanogenic-sedimentary and effusive-pyroclastic complexes of Permian and Triassic age, represented by rocks of andesite–basalt, trachyandesite–trachybasalt, and trachyrhyolite formations. The geological structure of the deposit is shown in Figure 1.

The volcanic sequences are intruded by numerous intrusive bodies of Permian and Triassic age, including sills of basaltic and andesitic porphyrites, trachyrhyolite porphyries, as well as dikes of diorites, lamprophyres, and felsite porphyries [9].

Intrusive rocks within the study area belong to the Katutau and South Dzhungarian complexes. The former is represented by gabbro-diorites and granite porphyries, while the latter, within the described area, includes gabbro, diorites, and granosyenites.

A more detailed formational and petrochemical characterization of the South Dzhungarian complex is provided in [10].

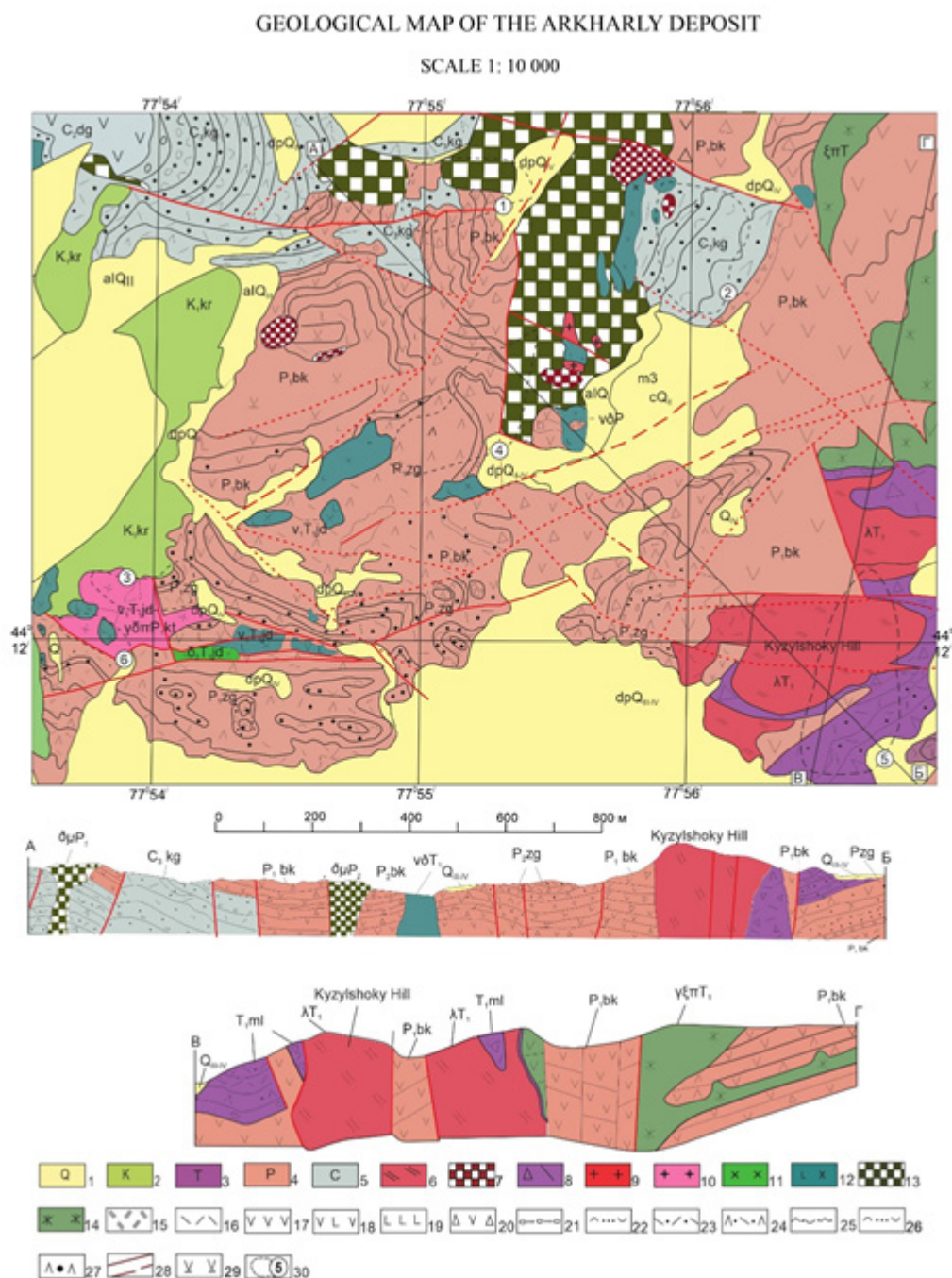


Figure 1 – Geological map of the Arkharly Au–Ag deposit showing lithological units, fault systems, and ore zones. Modified using data from Zhautikov T.M., Skrinnik L.I., as well as the authors' own observations.

- 1 – Quaternary deposits; 2 – Cretaceous deposits; 3 – Triassic deposits; 4 – Permian deposits; 5 – Carboniferous deposits; 6 – Kyzylshoky necks (λT_1); 7 – subvolcanic trachyandesite bodies (T); 8 – vent and near-vent breccias of Kyzylshoky; 9 – granite porphyries (Pkt); 10 – granodiorite porphyries ($\gamma\delta\pi P_{1kt}$); 11 – diorites ($\delta\pi T_{1jd}$); 12 – gabbro-diabases ($v\pi T_{1jd}$); 13 – subvolcanic trachyandesite bodies ($\delta\mu P$); 14 – syenodiorite and granodiorite porphyries ($\xi\pi T$); 15 – trachyrhyolite lavas; 16 – rhyolites; 17 – andesites; 18 – andesite-basalts; 19 – basalts; 20 – andesitic clastolavas; 21 – sandstones, conglomerates; 22 – tuffaceous sandstones; 23 – rhyolitic tuffs; 24 – rhyodacitic tuffs; 25 – thin interbedding of sandstones and tuffaceous sandstones; 26 – tuffaceous sandstones with tuffite interlayers; 27 – dacitic tuffs; 28 – faults: (a) exposed; (b) concealed beneath Quaternary deposits; 29 – pyroxene–plagioclase andesite-basalts; 30 – ore sites: 1 – Northern, 2 – Northeastern, 3 – Eastern, 4 – Eastern II, 5 – Kyzylshoky, 6 – Central.

The formation of the ore field is associated with the development of a dome-shaped volcano-plutonic structure controlling the localization of gold–silver mineralization.

Ore bodies are represented by a system of quartz veins, the total number of which reaches several hundred. The veins are up to 150–200 m long, 0.5 to 12 m thick, and dip at angles of 40–70°. The vertical extent of mineralization exceeds 600 m. Within the ore field, several economically significant ore zones are distinguished.

It should also be noted that breccia-type ores (explosive and eruptive breccia bodies) containing gold are present at the deposit. Their formation is associated with hydrothermal eruptions [11].

Metasomatic alterations of the host rocks are characterized by well-developed zonation. Zones of secondary quartzites (alunite-, kaolinite-, and sericite-bearing) and propylites of various compositions (chlorite–sericite–carbonate and chlorite–epidote–albite) are distinguished. Economically significant mineralization is predominantly localized within the sericite–chlorite–carbonate propylite zone.

Near-ore alterations include an outer propylitic zone and an inner zone of intense carbonatization, sericitization, and silicification. A characteristic feature of the deposit is the development of adularization associated with productive stages of ore formation [12].

Mineral zonation of the ore field is expressed by a transition from deep-seated pyrite–quartz assemblages to productive adularia–quartz and galena–sphalerite–quartz associations, followed by late chalcedony and carbonate assemblages, reflecting the evolution of the hydrothermal system.

Supergene processes are well developed in the oxidation zone and represent the final stage of deposit transformation [13]. At this stage, primary minerals are altered, iron oxides and hydroxides are formed, and secondary minerals develop. Under these conditions, gold is redistributed and secondary forms are formed, which is important for understanding the genesis of mineralization and evaluating the technological properties of ores.

Field observations confirm these patterns and allow for a more detailed characterization of the host rocks and hydrothermal alterations. Volcanic rocks of intermediate to felsic composition are developed, in which quartz veinlets and veins of varying thickness are widespread. Quartz veins often exhibit a crustiform-banded structure, reflecting the rhythmic precipitation of silica from pulsating hydrothermal fluids. Pronounced zonation of hydrothermal alteration is observed along the veins, typical of low-sulfidation deposits. From the quartz vein outward, there is a gradual transition to clay-altered rocks involving kaolinite and sericite.

Within the oxidation zones, quartz veins acquire a brownish-red coloration due to limonitization, forming halos of intense supergene alteration. In areas more distant from the veins, propylitization of andesites is observed, expressed by a greenish coloration of the rocks and the development of chlorite and epidote.

Field observations reflecting structural and metasomatic features are presented in Figure 2.

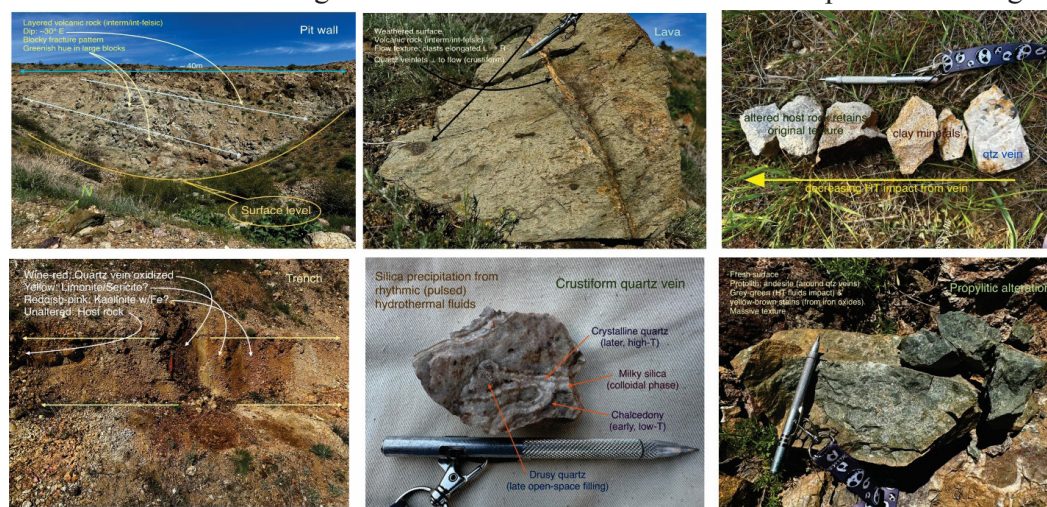


Figure 2 – Field relationships of quartz veins and hydrothermal alteration zones at the Arkharly deposit

Materials and methods

The object of the study was oxidized gold–silver ores from the Arkharly deposit, sampled from various ore zones and stratigraphic levels. Representative ore samples, as well as technological samples prepared by quartering composite material collected from mine workings and industrial dumps, were used for mineralogical, geochemical, and technological investigations.

Mineralogical studies were carried out using optical microscopy (thin and polished sections), which made it possible to determine the morphological features of minerals, the nature of their intergrowths, and the textural and structural characteristics of the ores. Particular attention was paid to the occurrence forms of gold and silver and their relationships with secondary minerals in the oxidation zone.

The chemical composition of minerals was determined using electron microprobe analysis, which provided quantitative data on major element contents and revealed compositional zoning within individual grains. Analyses were conducted at multiple points within single mineral grains, allowing assessment of compositional variations, including changes in gold fineness within a single grain.

Additionally, X-ray diffraction (XRD) analysis was employed to refine the mineral composition of selected samples. The results enabled identification of mineral phases and confirmed the presence of silver halides.

Technological studies included tests on gravity concentration, cyanidation, and sorption leaching of gold and silver. Prior to experimentation, the ore material was crushed and ground to a specified particle size (85–90% passing -0.074 mm). Gravity concentration was performed using a shaking table, allowing evaluation of free gold recovery.

Cyanidation processes were carried out under laboratory conditions with controlled concentrations of sodium cyanide and lime, as well as controlled leaching duration. To investigate process kinetics, a series of experiments was conducted with varying contact times and degrees of ore grinding.

Sorption leaching was performed using an anion-exchange resin, enabling evaluation of gold and silver recovery from the pulp. During the experiments, pulp parameters such as density, reagent concentration, and process duration were carefully monitored.

Analytical determination of gold and silver contents in the initial samples, processing products, and solutions was carried out using fire assay and chemical methods. The obtained data were used to calculate metal recovery and evaluate the efficiency of different processing flowsheets.

Results

Mineralogical Characteristics of the Ores

The ores of the Arkharly deposit are predominantly composed of quartz (up to 50–95%), with the presence of adularia, amethyst, chalcedony, chlorite, sericite, and carbonates. The sulfide content does not exceed 3%, which allows the ores to be classified as low-sulfide type.

The principal ore minerals include sphalerite, galena, pyrite, hematite, and native gold. Secondary minerals are represented by chalcopyrite, arsenopyrite, bornite, pyrrhotite, acanthite, and other phases. A distinctive feature of the deposit is the widespread occurrence of adularia, amethyst, and hematite.

Several paragenetic associations have been identified, among which the adularia–quartz and galena–sphalerite–quartz associations are productive [14]. These associations form the main gold–silver ores and determine their technological properties. The main mineral associations are summarized in Table 1.

Table 1 – Main mineral associations of the Arkharly deposit (modified after [14])

Association	Main minerals	Genetic type
Pyrite–quartz	Pyrite, quartz	Pre-ore
Adularia–quartz	Adularia, quartz, Au	Ore (productive)
Galena–sphalerite–quartz	Galena, sphalerite, Au, Ag	Ore (productive)
Chalcopyrite–quartz	Chalcopyrite	Post-ore
Chalcedony	Chalcedony	Late stage

Textural Characteristics of the Arkharly Deposit Ores

The ores of the Arkharly deposit are characterized by a wide variety of textural and structural types, reflecting the complex conditions of their formation. Breccia-like textures are the most widespread, represented by fragments of host rocks cemented by quartz, chalcedony, or carbonate material.

Colloform and colloform-banded (agate-like) textures are also widely developed and are often combined with rhythmic banding. This indicates a pulsating supply of hydrothermal fluids and periodic precipitation of silica.

Veinlet and stockwork textures are ubiquitous and reflect multiple stages of fracturing followed by infilling with quartz and chalcedony. In some cases, lens-shaped and nest-like siliceous bodies with internal zoning are observed.

Massive and disseminated textures are also present and are typical of relatively weakly altered ore zones. In certain samples, cataclastic structures are identified, indicating a significant role of tectonic processes in the formation of the ore body.

Overall, the predominance of breccia-like, banded, and colloform textures, as well as their combination with veinlet structures, indicates a complex interaction of hydrothermal, tectonic, and supergene processes during ore formation.

Representative ore textures are shown in Figure 3, while Table 2 demonstrates the relationship between gold and silver contents and ore type and texture.



Figure 3– Representative ore samples from the Arkharly deposit showing brecciated, banded and crustiform quartz textures.

Table 2 – Main ore types and their characteristics

Ore type	Texture	Au (g/t)	Ag (g/t)
Adularia–quartz	Banded	up to 280	up to 156
Amethyst–quartz	Disseminated	0.1–40	up to 125
Chalcedony	Colloform	<0.4	up to 20
Pyrite–quartz	Massive	<0.5	up to 5

Phase Forms of Gold

Gold in the ores of the deposit is represented by various morphological and genetic types. At least two generations of gold have been identified, differing in fineness and formation conditions.

Early-stage gold is characterized by relatively low fineness (approximately 596–635) and is associated with pyrite, chalcopyrite, and sphalerite. Later-stage gold exhibits higher fineness (up to 735–772) and is related to a galena–sphalerite association.

The morphology of gold particles shows considerable diversity, including dendritic, platy, irregular, and spongy forms [15]. Gold commonly occurs as monomineralic inclusions in quartz or as complex intergrowths with other minerals.

A characteristic feature is the presence of compositional zoning within individual grains, indicating multiple stages of mineral formation.

Supergene Alteration and Silver Halides

In the oxidation zone of the deposit, a wide development of secondary minerals has been identified, including hematite, goethite, jarosite, malachite, and azurite. A significant result of the study is the identification of silver halides forming complex intergrowths with native silver and gold [16]. The association of gold with silver halides is shown in Figure 4.

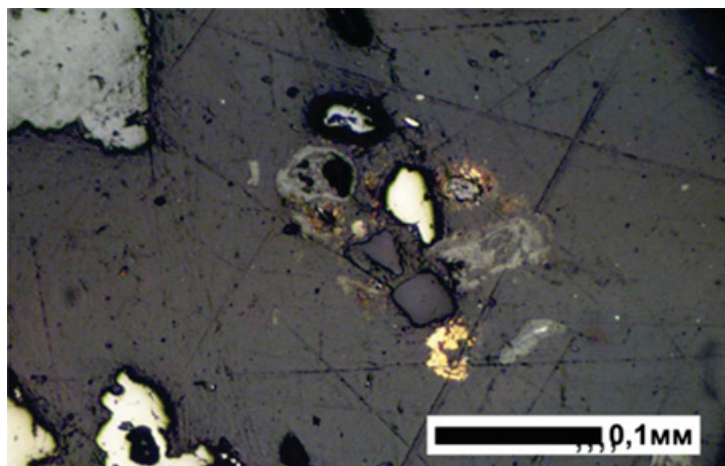


Figure 4 – Native gold associated with silver halides (grey) and pyrite in polished section.

Silver halides are represented by various phases (chlorides, bromides, and iodides) forming concentrically zoned structures. The most common are grains with a native silver core and halide rims containing microinclusions of gold [17].

Figure 5 shows chalcopyrite with a covellite rim and associated silver halides.

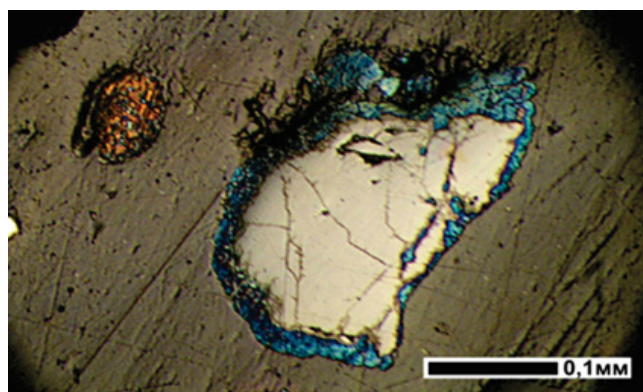


Figure 5 – Chalcopyrite with covellite rim and associated silver halides in polished section

Complex intergrowths of electrum, native silver, and high-fineness gold are also observed (Figure 6). In some cases, gold inclusions are identified within silver and halide phases, indicating their co-formation.



Figure 6 – BSE image showing intergrowth of electrum (1) and native silver (2) with inclusions of high-grade gold

Gold in the oxidation zone is characterized by significant variations in fineness (from ~67 to 98%), including within a single grain, which indicates its redistribution under supergene conditions.

Textural and Structural Characteristics of the Ores

The ores are characterized by the widespread development of colloform, banded, brecciated, and veinlet textures. Rhythmic zonal structures are frequently observed, indicating periodic precipitation of material.

Colloform and zonal textures are particularly typical of siliceous formations and are associated with finely dispersed gold. Breccia textures indicate the role of tectonic processes and repeated reopening of ore bodies.

The observed textural features point to a significant role of colloidal and hydrothermal–supergene processes in ore formation.

Technological Properties of the Ores

The results of technological tests show that the ores are characterized by low efficiency of gravity concentration. Gold recovery into gravity concentrates does not exceed ~18%, which is related to the predominance of finely dispersed and, in particular, structurally bound gold [18,19].

Therefore, the application of technologies ensuring more complete gold recovery is required. The average chemical composition of the deposit ore is presented in Table 3, while a comparison of different recovery methods is shown in Table 4. The average chemical composition of the ore is shown in Table 3.

Table 3 – Average chemical composition of ore

Component	Content
Au	8.17 g/t
Ag	78.73 g/t
Cu	0.012 %
Zn	0.02 %
Fe	1.91 %

Table 4 – Comparison of different methods for gold and silver recovery

Method	Au recovery (%)	Ag recovery (%)
Gravity concentration	18.5	15.3
Cyanidation	91.0	66.7
Sorption leaching	93.6	80.1
Combined scheme	92.1	74.8

Cyanidation provides gold recovery at the level of approximately 91%; however, a more efficient method is sorption leaching, which achieves recoveries of up to 93.5% for gold and up to 80% for silver.

It has been established that the optimal grinding size is about 85% passing -0.074 mm, while further increases in fineness do not lead to a significant improvement in recovery.

A combined processing flowsheet, including gravity concentration followed by sorption leaching, ensures a total gold recovery of approximately 92%.

Discussion

The obtained results indicate that the formation and redistribution of gold in the oxidation zone of the Arkharly deposit are controlled by a combination of supergene processes, including dissolution, migration, and subsequent precipitation of noble metals under varying physicochemical conditions.

One of the key factors is the involvement of chloride-bearing solutions formed as a result of sulfide oxidation and interaction with host rocks [6–20]. The presence and widespread development of silver halides in the oxidation zone indicate a significant role of halide migration, primarily for silver, and indirectly support the possibility of gold transport in the form of complex species.

The identified textural and structural features, such as concentrically zoned aggregates, inclusions of gold in silver and halides, as well as the occurrence of high-fineness gold as fine inclusions, indicate a staged character of supergene mineral formation. This suggests that gold precipitation occurred in multiple stages associated with changes in redox conditions and fluid chemistry.

Significant variations in gold fineness, including changes within a single grain, indicate processes of metal redistribution and recrystallization. High-fineness gold formed in the oxidation zone is likely the result of silver leaching from primary alloys and subsequent gold enrichment. Gold fineness reaching up to 98% in the oxidation zone confirms the leaching of silver from primary alloys and supports the supergene nature of gold redistribution [4].

Colloidal processes appear to play an important role in the formation of gold-bearing phases. The widespread development of colloform textures and the spongy morphology of gold grains indicate precipitation from gel-like media. Such conditions favor the formation of finely dispersed gold and its association with secondary minerals.

The established mineralogical features are directly reflected in the technological properties of the ores. The low efficiency of gravity concentration is explained by the fact that a significant portion of gold occurs in finely dispersed forms or as intergrowths with other minerals, which hinders its recovery by gravity methods.

At the same time, the high efficiency of cyanidation and sorption leaching is associated with the accessibility of gold for dissolution despite its dispersed state [21, 22]. This indicates that gold is not entirely locked within resistant mineral phases but occurs in forms amenable to chemical extraction.

Thus, the identified relationship between gold phase forms and technological recovery parameters confirms that mineralogical composition is a key factor controlling ore processing efficiency. Accounting for these features enables the rational selection of processing flowsheets and improves gold and silver recovery.

Conclusion

It has been established that the ores of the Arkharly deposit are characterized by a complex mineralogical composition and a wide variety of gold phase forms, represented by native gold of varying fineness, electrum, and finely dispersed occurrences.

In the oxidation zone of the deposit, a widespread development of silver halides has been identified, forming complex intergrowths with native silver and gold, indicating specific geochemical conditions of supergene mineral formation.

Gold redistribution in the oxidation zone is associated with migration in chloride-bearing and, likely, colloidal solutions, followed by precipitation as high-fineness gold and finely dispersed inclusions.

A staged character of gold formation has been established, expressed in zonal grain structures, variations in fineness, and morphological diversity of gold particles, reflecting changes in physicochemical conditions during mineral formation.

It has been shown that the textural and structural features of the ores (colloform, zonal, and brecciated textures) indicate a significant role of colloidal and supergene processes in the formation of gold–silver mineralization.

Low efficiency of gravity concentration is due to the finely dispersed nature of gold and its occurrence in intergrowths, whereas cyanidation and sorption leaching provide high recovery levels (up to 93.5% for gold).

It has been demonstrated that gold phase forms are a controlling factor in ore processing, and consideration of mineralogical features allows for the justification of an optimal gold and silver recovery flowsheet.

The obtained results have both theoretical and practical significance. On one hand, they refine the understanding of mechanisms of supergene gold redistribution and the role of silver halides in these processes. On the other hand, they can be applied to the development and optimization of processing technologies for similar oxidized gold–silver ores.

Funding

This research was funded by The Ministry of Science and Higher Education of the Republic of Kazakhstan, AP23487772 «Studying silver and gold halides in the Arkharly deposit (Jungar Alatau) for mining purposes».

REFERENCES

- 1 Hedenquist, J.W., Arribas, A., Gonzalez-Urien, E. Exploration for epithermal gold deposits. *Reviews in Economic Geology*, 13, 245–277 (2000). (in English).
- 2 Simmons, S.F., White, N.C., John, D.A. Geological characteristics of epithermal precious metal deposits. *Economic Geology*, 100th Anniversary Volume, 485–522 (2005). (in English).
- 3 Cooke, D.R., Simmons, S.F. Characteristics and genesis of epithermal gold deposits. *SEG Reviews*, 13, 221–244 (2000). (in English).
- 4 Thornber, M.R. Supergene processes in gold deposits. *Ore Geology Reviews*, 7, 349–364 (1992). (in English).
- 5 Webster, J.G., Mann, A.W. The influence of climate on the geochemistry of gold. *Geochimica et Cosmochimica Acta*, 48, 599–614 (1984). (in English).
- 6 Boyle, R.W. *The Geochemistry of Gold and Its Deposits* (Geological Survey of Canada, 1979), 584 p. (in English).
- 7 Zhautikov, T.M. et al. Gold-silver mineralization of Kazakhstan. *Geology of Kazakhstan*, 212–221 (2008). (in Russian).
- 8 Rafailovich, M.S., Egorov, S.A., Starova, M.M. et al. On a new type of epithermal gold-silver deposits in Kazakhstan. *Geology and Exploration of Subsurface of Kazakhstan*, No. 2, 7–12 (1996). (in Russian).
- 9 Zhautikov, T.M. Volcanic dome structures and their role in ore formation of Southern Jungaria. *Proceedings of the National Academy of Sciences of the Republic of Kazakhstan. Geological Series*, No. 4, 21–34 (2003). (in Russian).
- 10 Skrinnik, L., Gadeev, R., Umarbekova, Z., Tretyakov, A. Collisional and orogenic granitoids of Kazakhstan. *SGEM 2020 Proceedings*, 51–66 (2020). <https://doi.org/10.5593/SGEM2020/1.1/S01.006>. (in English).
- 11 Umarbekova, Z.T. et al. The role of hydrothermal eruptions in gold mineralization at the Arkharly epithermal gold-silver deposit. *SGEM 2025 Proceedings*, 97–100 (2025). <https://doi.org/10.5593/sgem2025/1.1/s01.13>. (in English).
- 12 Zhautikov, T.M. Regularities of Placement and Principles of Prediction of Gold Mineralization in Kazakhstan (Almaty, 1987), 448 p. PhD Thesis. (in Russian).
- 13 Corbett, G.J., Leach, T.M. Southwest Pacific Rim Gold-Copper Systems: Structure, Alteration and Mineralization (SEG, 1998). (in English).
- 14 Grebenchikov, A.M. Mineralogical and Geochemical Features of Near-Surface Gold Deposits of the Late Paleozoic Volcanic Belt of Kazakhstan (Moscow, 1976), 24 p. PhD Abstract. (in Russian).
- 15 Zhautikov, T.M. et al. Gold and silver of the hypogene zone of gold and gold-bearing deposits of Kazakhstan. *Bulletin of NAS RK. Geological and Technical Sciences Series*, No. 3, 15–33 (2011). (in Russian).
- 16 Umarbekova, Z.T. et al. Silver halides in the hypogene zone of the Arkharly gold deposit as indicators of formation in dry and hot climate (Dzungar Alatau, Kazakhstan). *International Journal of Engineering Research and Technology*, 13 (1) (2020). Available at: <http://www.irphouse.com>. (in English).
- 17 Umarbekova, Z.T. et al. The halides of silver in the hypogene zone of the Arkharly gold-silver deposit (South Dzhungaria). *News of NAS RK. Series of Geology and Technical Sciences*, 2(428), 242–250 (2018). (in English).
- 18 Reich, M., Kesler, S.E., Utsunomiya, S. Solubility of gold in arsenian pyrite. *Geochimica et Cosmochimica Acta*, 69, 2781–2796 (2005). (in English).
- 19 Palenik, C.S., Utsunomiya, S., Reich, M. “Invisible” gold revealed. *American Mineralogist*, 89, 1359–1366 (2004). (in English).
- 20 Saunders, J.A. Colloidal transport of gold. *Economic Geology*, 85, 171–181 (1990). (in English).
- 21 Marsden, J., House, I. *The Chemistry of Gold Extraction*, 2nd ed. (SME, 2006), 651 p. (in English).
- 22 Adams, M.D. (Ed.). *Advances in Gold Ore Processing* (Elsevier, 2005), 1076 p. (in English).

¹Умарбекова З.Т.,

PhD, ORCID ID: 0000-0001-7890-1851,

e-mail: zama7777@mail.ru

¹Гадеев Р.,

Ғылыми қызметкер, ORCID ID: 0000-0002-9428-8142,

e-mail: gadeev_ravil@mail.ru

^{1,2}Тогизов К.,

PhD, ORCID ID: 0000-0002-4830-405X,

e-mail: k.togizov@satbayev.university

¹Машрапова М.А.,

PhD, ORCID ID: 0000-0002-6009-9730

e-mail: moldir_m_m@mail.ru

^{1,2*}Минискул Ш.Д.,

Ғылыми қызметкер, ORCID ID: 0000-0001-8878-2008

*e-mail: shat_230393@mail.ru

¹Аманбаев Р.А.,

кіші ғылыми қызметкер, ORCID: 0000-0002-7798-1076,

e-mail: rustemamanbaev_85@mail.ru

¹Қ.И. Сәтбаев атындағы Геологиялық ғылымдар институты, Алматы қ., Қазақстан²Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті,
Алматы қ., Қазақстан

АРХАРЛЫ АЛТЫН-КҮМІС КЕН ОРНЫНЫҢ ТОТЫҒУ АЙМАҒЫНДАҒЫ АЛТЫННЫҢ ФАЗАЛЫҚ ТҮРЛЕРІ ЖӘНЕ ОЛАРДЫҢ АЛЫНУ ТИІМДІЛІГІНЕ ӘСЕРІ (ҚАЗАҚСТАН)

Аңдатпа

Осы зерттеудің мақсаты – Архарлы кен орнының тотығу аймағындағы алтынның фазалық түрлерін анықтау және олардың алтын мен күмісті алу тиімділігіне әсерін бағалау. Негізгі зерттеу әдістері ретінде минералогиялық талдау, электрондық-зондтық зерттеулер және технологиялық сынақтар қолданылды. Жүргізілген зерттеулер нәтижесінде күмістің галогенидтері анықталып, олардың алтынмен генетикалық байланысы көрсетілді, сондай-ақ алтынның қайта таралуындағы гипергендік процестердің рөлі негізделді. Архарлы алтын-күміс кен орны (Қазақстан) кендердің күрделі құрамымен және алтынның әртүрлі кездесуімен сипатталады, бұл олардың өңдеу тиімділігіне айтарлықтай әсер етеді. Жұмыста тотығу аймағындағы алтынның фазалық түрлері және олардың технологиялық көрсеткіштерге әсері қарастырылған. Минералогиялық, электрондық-зондтық және технологиялық зерттеулер негізінде алтынның әртүрлі түрде кездесетіні анықталды: әртүрлі сынамалы табиғи алтын, электрум, сондай-ақ күміс минералдарымен байланысқан ұсақ дисперсті бөлшектер. Ерекше назар тотығу аймағында кең таралған және табиғи күміс пен алтынмен күрделі бірігулер түзетін күміс галогенидтеріне аударылды. Минералдық агрегаттардың текстуралық-құрылымдық ерекшеліктері гипергендік жағдайларда алтынның қайта таралуын көрсетеді. Алтынды фазалардың қалыптасуы заттың хлоридтік және, мүмкін, коллоидтық миграциясымен байланысты, бұл зоналық құрылымдардың және жоғары сынамалы алтын қосындыларының түзілуіне әкеледі. Технологиялық сынақтардың нәтижелері алтынның ұсақ дисперсті сипатына байланысты гравитациялық байытудың төмен тиімділігін және циандау мен сорбциялық шаймалаудың жоғары тиімділігін көрсетті (алтынды алу 93,5%-ға дейін). Алтынның фазалық түрлері мен оны алу көрсеткіштері арасында тікелей байланыс анықталды. Алынған нәтижелер тотығу аймағындағы алтынның мінез-құлқын тереңірек түсінуге мүмкіндік береді және күрделі алтын-күміс кендерін өңдеудің тиімді технологияларын әзірлеуде қолданылуы мүмкін.

Түйін сөздер: Архарлы кен орны, алтын, тотығу аймағы, күміс галогенидтері, алтынның фазалық түрлері, гипергендік процестер, циандау, сорбциялық шаймалау.

¹Умарбекова З.Т.,

PhD, ORCID ID: 0000-0001-7890-1851,

e-mail: zama7777@mail.ru

¹Гадеев Р.,

научный сотрудник, ORCID ID: 0000-0002-9428-8142,

e-mail: gadeev_ravil@mail.ru

^{1,2}Тогизов К.,

PhD, ORCID ID: 0000-0002-4830-405X,

e-mail: k.togizov@satbayev.university

¹Машрапова М.А.,

PhD, ORCID ID: 0000-0002-6009-9730,

e-mail: moldir_m_m@mail.ru

^{1,2*}Минискул Ш.Д.,

научный сотрудник, ORCID ID: 0000-0001-8878-2008,

*e-mail: shat_230393@mail.ru

¹Аманбаев Р.А.,

младший научный сотрудник, ORCID ID: 0000-0002-7798-1076,

rustemamanbaev_85@mail.ru

¹Институт геологических наук им К.И. Сатпаева, г. Алматы, Казахстан

²Казахский национальный исследовательский технический университет им. К.И. Сатпаева,
г. Алматы, Казахстан

ФАЗОВЫЕ ФОРМЫ ЗОЛОТА И ИХ ВЛИЯНИЕ НА ИЗВЛЕКАЕМОСТЬ В ЗОНЕ ОКИСЛЕНИЯ НА ЗОЛОТО-СЕРЕБРЯННОМ МЕСТОРОЖДЕНИИ АРХАРЛЫ (КАЗАХСТАН)

Аннотация

Целью настоящего исследования является установление фазовых форм золота в зоне окисления месторождения Архарлы и оценка их влияния на эффективность извлечения золота и серебра. Основным методом исследования являлись минералогический анализ, электронно-зондовые и технологические исследования. Проведенным исследованием выявлены и охарактеризованы галогениды серебра, установлена их генетическая связь с золотом и обоснована роль гипергенных процессов в перераспределении золота. Золото-серебряное месторождение Архарлы (Казахстан) характеризуется сложным вещественным составом руд и разнообразием форм нахождения золота, существенно влияющих на эффективность их переработки. В работе рассмотрены фазовые формы золота в зоне окисления и их влияние на технологические показатели извлечения. На основе минералогических, электронно-зондовых и технологических исследований установлено, что золото представлено различными формами: самородным золотом разной пробы, электрумом, а также тонкодисперсными выделениями, ассоциированными с минералами серебра. Особое внимание уделено галогенидам серебра, широко развитым в зоне окисления и образующим сложные сростания с самородным серебром и золотом. Тектурно-структурные особенности минеральных агрегатов свидетельствуют о перераспределении золота в гипергенных условиях. Формирование золотоносных фаз связано с хлоридной и, вероятно, коллоидной миграцией вещества, что приводит к образованию зональных структур и включений высокопробного золота. Результаты технологических испытаний показали низкую эффективность гравитационного обогащения, обусловленную тонкодисперсным характером золота, и высокую эффективность цианирования и сорбционного выщелачивания (извлечение золота до 93,5%). Установлена прямая зависимость между фазовыми формами золота и показателями его извлечения. Полученные результаты расширяют представления о поведении золота в зоне окисления и могут быть использованы при разработке эффективных технологий переработки сложных золото-серебряных руд.

Ключевые слова: месторождение Архарлы, золото, зона окисления, галогениды серебра, фазовые формы золота, гипергенные процессы, цианирование, сорбционное выщелачивание.

UDC 553.31(574)

IRSTI 38.51.19

<https://doi.org/10.55452/1998-6688-2026-23-2-494-503>^{1*}**Nygmanova A.S.,**

PhD student, Senior-lecturer, ORCID ID: 0009-0003-3917-2707,

*e-mail: a.emls@kbtu.kz

¹**Korobkin V.V.,**

Candidate of Geological and Mineralogical Sciences, Professor,

ORCID ID: 0000-0002-1562-759X,

e-mail: korobkin_vv@kbtu.kz

²**Buslov M.M.,**

Doctor of Geological and Mineralogical Sciences, Professor, ORCID ID: 0000-0003-0606-2264,

e-mail: buslov@igm.nsc.ru

³**Chaklikov A.Ye.,**

PhD, Assistant Professor, ORCID ID: 0000-0001-8316-6599,

e-mail: a.chaklikov@kbtu.kz

^{1,3}Kazakh-British Technical University, Almaty, Kazakhstan²Sobolev Institute of Geology and Mineralogy, Siberian Branch of the Russian Academy
of Sciences, Novosibirsk State University, Novosibirsk, Russia

GEOLOGICAL CHARACTERISTICS AND ENVIRONMENTAL ASPECTS OF THE DEVELOPMENT OF THE ZHUANTOBE SKARN IRON ORE DEPOSIT

Abstract

Skarn iron ore deposits represent a significant component of the global iron resource base, yet their development is commonly associated with substantial environmental disturbance due to large-scale open-pit mining. This study presents an integrated geological and geoecological assessment of the Zhuantobe skarn iron ore deposit (Central Kazakhstan), focusing on the relationship between ore-forming processes and environmental response under active mining conditions. The research combines geological, mineralogical, geochemical, geophysical, and long-term environmental monitoring datasets (2018–2025). Ore bodies are characterized by complex morphology controlled by fault systems and multistage skarn evolution, including prograde (garnet–pyroxene) and retrograde (magnetite–sulfide) assemblages. Environmental monitoring indicates a systematic reduction in anthropogenic impact, including a decrease in dust load (>40%), soil iron content (~40%), and CO₂ emissions (~35%). The results demonstrate that the implementation of integrated environmental protection measures effectively stabilizes environmental parameters within permissible limits. The study provides a framework linking geological structure with environmental performance and proposes a set of transferable strategies for sustainable development of skarn iron ore deposits.

Keywords: skarn deposits, iron ores, geoecology, environmental monitoring, sustainable mining, Central Kazakhstan.

Received February 3, 2026; revised April 13, 2026; accepted April 30, 2026.

Introduction

Modern approaches to subsoil use require not only detailed geological investigation but also comprehensive assessment of environmental impacts associated with mineral resource development [1, 2]. This is particularly important for skarn iron ore deposits, which are typically located in tectonically complex zones and involve large-scale open-pit mining [3–5].

The Zhuantobe iron ore deposit, located within the Northwestern Balkhash metallogenic region of Central Kazakhstan (Fig. 1), represents a typical skarn system formed in contact zones between intrusive rocks and carbonate sequences.

The aim of this study is to provide an integrated geological and environmental assessment of the Zhuantobe deposit and evaluate the effectiveness of environmental protection measures implemented during its development. The objectives include characterization of geological structure and ore mineralization; analysis of mineralogical and textural features of ores; assessment of environmental conditions based on monitoring data (2018–2025); evaluation of mitigation measures and development of recommendations. Scientific novelty lies in the integration of geological and environmental datasets, enabling assessment of the relationship between ore-forming processes and environmental response during mining.

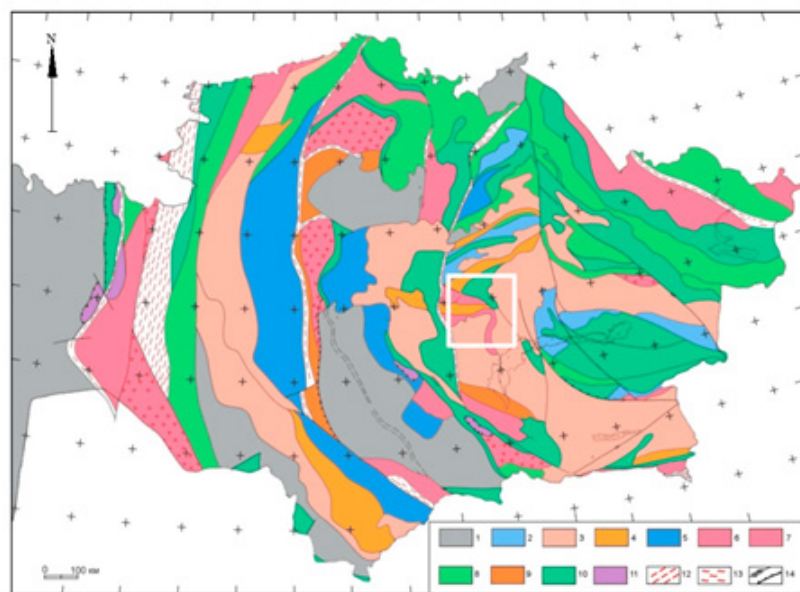


Figure 1 – Tectonic zoning scheme of the Paleozoic regions of Kazakhstan [6] (the iron ore district of the northwestern Balkhash area is highlighted by a white rectangle). The legend identifies 14 tectonic units:

- (1) Tectonic depressions, (2) Fragments of forearc terraces, (3) Volcanic and volcano-plutonic belts, (4) Rifts, (5) Folded strike-slip-thrust zones, (6) Cratonic terranes subjected to weak granitization, (7) Cratonic terranes with granite-gneiss domes, (8) Fragments of volcanic island arcs, (9) Fragments of riftogenic basins, (10) Collisional sutures, (11) Ophiolitic allochthons, (12) Zones of transform sutures and suture zones, (13) Schist zones, (14) Collisional and post-collisional faults.

Materials and methods

The geological investigation included field mapping, structural analysis, and drilling with continuous core recovery. A representative dataset of ore and host rock samples ($n \approx 300\text{--}400$) was analyzed.

Petrographic analysis was conducted using transmitted light microscopy (Leica DM750), while mineralogical composition was determined by X-ray diffraction (XRD). Geochemical analyses of major and trace elements were performed using spectrometric and ICP-based methods.

Magnetic surveys were used to delineate structural controls and ore zones. Gamma logging was applied to assess natural radiation background and identify potential anomalies [4].

Environmental assessment was based on long-term monitoring conducted from 2018 to 2025 [1, 2, 7–10]. Monitoring included:

- ♦ atmospheric parameters (dust load, CO_2 emissions);
- ♦ soil characteristics (pH, Fe content);
- ♦ surface and groundwater quality;
- ♦ biodiversity indicators.

Measurements were conducted at fixed monitoring stations on an annual (or quarterly) basis, ensuring temporal consistency of the dataset.

A semi-quantitative evaluation framework based on national environmental standards was applied. Baseline conditions (2018) were compared with post-mitigation data (2025), allowing assessment of environmental response to implemented measures.

Results

The Zhuantobe deposit is located within the Atasu–Mointy anticlinorium and is associated with the Dzhungar–Balkhash Hercynian folded system (Fig. 2). The geological structure is characterized by a combination of volcanic, sedimentary, and intrusive rocks, affected by fault systems of NW and sublatitudinal orientation.

Ore bodies exhibit complex morphology and are spatially controlled by fault zones. Mineralization is dominated by magnetite, with subordinate hematite, pyrite, and chalcopyrite [4, 6].

The evolution of the deposit reflects a multistage skarn-forming process:

- ♦ prograde stage: garnet–pyroxene assemblages;
- ♦ retrograde stage: magnetite, sulfides, and hydrothermal alteration.

This mineralogical zonation directly influences ore quality and environmental behavior.

During the development of the Zhuantobe deposit, a set of measures will be implemented to prevent contamination of soil–vegetation cover and water bodies (Fig. 3). In addition, the introduction of modern engineering solutions for waste storage and reclamation is planned in order to prevent their release into the environment. Regular environmental monitoring confirms that all parameters remain within the permissible limits established by the legislation of the Republic of Kazakhstan [1, 7–13].

The study area is characterized by semi-arid climatic conditions and low biological productivity. The absence of permanently established protected species reduces ecological sensitivity; however, mining activities remain a significant source of disturbance.

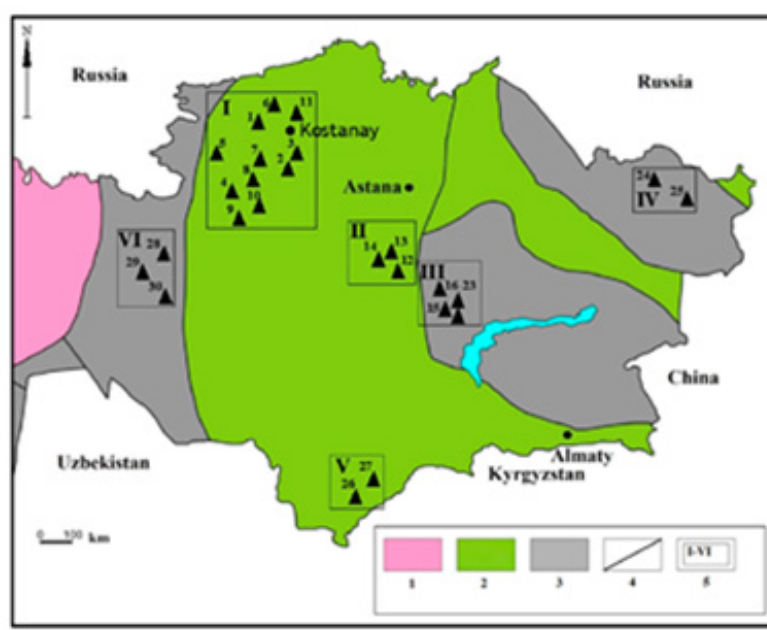


Figure 2 – Scheme of the main tectonic units and distribution of major industrial types of iron ore deposits in Kazakhstan: (1) East European Craton; (2, 3) Kazakh segment of the Central Asian Fold Belt: (2) Caledonian fold systems, (3) Hercynian fold systems; (4) geological boundaries; (5) outlines of iron ore districts. The deposits are grouped into six regions: I – Northern Torgay: 1. Kachar, 2. Sarbai, 3. Sokolov, 4. Lisakov, 5. Ayat, 6. Aleshino, 7. Lomonosov, 8. South Sarbai, 9. Sor, 10. Shagyrkol, 11. Korzhynkol; II – Central Kazakhstan: 12. Karazhal, 13. Ushkatyn, 14. Bolshoi Ktai; III – Northwestern Balkhash Region: 15. Bapy, 16. Zhuantobe, 17. Karaulken, 18. Akchagyl, 19. Ushtobe, 20. Kiyik, 21. Taitobe, 22. Tomashev, 23. Kyzyl-Sayak; IV – Eastern Kazakhstan: 24. Kholzun, 25. Rodionov Log; V – Southern Kazakhstan: 26. Irissu, 27. Abail; VI – Western Kazakhstan: 28. Velikhovskoye, 29. Kokbulak, 30. Taldy-Espe. III marks the position of the North Balkhash ore region, which hosts the Zhuantobe deposit

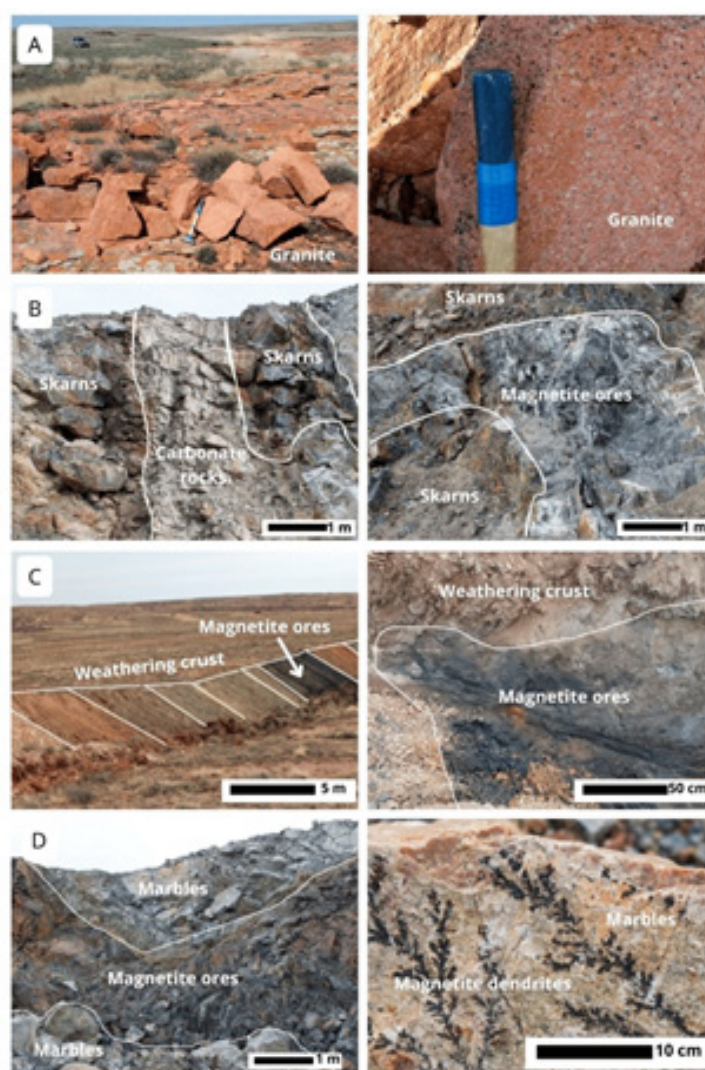


Figure 3 – Outcrops of rocks (A – granites) and mine workings (B–D) at the Zhuantobe deposit subject to reclamation after the completion of exploration and mining operations.
B, D – open pit; C – exploration trench

The main physical factor is noise from mining equipment and vehicles. However, the remoteness of the deposit from settlements (more than 50 kilometers) makes its impact negligible. Technological processes capable of causing vibration or electromagnetic pollution are absent [13–15]. Gamma logging data show that the radiation background of the site is within natural values (15–80 microrentgen/hour).

The Zhuantobe deposit area belongs to semi-desert steppes with extremely sparse vegetation (feather grass, cereals, shrubs, juniper). Red Book plant species are absent, which is confirmed by the conclusion of the Forestry and Wildlife Inspection. The fauna is represented by typical steppe and desert species: rodents, birds of prey, and small passerines. Near water bodies, rare species listed in the Red Book of Kazakhstan are occasionally encountered, such as saiga, golden eagle, and great bustard [1, 7, 8].

Radiation background values (15–80 $\mu\text{R/h}$) fall within natural limits.

The graph (Figure 1) demonstrates a reduction in all major pollution indicators. Of particular importance is the increase in the pH of water resources, which indicates a decrease in acidity previously caused by sulfide oxidation and the inflow of wastewater. The presented data show that: (1) the pH of

water resources has increased, confirming a reduction in acidity due to sulfide oxidation; (2) the iron content in soils has decreased by almost 40%; (3) dust load has been reduced by more than twofold; and (4) CO₂ emissions have declined by 35% relative to 2018.

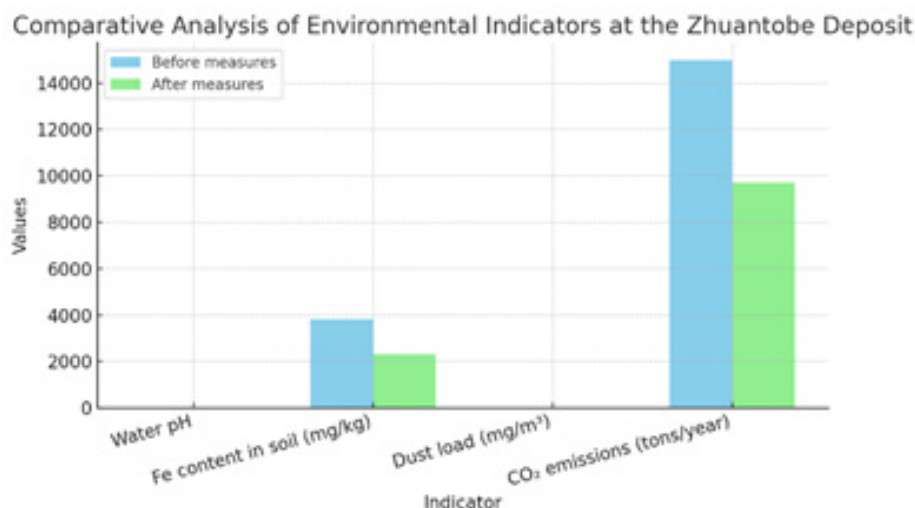


Figure 4 – Comparative analysis of environmental indicators at the Zhuantobe deposit

The environmental aspect associated with geological and mining operations for iron ore development is regulated by the legislation of the Republic of Kazakhstan. In this regard, all relevant regulatory documents were applied in the present study. These include works addressing problems of ecology and geoecology, as outlined in [1, 2, 16–21].

The development of the Zhuantobe skarn iron ore deposit represents an example of a comprehensive approach to involving medium and small-scale mineral resource objects of the Northwestern Balkhash ore region into exploitation, taking into account both economic and environmental factors. The geological exploration and geoecological studies conducted allow a number of conclusions to be drawn, significant both for assessing anthropogenic impacts on the environment and for optimizing design solutions at all stages of development.

From a geological perspective, the Zhuantobe deposit possesses a number of favorable characteristics: a thin overburden, stability of host rocks, and the possibility of open-pit mining with minimal volumes of capital works [7, 8, 22–24]. The physical and mechanical properties of the ores and host rocks require drilling and blasting preparation, which, in turn, necessitates strict control of dust emissions and vibration impacts on the environment [25–27].

The mineralogical composition of the ores (magnetite with admixtures of pyroxene, garnet, serpentine, biotite, and other minerals) and their monometallic nature simplify the processing scheme but require monitoring of associated elements (Ti, V, Mn, Cr, etc.), which, although not exceeding critical thresholds, may affect the environmental characteristics of waste [28–30].

Geoecological studies confirmed that the background concentrations of mobile forms of heavy metals, as well as the content of natural and technogenic radionuclides within the deposit area, do not exceed established standards. This provides grounds to conclude that, if project solutions are followed, the development of the deposit will not lead to significant contamination of environmental components.

It should be noted, however, that during operation, 542 hectares of land previously used as low-productive pastures will be disturbed. This requires reclamation with a sanitary–hygienic orientation, reflecting the responsible attitude of the subsoil user to ecosystem restoration. The planned pit depth (up to 266 m) and area (75.6 ha) dictate the necessity of a well-designed reclamation system upon completion of mining activities.

In terms of the area's water balance, the low water content of the productive horizon, poor surface runoff development, and arid climate minimize the risk of water contamination. Nevertheless, an important condition remains strict control of the quality of discharged pit waters and compliance with measures for water resource protection.

The flora and fauna of the area do not include ecologically valuable or protected species, eliminating risks for rare and endangered biodiversity. This simplifies environmental justification procedures and reduces biodiversity-related risks.

Thus, analysis of the obtained data makes it possible to conclude that, provided all project-based environmental measures are implemented, the development of the Zhuantobe deposit can proceed with minimal environmental impact. The mining and exploitation of the object are technically and environmentally justified, and the established ecological database provides a basis for subsequent monitoring of environmental conditions during the deposit's operation.

Table 1 summarizes environmental parameters before and after the implementation of mitigation measures.

Table 1 – Environmental parameters before (2018) and after (2025) mitigation measures

Parameter	2018	2025
Soil pH	6.2	7.1
Dust load (mg/m ² /day)	3800	2300
Fe in soil (%)	0.42	0.18
CO ₂ emissions (t/year)	15000	9700

A consistent decrease in pollution indicators is observed, indicating the effectiveness of environmental management strategies.

Discussion

The results demonstrate a strong relationship between geological structure and environmental response. Fault-controlled ore bodies influence:

- ♦ spatial distribution of mining operations;
- ♦ localization of environmental impacts;
- ♦ pathways of contaminant migration.

The observed reduction in environmental impact is attributed to:

- ♦ improved waste management systems;
- ♦ dust suppression technologies;
- ♦ controlled water discharge;
- ♦ progressive land reclamation.

A comparative analysis with the Zhuantobe skarn deposit confirms that similar geological settings exhibit comparable environmental responses, supporting the general applicability of the proposed approach.

Importantly, the integration of geological and environmental datasets allows prediction of environmental risks based on geological parameters, which represents a significant advancement in geoecological assessment.

Conclusion

This study demonstrates that the sustainable development of skarn iron ore deposits requires the integration of geological understanding with comprehensive environmental management approaches. The results indicate that geological structure plays a fundamental role not only in controlling ore

formation but also in influencing the spatial distribution and intensity of environmental impacts. Long-term environmental monitoring confirms a significant reduction in anthropogenic pressure, particularly in terms of dust load, soil contamination, and atmospheric emissions, reflecting the effectiveness of the implemented mitigation measures. These findings highlight that properly designed environmental protection strategies ensure compliance with regulatory standards and contribute to the stabilization of ecosystem conditions in mining areas.

From a practical perspective, sustainable mining of skarn deposits can be achieved through the implementation of continuous environmental monitoring throughout the entire mining lifecycle, the application of advanced dust suppression and emission control technologies, strict management and treatment of mine and waste waters, and systematic reclamation of disturbed lands. Furthermore, the integration of environmental considerations at the early stages of geological exploration is essential for minimizing long-term impacts. The proposed approach provides a transferable and scalable framework for environmentally responsible mining, applicable to skarn deposits in Central Kazakhstan and analogous geological settings worldwide.

REFERENCES

- 1 Environmental Code of the Republic of Kazakhstan No. 212-III ZRK (January 9, 2007, as amended).
- 2 Sbornik metodik po raschetu vybrosov vrednykh veshchestv v atmosferu razlichnymi proizvodstvami [Collection of Methods for Calculating Emissions of Harmful Substances into the Atmosphere by Various Industries] (Almaty, 1996). (in Russian).
- 3 Satpaev, K.I. Main patterns of spatial distribution of endogenous ore zones in Central Kazakhstan. *Sovetskaya Geologiya* [Soviet Geology], 58, 93–109 (1957). (in Russian).
- 4 Korobkin, V., Nygmanova, A., Tulemissova, Z., and Chaklikov, A. Geological and mineralogical analysis of Zhuantobe skarns in Central Kazakhstan considering the influence of textural features on iron ore quality. *Applied Sciences*, 15(17), 9816 (2025). <https://doi.org/10.3390/app15179816>
- 5 Daukeev, S.Zh., Uzhkenov, B.S., Abdulin, A.A., Bespaev, Kh.A., Votsalevskiy, E.S., Lyubetskiy, V.N., Mazurov, A.K., and Miroschnichenko, L.A. *Glubinnoye stroeniye i mineralnyye resursy Kazakhstana. T. 1: Glubinnoye stroeniye i geodinamika* [Deep Structure and Mineral Resources of Kazakhstan. Vol. 1: Deep Structure and Geodynamics] (Almaty, 2002). (in Russian).
- 6 Korobkin, V.V., and Buslov, M.M. Tectonics and geodynamics of the western Central Asian Fold Belt (Kazakhstan Paleozooids). *Russian Geology and Geophysics*, 50(12), 1585–1603 (2011).
- 7 Edilbaev, A.I. *Razrabotka tekhnologiy obogashcheniya bednykh zheleznykh rud* [Development of Beneficiation Technologies for Low-Grade Iron Ore]. *Gornyy informatsionno-analiticheskiy byulleten* [Mining Informational and Analytical Bulletin], (10), 247–251 (2011). (in Russian).
- 8 Kubeeva, L.V., Zيابkin, V.F., Muzgina, V.S., et al. *Geoekologicheskiye issledovaniya pri osvoenii zhelezorudnogo mestorozhdeniya Bapy Respubliki Kazakhstan* [Geocological Research in the Development of the Bapy Iron Ore Deposit, Republic of Kazakhstan]. *Nedropolzovaniye XXI vek* [Subsoil Use XXI Century], (1), 81–83 (2010). (in Russian).
- 9 Butakov, A.V. *Nauchno-metodicheskiye aspekty obespecheniya ekologicheskoy bezopasnosti pri osvoenii novykh mestorozhdeniy* [Scientific and Methodological Aspects of Ensuring Ecological Safety in the Development of New Deposits]. *Evraziyskiy Soyuz Uchenykh* [Eurasian Union of Scientists], 7–6(16), 135–137 (2015). (in Russian).
- 10 Zenkov, I.V., Anishchenko, Yu.A., Fedorov, V.A., et al. *Informatsionnoye obespecheniye otsenki ekologii narushennykh zemel zhelezorudnykh karierov Srednego i Yuzhnogo Urala* [Information Support for Assessing the Ecology of Disturbed Lands from Iron Ore Quarries in the Middle and Southern Urals]. *Ekologiya i promyshlennost Rossii* [Ecology and Industry of Russia], 25 (1), 38–43 (2021). <https://doi.org/10.18412/1816-0395-2021-1-38-43> (in Russian).
- 11 Badmaeva, Z.B. *Ekologicheskaya otsenka pochvenno-rastitelnogo pokrova karierov Respubliki Kalmykiya* [Ecological Assessment of Soil and Vegetation Cover of Quarries in the Republic of Kalmykia]. Doctoral dissertation (Elista, 2012). (in Russian).
- 12 *Problemy otkrytoy razrabotki glubokikh karierov: Mezhdunarodnyy simpozium “Mirny-91”* [Problems of Open-Pit Mining of Deep Quarries: International Symposium “Mirny-91”]. Vol. 1 (Udachny: NITs Master, 1991). (in Russian).

13 Alekseeva, Yu.Yu., and Buzyakova, I.V. Geoekologicheskkiye tekhnologii rekultivatsii deystvuyushchikh karierov poleznykh iskopaemykh [Geoecological Technologies for Reclamation of Operating Mineral Quarries]. *Geologiya, geografiya i globalnaya energiya* [Geology, Geography and Global Energy], 1 (76), 139–148 (2020). (in Russian).

14 Kosolapov, O.V., and Ignatieva, M.N. Ekologicheski ustoychivoye nedropolzovaniye: kontseptsiya i osnovnyye printsipy [Ecologically Sustainable Subsoil Use: Concept and Basic Principles]. *Izvestiya vysshikh uchebnykh zavedeniy. Gornyy zhurnal* [Proceedings of Higher Educational Institutions. Mining Journal], (2), 79–89 (2019). <https://doi.org/10.21440/0536-1028-2019-2-79-89> (in Russian).

15 Golik, V.I., and Titova, A.V. Efficiency of mining and processing waste utilization in mining processes. *Gornyy zhurnal* [Mining Industry Journal], 110–114 (2025). <https://doi.org/10.30686/1609-9192-2025-4-110-114>

16 Kretschmann, J. Post-mining: A holistic approach. *Mining, Metallurgy & Exploration*, 37, 1–11 (2020). <https://doi.org/10.1007/s42461-020-00265-y>

17 Offenbergl, I., Frătilă, C., Ilias, T., Ilias, N., Offenbergl, A., Radu, S., and Fetiniuc, V. Post-industrial mining communities: Aspects on geoecology of quality of life. *Bulletin of the National Technical University KhPI. Series: Actual Problems of Ukrainian Society Development*, 55–61 (2023). <https://doi.org/10.20998/2227-6890.2023.1.10>

18 Sayfulloeva, K.G. Geoecological characteristics of the Adrasman mining and processing plant. *Science and Innovation. Series of Geological and Technical Sciences*, (1), 18–20 (2021).

19 Dolyeva, A.G. Some geoecological aspects of deep drilling. In: *Glaciers of Tajikistan—Source of Water Resources of Central Asia under Climate Change*, pp. 113–118 (Dushanbe: Tajik National University, 2025).

20 Massel, L.V., Massel, A.G., and Zorina, T.G. Methods and technologies for assessing the impact of energy on the geoecology of a region (examples from the Baikal region and Belarus). *Open Semantic Technologies for Designing Intelligent Systems*, (3), 209–214 (2019).

21 Korelskiy, D., Mentsiev, A., Dengayev, A., et al. Land assessment in mining regions considering ecology. *International Journal of Engineering*, 37(11), 2344–2353 (2024). <https://doi.org/10.5829/ije.2024.37.11b.19>

22 Shpilkina, T., Zhidkova, M., and Filimonova, N. Green economy investment program as a way to improve the ecology of the country's mining regions. *E3S Web of Conferences*, 278, 03012 (2021). <https://doi.org/10.1051/e3sconf/202127803012>

23 Purwins, S. Bauxite mining at Atewa Forest Reserve, Ghana: A political ecology of a conservation–exploitation conflict. *GeoJournal*, 87(2), 1085–1097 (2022). <https://doi.org/10.1007/s10708-020-10303-3>

24 Verkhoturov, A.G., and Verkhoturov, G.G. Vliyaniye razrabotki mestorozhdeniy Streltsovskogo rudnogo polya na ekologiyu regiona [Influence of Mining of the Streltsov Ore Field Deposits on the Ecology of the Region]. *Gornyy informatsionno-analiticheskiy byulleten* [Mining Informational and Analytical Bulletin], (7), 263–267 (2014). (in Russian).

25 Barmakova, T.V., Malyutina, N.M., Perepelkina, A.M., and Cheychuk, A.A. Iron: Mining, beneficiation, metallurgy, economics and their application in modern engineering. In: *Development of the Economy of the Russian Federation under Technological Sovereignty: Current Issues of Theory and Practice*, pp. 15–32 (Moscow: Nauchnyy Konsultant, 2023).

26 Nicolae, I., Sorin, M.R., Iulian, O., et al. Mining geoecology and geotechnologies: Conceptualization and evaluation elements. *Mining Informational and Analytical Bulletin (Scientific and Technical Journal)*, (3–1), 338–349 (2021). https://doi.org/10.25018/0236_1493_2021_31_0_338

27 Sokolov, I.V., Gobov, N.V., and Solomein, Yu.M. Opredeleniye parametrov eko-oriyentirovannoy geotekhnologii kompleksnogo osvoyeniya i pererabotki zheleznykh rud [Determining Parameters of Eco-Oriented Geotechnology for Complex Mining and Processing of Iron Ores]. In: *Innovatsionnyye geotekhnologii pri razrabotke rudnykh i nerudnykh mestorozhdeniy* [Innovative Geotechnologies in Ore and Non-Ore Mining], pp. 64–68 (Yekaterinburg: Ural State Mining University, 2015). (in Russian).

28 Andreeva, A.V., Alekseeva-Popova, N.V., Buznikov, A.A., and Drozdova, I.V. Patent No. 2411502 C2. Method for Assessing Environmental Pollution by Iron, Zinc, and Manganese. Russian Federation (2011).

29 Andreeva, A.V., Alekseeva-Popova, N.V., Buznikov, A.A., and Drozdova, I.V. Patent No. 2410670 C2. Method for Assessing the Ecological State of the Environment. Russian Federation (2011).

30 Askarova, N.M. Kompleksnaya pererabotka tekhnogennykh otkhodov (khvostov) posle flotatsii mednykh shlakov [Complex Processing of Man-Made Wastes (Tailings) after Flotation of Copper Slags]. *Universum: Tekhnicheskkiye nauki* [Universum: Technical Sciences], 3–1(108), 33–35 (2023). <https://doi.org/10.32743/UniTech.2023.108.3.15095> (in Russian).

^{1*}**Ныгманова А.С.,**

докторант, аға оқытушы, ORCID ID: 0009-0003-3917-2707,

*e-mail: a.emls@kbtu.kz

¹**Коробкин В.В.,**

Г.-М.Ғ.К., профессор, ORCID ID: 0000-0002-1562-759X,

e-mail: korobkin_vv@kbtu.kz

²**Буслов М.М.,**

Г.-М.Ғ.Д., профессор, ORCID ID: 0000-0003-0606-2264,

e-mail: buslov@igm.nsc.ru

³**Чакликов А.Е.,**

PhD, ассистент-профессор, ORCID ID: 0000-0001-8316-6599,

e-mail: a.chaklikov@kbtu.kz

^{1,3}Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

²В.С. Соболев атындағы Геология және минералогия институты, Ресей ғылым академиясының Сібір бөлімшесі, Новосібір мемлекеттік университеті, Новосібір қ., Ресей

ЖУАНТӨБЕ СКАРНДЫ ТЕМІР КЕН ОРНЫНЫҢ ГЕОЛОГИЯЛЫҚ СИПАТТАМАСЫ ЖӘНЕ ОНЫ ИГЕРУДІҢ ЭКОЛОГИЯЛЫҚ АСПЕКТІЛЕРІ

Аңдатпа

Скарнды темір кен орындары әлемдік темір ресурстарының маңызды құрамдас бөлігі болып табылады. Алайда оларды игеру көбінесе ашық әдіспен жүргізілетін ауқымды тау-кен жұмыстарына байланысты қоршаған ортаға елеулі әсер етеді. Осы зерттеуде Орталық Қазақстандағы Жуантөбе скарнды темір кен орнының геологиялық және геоэкологиялық жағдайына кешенді баға беріліп, кен түзілу процестері мен тау-кен жұмыстары жүргізілетін жағдайдағы қоршаған орта өзгерістері арасындағы өзара байланыс қарастырылады. Зерттеу барысында геологиялық, минералогиялық, геохимиялық және геофизикалық деректер, сондай-ақ 2018–2025 жж. аралығында жүргізілген ұзақ мерзімді экологиялық мониторинг нәтижелері кешенді түрде талданды. Рудалы денелер күрделі морфологиясымен сипатталады және олардың қалыптасуы жарылымдық құрылымдармен, сондай-ақ скарн түзілуінің көпсатылы эволюциясымен байланысты. Бұл эволюция барысында проградты (гранат–пироксен) және ретроградты (магнетит–сульфид) кезеңдер ажыратылады. Экологиялық мониторинг нәтижелері антропогендік жүктеменің жүйелі түрде төмендегенін көрсетті. Атап айтқанда, шаң жүктемесінің 40%-дан астамға, топырақтағы темір мөлшерінің шамамен 40%-ға және көмірқышқыл газы шығарындыларының шамамен 35%-ға азайғаны анықталды. Зерттеу нәтижелері табиғатты қорғау шараларын кешенді түрде енгізу қоршаған орта көрсеткіштерін рұқсат етілген шектерде тұрақтандыруға мүмкіндік беретінін дәлелдейді. Сонымен қатар, жұмыста геологиялық құрылым мен экологиялық жағдай арасындағы өзара байланысты сипаттайтын тұжырымдамалық негіз ұсынылып, скарнды темір кен орындарын тұрақты игеру тәжірибесінде қолдануға болатын әмбебап тәсілдер негізделген.

Түйін сөздер: скарнды кен орындары, темір кендері, геоэкология, экологиялық мониторинг, тұрақты тау-кен өндірісі, Орталық Қазақстан.

^{1*}**Ныгманова А.С.,**

докторант, ст. преподаватель, ORCID ID: 0009-0003-3917-2707,

*e-mail: a.emls@kbtu.kz

¹**Коробкин В.В.,**

к.г.-м.н., профессор, ORCID ID: 0000-0002-1562-759X,

e-mail: korobkin_vv@kbtu.kz

²**Буслов М.М.,**

д.г.-м.н., профессор, ORCID ID: 0000-0003-0606-2264,

e-mail: buslov@igm.nsc.ru

³**Чакликов А.Е.,**

PhD, ассистент-профессор, ORCID ID: 0000-0001-8316-6599,

e-mail: a.chaklikov@kbtu.kz

^{1,3}Казахстанско-Британский технический университет, г. Алматы, Казахстан

²Институт геологии и минералогии им. В.С. Соболева СО РАН,

Новосибирский государственный университет, г. Новосибирск, Россия

ГЕОЛОГИЧЕСКАЯ ХАРАКТЕРИСТИКА И ЭКОЛОГИЧЕСКИЕ АСПЕКТЫ ОСВОЕНИЯ СКАРНОВОГО ЖЕЛЕЗОРУДНОГО МЕСТОРОЖДЕНИЯ ЖУАНТОБЕ

Аннотация

Скарновые железорудные месторождения являются важной составляющей мировой сырьевой базы железа, однако их освоение, как правило, сопровождается значительным воздействием на окружающую среду вследствие масштабных открытых горных работ. В настоящем исследовании представлена комплексная геологическая и геоэкологическая оценка Жуантобинского скарнового железорудного месторождения (Центральный Казахстан) с акцентом на взаимосвязь процессов рудообразования и реакции окружающей среды в условиях действующего горного производства. В работе объединены геологические, минералогические, геохимические, геофизические данные, а также результаты долгосрочного экологического мониторинга за период 2018–2025 гг. Рудные тела характеризуются сложной морфологией, контролируемой разрывными нарушениями и многостадийной эволюцией скарнообразования, включая проградную (гранат–пироксеновую) и ретроградную (магнетит–сульфидную) стадии. Результаты экологического мониторинга свидетельствуют о систематическом снижении антропогенной нагрузки, в том числе уменьшении пылевой нагрузки (>40%), содержания железа в почвах (~40%) и выбросов CO₂ (~35%). Показано, что реализация комплекса природоохранных мероприятий обеспечивает стабилизацию экологических показателей в пределах допустимых нормативов. В работе предложена концептуальная модель, связывающая геологическое строение месторождения с экологическими характеристиками, а также сформулирован комплекс универсальных подходов к устойчивому освоению скарновых железорудных месторождений.

Ключевые слова: скарновые месторождения, железные руды, геоэкология, экологический мониторинг, устойчивое недропользование, Центральный Казахстан.

ЭКОНОМИКА ЖӘНЕ БИЗНЕС
ECONOMY AND BUSINESS
ЭКОНОМИКА И БИЗНЕС

UDC 005.8:005.334
IRSTI 06.52.13

<https://doi.org/10.55452/1998-6688-2026-23-2-504-523>

¹***Aralbayeva Zh.M.,**

Collateral Officer, ORCID ID: 0009-0002-8449-9437,

*e-mail: Zhamila.aralbayeva@adcb.com

²**Bakdaulet A.,**

Project Manager, ORCID ID: 0009-0004-4341-2176,

e-mail: aknar.bakdaulet@gmail.com

¹“ADCB” Islamic Bank JSC, Almaty, Kazakhstan

²Dream Space Decoration Cooperation, Almaty, Kazakhstan

ANALYZING CRITICAL SUCCESS FACTORS
AND RISKS IN PROJECT MANAGEMENT PRACTICES
IN DEVELOPING COUNTRIES

Abstract

Project management has primarily developed within high-income countries. However, over the last 15 years, capital-intensive projects have been shifting toward developing nations where conventional frameworks often underperform. Despite this, research on the critical success factors (CSFs) and risks influencing project outcomes in developing countries remains fragmented. This paper conducts a systematic literature review of 41 peer-reviewed articles published in the Scopus database in the last decade, covering 20 developing countries and 8 cross-country studies. Following the PRISMA protocol, this study identifies CSFs and risks in project management (PM) in developing countries through inductive review and deductive coding based on two theoretical lenses, namely, institutional and contingency theory. Furthermore, this paper analyzes the mechanisms and contextual dependency of these factors. The inductive findings identify 22 CSFs and 21 risk factors. In order to report the most critical ones, 8 CSFs and 8 risks are retained based on their frequency of occurrence. The results reveal that they form systematic corresponding pairs, with six of the eight CSFs directly mapping onto six risks. Moreover, institutional theory (IT) explains the dominance of external influences through coercive, normative, and mimetic mechanisms, as well as institutional voids, while the contingency theory (CT) shows variable outcomes across different project characteristics. Overall, the findings provide practical implications for project managers, policy makers and business organizations on managing projects in emerging countries. It is noted that project strategies should be adapted to project scale, with stakeholder-focused approaches in small projects, institutional building in medium projects, and macroeconomic and political risk management in large projects.

Keywords: Contingency theory, Critical success factors, Developing countries, Institutional theory, Project management, Risk factors, Risk management, Systematic literature review.

Received May 19, 2026; revised June 8, 2026; accepted June 9, 2026.

Introduction

Project management emerged as a formal professional discipline in the middle decades of the 20th century. For much of its history, PM theory and practice were developed in mature economies which have reliable contract enforcement, stable regulatory systems, specialized labor markets, and effective dispute-resolution mechanisms. However, these conditions are often underdeveloped in emerging economies.

Over the last fifteen years, there has been a growing shift of large-scale projects toward developing countries through initiatives such as the Belt and Road Initiative (BRI), public–private partnerships (PPP), and multilateral development programs. These programs accelerate infrastructure and development projects in Asia, Africa and other emerging regions, which are frequently exposed to high geopolitical risks and subsequent construction delays [1]. Despite this notable growth, empirical studies across construction, transport, energy, and information technology consistently report higher rates of cost overrun, scope changes, schedule delays, and project abandonment in developing countries than in developed countries [2]. This study employs a systematic literature review (SLR) approach to develop an integrated cross-country synthesis of CSFs and risks in PM in developing countries [3]. So, project outcomes are strongly influenced by contextual conditions such as political volatility, weak regulatory enforcement, volatile exchange rates, and limited access to finance.

The existing literature on studying the CSFs and risk factors in PM practices in developing countries demonstrates several consistent limitations. On the one hand, CSFs and risks are often studied separately. In practice, CSFs and risks are frequently mirror images of the same underlying phenomena. For example, effective stakeholder engagement and stakeholder conflict, or skilled personnel and shortages of skilled personnel. Thus, studying them as isolated research terms neglects their structural relationship, produces redundant data, and limits the development of integrated PM frameworks.

On the other hand, the field is theoretically underdeveloped. Many studies report CSFs and risks without explaining their mechanisms. Even when frameworks such as contingency theory and institutional theory are applied, they still lack analytical depth. As a result, the significance of this study lies in the following perspectives. From a theoretical perspective, this study addresses the limitation of separated research on CSFs and risks in existing literature by integrating them into a unified analytical framework. It provides cross-country evidence by applying institutional theory and contingency theory to interpret project outcomes systematically in emerging countries. From a practical perspective, this paper guides project managers, policymakers, and development agencies to align project strategies with contextual conditions and project scale, and improve risk management in resource-constrained environments.

Therefore, this study is designed to achieve the following research aims:

First, identify the CSFs and risk factors reported in the reviewed studies;

Second, explain the observed patterns using institutional theory as the primary theoretical lens and contingency theory as the complementary lens;

Third, provide practical implications and valuable insights for project governance in developing-country environments.

This analysis is guided by the following three research questions:

1. Which critical success factors are most frequently reported in empirical studies of project management in developing countries published between 2015 and 2025?

2. What risk factors are most commonly reported as constraining project outcomes in the reviewed literature?

3. How do these success and risk factors vary across different projects and institutional contexts in emerging countries?

Materials and methods

Research design

This study conducts a systematic literature review (SLR) approach to develop an integrated cross-country synthesis of CSFs and risks in PM in developing countries. It supports transparent study selection, systematic evidence synthesis, and comparative analysis across multiple developing-country contexts.

After collecting data from the Scopus database, this paper follows an inductive–deductive analytical logic to identify and code CSFs and risks from the reviewed studies. Then a descriptive frequency analysis is applied to prioritize the most frequently reported factors and examine their cross-country consistency. These patterns are further interpreted through a deductive lens informed by PM characteristics, institutional theory, and contingency theory.

Material collection

The literature search was conducted in the Scopus database. The search was limited to peer-reviewed journal articles and conference proceedings published in English between 2015 and 2025. During this period, project activity increased in developing countries, such as the Belt and Road Initiative, the digitization of project delivery, and the disruption and recovery phases associated with the COVID-19 pandemic. Book chapters, editorials, technical reports, and unpublished materials were excluded to ensure uniform peer review standards.

The following Boolean string was applied to the title, abstract, and keyword field: TITLE-ABS-KEY ((“critical success factor*” OR “CSF*” OR “enabler*” OR “risk*” OR “challenge*” OR “barrier*”) AND (“project management”) AND (“emerging econom*” OR “developing countr*” OR “emerging market*”)) AND PUBYEAR > 2015 AND PUBYEAR < 2026 AND (LIMIT-TO (SRCTYPE , “j”)) AND (LIMIT-TO (DOCTYPE , “ar”) OR LIMIT-TO (DOCTYPE , “cp”)) AND (LIMIT-TO (LANGUAGE , “English”)) .

Screening and selection

Article selection was conducted according to the PRISMA protocol [4]. As illustrated in Figure 1, the article selection consisted of four sequential steps. At the first step, the initial retrieval from the Scopus database returned 298 articles. In the second stage, titles and abstracts were screened, and returned articles were classified into “Include,” “Possible,” or “Exclude” based on their relevance to the research scope. This step excluded 62 articles as being outside the scope of the study, leaving 236 articles for full-text analysis. In Step 3, full-text reading was conducted and excluded 195 articles based on predefined criteria, including studies not related to developing countries ($n = 78$), studies lacking empirical evidence ($n = 42$), studies not addressing critical success and risk factors ($n = 51$), and those without accessible full text ($n = 24$). This step resulted in 41 articles for final inclusion. Finally, 41 studies were analyzed through cross-validation of screening decisions among the authors to minimize selection bias.

Data extraction and Data analysis:

After reading the full text of 41 articles, the authors constructed a standardized data extraction grid in Microsoft Excel, including baseline metadata such as title, authors, publication year, and journal name. To deepen the analysis, the grid also documented the empirical methodologies, analytical procedures, and country classifications. The findings from this analysis are presented in the first section of the Results and Discussion part. Then, texts, phrases, keywords, and sentences were extracted from the chosen articles, further analyzed, and coded into themes by the two authors based on common patterns across the studies. A certain level of conceptual interpretation was applied during this stage, where semantically similar factors reported across different studies were consolidated into common CSFs and risk categories.

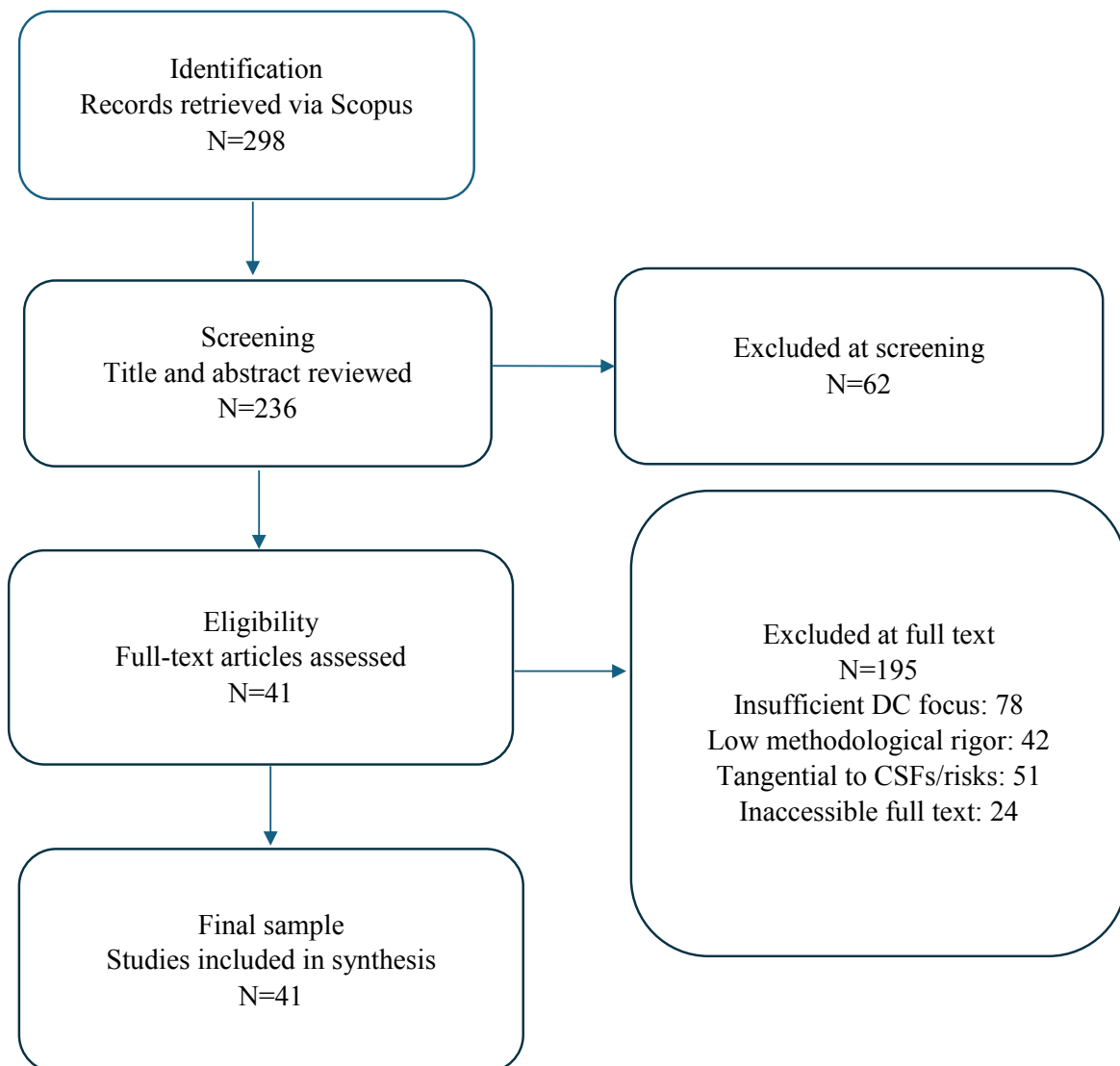


Figure 1 – PRISMA flow diagram

After that, the authors combined deductive and inductive approaches to identify structural dimensions and analytical categories of the final 41 articles. Table 1 presents the results of this step. The inductive approach was applied to identify recurring patterns of CSFs and risks in the reviewed articles. To select the most critical factors, the identified recurring CSFs and risk factors are further ranked by their frequency of occurrence across the reviewed studies. Ultimately, only factors reported in at least 20 papers were counted as dominant factors and were included in the detailed discussion. In addition, the deductive approach drew on three dimensions. Firstly, PM-specific structural dimensions were identified from reviewing previous research in PM [5, 6, 7]. Secondly, Institutional theory [8, 9] explains how the external environment shapes critical success factors and risks. Finally, Contingency theory [10, 11, 12] explains why these factors vary across project phases, sectors, and scales.

Table 1 – Structural dimensions and their analytical categories

Methods	Structural Dimensions	Analytical Categories
Inductive, based on the literature being reviewed	CSF categories (drivers)	Effective planning; project manager competence; effective risk management; stakeholder engagement; clear definition of project objectives; government support; monitoring and control; adequate financial capacity
	Risk categories (barriers)	Weak institutional capacity; weak regulatory framework; communication failures; lack of skilled personnel; bureaucracy; inadequate funding; political instability; corruption
Deductive, related to project management (PM)	Project type	Construction and infrastructure (incl. PPP, transport, housing, building); energy, oil&gas and utilities; IT and software, research and development; international development/donor-funded aid projects; not project specific
	Project phase	Feasibility; planning; bidding; execution; post-project
	Project scale	Small scale; medium scale; large scale; mega-project
Deductive, related to institutional theory (InstT) and contingency theory (CT)	Institutional environment (InstT)	Coercive pressures; normative pressures; mimetic pressures; institutional voids; institutional conformity mechanisms
	Contingency fit (CT)	Organizational-PM fit; Environment-PM fit; Strategy-PM fit; Multi-contingency alignment

Results and discussion

Characteristics of selected articles

It is clear from Figure 2, despite the Scopus search starting in 2015, the reviewed 41 articles only covered the period from 2017 to 2025. From around 2017 onwards, the integration of CSFs and risks in PM within emerging country contexts became more visible. Also, a notable concentration in 2021–2023 showed a broader expansion of PM research in developing countries during this timeframe. On the other hand, 33 of 41 articles were single-country studies covering 20 developing countries, while the remaining 8 articles adopt a multi-country or cross-regional approach. Although classified as single-country studies, many of them include cross-country references in their discussion sections, indicating a partially comparative interpretation of findings.

In terms of geographical distribution, Sub-Saharan Africa is the most represented region, with 10 articles (Ghana, Nigeria, Tanzania, Ethiopia, South Africa). The Middle East, North Africa, Afghanistan, and Pakistan (MENAAP) region is accountable for 12 articles (Pakistan, Yemen, Afghanistan, Egypt, Iran, Iraq, Jordan). Multi-regional studies are represented by 8 articles; East Asia and the Pacific – 4; Europe and Central Asia – 3 (including 2 from Kazakhstan); Latin America and the Caribbean – 2; South Asia – 2.

The selected 41 articles applied various research methodologies. By reviewing the primary research design of each study, the authors classified five types of them in Table 2. First, the most common approach is quantitative surveys and questionnaires using descriptive and inferential statistical techniques, such as relative-importance index(RII) and mean-item-score (MIS), chi-square tests, and other non-parametric methods. This approach accounts for 12 of the 41 articles.

Second, regression and structural equation modelling (SEM) also account for 12 studies. It includes Partial Least Squares Structural Equation Modelling (PLS-SEM), Confirmatory Factor Analysis (CFA), and factor analysis methods such as Exploratory Factor Analysis (EFA) and Principal Component Analysis (PCA).

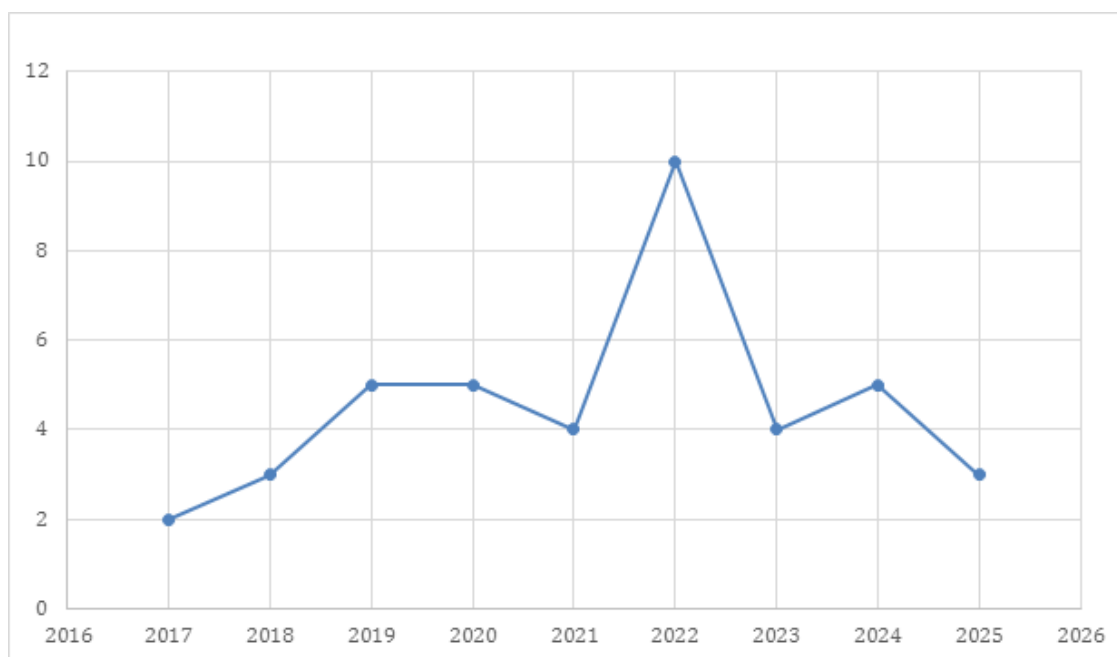


Figure 2 – Articles distribution by time period

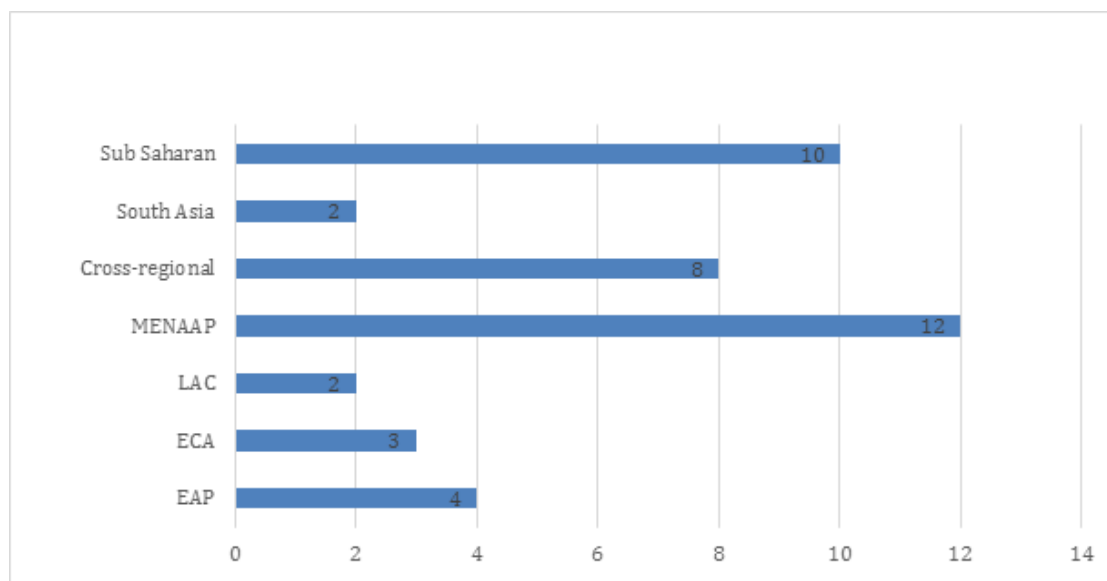


Figure 3 – Distribution of articles by region

Third, 8 articles adopt mixed-methods designs that combine qualitative and quantitative approaches in sequential or convergent configurations. Moreover, six articles solely use qualitative tools such as semi-structured interviews, content analysis, and in-depth case studies. Finally, multi-criteria decision-making and mathematical-modelling techniques appear in 3 studies, including the Analytic Hierarchy Process, the best-worst method, fuzzy comprehensive evaluation, and fuzzy-set qualitative comparative analysis.

Table 2 – Methodological Distribution of the Reviewed Articles

Methodology types	Reference article No.	Frequency
Quantitative Surveys and Questionnaires	S1, S5, S6, S9, S13, S14, S17, S20, S24, S26, S39, S41	12
Regression and SEM	S3, S8, S12, S18, S23, S25, S28, S30, S33, S35, S36, S38	12
Mixed Methods Design	S2, S4, S11, S15, S19, S27, S32, S34	8
Qualitative Studies	S16, S22, S29, S31, S37, S40	6
MCDM and Mathematical Modelling	S7, S10, S21	3

Critical Success Factors

Inductive coding of the 41 articles identified 22 CSFs. The number of references and the occurrence frequencies are shown in Table 3. Among the 22 CSFs, eight were mentioned in more than 20, including effective planning (29 of 41 articles), project manager competence (28), effective risk management (27), stakeholder engagement (26), clear definition of project objectives (25), government support (24), monitoring and control (23), and adequate financial capacity (21). These eight CSFs are discussed below.

Effective planning is the primary driver of project success, as it requires thorough upfront preparation, a realistic definition of the scope, time and cost considerations, and a clearly defined work breakdown structure, particularly in developing countries. Shafiei and Puttanna [13] demonstrate this in the context of Afghan international development projects. Their five-factor EFA shows Project Cycle Management as the dominant dimension (29.08% of variance explained), with discipline on time/cost/quality and clear project goals ranking highest among the 31 items. Similarly, Durna et al. [14] support this in Turkish PPP infrastructure, where Detailed and Clear Project Identification is also ranked as a top CSF.

The project manager's competence ranked as the second most critical CSFs. Evidence from empirical studies validate this in both Vietnamese construction projects ($n=216$) [15], and Turkish infrastructure sectors (2–130 experts) [14]. These studies focus on specific rather than generic project manager competencies, particularly leadership ability, human resource management skills, and technical planning skills.

Effective risk management is a crucial connection between the CSFs and risk factors in this study. Saddiqa et al. [16] formally validated this in their study of an IT project in Pakistan. Their findings showed PM risk strongly affects project performance ($\beta = 0.499$, $p < 0.001$), and plays a fully mediating role between contract governance and project performance. This suggests that contractual terms do not improve project performance directly, instead, they operate indirectly by reducing or reallocating project risks. Conversely, Teklemariam and Mnkanla [17] highlight inadequacies in risk management systems in a software project in Ethiopia. The study showed that only 16% of organizations adopted a formal risk management model, and just 27% fully implemented the four steps of risk management. These findings indicate that simply having risk management software or tools is insufficient to guarantee effective risk control. Risk management is only effective when it is involved in regular project procedures and organizational practices.

Stakeholder engagement ranked as the fourth with 26 occurrences. In 2022, Sepúlveda-Rivillas et al. [18] demonstrated a very strong relationship between team management and stakeholder management ($r = 0.91$). In the same year, Alade et al. [19] explained that stakeholder engagement is not a one-time event but an ongoing process. By tracking the Addis Ababa Light Rail project, this project emphasizes the importance of sustained collaboration among key stakeholders for delivering projects successfully.

Clear definition of project objectives was placed 5th. It serves as a foundational front-end enabler. For example, Durna et al. [14] rank clear project identification among the top success drivers in

Turkish PPP projects (mean 4.33), while Shafiei and Puttanna [13] find “clear goals” highest-rated in Afghan development projects (all $M > 3.5$). This early alignment on scope provides a stable reference point for planning and monitoring.

Government support was placed 6th. As a vital institutional factor for project success, it operates in two distinct ways. First, it offers direct aid such as permits, subsidies, and concessional financing. Second, it creates the political and institutional stability needed for other CSF to work effectively. In the case of Kazakhstan, Kambar et al. [20] found similar evidence in the public sector, out of 40 analyzed factors, state policy and digital platforms received the highest weighted score ($W = 2.38$), followed closely by the legislative framework for PM. ($W = 2.30$).

Monitoring and control were ranked as 7th. Based on studying Tanzanian PPP housing projects, Kavishe and Chileshe [21] document the operational form of monitoring. Of the 14 PM practices identified, the three most frequently mentioned were site visits/inspections, inspection records, and site meetings. The study from Ghana in 2020 shows that contract monitoring is considered one of the most effective anti-corruption tools. However, none of these 26 anti-corruption measures reach the required effectiveness level. It reveals that although monitoring tools exist, they are not used effectively in practice.

The eighth CSFs in the developing-country context is the adequate financial capacity. This factor should be considered at the macroeconomic level, not at the project level. Because exchange rate volatility, inflation, timing of funding, and delays in government payments are outside the control of an individual project. This is evidenced by Nguyen et al. [22] in the Vietnamese PPP project studies. It shows that 60% of Vietnamese PPP projects have failed or been delayed partly because of the inadequate compensation mechanisms. Alawneh et al. [23] confirm this pattern for Jordanian green building context, with insufficient client funds is a major cause of cost and time overruns.

Taken together, the 8 retained CSFs work as an integrated system rather than as 8 independent levers [24]. The operational core includes planning, competence, monitoring and control. The coordination layer refers to objectives, stakeholder engagement and risk management. Government support and financial capacity form the institutional and macroeconomic environment within which the operational core operates.

Table 3 – Critical success factors of PM practices in emerging countries

Code	Critical Success Factors	References	Occurrence frequency
C1	Effective planning	S1, S2, S3, S4, S5, S6, S9, S10, S11, S13, S14, S15, S17, S18, S19, S20, S23, S26, S27, S28, S29, S31, S32, S33, S34, S35, S37, S38, S40	29
C2	Project manager competence	S1, S2, S3, S4, S5, S7, S9, S10, S11, S13, S14, S15, S17, S18, S19, S20, S23, S26, S27, S29, S30, S31, S32, S34, S36, S37, S38, S40	28
C3	Effective risk management	S1, S2, S4, S5, S8, S11, S12, S13, S14, S16, S17, S18, S19, S20, S21, S23, S26, S27, S28, S31, S32, S33, S36, S37, S38, S39, S40	27
C4	Stakeholder engagement	S1, S2, S3, S4, S5, S7, S15, S16, S18, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S34, S36, S38, S40	26
C5	Clear definition of project objectives	S1, S2, S4, S5, S6, S8, S9, S10, S13, S14, S15, S16, S17, S18, S19, S20, S23, S29, S31, S32, S33, S36, S38, S39, S40	25

Continuation of table 3

C6	Government support	S2, S3, S4, S5, S6, S8, S12, S15, S16, S19, S21, S22, S23, S24, S25, S26, S28, S29, S30, S32, S34, S37, S39, S41	24
C7	Monitoring and control	S1, S2, S3, S5, S6, S9, S13, S16, S17, S19, S20, S21, S26, S29, S30, S31, S32, S33, S35, S36, S37, S38, S40	23
C8	Adequate financial capacity	S1, S2, S3, S4, S5, S6, S7, S10, S17, S18, S19, S23, S28, S29, S31, S32, S35, S38, S39, S40, S41	21
C9	Competent project team	S1, S2, S3, S5, S6, S13, S15, S18, S19, S20, S22, S23, S26, S31, S32, S34, S38, S40	18
C10	Effective contract management	S2, S4, S6, S16, S20, S21, S22, S23, S26, S27, S28, S29, S30, S32, S33, S35, S38, S40	18
C11	Top management support	S1, S3, S5, S6, S8, S15, S17, S19, S20, S21, S23, S24, S25, S30, S31, S32, S38, S41	18
C12	IT and technology adoption	S1, S2, S3, S4, S7, S14, S19, S21, S22, S23, S24, S26, S29, S30, S37, S38, S39, S41	18
C13	Sound legal framework	S2, S4, S5, S6, S7, S16, S19, S21, S22, S23, S30, S31, S32, S38, S39, S40, S41	17
C14	Capacity building	S2, S15, S19, S22, S23, S24, S26, S29, S30, S31, S32, S37, S38, S40, S41	15
C15	Organizational capacity	S2, S3, S4, S5, S6, S7, S19, S21, S23, S25, S26, S30, S31, S32, S38	15
C16	Innovation	S2, S7, S8, S15, S19, S21, S22, S23, S24, S25, S21, S31, S37, S40, S41	15
C17	Feasibility analysis	S2, S4, S5, S13, S16, S20, S23, S26, S28, S29, S31, S32, S34, S40	14
C18	Continuous learning	S1, S2, S22, S23, S24, S25, S29, S30, S34, S37, S40	11
C19	Risk allocation	S4, S5, S8, S12, S16, S21, S27, S28, S32, S40	10
C20	Anti-corruption measures	S6, S16, S19, S22, S27, S30, S32	7
C21	Cross-cultural management	S1, S23, S27, S28, S29, S34	6
C22	Local participation	S16, S23, S28, S29, S32	5

Risk Factors

Inductive coding resulted in 21 risk factors for the 41 articles. (See in Table 4) Since the minimum occurrence threshold was set at 20, 10 of 21 risk factors met this requirement. Hillson [25] identifies “being late for a milestone” and “exceeding the authorized budget” as examples of risk effects. Accordingly, among these ten factors, schedule delays and cost overruns can be treated as the outcomes of risk occurrence rather than risk factors, indicating project performance deficiencies instead of underlying risk drivers.

Therefore, they are included in the appendix for reporting purposes but excluded from the ten risk factors selected for detailed analysis. Following the exclusion, the remaining eight risk factors were ranked according to their occurrence frequency: weak institutional capacity (39 of 41 articles), weak regulatory framework (34 articles), communication failures (31 articles), lack of skilled

personnel (30 articles), bureaucracy (27 articles), inadequate funding (25 articles), political instability (24 articles) and corruption (23 articles).

Weak institutional capability and weak regulatory framework are the most frequently reported key risk factors, appearing in 39 and 36 of 41 the analyzed articles, respectively. It refers to the gap between institutional structural formalism and their actual project performance, including inefficient regulatory bodies and inconsistent implementation of procurement policies. The existence of such disparities is described by Kambar et al. [20] for the Kazakhstani public sector. The authors reveal that the two most prevalent political risks in their case study are bureaucratic inertia and weaknesses in anti-corruption mechanisms. Moreover, about three-quarters of respondents evaluated the implementation of a national program across central and regional administrations at level 2–3 on a five-point scale, which demonstrates organizational disintegration despite a formal structure.

Communication failures are ranked as 3rd. In a PLS-SEM analysis of delay factors in Pakistani construction projects, information and communication issues are the strongest predictor of time overrun, and lack of coordination within it records the highest single factor loading (0.87) in the entire model [26]. Kwofie et al. [27] refine this pattern in PPP contexts across Ghana and South Africa. Misunderstanding is identified as a universal issue in these two countries, although other communication-related problems vary by context. For example, Ghana is more affected by delays and conflicting information, whereas South Africa faces more challenges related to poorly defined roles among project participants.

The fourth most critical risk factor is lack of skilled personnel. This issue is observed in the Tanzanian case of PPP housing projects by Kavishe et al. [28]. The lack of adequate PPP skills and knowledge emerges as the first out of 19 assessed risks with a mean score of 4.82. 69.2% of respondents in this research have received no formal training in PPPs, even being actively involved in PM, representing a gap between formal qualifications and practical competencies. This finding is supported by Shafiei and Puttanna [29] in Afghan IDP projects, where improper appointment of project managers is identified as the fifth most important failure variable among 30 factors.

Bureaucracy refers to administrative delays, multilayered approvals, and procedural rigidity. Evidence from multiple studies confirms its negative impact on PM. In Ghana, Amoah et al. [30] find that span of control, as a bureaucratic indicator, significantly reduces project management success. In Nigeria, Owusu et al. [31] show that bureaucratic procedures in corruption reporting rank second among 17 anti-corruption barriers, with an impact index of 3.75. This risk remains predominantly procedural. It disrupts projects most heavily during procurement and permitting phases, creating direct vulnerabilities to corruption.

Inadequate funding was placed 6th. At the project level, insufficient client liquidity leads to cost overruns and schedule delays, particularly when combined with additional fiscal pressures such as the higher costs of sustainable materials [23]. Vietnamese PPP projects also present that limited private-sector financial capacity results in failures in approximately 60% of initiatives [22].

Political instability and corruption represent two distinct but interrelated institutional risk categories and were ranked 7th and 8th respectively. Political instability is a highly context-dependent mechanism. For instance, In the Yemeni case, its direct statistical effect on project outcomes is unexpectedly weak, which is attributed to the normalization of chronic instability [32]. In high conflict settings such as Afghanistan, political instability results in operational barriers, such as demining requirements that directly stop the project execution [1]. Cross-country comparative evidence also suggests that only a very small subset of political risk factors remains consistent across emerging markets [33], indicating that “political instability” functions more as an umbrella term than as a single stable mechanism.

On the other hand, corruption demonstrates a different but equally systemic pattern. In Afghan IDP projects, it is identified as one of the top three failure factors ($R^2 = 0.876$) [29]. In road construction projects, political interference in consultant selection and procurement processes directly contributes to project failure [35]. In Nigeria, corruption-related risks are further amplified by external disruptions such as kidnapping, banditry, fuel subsidy removal, currency volatility, and cybercrime, creating a highly unstable operational environment [36].

Table 4 – Risk factors of PM practices in emerging countries

Code	Risks	References	Occurrence frequency
R1	Weak institutional capacity	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S28, S29, S30, S31, S32, S33, S35, S36, S37, S38, S39, S40, S41	39
R2	Weak regulatory framework	S2, S3, S4, S5, S6, S7, S8, S9, S10, S12, S13, S15, S16, S18, S19, S21, S22, S23, S24, S25, S26, S27, S28, S30, S31, S32, S34, S35, S36, S37, S38, S39, S40, S41	34
R3	Communication failures	S1, S3, S4, S5, S6, S7, S8, S9, S11, S13, S14, S15, S17, S18, S19, S20, S21, S22, S23, S27, S28, S29, S30, S32, S33, S34, S35, S36, S38, S39, S40	31
R4	Schedule delays	S1, S4, S5, S6, S7, S8, S9, S10, S11, S12, S14, S15, S17, S18, S19, S20, S21, S22, S23, S26, S27, S28, S29, S30, S31, S32, S34, S35, S38, S40	30
R5	Lack of skilled personnel	S1, S2, S3, S4, S7, S8, S9, S10, S11, S13, S14, S19, S20, S21, S22, S23, S24, S26, S27, S28, S29, S30, S31, S32, S34, S35, S37, S38, S40, S41	30
R6	Cost overruns	S1, S4, S5, S7, S9, S10, S11, S12, S14, S15, S17, S18, S19, S20, S21, S22, S23, S26, S27, S28, S29, S30, S31, S32, S35, S38, S40	27
R7	Bureaucracy	S1, S3, S4, S5, S6, S7, S8, S9, S12, S13, S16, S18, S19, S20, S21, S22, S26, S27, S28, S30, S31, S32, S34, S35, S36, S37, S38	27
R8	Inadequate funding	S2, S4, S5, S6, S9, S10, S11, S13, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S27, S29, S31, S32, S35, S37, S38	25
R9	Political Instability	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S12, S13, S16, S19, S20, S21, S22, S23, S27, S28, S30, S36, S37, S38	24
R10	Corruption	S1, S3, S4, S6, S8, S9, S11, S12, S13, S16, S18, S19, S20, S21, S22, S25, S27, S30, S32, S34, S35, S37, S38	23
R11	Procurement failures	S1, S6, S7, S8, S9, S20, S21, S22, S23, S27, S28, S29, S30, S32, S34, S35, S40	17
R12	Poor project design	S1, S8, S9, S10, S11, S13, S20, S22, S23, S31, S32, S34, S35, S38, S40	15
R13	Scope changes	S4, S5, S11, S13, S16, S20, S21, S22, S23, S27, S29, S32, S35, S38	14
R14	Cultural barriers	S1, S12, S17, S20, S22, S23, S28, S29, S30, S31, S34, S39, S41	13
R15	Security threats	S1, S9, S12, S13, S16, S22, S27, S28, S30, S34, S37, S38	12
R16	Resistance to change	S2, S19, S22, S23, S24, S25, S29, S30, S31, S37, S38, S41	12
R17	Inflation	S9, S20, S21, S22, S23, S28, S29, S35, S37, S41	10
R18	Lack of transparency	S4, S6, S18, S19, S22, S23, S25, S27, S30, S32	10
R19	Digital infrastructure gaps	S2, S17, S19, S22, S24, S26, S29, S37, S39, S41	10
R20	Force majeure	S9, S12, S16, S20, S21, S27, S31, S34, S37	9
R21	Communication opposition	S12, S13, S16, S21, S27, S29, S34, S36	8

Based on the findings above, a partial structural correspondence between CSFs and risk factors across the reviewed literature can be identified. According to Table 1, six of the eight CSFs have direct or indirect counterparts in risk factors, revealing that many CSFs and risks represent opposite sides of the same underlying conditions. For example, government support corresponds to political instability, a strong legal framework corresponds to weak regulatory conditions, adequate financial capacity corresponds to insufficient funding, and competent project teams and managers correspond to a shortage of qualified personnel. Similarly, effective risk management represents a positive practice, the absence of which is reflected in a broader diffusion of operational risks.

However, this mirror relationship is not fully symmetrical. Remaining risk factors, like corruption, cultural barriers, and inflation, do not have clear positive equivalents among CSFs. This asymmetry is analytically important, because it suggests that CSFs research tends to emphasize desirable enabling conditions, whereas implicitly neglecting adverse conditions that are directly addressed in risk literature. Theoretically, this relationship suggests that CSF and risk literature should not be treated as separate analytical streams but should be integrated into a unified framework for understanding project performance.

Project types

Construction and infrastructure projects dominate the corpus, accounting for over half of the 41 articles (28 articles). It includes Public-Private Partnership (PPP), transport, housing, and general building projects. Such projects face a full risk profile, such as weak institutions, funding gaps, communication breakdowns, and political exposure. However, these risks are balanced by effective government support and risk management. Within this context, PPP variants introduce additional risks, including concession, guarantee obligations and contract renegotiation risk [14, 27, 28].

Energy, oil and gas projects form the next group (3 articles). External risk and stakeholder conflict are central here with financial capacity and risk allocation as the main enablers [37, 32]. This cluster maps primarily onto political instability and regulatory weakness as dominant risks. Consequently, adequate financial capacity and robust risk management emerge as the critical enabling CSFs.

IT and software projects (3 articles) change the picture. Requirements volatility, contractual governance and project management risk lead, while skilled staff and clear objectives drive success [17, 16]. Research and development and donor-funded projects make up a smaller group. There, stakeholder alignment, capacity building and steady funding matter most [38, 20]. Ugonna et al. [39] further extend this scope to public R&D organizations in Nigeria. Unlike infrastructure projects, this cluster identifies communication failures and a lack of skilled personnel as the primary risks; consequently, project manager competence and the precise definition of project objectives emerge as the critical enabling CSFs.

International development and donor-funded aid projects form the fourth cluster (3 articles). The dominant top risks in this project type are weak institutional capacity and lack of skilled personnel. The corresponding enabling CSFs are stakeholder engagement, government support, and effective planning, since donor frameworks presuppose a delivery capacity that the local environment must be built up to supply.

At last, the other cluster comprises non-project-specific research (4 articles), including practitioner surveys, value-management analyses, and innovation-adoption studies. These studies provide essential input for the inductive factor extraction, although they do not align with any single project type.

Project phases

The reviewed literature covers five distinct project phases: feasibility, planning, bidding, execution, and post-project. Execution-stage studies dominate due to the data availability and measurable performance indicators, particularly within quantitative survey-based research. Moreover, early project stages are mainly associated with decision-making and structuring processes. At the bidding stage, a substantial proportion of PPP challenges originate in procurement and preparation, showing that early contractual arrangements shape downstream performance outcomes. In contrast, post-project evaluation remains underexplored. This indicates a methodological bias toward implementation-focused research, potentially limiting insights into full lifecycle learning.

Project scale

The corpus distinguishes between small-scale, medium-scale, large-scale, and mega projects. The risk profile changes systematically across these scales (See Table 5).

First, small-scale projects are typically valued below USD 10 million. These projects rely heavily on stakeholder alignment, community acceptance, and funding continuity. Research by Kavishe et al. [28, 21] on Tanzanian PPP housing, Shih et al. [40] on construction in Saint Lucia, and Sarvari et al. [41] on Iranian construction SMEs confirms this pattern.

Medium-scale projects are characterized by institutional voids and cross-border coordination challenges. This cluster is exemplified by Durna et al. [14] on Turkish PPPs, Nguyen et al. [22] on Vietnamese PPP infrastructure, and Kassem et al. [37] on Yemeni oil-and-gas operations, where institutional logics begin to dominate. Large-scale projects span larger budgets while still being managed as bounded undertakings.

Mega projects (above USD 1 billion) are structurally different due to their duration, technical complexity, and multi stakeholder environments. Evidence from Yemeni mega projects [32], Chinese international projects [42], and the Addis Ababa LRT [19] shows that complexity and delivery system design are the primary drivers of performance outcomes in mega projects [43]. Consistent with this, De Marco et al. [44] demonstrate that cost contingency management directly shapes final estimates. This confirms that mega project cost outcomes depend on project-level delivery rather than the macro context.

Overall, these findings suggest that risk governance must adapt to various project scales. Small projects require attention to stakeholders, community, and funding continuity, while medium projects depend on institutional coordination. However, medium-scale projects are particularly challenging as they operate across both institutional and operational domains. Finally, large and mega projects require attention to political, institutional, and macroeconomic exposure.

Table 5 – Synthesis of dominant CSFs and risks across different project scales

Project scale	Dominant CSFs	Dominant risks
Small	F4 Stakeholder engagement; F5 Clear definition of project objectives; F8 Adequate financial capacity	R3 Communication failures; R8 Inadequate funding
Medium	F6 Government support; F2 Project Manager Competence; F7 Monitoring and Control	R1 Weak institutional capacity; R2 Weak regulatory framework; R7 Bureaucracy; R5 Lack of skilled personnel
Large	F1 Effective planning; F3 Effective risk management; F6 Government support	R9 Political instability; R10 Corruption; R2 Weak regulatory framework; R1 Weak institutional capacity
Mega	F1 Effective planning; F3 Effective risk management; (emphasis on risk allocation / contractual governance)	R9 Political instability; R2 Weak regulatory framework; R1 Weak institutional capacity

Institutional theory patterns

In this part, the authors used institutional theory to explain how the external environment shapes critical success factors and risks. Institutional theory functions as the primary theoretical lens. Each mechanism is treated below as a standalone analytical unit before any synthesis.

Coercive isomorphism occurs due to coercion by dependent organizations and societal expectations. Based on the reviewed articles, it can be observed that state regulatory capacity is the main source of coercion. It is important to note that state regulatory capacity acts variably in developing countries due to differences in institutional strength. Ullah et al. [2] provide empirical

evidence in the case of Pakistani construction projects, revealing that coercion acts through multinational clients' demands and not domestic regulations. Multinational clients, international aid donors, and multilateral institutions apply contract and lending conditions that act as a de facto compliance policy. In contrast, Kambar et al. [20] observe the opposite trend. Coercive pressure originates from the state, but its substantive effectiveness depends on unequal enforcement.

Normative isomorphism is the result of professionalization, it conveys expectations through educational systems, professional bodies, and certification programs [5]. Ullah et al. [2] report the effect of this normative pressure on adopting sustainable project management practices, which is statistically significant but much lower than mimetic pressure. Evidence from Cambodia regarding awareness is provided by Kang et al. [45] shows that 48.6% of surveyed people were aware of construction management, whereas the other part was either unfamiliar with the concept or just heard about it, but the importance of the field was recognized by 98.6% of participants after being explained. Normative pressure does exist, but the lack of diffusion channels hinders its implementation on a population level.

Mimetic isomorphism is defined as the imitation of practices deemed legitimate by other organizations under uncertainty. It is the most frequently observed institutional channel among those studied. According to Ullah et al. [2], mimetic pressures exert the highest standardized path coefficient of sustainable project management in Pakistan. Mimetic pressure is significantly higher than normative pressure ($\beta = 0.276$, $f^2 = 0.101$) used in the same context and surpasses all the possible coercive paths of the host-state institution. The rationale behind the mechanism lies in the imitation of competitor practices rather than following the industry norms. As Pakistani businesses treat sustainability as a competitive requirement than regulatory obligations.

Institutional voids refer to the absence or failure of institutions required for effective market operation [34]. It is the most frequently coded theme in the corpus. The construct supplements the original DiMaggio–Powell framework by incorporating North's [9] institutional-quality dimension.

Four different void types can be identified from the reviewed literature. Procurement-rule voids are most common in the sub-Saharan African region. The research of Barajei et al. [35] on consultants' recruitment in Ghana, Owusu et al. [31] on anti-corruption approaches in Ghana, and Aghimien et al. [46] on sustainable construction in Nigeria and South Africa support this finding. Void of enforcement is common in cross-border jurisdiction studies. Examples include analysis carried out by Ahmad et al. [47] between Nepal and Pakistan, and Kambar et al. [20] in Kazakhstan for public projects. On the other hand, dispute-resolution voids have been identified in studies on PPP projects in Tanzania and Vietnam. Also, Kavishe et al. [28] examine this in 2018 through PPP contract bias favoring the private sector, while Nguyen et al. [22] linking dispute-resolution voids to compensation disputes as a cause of project failure. Regulatory-competence voids surface in IT and emerging-technology adoption studies. This pattern is documented by Olusola and Phung [48] regarding smart-contract adoption in Nigeria, and Teklemariam and Mnkandla [17] concerning Ethiopian software risk management.

The fifth group of institutional categories describes how organizations respond after receiving the institutional pressure. The corpus presents three different types of responses. First, active conformity occurs when organizations embrace institutionalized practices in reality. Sepúlveda-Rivillas et al. [18] document this in a Colombian sample. Second, ceremonial conformity refers to organizations adopting the structural form of a practice but not its substantive content. Aghimien et al. [46] identified this pattern in Nigeria and South Africa concerning sustainable construction. This research reveals a gap between the awareness and application of the practice, where awareness levels in sustainable construction greatly surpass those in actual sustainable material use. Third, resistance is organizations actively resisting institutional pressure. This pattern is revealed in the research by Ebekozi et al. [36] on Nigerian cultural inertia and Olusola & Phung [48] on Nigerian smart-contract resistance, respectively.

Two main themes can be identified across these five institutions' categories. The use of institutional pressure in the developing-country context is multi-channel and imbalanced. Organizations prefer to adopt institutional labels at a much higher rate than implementing institutional substance.

Contingency theory patterns

Although applying institutional theory clarifies the broader macro-environment, it fails to capture the reasons for various project outcomes within the same national borders. Contingency theory can solve this issue by analyzing data points to four distinct dimensions of fit, which is explained in the following.

Organizational-PM Fit. This refers to the extent of consistency between the complexity of the project management process and organizational capacity necessary for implementing that process. If the PM requirement matches the organizational capabilities, projects are more likely to succeed. Saddiqa et al. [16] tested that the PM risk has a significant impact on project performance ($\beta = 0.499$, $p < 0.001$), and completely mediates the effect of the dimensions of contract-based governance on performance. Another complementary piece of evidence comes from the findings of Sun et al. [42]. These researchers discovered that Chinese construction companies achieve better results in collaboration with Chinese designers than with foreign designers when implementing complex projects. This can be explained by organizational-PM fit between design and execution is inherently stronger within a shared institutional template.

Environment-PM Fit. Environment-PM fit is that PM procedures are aligned with external macro-environmental factors. External factors directly link to our findings, including political stability, regulatory framework, institutional capacity, market volatility, and cultural norms. Many of the identified risks in this paper can be interpreted as manifestations of Environment-PM misfit, conversely, CSFs function as adaptive mechanisms that reduce the degree of misfit. Kassem et al. [32] validate the strong influence of external environments for mega projects by revealing that 74.3% of the variation in project success in Yemen can be explained by Environment-PM fit. This challenge is structurally driven. On the grounds that, most PM models originating in developed countries often assume specific institutional foundations and organizational conditions, which may not be present in developing countries. Therefore, when PM practices are mismatched with the local institutional environment, the project effectiveness can be significantly limited.

Strategy-PM Fit. Among these four contingency dimensions, Strategy-PM fit is the least developed dimension. The possible reason is that strategic alignment is a higher-level decision-making process, while PM has traditionally focused on the operational level. Besides this, few studies measure it directly because of its abstract nature and the difficulty in collecting data. In emerging economies, a higher degree of strategic ambiguity is observed, due to the fact that Strategic-PM alignment is often constrained by institutional instability, shifting policy priorities, and external stakeholder influence. Consequently, project teams frequently operate without clear objectives, because projects in emerging countries are not purely strategy-driven, but may be government-driven and donor-driven, or highly influenced by political decisions. Still, the importance of Strategy-PM fit cannot be neglected. Ugonna et al. [39] reveal that the alignment between the strategic positions of project teams and organizational strategy is the most highly evaluated element, based on a sample of Nigerian R&D projects. As a result, this misfit leads to inefficient resource allocation, unclear priorities, and inconsistent performance evaluation criteria, which directly result in project failure.

Multi-contingency alignment. The current dimension refers to the scenario where all contingency factors interact to influence project outcomes. This brings a cumulative effect, whereby project success depends on the combination and compatibility of multiple contextual variables. This finding is supported by Sun et al. [42]. Through multiple correspondence analysis of 112 international architectural projects, they identified two potential structural dimensions, namely an adaptability dimension and a complexity dimension. The former includes technical standards, project location, and the nationality of the design firm; the latter includes project cost and project type.

In addition, significant contextual differences were also observed at the institutional and risk levels. Based on the findings of Ali et al. [33], only 4 out of 20 political risks were cross-nationally universal across five emerging markets, and the remaining risks differ across institutional realities. It can be concluded that risk exposure is not static but highly depends on interaction of the institutional environment, project characteristics, and industry conditions.

Furthermore, failing to adequately consider the integration of multiple contingency structures during cross-context transfer leads to mismatches. For example, mimetic institutional adoption in international construction and PPP projects often results in scale-related systemic mismatches. Therefore, replicated PM models are often built on institutional foundations which may be inconsistent with the local environment. Moreover, communication failure is identified as a universal risk across all projects, but its manifestations and severity depend on project size, industry complexity, and institutional environment.

Conclusion

This study has explored critical success factors and risks in PM practices in emerging countries. Following the PRISMA protocols, 41 articles were retrieved and screened from the Scopus database that are highly relevant to the research field. Based on the findings, 22 CSFs and risks were identified and further ranked by their occurrence frequencies. As a result, top 8 CSFs and 8 risks were explained and compared through deductive dimensions, including project characteristics, institutional theory and contingency theory. Top 8 CSFs are effective planning; project manager competence; effective risk management; stakeholder engagement; clear definition of project objectives; government support; monitoring and control; and adequate financial capacity. While top 8 risks include Weak institutional capacity; Weak regulatory framework; Communication failures; Lack of skilled personnel; Bureaucracy; Inadequate funding; Political Instability; and Corruption.

However, this paper has several limitations. The major one is that Scopus was the only database used for data extraction. Thus, several regional journals and studies from other sources may have been underrepresented. Second, only English-language articles were included in the search, which may have led to the exclusion of some relevant articles in other languages, especially in emerging-country contexts. Third, the coding process involved only two coders, so the findings may contain a certain level of subjectivity and bias despite a structured protocol.

Based on the synthesized data, this paper provides the following implications from four various perspectives, which are practical, policy, development agencies and multilateral financiers, and future research directions.

From a practical perspective, four guidelines emerged for managing risks in developing countries. First, selecting a risk management strategy should consider the size of the projects. Second, communication should be improved as it acts as the strongest indicator of time delays in the literature. It implies that the corresponding stakeholder engagement critical success factor becomes more important. Third, the competence of the project manager is particularly vital when there are significant institutional voids in the country, especially in donor-financed development projects [28, 29, 39]. Fourth, the emulation of PPPs and sustainability systems from advanced nations should come with clear local adaptation. [22, 47].

From a policy perspective, the institution voids, rather than the existing formal institutions, seem to be the critical factor that needs to be addressed. The procurement-rule void in sub-Saharan African PPPs [35, 31], the enforcement void in inter-jurisdictional initiatives [47, 20], and regulatory competence void in emerging technologies [48, 17] have been evident throughout the literature.

For development agencies and multilateral financiers, the findings suggest that con dictionary PM frameworks often assume organizational capabilities which the local environment may not have at its disposal. Thus, investments made in terms of capacity building efforts prior to project execution aimed at building the capacity in terms of the skilled manpower required could potentially yield greater results than making incremental adjustments to frameworks.

For future researchers, considering several limitations in this paper, the authors suggest the following four research directions. First, extending the databases by involving articles from sources other than Scopus, such as Web of Science and Google scholar, to ensure the richness and reliability of extracted data. Next, the use of multi-coder reliability measures will enhance the procedural integrity of the synthesis and enable the calculation of inter-rater statistics. Third, further testing of

the integrated framework combining institutional and contingency perspectives is necessary. Future studies could use surveys or case studies in regions that are less studied in the current literature to validate the framework and conduct deeper analysis. Finally, adopting longitudinal approaches to study the evolution of CSFs and risks can address one of the most crucial limitations in cross-sectional approaches.

REFERENCES

- 1 Kadry, M., Osman, H., and Georgy, M. Causes of construction delays in countries with high geopolitical risks. *Journal of Construction Engineering and Management*, 143 (2), 04016095 (2017). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001222](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001222)
- 2 Ullah, M., Khan, M.W.A., Kuang, L.C., Hussain, A., Rana, F., Khan, A., and Sajid, M.R. A structural model for the antecedents of sustainable project management in Pakistan. *Sustainability*, 12 (19), 8013 (2020). <https://doi.org/10.3390/su12198013>
- 3 Tranfield, D., Denyer, D., and Smart, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14 (3), 207–222 (2003). <https://doi.org/10.1111/1467-8551.00375>
- 4 Page, M.J., McKenzie, J.E., Bossuyt, P.M., et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71 (2021). <https://doi.org/10.1136/bmj.n71>
- 5 Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*, 8th ed. (Newtown Square, PA: Project Management Institute, 2025).
- 6 Kerzner, H. *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*, 12th ed. (Hoboken, NJ: Wiley, 2017).
- 7 Flyvbjerg, B. What you should know about megaprojects and why: An overview. *Project Management Journal*, 45 (2), 6–19 (2014). <https://doi.org/10.1002/pmj.21409>
- 8 DiMaggio, P.J., and Powell, W.W. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48 (2), 147–160 (1983). <https://doi.org/10.2307/2095101>
- 9 North, D.C. *Institutions, Institutional Change and Economic Performance* (Cambridge: Cambridge University Press, 1990). <https://doi.org/10.1017/CBO9780511808678>
- 10 Lawrence, P.R., and Lorsch, J.W. *Organization and Environment: Managing Differentiation and Integration* (Boston: Harvard Business School Press, 1967).
- 11 Donaldson, L. *The Contingency Theory of Organizations* (Thousand Oaks, CA: Sage Publications, 2001). <https://doi.org/10.4135/9781452229249>
- 12 Wadongo, B., and Abdel-Kader, M. Contingency theory, performance management and organisational effectiveness in the third sector: A theoretical framework. *International Journal of Productivity and Performance Management*, 63 (6), 680–703 (2014). <https://doi.org/10.1108/IJPPM-09-2013-0161>
- 13 Shafiei, N.A., and Puttanna, K. Critical success factors for international development projects in Afghanistan: An exploratory factor analysis. *Journal of Project Management*, 7 (3), 173–188 (2022). <https://doi.org/10.5267/j.jpm.2022.5.001>
- 14 Durna, E., Ozorhon, B., and Çağlayan, S. Identifying critical success factors of public-private partnership projects in Türkiye. *Sakarya University Journal of Science*, 28 (2), 1–14 (2024). <https://doi.org/10.16984/saufenbilder.1334483>
- 15 Pham, V.Q., Nguyen, B.K.Q., Van Tu, B., Pham, H.T.T., and Le, T.Q. Critical success factors of project management: The case of construction related projects in Vietnam. *Journal of Asian Finance, Economics and Business*, 6 (2), 223–230 (2019). <https://doi.org/10.13106/jafeb.2019.vol6.no2.223>
- 16 Saddiqa, A., Shehzad, M.U., and Mohiuddin, M. Effects of contractual governance on IT project performance under the mediating role of project management risk: An emerging market context. *Information*, 14 (9), 490 (2023). <https://doi.org/10.3390/info14090490>
- 17 Teklemariam, M.A., and Mnkandla, E. Software project risk management practice in Ethiopia. *Electronic Journal of Information Systems in Developing Countries*, 79 (1), 1–14 (2017). <https://doi.org/10.1002/j.1681-4835.2017.tb00583.x>
- 18 Sepúlveda-Rivillas, C.I., Alegre, J., and Oltra, V. The core functions of project management in an emerging economy: A new measurement scale. *International Journal of Project Organisation and Management*, 14 (4), 401–428 (2022). <https://doi.org/10.1504/IJPOM.2022.125878>

- 19 Alade, T.A., Bukoye, O.T., Roehrich, J.K., and Edelenbos, J. Cross-national collaboration in strategic transport projects: The impact on benefits realization. *International Journal of Project Management*, 40 (4), 347–361 (2022). <https://doi.org/10.1016/j.ijproman.2022.03.009>
- 20 Kambar, R., Ismailova, R., Aitbayeva, A., Mukhamejanuly, S., and Zharov, Y. Assessing the implementation of project management in public administration: A PEST analysis approach. *Journal of Governance and Regulation*, 14 (4), 220–231 (2025). <https://doi.org/10.22495/jgrv14i4art20>
- 21 Kavishe, N., and Chileshe, N. Identifying project management practices and principles for public-private partnerships in housing projects: The case of Tanzania. *Sustainability*, 10 (12), 4609 (2018). <https://doi.org/10.3390/su10124609>
- 22 Nguyen, P.T., Likhitrungsilp, V., and Onishi, M. Success factors for public-private partnership infrastructure projects in Vietnam. *International Journal on Advanced Science, Engineering and Information Technology*, 10 (2), 631–637 (2020). <https://doi.org/10.18517/ijaseit.10.2.5839>
- 23 Alawneh, R., Jannoud, I., Rabayah, H., Asaad, S., Almasaeid, H., Imam, R., and Ghazali, F.E.M. Development of a new method for assessing project risks in sustainable building construction projects in developing countries: The case of Jordan. *Buildings*, 14 (6), 1573 (2024). <https://doi.org/10.3390/buildings14061573>
- 24 Ottaviani, F.M., Zenezini, G., Saba, F., De Marco, A., and Gavinelli, L. System-level critical success factors for BIM implementation in construction management: An AHP approach. *Systems*, 13 (2), 94 (2025). <https://doi.org/10.3390/systems13020094>
- 25 Hillson, D. Project risks, causes, and effects. Project Management Institute (2000). <https://www.pmi.org/learning/library/2019/04/07/15/14/project-risks-causes-risks-effects-4663>
- 26 Memon, A.H., Memon, A.Q., Khahro, S.H., and Javed, Y. Investigation of project delays: Towards a sustainable construction industry. *Sustainability*, 15 (2), 1457 (2023). <https://doi.org/10.3390/su15021457>
- 27 Kwofie, T.E., Aigbavboa, C.O., and Thwala, W.D. Communication performance challenges in PPP projects: Cases of Ghana and South Africa. *Built Environment Project and Asset Management*, 9 (5), 628–641 (2019). <https://doi.org/10.1108/BEPAM-11-2018-0137>
- 28 Kavishe, N., Jefferson, I., and Chileshe, N. An analysis of the delivery challenges influencing public-private partnership in housing projects: The case of Tanzania. *Engineering, Construction and Architectural Management*, 25 (2), 202–240 (2018). <https://doi.org/10.1108/ECAM-12-2016-0261>
- 29 Shafiei, N.A., and Puttanna, K. An investigation into the factors causing international development project failure in developing countries: Focus on Afghanistan. *Journal of Project Management*, 6 (3), 133–142 (2021). <https://doi.org/10.5267/j.jpm.2021.2.002>
- 30 Amoah, A., Berbegal-Mirabent, J., and Marimon, F. What drives project management success in developing countries? The case of Ghana. *Tec Empresarial*, 16 (2), 22–46 (2022). <https://doi.org/10.18845/te.v16i2.6186>
- 31 Owusu, E.K., Chan, A.P.C., and Hosseini, M.R. Impacts of anti-corruption barriers on the efficacy of anti-corruption measures in infrastructure projects: Implications for sustainable development. *Journal of Cleaner Production*, 246, 119078 (2020). <https://doi.org/10.1016/j.jclepro.2019.119078>
- 32 Kassem, M.A., Khoiry, M.A., and Hamzah, N. Assessment of the effect of external risk factors on the success of an oil and gas construction project. *Engineering, Construction and Architectural Management*, 27 (7), 1683–1715 (2020). <https://doi.org/10.1108/ECAM-10-2019-0573>
- 33 Ali, T., Butt, A., Arslan, A., Tarba, S.Y., Sniazhko, S.A., and Kontkanen, M. International projects and political risk management by multinational enterprises: Insights from multiple emerging markets. *International Marketing Review*, 38 (5), 957–983 (2021). <https://doi.org/10.1108/IMR-03-2020-0060>
- 34 Khanna, T., and Palepu, K.G. *Winning in Emerging Markets: A Road Map for Strategy and Execution* (Boston: Harvard Business Press, 2010).
- 35 Barajei, C., Kusi, E., Ackon, F., Osman, A.M., Mohammed, A.M.Z., Simpeh, F., and Gyimah, F. Success factors of the consultant selection stage of the Ghanaian public construction projects: The road sector's perspective. *Construction Economics and Building*, 24 (1/2), 1–20 (2024). <https://doi.org/10.5130/AJCEB.v24i1/2.8660>
- 36 Ebekozi, A., Aigbavboa, C., Samsurijan, M.S., Ahmed, M.A.H., Akinradewo, O., and Omoh-Paul, I. Managing construction project risks in turbulent times: A stakeholders' perspective. *International Journal of Building Pathology and Adaptation*, 42 (3), 445–463 (2024). <https://doi.org/10.1108/IJBPA-01-2024-0003>
- 37 Kassem, M.A., Khoiry, M.A., and Hamzah, N. Risk factors in oil and gas construction projects in developing countries: A case study. *International Journal of Energy Sector Management*, 13 (4), 846–861 (2019). <https://doi.org/10.1108/IJESM-11-2018-0002>

- 38 Dufková, G. Time and budget overruns on Czech international development projects. *Ekonomický časopis*, 70 (7–8), 651–671 (2022). <https://doi.org/10.31577/ekoncas.2022.07-8.04>
- 39 Ugonna, D.C.U., Ochieng, E.G., and Zuofa, T. Augmenting the delivery of public research and development projects in developing countries. *Journal of Building Pathology and Adaptation*, 42 (3), 445–463 (2024). <https://doi.org/10.1108/IJBPA-01-2024-0003>
- 40 Shih, H.-S., Chen, I.-F., Munier, N., and Alcide, Z. Investigating risk-constraint nexus of construction projects in Caribbean small island developing states. *SAGE Open* (2023). <https://doi.org/10.1177/21582440231158023>
- 41 Sarvari, H., Chan, D.W.M., Alaeos, A.K.F., Olawumi, T.O., and Abdalridah Aldaud, A.A. Critical success factors for managing construction small and medium-sized enterprises in developing countries of Middle East: Evidence from Iranian construction enterprises. *Journal of Building Engineering*, 43, 103152 (2021). <https://doi.org/10.1016/j.jobe.2021.103152>
- 42 Sun, H., Tang, W., Duffield, C.F., Zhang, L., and Hui, F.K.P. How to get international construction projects delivered on time: From Chinese contractors' perspective. *Journal of Civil Engineering and Management*, 28 (5), 365–377 (2022). <https://doi.org/10.3846/jcem.2022.16381>
- 43 De Marco, A., and Narbaev, T. Factors of schedule and cost performance of tunnel construction megaprojects. *The Open Civil Engineering Journal*, 15 (1), 15–38 (2021). <https://doi.org/10.2174/1874149502115010038>
- 44 De Marco, A., Narbaev, T., Ottaviani, F.M., and Vanhoucke, M. Influence of cost contingency management on project estimates at completion. *International Journal of Construction Management*, 24 (9), 935–945 (2024). <https://doi.org/10.1080/15623599.2023.2239487>
- 45 Kang, Y., Jin, Z., Hyun, C., and Park, H. Construction management functions for developing countries: The case of Cambodia. *Journal of Management in Engineering*, 34 (4), 04018017 (2018). [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000609](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000609)
- 46 Aghimien, D.O., Aigbavboa, C.O., and Thwala, W.D. Microscoping the challenges of sustainable construction in developing countries. *Journal of Engineering, Design and Technology* (2019). <https://doi.org/10.1108/JEDT-01-2019-0002>
- 47 Ahmad, M.I., Sukamani, D., Wang, J., and Kusi, M. Risk factors and their impact on the success of construction firms: A comparative study between Pakistan and Nepal. *IAENG International Journal of Applied Mathematics*, 52 (3), 518–527 (2022).
- 48 Olusola, J., and Phung, Q. A framework for adopting smart contracts in the Nigerian construction industry. *International Journal of Construction Management*, 25 (1), 45–58 (2025). <https://doi.org/10.1080/15623599.2025.2550478>

¹*Аралбаева Ж.М.,

кепілзат жөніндегі маман, ORCID ID: 0009-0002-8449-9437,

*e-mail: Zhamila.aralbayeva@adcb.com

²Бакдәулет А.,

жоба менеджері, ORCID ID: 0009-0004-4341-2176,

e-mail: aknar.bakdaulet@gmail.com

¹АҚ «ADCB» Ислам Банкі», Алматы қ., Қазақстан

²«Dream Space Decoration» ЖК, Алматы қ., Қазақстан

ДАМУШЫ ЕЛДЕРДЕГІ ЖОБАЛАРДЫ БАСҚАРУ ТӘЖІРИБЕСІНДЕГІ МАҢЫЗДЫ ТАБЫС ФАКТОРЛАРЫ МЕН ТӘУЕКЕЛДЕРДІ ТАЛДАУ

Аңдатпа

Жобаларды басқару саласындағы зерттеулер жоғары табысты институционалдық ортада кеңінен дамыған. Алайда соңғы 15 жыл ішінде капиталға бай жобалық қызмет дамушы елдерге ауысты, мұнда стандартты басқару шеңберлері көбінесе өз тиімділігін көрсете бермейді. Бұл зерттеу 2015–2025 жж. аралығында жарияланған, 20 дамушы ел мен 8 көпұлтты ортадағы жобаларды басқарудағы маңызды табысқа жету факторлары (Critical Success Factors – CSF) мен тәуекелдерді қарастырған 41 рецензияланған эмпирикалық мақаланы жүйелі түрде талдауға негізделген. Шолу әдебиеттегі үш тұрақты олқылықты жоюға бағытталған: табысқа жетудің маңызды факторлары мен тәуекелдердің ұғымдық тұрғыдан ажыратылмауы, дамушы елдерге бағытталған зерттеулердің шектеулі болуы және елдер арасындағы тәжірибені

біртұтас теориялық шеңберде біріктіретін интеграциялық синтездің болмауы. PRISMA хаттамасына сәйкес зерттеуде институционалдық теория мен контингенттік теорияға негізделген дедуктивті жіктеу тәсілі эмпирикалық материалды индуктивті кодтаумен ұштастырылды. Индуктивті кодтау нәтижесінде 22 маңызды табыс факторы және 21 тәуекел факторы анықталды. Олардың ішінен 8 маңызды табыс факторы мен 8 негізгі тәуекел факторы іріктеліп, екі теориялық көзқарас тұрғысынан синтезделді. Институционалдық теория саяси және институционалдық факторлардың басымдығын мәжбүрлеуші, нормативтік және миметикалық механизмдер, сондай-ақ институционалдық олқылықтар арқылы түсіндіреді. Ал контингенттік теория жобаның көлеміне, секторына және іске асыру әдісіне байланысты жүйелі айырмашылықтарды ескеруге мүмкіндік береді. Зерттеу нәтижелері жоба тәуекелдерінің сипаты жобаның көлемі ұлғайған сайын жүйелі түрде өзгеретінін көрсетті: шағын жобаларда негізгі тәуекелдер мүдделі тараптардың үйлесімділігімен байланысты болса, орта көлемді жобаларда институционалдық әлеует мәселелері алдыңғы орынға шығады, ал ірі жобаларда макроэкономикалық және саяси тәуекелдер басым болады. Зерттеу қорытындылары жоба менеджерлеріне, мемлекеттік институттарға және даму қаржы институттарына практикалық ұсыныстар ұсынады.

Түйін сөздер: дамушы елдер, жоба менеджменті, жүйелі әдеби шолу, институционалдық теория, табысқа жетудің түйінді факторлары, тәуекел факторлары, тәуекелдерді басқару, шұғыл жағдайлар теориясы.

¹*Аралбаева Ж.М.,

специалист по залоговому обеспечению, ORCID ID: 0009-0002-8449-9437,

*e-mail: Zhamila.aralbayeva@adcb.com

²Бакдаулет А.,

менеджер проекта, ORCID ID: 0009-0004-4341-2176,

e-mail: aknar.bakdaulet@gmail.com

¹АО «Исламский Банк «ADCB», г. Алматы, Казахстан

²ИП «Dream Space Decoration», г. Алматы, Казахстан

АНАЛИЗ КЛЮЧЕВЫХ ФАКТОРОВ УСПЕХА И РИСКОВ В ПРАКТИКЕ УПРАВЛЕНИЯ ПРОЕКТАМИ В РАЗВИВАЮЩИХСЯ СТРАНАХ

Аннотация

Исследования в области управления проектами широко развивались в институциональной среде стран с высоким уровнем дохода. Однако за последние 15 лет капиталоемкая проектная деятельность сместилась в сторону развивающихся стран, где стандартные рамки зачастую не дают ожидаемых результатов. За основу взят систематический анализ 41 рецензируемой эмпирической статьи, опубликованных в период с 2015 по 2025 гг. и посвященных критическим факторам успеха (КФУ) и рискам в управлении проектами в 20 развивающихся странах и 8 странах с многонациональной средой. В данном обзоре рассматриваются три постоянных пробела в литературе: концептуальное разделение КФУ и рисков, ограниченность исследований, посвященных развивающимся странам, и отсутствие интегративного межстранового обобщения в рамках единой теоретической концепции. В соответствии с протоколом PRISMA в статье применяется качественный контент-анализ, сочетающий дедуктивную категоризацию на основе институциональной теории и теории непредсказуемости с индуктивным кодированием эмпирического корпуса. В результате индуктивного кодирования было выявлено 22 КФУ и 21 фактор риска. 8 КФУ и 8 факторов риска были отобраны и проанализированы через призму двух теорий. Институциональная теория объясняет доминирование политических и институциональных факторов с помощью механизмов принуждения, нормативности и миметизма, а также концепции институциональных пробелов, в то время как теория контингенции учитывает систематические различия в зависимости от масштаба проекта, сектора и метода реализации. Результаты исследования показывают, что проектные риски систематически меняются в зависимости от масштаба: в небольших проектах они связаны с согласованностью действий заинтересованных сторон, в проектах среднего масштаба – с институциональным потенциалом, а в крупных проектах – с макроэкономическими и политическими рисками. Сделаны выводы для руководителей проектов, государственных учреждений и институтов финансирования развития.

Ключевые слова: институциональная теория, критические факторы успеха, развивающиеся страны, систематический обзор литературы, теория непредвиденных обстоятельств, управление рисками, факторы риска.

UDC 658.046
IRSTI 06.52.13

<https://doi.org/10.55452/1998-6688-2026-23-2-524-533>

¹***Ryskulbekova M.M.**,
Researcher, ORCID ID: 0009-0008-2979-7857,
*e-mail: maralryskulbekova.98@gmail.com

¹Nurstroy Yug LLP, Shymkent, Kazakhstan

PROJECT-LEVEL DETERMINANTS OF TOTAL INVESTMENT IN GREENFIELD INFRASTRUCTURE PROJECTS

Abstract

This study examines the relationship between project-level characteristics and total investment in greenfield infrastructure projects. While existing foreign direct investment literature primarily focuses on macroeconomic and institutional determinants, project-specific characteristics remain comparatively underexplored. Using a dataset of 131 infrastructure projects obtained from the World Bank Private Participation in Infrastructure (PPI) database, the study applies multiple linear regression analysis to evaluate the effects of physical assets, project capacity, procurement conditions, financial structure, and country-level characteristics on investment size. The results show that physical assets have a positive and statistically significant effect on total investment, confirming the capital-intensive nature of greenfield infrastructure development. In contrast, project capacity demonstrates a significant negative relationship, suggesting the presence of operational efficiency or economies of scale. Other variables, including procurement competition, financial structure, and country-level conditions, do not demonstrate statistically significant effects. The findings provide empirical evidence that investment size in greenfield infrastructure projects is influenced more strongly by internal project characteristics than by broader external conditions. The study contributes to project-level infrastructure investment literature and highlights the importance of technical planning, infrastructure configuration, and operational assessment in investment decision-making.

Keywords: greenfield investment, infrastructure investment, investment determinants, multiple linear regression, project design; project-level characteristics.

Received May 1, 2026; revised May 20, 25, 2026; accepted May 26, 2026

Introduction

Greenfield investment plays an important role in infrastructure development because it supports the creation of new assets, production facilities, and operational capacity [1]. Unlike mergers and acquisitions, greenfield projects require the construction of new infrastructure systems and facilities, which makes them highly capital-intensive and closely related to project financing decisions [2]. Some research also shows that accuracy of cost estimations in capital-intensive construction projects depends not only on the consumption of the initial investment but also on the project-level risks which impact the final cost at project completion [3].

In recent years, greenfield infrastructure investment has become increasingly important in developing economies due to urbanization, industrial expansion, and rising demand for large infrastructure projects. Amendolagine et al. [4] additionally emphasize the importance of regional institutional conditions and project geography in shaping greenfield investment allocation across countries and regions. Therefore, understanding the factors influencing investment decisions in greenfield infrastructure projects has become important for governments, investors, and infrastructure planners.

The literature on foreign direct investment mainly focuses on macroeconomic and institutional determinants of investment activity. Previous studies emphasize the role of governance quality,

political stability, institutional conditions, and market attractiveness in shaping foreign investment flows [5]. Other studies also show that economic stability and institutional quality influence the attractiveness of greenfield investment across countries and industries [6]. Political and regulatory conditions are also frequently identified as important determinants of foreign direct investment decisions [7].

Several theoretical studies explain greenfield investment through broader multinational enterprise and foreign direct investment frameworks. Dunning [8] argues that foreign investment decisions are influenced by ownership, location, and internalization advantages. Helpman et al. [9] explain that firms choose foreign direct investment when international production becomes more efficient than export-based strategies. Markusen and Maskus [10] also show that differences in factor endowments and market conditions may influence multinational investment decisions across countries. However, existing literature pays comparatively less attention to project-level characteristics in infrastructure investment research.

Kazakhstan has recently increased the implementation of large-scale greenfield infrastructure projects through public-private partnership mechanisms. One of the largest infrastructure projects in Central Asia is the Big Almaty Ring Road (BAKAD), which demonstrates the growing importance of infrastructure financing, project governance, and investment planning in the region. The expansion of transport, energy, and infrastructure projects in Kazakhstan additionally highlights the practical importance of understanding the factors associated with investment allocation and project scale in greenfield infrastructure development.

This study examines the relationship between project characteristics and investment size in greenfield infrastructure projects using project-level data from the World Bank Private Participation in Infrastructure database. The analysis focuses on how internal project characteristics and external contextual conditions are associated with differences in investment size across infrastructure projects.

The study contributes to the literature by providing project-level evidence on greenfield infrastructure investment. The findings may also support infrastructure planning, project evaluation, and investment decision-making in developing economies where infrastructure expansion continues to play an important role in economic development.

Materials and methods

Recent project-level studies show that greenfield investment decisions may differ according to project structure, implementation conditions, financing mechanisms, and operational characteristics. Moghadam et al. [11] explain that greenfield projects differ from acquisition-based investments in terms of investment behavior and project implementation. Ti et al. [12] show that investment decisions in greenfield projects may also depend on firm-specific motivations, ownership structure, and operational considerations. Wu et al. [13] further emphasize the importance of location conditions and knowledge-intensive clusters in shaping greenfield investment decisions.

Other infrastructure-related studies highlight the importance of project configuration, technical characteristics, and sector-specific investment conditions. Patala et al. [14] demonstrate that infrastructure investment drivers may differ across sectors depending on project configuration and operational requirements. Li et al. [15] show that infrastructure location decisions and project characteristics may significantly influence investment allocation across projects and regions. These studies suggest that project-level characteristics may influence infrastructure investment decisions differently across sectors and project environments.

Infrastructure and project finance literature also emphasizes the importance of project governance, financial planning, and risk allocation in large-scale infrastructure development. Narbaev [16] highlights the growing importance of governance mechanisms in public-private partnership projects. De Marco et al. [17] emphasize the role of cost forecasting and financial planning in project implementation. Other studies also show that infrastructure projects may be affected by external risks, institutional coordination, and financing conditions across countries [18].

Contemporary project management research also highlights the increasing role of data analytics and AI-based approaches in infrastructure decision-making and project evaluation [19]. These approaches may improve forecasting accuracy, analytical assessment, and decision-making efficiency in large-scale infrastructure projects.

The present study uses project-level data obtained from the World Bank Private Participation in Infrastructure database. The analysis focuses exclusively on greenfield infrastructure projects to maintain consistency in project structure and investment characteristics across observations. The initial dataset included 159 projects. After removing observations with incomplete information, the final sample consisted of 131 projects.

The empirical model includes both internal and external variables. Internal project characteristics include Physical Assets and Capacity, which reflect project scale and operational structure. External and contextual variables include Total Debt Funding, Debt–Equity–Grant Ratio, Number of Bids, and Income Group classification. Total Investment is used as the dependent variable representing project investment size and capital intensity.

To examine the relationship between project characteristics and investment size, the study applies multiple linear regression analysis. The regression model evaluates how internal project characteristics and external contextual factors are associated with variation in investment size across infrastructure projects. Descriptive statistics and correlation analysis are additionally used to examine relationships among variables and identify potential multicollinearity issues before model estimation.

Based on the existing literature and empirical framework, the following hypotheses are proposed:

H1: Internal project characteristics, particularly physical assets and project capacity, have a statistically significant effect on investment size.

H2: External and contextual factors, including financial structure, procurement competition, and country-level economic classification, have a statistically significant effect on investment size.

Data and Variables

The empirical analysis is based on project-level data obtained from the World Bank Private Participation in Infrastructure database. The analysis focuses only on greenfield infrastructure projects to ensure consistency in project structure and investment characteristics.

The initial dataset included 159 projects. Observations with missing values in key variables were excluded from the analysis. The dataset includes greenfield infrastructure projects implemented between 2000 and 2024 and obtained from the World Bank Private Participation in Infrastructure (PPI) database. As a result, the final sample consisted of 131 greenfield infrastructure projects.

The dependent variable is Total Investment, measured in million USD. In this study, it is treated as an indicator of project scale and capital intensity rather than a direct measure of project performance.

The independent variables represent the main structural, operational, and financial characteristics of infrastructure projects. Internal project variables include Physical Assets and Capacity, which reflect project scale and operational structure. External and contextual variables include Number of Bids, Total Debt Funding, Debt–Equity–Grant Ratio, and Income Group classification. These variables are consistent with project-level and infrastructure investment studies that emphasize project structure, financing conditions, and location-related characteristics [11, 14, 15, 17].

The sample includes projects from both upper middle income and lower middle income countries based on the World Bank classification. Upper middle income countries include Brazil, Turkey, Mexico, and South Africa, while lower middle income countries include India, Indonesia, Morocco, and Pakistan.

A summary of the variables used in the analysis is presented in Table 1.

To examine the relationship between project characteristics and investment size, the study applies multiple linear regression analysis. This method is appropriate for evaluating the simultaneous effect of several explanatory variables on a continuous dependent variable. Similar approaches are widely used in infrastructure investment and project finance research [16, 17].

Table 1 – Variable Description

Variable	Description
Total Investment	Total investment value of the project (million USD)
Physical Assets	Value of physical infrastructure assets associated with the project (million USD)
Capacity	Operational or production capacity of the project (project-specific units)
Number of Bids	Number of bidders participating in the procurement process (count)
Total Debt Funding	Total debt financing allocated to the project (million USD)
Debt–Equity–Grant Ratio	Composition of debt, equity, and grant financing (ratio)
Income–Group Dummy	Country income classification (1 = upper middle income; 0 = lower middle income)

Analytical Approach

The regression model evaluates how internal project characteristics and external contextual factors are associated with variation in investment size across infrastructure projects. The model includes variables related to project scale, operational capacity, financing structure, procurement conditions, and country-level characteristics.

The regression model is specified as follows:

$$TI = \beta_0 + \beta_1 PA + \beta_2 CAP + \beta_3 NB + \beta_4 TDF + \beta_5 DEG + \beta_6 IGD + \varepsilon$$

where β represents the estimated coefficients and ε is the error term.

The analysis begins with descriptive statistics to summarize the main characteristics of the dataset and identify variation in project size and project structure. Correlation analysis is then conducted to examine relationships between variables and identify potential multicollinearity issues before model estimation.

Regression models are estimated using SPSS software. Alternative model specifications are additionally considered to evaluate the stability of the results and the influence of highly correlated variables. Collinearity diagnostics are also applied to assess model reliability and ensure that the estimated relationships satisfy acceptable statistical thresholds.

Results and discussion

The empirical analysis examines the relationship between project characteristics and investment size in greenfield infrastructure projects. Descriptive statistics, correlation analysis, and regression estimation are used to evaluate the proposed hypotheses and identify the main factors associated with variation in total investment across projects.

Table 2 – Descriptive Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
Physical Assets (million USD)	660.03	1668.53	6	14800
Capacity (project-specific units)	418.33	1193.30	1	11200
Number of Bids (count)	27.84	45.15	1	180
Total Debt Funding (million USD)	487.46	1386.49	3	13415
Debt–Equity–Grant Ratio (ratio)	0.68	0.176	0	1
Total Investment (million USD)	885.81	3443.14	6	35587
Income Group Dummy (binary variable)	~0.5	-	0	1

The descriptive statistics show substantial variation across projects included in the sample. Total investment and physical assets demonstrate particularly high standard deviations, indicating the presence of both small-scale and large-scale infrastructure projects. This reflects the heterogeneous nature of greenfield infrastructure investment and project structure [11, 15].

Table 3 – Correlation Matrix

	Variable	1	2	3	4	5	6	7
1	Physical Assets	1						
2	Capacity	0.686***	1					
3	Number of Bids	-0.188*	-0.168	1				
4	Total Debt Funding	0.981***	0.745***	-0.168	1			
5	Debt–Equity–Grant Ratio	0.159	0.113	0.023	0.166	1		
6	Total Investment	0.712***	0.311***	-0.126	0.682***	0.120	1	
7	Income Group Dummy	0.084	0.017	-0.286	0.074	-0.204	0.094	1

Note: Physical Assets, Total Debt Funding, and Total Investment are measured in million USD. Capacity is measured in project-specific units. Number of Bids is measured as count data. Income Group Dummy is a binary variable. * indicates significance at the 5% level, *** indicates significance at the 1% level.

The correlation analysis indicates a strong positive relationship between Physical Assets and Total Investment, suggesting that projects with larger infrastructure components generally require higher levels of investment. A very strong positive relationship is also observed between Physical Assets and Total Debt Funding ($r = 0.981$), reflecting the financing structure commonly associated with capital-intensive infrastructure projects [11, 17].

ANOVA testing was conducted to evaluate the overall statistical significance of the regression models. The results indicate that the models are statistically significant at the 1% level ($F = 27.27$, $p < 0.001$), confirming that the explanatory variables jointly contribute to explaining variation in Total Investment across infrastructure projects.

Variance Inflation Factor (VIF) analysis and correlation diagnostics were additionally used to assess potential multicollinearity among the explanatory variables. The results suggest potential multicollinearity between Physical Assets and Total Debt Funding due to their very strong correlation. To further evaluate the proposed hypotheses, multiple linear regression analysis was conducted using two alternative model specifications. Model 1 focuses on the core project-level variables, while Model 2 additionally incorporates financing-related and country-level contextual variables. This specification approach allows the robustness of the estimated relationships to be evaluated across alternative explanatory structures.

Capacity demonstrates a moderate correlation with Total Investment, while Number of Bids and Debt–Equity–Grant Ratio show relatively weak relationships. The Income Group Dummy also demonstrates only a weak association with investment size, suggesting that project-level characteristics may play a more important role than broader external conditions in explaining differences across projects [14, 15].

The regression results show that Physical Assets have a positive and statistically significant effect on total investment in both models. This confirms the capital-intensive nature of greenfield infrastructure projects and highlights the importance of infrastructure scale in determining investment requirements [11, 15].

Table 4 – Regression Results

Variable	Model 1	Std. Error	p-value	Model 2	Std. Error	p-value
Physical Assets	1.940***	0.169	<0.001	1.409**	0.679	0.040
Capacity	-0.972***	0.233	<0.001	-1.078***	0.275	<0.001
Number of Bids	-0.479	4.570	0.917	-0.514	4.821	0.915
Debt–Equity–Grant Ratio	158.18	1165.11	0.892	201.44	1200.46	0.867
Total Debt Funding	–	–	–	0.711	0.889	0.426
Income Group Dummy	–	–	–	154.18	443.30	0.729
Constant	-82.65	824.14	0.919	-154.90	955.45	0.871
R ²	0.566			0.569		
Adjusted R ²	0.552			0.548		
F-statistic	41.11			27.27		
N	131			131		

Note: Dependent variable: Total Investment (million USD). Physical Assets and Total Debt Funding are measured in million USD. Capacity is measured in project-specific units. Number of Bids is measured as count data. Income Group Dummy is a binary variable. *** indicates significance at the 1% level, ** indicates significance at the 5% level.

This result is reasonable because greenfield infrastructure projects usually require large upfront investment in physical facilities, construction works, and technical systems. Projects with larger asset bases naturally require higher capital allocation, especially when investment is related to the creation of new infrastructure rather than the expansion of existing assets. This finding is consistent with project-level studies showing that greenfield investment decisions are closely linked to project structure, location, and infrastructure configuration [11, 15]. It also supports infrastructure finance research, where cost planning and project scale are treated as important elements of investment estimation.

Capacity is also statistically significant, although the coefficient is negative. This may indicate the presence of operational efficiency or economies of scale, where higher-capacity projects are able to generate greater output without a proportional increase in investment [14].

The negative relationship may also reflect differences in infrastructure utilization and project design across sectors. Some large-capacity projects may operate more efficiently due to standardized infrastructure systems, technological optimization, or lower marginal expansion costs. In infrastructure projects, capacity growth does not always require proportional increases in physical investment, particularly in sectors where operational efficiency and technological integration play an important role. Similar observations are discussed in infrastructure investment studies emphasizing the relationship between project configuration and operational performance [14].

In contrast, Number of Bids, Debt–Equity–Grant Ratio, Total Debt Funding, and Income Group Dummy do not demonstrate statistically significant effects. These findings suggest that external and contextual factors may play a more limited role in explaining differences in investment size across projects.

One possible explanation is that investment size in greenfield infrastructure projects is determined primarily during the project planning and design stage, where technical structure and infrastructure requirements have a stronger influence than external market conditions. Variables such as procurement competition or country income classification may affect implementation conditions, financing accessibility, or project risk, but they do not necessarily determine the overall scale of infrastructure investment. This may explain why project-level characteristics demonstrate stronger statistical relationships compared to broader contextual variables. Similar conclusions are observed in studies emphasizing the importance of project configuration and infrastructure structure in investment allocation decisions [11, 14, 15].

The explanatory power of both models remains relatively stable, with R^2 values close to 0.57, indicating satisfactory model performance. Overall, the results support H1, confirming that internal project characteristics significantly affect investment size. However, H2 is not supported, as external and contextual variables do not demonstrate statistically significant relationships within the sample. The relatively small difference in explanatory power between the two models additionally suggests that financing-related and contextual variables contribute only marginally to explaining variation in investment size within the analyzed sample. This further supports the importance of internal project characteristics in greenfield infrastructure investment decisions.

The findings are generally consistent with previous studies emphasizing the importance of project configuration, infrastructure scale, and operational characteristics in shaping greenfield investment decisions. Moghadam et al. [11] argue that greenfield projects differ substantially from acquisition-based investments due to their higher dependence on infrastructure development and project implementation requirements. Similarly, Patala et al. [14] show that investment allocation in infrastructure projects may vary depending on project configuration and sector-specific operational conditions. The results are also consistent with Li et al. [15], who emphasize the importance of infrastructure location and project characteristics in shaping investment patterns across projects and regions. In addition, the findings support project management research highlighting the importance of cost forecasting and infrastructure planning in large-scale construction projects [17].

From a practical perspective, the findings highlight the importance of technical planning, infrastructure configuration, and operational assessment in greenfield infrastructure investment decisions. The results suggest that project-specific characteristics may provide stronger explanatory value for investment allocation than broader country-level indicators within the analyzed sample. These findings may assist policymakers, infrastructure planners, and investors in improving project evaluation and investment planning processes in developing economies.

Conclusion

This study examined the relationship between project characteristics and investment size in greenfield infrastructure projects using project-level data from the World Bank Private Participation in Infrastructure database. The analysis focused on how internal project characteristics and external contextual factors are associated with differences in total investment across infrastructure projects.

The findings lead to several important conclusions. First, the results show that internal project characteristics play the most important role in explaining investment size. Physical Assets demonstrated a positive and statistically significant relationship with total investment in both regression models, confirming the capital-intensive nature of greenfield infrastructure projects. This suggests that infrastructure scale remains one of the main factors influencing investment requirements.

Second, the analysis showed that Capacity also has a statistically significant effect on investment size. The negative relationship may indicate the presence of operational efficiency or economies of scale, where projects with higher capacity are able to generate greater output without proportional increases in investment. This demonstrates that different dimensions of project scale may influence investment decisions in different ways.

Third, external and contextual variables, including Number of Bids, Debt–Equity–Grant Ratio, Total Debt Funding, and Income Group classification, did not demonstrate statistically significant effects in the regression models. These findings suggest that project-level structure and operational characteristics may have a stronger influence on investment decisions than broader country-level conditions or financing composition.

The study contributes to the existing literature by providing project-level empirical evidence on greenfield infrastructure investment. While many previous studies focus primarily on macroeconomic and institutional determinants of foreign direct investment, the present analysis highlights the importance of infrastructure configuration, project scale, and operational characteristics in explaining differences in investment size across projects.

The findings additionally demonstrate the importance of combining project-level analysis with infrastructure finance perspectives when examining investment behavior in greenfield projects. The results suggest that infrastructure investment decisions may depend more strongly on technical and operational project characteristics than on broader macroeconomic conditions alone. This highlights the need for greater attention to project structure, infrastructure configuration, and operational planning in future greenfield investment research.

From a practical perspective, the findings suggest that policymakers and investors should place greater attention on project-specific characteristics during infrastructure planning and investment evaluation. In developing economies, where infrastructure expansion often requires substantial financial resources, accurate project assessment becomes particularly important. Evaluating infrastructure scale, operational capacity, and technical structure during the planning stage may improve investment allocation and reduce uncertainty during project implementation. The findings also suggest that infrastructure projects should not be evaluated only through broader economic indicators, since project-level characteristics may provide more direct information about investment requirements and project scale. Improving project design, infrastructure configuration, and operational assessment may support more efficient allocation of investment resources in large-scale infrastructure projects.

Despite its contributions, the study has several limitations. First, the sample size of 131 observations may limit the generalizability of the findings across different infrastructure sectors and regions. Second, the analysis relies on secondary data obtained from the PPI database and is therefore limited to available project-level variables. In addition, total investment is used as an indicator of project scale rather than a direct measure of project performance, which limits the ability to evaluate efficiency, profitability, or long-term project outcomes. In addition, the external environment is represented mainly by the Income Group Dummy variable, which may not fully capture broader institutional, macroeconomic, and political conditions affecting greenfield infrastructure investment.

Future research may extend the analysis by incorporating larger datasets, additional financial and operational variables, and alternative performance measures. Further studies may also examine sector-specific differences and compare infrastructure investment patterns across countries and regions. Future studies may also examine the relationship between project characteristics and alternative dimensions of project performance, including operational efficiency, profitability, implementation risk, and long-term infrastructure sustainability. Additional comparative studies across sectors and regions may further improve understanding of investment behavior in greenfield infrastructure projects.

Overall, the findings suggest that investment size in greenfield infrastructure projects depends more strongly on internal project structure and infrastructure characteristics than on broader country-level conditions.

REFERENCES

- 1 Harms, P., Méon, P.-G. Good and useless FDI. *Review of International Economics*, 26(1), 37–59 (2018). <https://doi.org/10.1111/roie.12302>
- 2 Kumar, A., Srivastava, V., Tabash, M.I. Infrastructure project finance: a systematic literature review and directions for future research. *Qualitative Research in Financial Markets*, 13(3), 295–327 (2021). <https://doi.org/10.1108/QRFM-07-2020-0130>
- 3 De Marco, A., Rosso, M., Narbaev, T. Nonlinear cost estimates at completion adjusted with risk contingency. *Journal of Modern Project Management*, 4(2), 24–33 (2016). <https://doi.org/10.19255/JMPM01102>
- 4 Amendolagine, V., Crescenzi, R., Rabellotti, R. The geography of acquisitions and greenfield investments: Firm heterogeneity and regional institutional conditions. *Journal of Regional Science*, 64(3), 1–30 (2024). <https://doi.org/10.1111/jors.12705>

- 5 Alon, I., Bretas, V.P.G., Schlip, A., Paltrinieri, A. Greenfield FDI attractiveness index: A machine learning approach. *Multinational Business Review*, 30(4), 620–642 (2022). <https://doi.org/10.1108/MBR-11-2020-0155>
- 6 Antonietti, R., Cainelli, G., Lupi, C. Foreign direct investment and institutional quality: Evidence from emerging economies. *Review of World Economics*, 159, 421–445 (2023). <https://doi.org/10.1007/s10290-022-00473-7>
- 7 Busse, M., Hefeker, C. Political risk, institutions and foreign direct investment. *European Journal of Political Economy*, 23(2), 397–415 (2007). <https://doi.org/10.1016/j.ejpoleco.2007.02.003>
- 8 Dunning, J.H. The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review*, 9(2), 163–190 (2000). [https://doi.org/10.1016/S0969-5931\(99\)00035-9](https://doi.org/10.1016/S0969-5931(99)00035-9)
- 9 Helpman, E., Melitz, M., Yeaple, S. Export versus FDI with heterogeneous firms. *American Economic Review*, 94(1), 300–316 (2004). <https://doi.org/10.1257/000282804322970814>
- 10 Markusen, J.R., Maskus, K.E. Discriminating among alternative theories of the multinational enterprise. *Review of International Economics*, 10(4), 694–707 (2002). <https://doi.org/10.1111/1467-9396.00359>
- 11 Moghadam, A.T., Zali, M.R., Rezaei, S. M&A and greenfield FDI: location choice and investment behavior. *Journal of Economic Studies*, 46(5), 1037–1054 (2019). <https://doi.org/10.1108/JES-10-2017-0280>
- 12 Ti, L.-L., Ng, B.-K., Rasiyah, R. Motivations behind SME greenfield investment in emerging markets. *Management Decision*, 61(6), 1656–1676 (2023). <https://doi.org/10.1108/MD-04-2022-0482>
- 13 Wu, L., Sutherland, D., Peng, X., Anderson, J. Are Chinese MNEs more strongly attracted to global cities and knowledge intensive city clusters than developed market MNEs when undertaking greenfield strategic asset seeking related FDI? *Competitiveness Review*, 33(3), 534–556 (2023). <https://doi.org/10.1108/CR-02-2022-0026>
- 14 Patala, S., Jalkala, A., Keränen, J. Configurational drivers of renewable energy investment. *Journal of Cleaner Production*, 281 (2021). <https://doi.org/10.1016/j.jclepro.2020.124752>
- 15 Li, S., Li, X., Qian, Z. Geography of greenfield investment and infrastructure location decisions. *Journal of Regional Science*, 64(1), 55–78 (2024). <https://doi.org/10.1111/jors.12688>
- 16 Narbaev, T. A meta-analysis of the public-private partnership literature reviews: Exploring the identity of the field. *International Journal of Strategic Property Management*, 26(4), 318–331 (2022). <https://doi.org/10.3846/ijspm.2022.17860>
- 17 De Marco, A., Ottaviani, F.M., Bolognesi, F. Time series-based project cost forecasting framework. *Procedia Computer Science*, 239, 105–113 (2024). <https://doi.org/10.1016/j.procs.2024.06.152>
- 18 Castelblanco, G., Narbaev, T., Osei-Kyei, R., Serikbay, D., Mukashev, Y., Guevara, J. Pandemic risks in PPPs: comparative analysis within developing countries. *International Journal of Construction Management*, 25(4), 456–468 (2025). <https://doi.org/10.1080/15623599.2024.2336660>
- 19 Kalidullin, B., Samoilov, A., De Marco, A. AI Applications in Project Management: Trends, Challenges and Future Perspectives. *Herald of the Kazakh-British Technical University*, 22(3), 390–403 (2025). <https://doi.org/10.55452/1998-6688-2025-22-3-390-403>

^{1*}Рыскулбекова М.М.,
зерттеуші, ORCID ID: 0009-0008-2979-7857,
*e-mail: maralryskulbekova.98@gmail.com

¹«Nurstroy Yug» ЖШС, Шымкент қ., Қазақстан

ГРИНФИЛД ИНФРАҚҰРЫЛЫМДЫҚ ЖОБАЛАРДАҒЫ ИНВЕСТИЦИЯ КӨЛЕМІНІҢ ЖОБАЛЫҚ ДЕҢГЕЙДЕГІ ДЕТЕРМИНАНТТАРЫ

Аңдатпа

Бұл зерттеу гринфилд инфрақұрылымдық жобалардағы жоба сипаттамалары мен инвестиция көлемі арасындағы байланысты талдауға бағытталған. Қолданыстағы тікелей шетелдік инвестициялар жөніндегі зерттеулер негізінен макроэкономикалық факторларға назар аударса, жобаға тән сипаттамалардың рөлі жеткілікті деңгейде зерттелмеген. Зерттеуде Дүниежүзілік банктің Private Participation in Infrastructure (PPI) дерекқорынан алынған 131 инфрақұрылымдық жоба негізінде көптік сызықтық регрессия әдісі қолданы-

лады. Талдау барысында физикалық активтер, жоба қуаттылығы, бәсекелестік шарттары, қаржыландыру құрылымы және елдік факторлардың инвестиция көлеміне әсері бағаланады. Нәтижелер физикалық активтердің инвестиция көлеміне оң және статистикалық маңызды әсер ететінін көрсетеді, бұл инфрақұрылымдық жобалардың капитал сыйымдылығын растайды. Ал жоба қуаттылығы кері бағыттағы маңызды байланыс көрсетеді, бұл ауқым тиімділігі немесе өнімділіктің артуымен түсіндірілуі мүмкін. Қалған айнымалылар, соның ішінде бәсекелестік деңгейі, қаржылық құрылым және елдік факторлар статистикалық маңызды әсер көрсетпейді. Зерттеу нәтижелері гринфилд жобалардағы инвестициялар көбіне сыртқы жағдайларға емес, жобаның ішкі құрылымына тәуелді екенін көрсетеді. Бұл макроэкономикалық факторлардан гөрі жобалық деңгейдегі талдаудың маңыздылығын арттырады. Практикалық тұрғыдан алғанда, нәтижелер мемлекеттік органдар мен инвесторларға инвестиция тартуда макроэкономикалық жағдайларға ғана емес, жобаның техникалық құрылымы мен сапасына басымдық беру қажеттігін көрсетеді.

Түйін сөздер: гринфилд инвестициялар, жоба сипаттамалары, инфрақұрылымдық инвестициялар, инвестиция детерминанттары, жобаны жобалау, көптік сызықтық регрессия.

^{1*}**Рыскулбекова М.М.,**
исследователь, ORCID ID: 0009-0008-2979-7857,
*e-mail: maralryskulbekova.98@gmail.com

¹ТОО «Nurstroy Yug», г. Шымкент, Казахстан

ДЕТЕРМИНАНТЫ ОБЩЕГО ОБЪЕМА ИНВЕСТИЦИЙ В ГРИНФИЛД ИНФРАСТРУКТУРНЫХ ПРОЕКТАХ НА УРОВНЕ ПРОЕКТА

Аннотация

Данное исследование рассматривает взаимосвязь между характеристиками проектов и общим объемом инвестиций в greenfield инфраструктурные проекты. В то время как существующая литература по прямым иностранным инвестициям в основном сосредоточена на макроэкономических и институциональных факторах, проектные характеристики остаются сравнительно недостаточно изученными. На основе выборки из 131 инфраструктурного проекта, полученной из базы данных Всемирного банка Private Participation in Infrastructure (PPI), в исследовании применяется множественный линейный регрессионный анализ для оценки влияния физических активов, мощности проекта, условий закупок, финансовой структуры и страновых характеристик на объем инвестиций. Результаты показывают, что физические активы оказывают положительное и статистически значимое влияние на общий объем инвестиций, что подтверждает капиталоемкий характер greenfield инфраструктурных проектов. В то же время мощность проекта демонстрирует значимую отрицательную взаимосвязь, что может свидетельствовать о наличии операционной эффективности или эффекта масштаба. Другие переменные, включая конкурентность закупок, финансовую структуру и страновые условия, не показали статистически значимого влияния. Полученные результаты подтверждают, что объем инвестиций в greenfield инфраструктурные проекты в большей степени определяется внутренними характеристиками проекта, чем внешними условиями. Исследование вносит вклад в литературу по инвестициям в инфраструктурные проекты и подчеркивает важность технического планирования, конфигурации инфраструктуры и операционной оценки при принятии инвестиционных решений.

Ключевые слова: greenfield инвестиции, инвестиции в инфраструктуру, множественная линейная регрессия, проектирование проекта, проектные характеристики, факторы инвестиционной привлекательности.

UDC 658.046
IRSTI 06.52.13

<https://doi.org/10.55452/1998-6688-2026-23-2-534-550>

¹***Israzhdinova T.S.**,

Master's student, ORCID ID: 0009-0005-7203-6058,

*e-mail: t_israzhdinova@kbtu.kz

¹Kazakh-British Technical University, Almaty, Kazakhstan

DRIVERS AND BARRIERS TO ESG IMPLEMENTATION IN ORGANISATIONAL AND PROJECT-RELATED MANAGEMENT PRACTICES IN EMERGING COUNTRIES

Abstract

This article examines the drivers and barriers to implementing Environmental, Social and Governance (ESG) principles in organisational and project-related management practices in emerging countries. A systematic literature review based on the Scopus database was conducted using PRISMA logic and interpreted through institutional theory. The search identified 605 records, from which 40 peer-reviewed studies were included after screening and eligibility assessment. The findings show that ESG implementation is shaped by the interaction of regulatory pressure, organisational readiness, financial feasibility, stakeholder coordination and implementation capability. The main drivers include supportive regulation, stakeholder pressure and involvement, training, leadership commitment, governance arrangements and technological capability. The main barriers include high upfront costs, restricted finance, fragmented regulation, insufficient professional competence, low awareness, cultural resistance and weak stakeholder engagement. The results reveal an imbalanced distribution of attention across ESG dimensions. Governance emerges as the most visible, appearing in 62.5% of the reviewed sources. Environmental follows with 55.0%. Social receives the least attention, at only 32.5%. The article identifies a persistent translation failure between ESG adoption and its operational implementation across organisational and project-related contexts. The findings contribute to understanding ESG implementation as an institutionally conditioned and capability-dependent process and offer implications for project managers, organisations and policymakers.

Keywords: barriers; drivers; emerging countries; ESG; institutional theory; project management; systematic literature review.

Received May 20, 2026; accepted June 11, 2026.

Introduction

ESG principles are increasingly important for organisations. They support sustainability performance, accountability, and long-term value creation.

In project management, ESG is particularly relevant. Project delivery translates sustainability commitments into practical decisions. For example, in the case of Nepal, the objectives of ensuring the sustainability of a water supply and sanitation project in rural areas were achieved through the coordination of stakeholders, the involvement of the local population, and the continuous refinement of management decisions throughout the project's implementation. These management practices helped translate the project's sustainability objectives into concrete outcomes, including improved sanitation, public health and gender equality [1].

ESG principles also produce measurable outcomes. This is visible in the Ghanaian construction context, where construction accounts for more than 40% of global energy consumption and greenhouse gas emissions, while green building technology adoption is shaped by government-related barriers, company-level drivers, regulation, incentives, and R&D support [2].

Formal ESG expectations often develop faster than the capacity needed for implementation. This problem is especially important in emerging and developing-country contexts. The Vietnamese

case provides direct evidence of this gap: although 80% of enterprises had committed to or planned ESG practices within the next two to four years, 71% had not planned ESG disclosure or had not considered its implications, and only 30% had taken action on ESG disclosure. This shows that formal ESG intention does not necessarily lead to practical disclosure or implementation action [3].

Other studies also link implementation difficulties to several factors. These include high investment costs and uncertain returns. Weak legislative support, training gaps, and employee awareness problems are also mentioned [4, 5].

The central argument advanced in this study is that emerging countries remain predominantly at the adoption stage of ESG. Organisations in these contexts recognise ESG expectations and encounter external pressure. Nevertheless, the translation of such commitments into operational practice across organisational and project-related contexts remains limited, selective, and inconsistent.

Existing systematic reviews address related but distinct areas of the literature. Daugaard [6] maps the development of ESG investing research, while Govindan and Hasanagic [7] examine circular-economy drivers, barriers and practices from a supply-chain perspective. Reviews of sustainable supply-chain management have also developed conceptual frameworks centred on supply-chain triggers and strategies [8]. In project-related research, Aarseth et al. [9] review project sustainability strategies, while Osei-Kyei et al. [10] focus specifically on net-zero-carbon procurement in construction projects. These reviews provide valuable insights, but their scopes remain concentrated on investment, particular sustainability models, supply chains, or specific project-related practices.

While most literature in this domain looks at a particular industry, a certain project implementation mechanism, or the determinants of either of these two variables, this study seeks to understand the link between the commitment to environmental, social, and governance issues and their corresponding implementations in terms of organisational and project management systems in emerging nations. Specifically, this study combines drivers, barriers, implementation mechanisms, environmental, social, and governance dimensions, institutional pressure and project management categories. Institutional theory explains why the adoption of a practice does not necessarily equate to the implementation of the same.

Three research questions were addressed:

1. What are the main drivers and barriers to implementing ESG principles in organisational and project-related management practices in emerging countries?
2. How is ESG implementation operationalised through concrete mechanisms such as tools, routines, systems and coordination arrangements?
3. How do emerging-country conditions shape the translation of ESG from organisational adoption to project-management practice?

The purpose of this study is to systematically review the factors facilitating and hindering the implementation of ESG principles in organisational and project-related management practices in emerging countries, and to interpret these patterns through institutional theory.

Materials and methods

A systematic literature review was conducted using PRISMA logic [10, 11, 12]. The Scopus database was selected because it provides broad coverage of peer-reviewed publications in business, management, sustainability and related fields [13]. The search was conducted in title, abstract and keyword fields and combined ESG/sustainability terms, implementation/adoption terms, driver/barrier terms and emerging/developing-country terms.

The final Scopus search string was:

(TITLE-ABS-KEY ((«ESG» OR «environmental, social, and governance» OR «sustainability» OR «sustainable development») AND («driver*» OR «barrier*» OR «challenge*» OR «enabler*» OR «success factor*» OR «factor*» OR «obstacle*»)) AND («implementation» OR «adoption») AND («emerging econom*» OR «emerging market*» OR «developing countr*»))) AND PUBYEAR

> 2015 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE, «ar»)) OR LIMIT-TO (DOCTYPE, «re»)) AND (LIMIT-TO (SUBJAREA, «BUSI»)) AND (LIMIT-TO (LANGUAGE, «English»))

The search was limited to English-language Scopus-indexed articles and reviews published between 2016 and 2025 in the Business, Management and Accounting area. The initial search produced 605 records. After removing 4 duplicates, 601 records were screened by title and abstract. At this stage, 535 records were excluded because they did not meet one or more of the inclusion criteria. The main reasons were the absence of a clear emerging- or developing-country context, the lack of a focus on ESG or sustainability implementation and adoption, a purely technical or engineering orientation, or the absence of a meaningful connection with organisational or project-related management practices.

The remaining 66 reports were assessed through full-text review. Of these, 24 were excluded because their relevance to the implementation focus of the study was indirect or only partial. A further 2 articles were excluded during final coding verification because the available evidence was insufficient for consistent inclusion in the analytical framework. As a result, 40 studies were included in the final review.

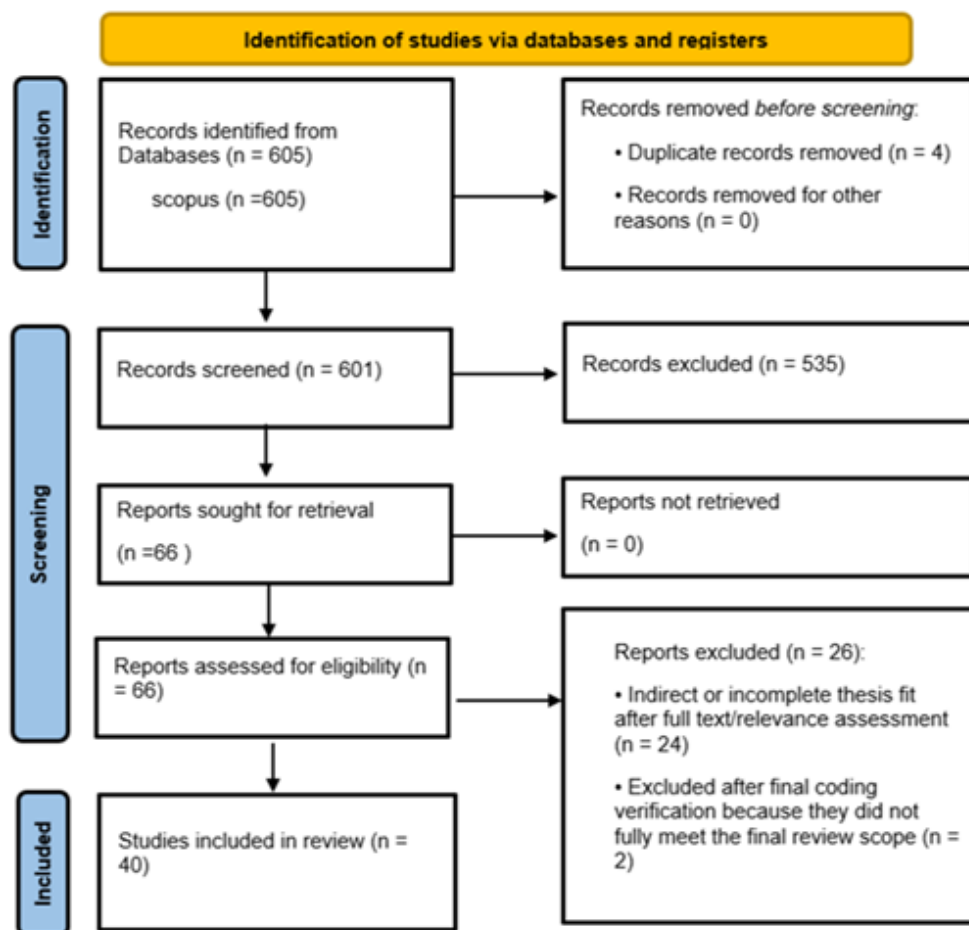


Figure 1 – PRISMA flow diagram of the study selection process

The inclusion criteria required studies to address ESG or sustainability issues; focus on implementation, adoption, drivers, barriers or related factors; examine emerging-market economies or developing countries; and relate to organisational or project-related management practices. Studies were excluded if they lacked practical relevance to implementation, focused primarily on technical or engineering aspects without a management-related dimension, or did not provide sufficiently clear evidence for inclusion in the coding framework.

Data from the 40 included articles were extracted into an Excel-based coding matrix. The analysis combined inductive coding of drivers, barriers and implementation approaches with deductive coding of ESG dimensions, institutional pressures, level of analysis, project type and project-management domains. Coding was applied conservatively; where evidence was insufficient, “No clear evidence” was recorded.

The review dimensions were developed through a combination of inductive and deductive logic. This approach is consistent with prior systematic review studies that classify literature through analytical categories, thematic synthesis and frequency-based interpretation of drivers and barriers [14]. Inductive and deductive dimensions were identified based on recurring patterns identified in the analysed studies. These dimensions capture the factors that facilitate, constrain, or operationalise ESG implementation. The deductive dimensions were used to connect the extracted evidence with the theoretical and project-management focus of the article. ESG dimensions were included to classify whether implementation was mainly related to environmental, social, governance, or combined ESG concerns. Institutional pressures were included to explain why organisations in emerging countries move toward ESG adoption. Finally, project type and PM domains were used as project-management-related lenses to assess where ESG implementation is located, whether the evidence refers to clearly bounded project contexts, and which areas of project management are most visible in the reviewed literature. This structure allowed the review to move beyond listing drivers and barriers and to analyse how ESG implementation is shaped by different conditions.

Table 1 – Review dimensions and categorisation framework

Methods	Structural dimensions	Analytical categories
Inductive	Drivers of ESG implementation	Regulatory support and policy incentives; Financial and economic benefits; Technological innovation; Stakeholder demand and pressure; Organisational commitment and strategic orientation
	Barriers to ESG implementation	Lack of regulatory frameworks; Financial constraints and high implementation costs; Technical and operational complexity; Lack of knowledge and expertise; Coordination and integration challenges; Site-specific and logistical constraints
	Implementation approaches	Technology-based approaches; Process-based approaches; Management and organisational approaches; Capability-building approaches
Deductive	ESG dimensions	Environmental; Social; Governance; combined ESG patterns
	Institutional pressures	Regulatory pressure; Normative pressure; Mimetic pressure; combined patterns
	Project type	Infrastructure; Building / construction; Green building; Energy; Public sector; No clearly bounded project type / organisational-level implementation context
	PM domains	Cost; Time; Resource; Risk; Procurement; Stakeholder; Integration
Source: compiled by the author based on the review framework and coded article data.		

The coded data were analysed through thematic synthesis and frequency analysis. Frequencies were counted only where clear evidence was available and were interpreted as recurring patterns in the reviewed literature, not as measures of causal strength [10, 14, 15]. The results of this synthesis are presented in the next section.

Results and discussion

Dominant drivers of ESG implementation

The inductive findings show that ESG implementation is supported by several interrelated conditions rather than by a single dominant factor.

Table 2 – Drivers of ESG implementation identified in the reviewed literature

Code	Drivers/success factors	Supporting references	% of 40 articles	Rank
D1	Supportive regulatory framework for ESG implementation	[2], [35], [4], [36], [16], [37], [25], [32]	20.0%	1
D2	Professional training provision for implementation practice	[1], [26], [38], [39], [40], [41], [32]	17.5%	2
D3	Strong external stakeholder pressure for ESG adoption	[16], [29], [21], [22], [42], [39]	15.0%	3
D4	Active stakeholder involvement in implementation processes	[1], [31], [38], [43], [44], [32]	15.0%	3
D5	Strategic leadership commitment to sustainability implementation	[3], [16], [29], [37], [22]	12.5%	4
D6	Technological capability for sustainability implementation	[17], [45], [41], [25], [46]	12.5%	4
D7	Strategic orientation toward long-term sustainability value	[19], [25], [46], [41], [37]	12.5%	4
D8	Effective governance architecture for implementation oversight	[3], [16], [46], [47]	10.0%	5
D9	Long-term cost efficiency of sustainability measures	[21], [22], [19], [37]	10.0%	5
D10	Practical knowledge base for ESG implementation	[3], [40], [43], [47]	10.0%	5
D11	Supportive organisational culture for sustainability practice	[38], [43], [40]	7.5%	6
D12	Risk-management capability in implementation processes	[17], [46]	5.0%	7
D13	Enabling institutional environment for implementation uptake	[16], [42]	5.0%	7
D14	Client demand for sustainable project solutions	[35], [4]	5.0%	7
D15	Favourable market environment for sustainable delivery	[35], [29]	5.0%	7
D16	Organisational monitoring systems for implementation progress	[1], [46]	5.0%	7
Note: Categories are not mutually exclusive; frequencies indicate recurrence in the reviewed literature, not causal strength. Source: compiled by the author based on existing data.				

The drivers in Table 2 cannot be read as separate and equal factors. They form an implementation chain. External pressure creates the reason for organisations to act. Internal governance determines how this pressure is interpreted. Practical capability then affects whether ESG-related commitments become part of organisational and project management routines. These three clusters were derived

inductively from frequency patterns in the reviewed literature. They reflect repeated groupings of favourable conditions within the framework of the reviewed studies.

Cluster 1: External institutional and stakeholder pressure. External pressure is the first reason organisations act. This pressure comes from institutions and stakeholders (D1, D3, D4, D13, D14, D15). For example, in Ghana’s mining sector regulatory compliance, voluntary compliance, and stakeholder demands encouraged multinational mining companies to adopt sustainability practices. Regulatory requirements had clear goals. They focused on environmental protection, impact mitigation, and post-closure land rehabilitation. This example illustrates a key point. External pressure can turn sustainability from a general corporate aspiration into formal organisational requirements [16].

Cluster 2: Internal governance and strategic commitment. This translates external expectations into organisational priorities and responsibilities (D7, D8, D11, D16). Amoah and Eweje (2023) show that sustainability implementation in Ghanaian mining companies was shaped by external institutional pressure and internal organisational features such as managerial cognition or organisational size [16]. This suggests that organisations do not simply react to ESG pressure. They interpret it through their own values, strategic interests, and other conditions.

Cluster 3: Practical implementation capability. This cluster reflects the capability-related drivers (D2, D6, D10, D12). Capability is the condition that allows ESG-related implementation to move beyond policy language. The I5.0 drones study in Nigeria supports this point. It shows that drone use can support cleaner production through specific operational tasks, such as reducing plant diseases and supporting more precise agricultural operations. The study also notes the need for awareness and knowledge transfer on the potential of I5.0 technologies in emerging economies. This means that technology alone is not enough. It must be supported by user knowledge and practical understanding of where it can improve implementation [17]. The RTP study in Nigeria and South Africa gives a similar operational example. It conceptualises returnable transport packaging as an environmental technology and resource, but also shows that effective use requires tracking, control systems and management capabilities [18]. Together, these studies show that implementation capability is the practical bridge between sustainability goals and operational change.

Overall, the driver findings indicate that ESG implementation is sequential rather than additive. External pressure may initiate adoption, but it does not by itself ensure operational change. Its effect depends on whether organisations translate expectations into governance arrangements, assign responsibilities and develop the skills, systems and technologies needed for implementation. This helps explain why similar regulatory or stakeholder pressures can lead to uneven outcomes across organisations. In emerging-country contexts, the decisive issue is therefore not the presence of an individual driver, but the alignment between external pressure, internal governance and practical capability.

Dominant barriers of ESG implementation

Table 3 – Barriers to ESG implementation identified in the reviewed literature

Code	Barriers	References	% of 40 articles	Rank
B1	High upfront cost burden of sustainability uptake	[35], [4], [5], [23], [29], [21], [22], [24], [18], [34], [32]	27.5%	1
B2	Restricted access to implementation finance	[48], [5], [19], [49], [24], [18], [37], [50]	20.0%	2
B3	Fragmented regulatory framework for sustainability uptake	[1], [2], [5], [23], [29], [21], [33], [32]	20.0%	2
B4	Insufficient professional competence for sustainability implementation	[35], [4], [21], [19], [26], [34], [33], [32]	20.0%	2

Continuation of table 3

B5	Low awareness of implementation benefits	[21], [49], [31], [39], [33], [32]	15.0%	3
B6	Inadequate digital readiness for ESG integration	[48], [5], [21], [34], [44]	12.5%	4
B7	Entrenched cultural resistance to sustainable practice	[1], [49], [31], [26], [33]	12.5%	4
B8	Weak stakeholder engagement in implementation processes	[5], [49], [31], [33], [32]	12.5%	4
B9	Thin practical knowledge base for ESG uptake	[22], [49], [51], [33]	10.0%	5
B10	Fragile executive commitment to sustainability goals	[48], [21], [26], [34]	10.0%	5
B11	Underdeveloped implementation infrastructure	[48], [24], [44], [33]	10.0%	5
B12	Insufficient workforce training for ESG practice	[4], [19], [26], [33]	10.0%	5
B13	Weak transition management during organisational change	[1], [35], [32]	7.5%	6
B14	Weak governance architecture for ESG oversight	[1], [48]	5.0%	7
B15	Constraining institutional environment for implementation	[24], [39]	5.0%	7
B16	Unfavourable market conditions for sustainable delivery	[24], [44]	5.0%	7
Note: Categories are not mutually exclusive; frequencies indicate recurrence in the reviewed literature, not causal strength. Source: compiled by the author based on existing data.				

The barrier findings suggest that ESG-related implementation in emerging-country contexts is constrained by interacting financial, institutional and capability-related conditions. This pattern indicates that financial limits, weak institutional support and capability gaps often reinforce one another rather than operating as isolated obstacles.

Financial feasibility is the first major barrier cluster (B1, B2). ESG-related implementation often requires investment in technologies, reporting procedures and training. Under conditions of limited resources, these requirements can constrain implementation even where organisations are willing to adopt sustainability-related practices. It is visible in the Thai manufacturing sector [19]. This study found that environmental sustainability approaches were commonly used, but implementation remained moderate. Full implementation required large investment capacity and proper training and knowledge, which were among the weakest elements of the firms studied. This suggests that financial barriers are not only about cost. They also restrict the organisational capacity needed to build implementation routines.

The second cluster concerns external and market-environment barriers (B3, B15, B16). Fragmented regulation and unfavourable market conditions refer to external conditions that make ESG implementation harder. They include unclear rules or ambiguous rules, weak monitoring, limited external support and challenging market conditions. Organisations face uncertainty about what is required and lack the external pressure needed to sustain implementation after formal ESG commitments are made [16]. In emerging-country contexts, this institutional uncertainty is particularly significant because it can leave organisations without clear guidance or monitoring mechanisms. Even in resource-rich contexts such as Kazakhstan, the implementation of ESG faces structural constraints. Akhmedov et al. [20] show that the transition to renewable energy is hindered by economic, financial, institutional, technical and regulatory barriers, including investment risks and

the need for regulatory and institutional improvements. This suggests that institutional uncertainty is not limited to low-income developing countries but extends to emerging economies with significant resource endowments, where structural dependencies on existing industries can delay sustainability transitions regardless of financial capacity.

The third cluster concerns organisational capability (B4, B5, B6, B9, B12). Low awareness, weak training and limited technical readiness affect whether sustainability goals can be translated into operational practice. A clear example comes from Pakistan's automobile industry [21]. The main barriers to circular-economy implementation were unawareness, high costs, financial constraints. The study also shows that firms lacked the technical expertise needed to convert waste materials into useful remanufactured products. This evidence supports the interpretation that implementation depends not only on willingness, but also on knowledge, technical competence and learning capacity.

A similar capability problem appears in service and supply-chain contexts. Local managers in Malawi's hospitality sector lacked knowledge about sustainability practices and how these practices could be integrated into business operations [22]. Some also viewed sustainability as a cost burden because they did not clearly understand its business benefits. Similarly, evidence from Pakistan's textile sector shows that sectoral-economic, managerial and supplier-related hindrances influence the adoption of sustainable supply-chain management practices [23]. These findings suggest that ESG-related implementation depends on both internal readiness and coordination across organisational and supply-chain actors.

Overall, the barrier findings show that ESG-related implementation in emerging countries is financially constrained, institutionally uncertain and capability-dependent. High upfront costs limit investment in systems and technologies. Weak institutions reduce enforcement and accountability. Capability gaps then make it difficult to convert sustainability expectations into operational routines. Therefore, the main finding is not simply that some barriers appear more often than others. The deeper point is that these barriers interact and create a translation failure between ESG-related adoption and practical implementation.

The driver and barrier findings point to a common pattern: ESG implementation is less a question of formal commitment than of organisational translation. The same factor can operate differently depending on context. Regulation is a clear example. Coherent standards, monitoring and policy incentives support implementation, whereas fragmented rules and weak enforcement increase uncertainty and reduce accountability. Financial constraints have a similar multiplier effect: they limit not only investment in technologies but also training, monitoring and coordination capacity.

This pattern creates a risk of decoupling between adoption and practice. Organisations may respond to external pressure by adopting ESG language, policies or reporting commitments, while implementation remains selective and uneven. The findings therefore suggest that the central challenge in emerging-country contexts is not simply to increase ESG awareness, but to build the institutional and organisational conditions that allow commitments to become stable routines.

Implementation approaches through which ESG is operationalised

This subsection examines how ESG-related principles are operationalised. The focus is on the reviewed literature. To avoid duplication with drivers and barriers, implementation approaches are interpreted in different ways. They are considered as practical mechanisms. Such mechanisms include tools and methods of coordination.

Table 4 shows that the implementation of ESG principles depends on an organisation's ability to put its sustainability goals into practice.

Technology-based approaches (IA1) show that technology can support implementation when it improves operational control. For instance, a case study of a sub-Saharan African firm operating in the agricultural industry shows that digital technologies can support operational optimisation and more efficient resource use [24]. However, the same study also shows that technology alone is insufficient. Wider adoption was constrained by limited government support and access to foreign markets, as well as economic, technological and cultural barriers. This means that technology-based implementation depends on whether firms operate in conditions that allow these tools to be used effectively.

Table 4 – Implementation approaches and mechanisms through which ESG is operationalised

Code	Broad implementation approach	Representative mechanisms	References	% of 40 articles
IA1	Technology-based approaches	Digital tools; drones; KPI/carbon monitoring	[35], [17], [24], [25]	10.0%
IA2	Process-based approaches	Environmental management routines; ISO 14001/PDCA routines; circular-economy practices; recycling; reverse logistics;	[4], [5], [19], [51], [18], [27], [30]	17.5%
IA3	Management and organisational approaches	Local facilitation; coordination; supplier partnerships; monitoring systems; carbon management accounting;	[1], [29], [49], [31], [38], [46], [30]	17.5%
IA4	Capability-building approaches	Training; organisational learning; CSR communication	[1], [29], [40], [26], [38], [43], [39]	17.5%
Note: Categories are not mutually exclusive; frequencies indicate recurrence in the reviewed literature, not causal strength. Source: compiled by the author based on existing data.				

Process-based approaches (IA2) show that implementation depends on repeatable practice, not only on formal sustainability goals. For example, 70.9% of firms in Nigeria and South Africa had adopted RTP, but only 4.3% of non-adopters were certain that they would implement it in the future [18]. This suggests that process-based mechanisms can make sustainability practical, but their diffusion may remain disparate in developing-country contexts.

Management and organisational approaches (IA3) indicate that implementation hinges on how organisations allocate responsibilities and coordinate internal decisions. Evidence from the hotel sector across Africa and Asia suggests that sustainability practices gain traction when technological advancement, organisational commitment, and environmental standards are aligned [25]. The issue is not the introduction of a single tool. Rather, it lies in connecting technology, internal policies, and external standards into a functioning system. In the absence of such alignment, sustainability commitments tend to remain formal declarations rather than operational realities.

Capability-building approaches (IA4) show that training and learning do not automatically produce implementation. For example, in Brazilian public universities environmental training was viewed as useful, but it produced limited accomplishments [26]. This suggests that capability-building is not only about providing training. It requires internal conditions that allow learning to change practice.

The implementation-approach findings suggest that operationalising ESG is not merely a technical challenge. Process routines, stakeholder coordination, and capacity-building appear more often in the reviewed literature than technology-centric tools. This distinction is important: it moves the question away from which technologies an organisation adopts and toward whether it has the organisational preconditions to use them well.

In emerging-country contexts, limited resources, weak institutional support and infrastructure constraints mean that even well-designed mechanisms may be applied unevenly. The gap between formally adopting an implementation approach and embedding it into routine practice reflects the same translation failure identified in the driver and barrier findings. External commitments, on their own, do not automatically translate into operational change.

ESG dimensions

The ESG-dimension findings show that ESG implementation in the reviewed literature is not presented as an equal balance of environmental, social and governance concerns. The distribution is not consistent: Governance dominates, Environmental follows closely, while Social appears in a smaller share of articles.

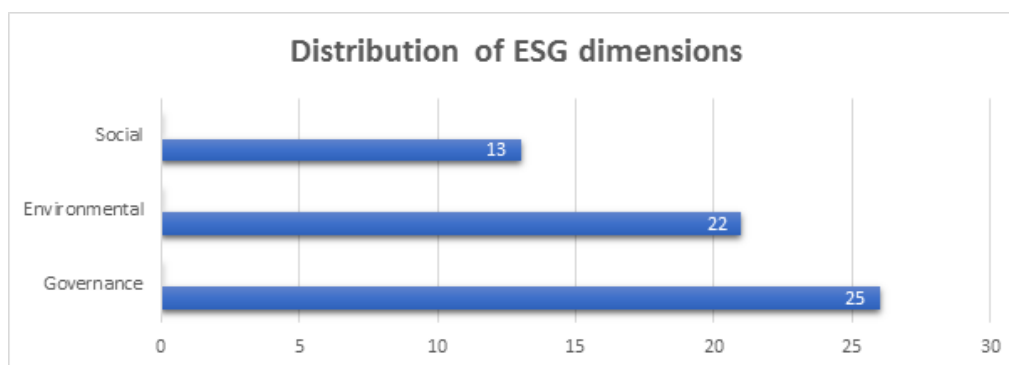


Figure 2 – Distribution of ESG dimensions

The ESG-dimension findings reveal an imbalanced distribution of attention across the three pillars that warrants critical interpretation rather than simple description. Governance emerges as the most visible dimension, appearing in 62.5% of the reviewed sources, followed by Environmental at 55.0%, while Social receives the least attention at only 32.5%.

This hierarchy is not arbitrary. It reflects a structural feature of how ESG translates into observable organisational practice. Governance and environmental concerns are more readily codified: they produce measurable artefacts such as board-level policies, environmental management systems, carbon accounting routines and ISO-certified procedures. These artefacts are easier to study, report and verify, which means they are more likely to appear in peer-reviewed implementation research. The dominance of governance in the findings therefore reflects not its greater practical importance, but its greater institutional visibility. This is consistent with Lucas et al. [25], who show that sustainability practices in the hospitality sector cluster around technological, organisational and environmental conditions, precisely the dimensions that lend themselves to formal measurement and compliance reporting.

The underrepresentation of the social dimension is more consequential than a bibliometric curiosity. Social concerns such as worker welfare, community impact, health and safety, gender equity, stakeholder participation are harder to standardise, harder to measure and more politically contested than environmental or governance metrics. As Esfahbodi et al. [27] note, direct social measures remain less developed in sustainable supply chain management research. In emerging-country contexts, this gap has practical implications: if social concerns are less visible in the literature, they are also less likely to be integrated into implementation frameworks, monitoring systems or project-level decision-making. The result is an ESG model that, in practice, skews toward the dimensions that are easiest to report rather than those that may matter most to affected communities. Future research and practice should treat this imbalance as a substantive problem, not a measurement gap to be corrected at some later stage.

Institutional pressures

The institutional-pressure findings explain the external conditions shaping ESG implementation in the reviewed literature [28].

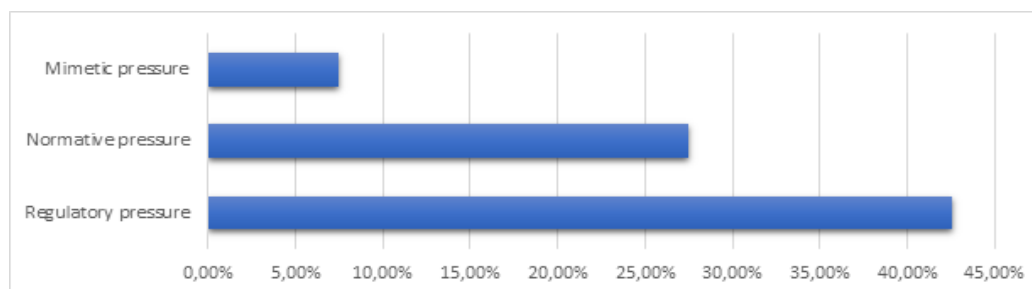


Figure 3 – Institutional pressure pattern

Regulatory pressure is the most visible institutional pattern in the coding results. This indicates that ESG implementation in emerging-country contexts is often explained through compliance with various requirements. Manufacturing firms in emerging markets provide relevant evidence. Regulatory tightening and international legislation can drive the adoption of SSCM [27].

Normative pressure is visible when organisations respond to external expectations. These expectations concern responsible and legitimate behaviour. In Malawi's hospitality sector, local societal expectations and the pursuit of community legitimacy shaped sustainability practices; firms faced pressure to contribute to social well-being [22]. This suggests that ESG-related adoption may be shaped by formal regulation and legitimacy concerns.

In the reviewed literature, mimetic pressure does not emerge as consistently. This does not mean that organisations do not imitate one another. It means that imitation requires clear examples to follow. In many emerging economies, the adoption of ESG practices remains constrained by weak regulation and limited organisational capacity. These conditions can hinder the development of robust practices at the sectoral level. However, mimetic pressure may still arise when suppliers compete for contracts or employees [29]. Therefore, mimetic pressure is best viewed as a conditional factor. Taken together, these findings show that ESG implementation is externally encouraged and internally determined. Governance capacity, resources and coordination determine whether external pressure becomes embedded practice.

The findings also distinguish stakeholder pressure from stakeholder involvement. Stakeholder pressure may push organisations toward ESG adoption, but implementation requires practical coordination, dialogue, partnership and participation. Customer and stakeholder expectations can create adoption pressure, while supplier partnerships, community engagement, collaborative work and knowledge transfer help turn ESG expectations into implementation mechanisms [30].

This distinction matters especially in emerging-country contexts. Where formal institutions are weak or enforcement is limited, stakeholder involvement becomes a practical substitute for missing regulatory support [31, 32]. Implementation therefore depends not only on whether stakeholders demand ESG, but on whether they are actively integrated into implementation processes.

Project-management domains and level of analysis

This section examines the project-management relevance of ESG implementation through two indicators: PM domains and project type. PM domains show which project-management concerns are most visible among studies. Project type clarifies whether the evidence is based on clearly bounded projects or broader organisational and sectoral settings. Together these factors shift the focus from the question of why organisations adopt ESG principles to where and how they are put into practice in project management.

Table 5 – PM domains associated with ESG implementation in the reviewed literature

Code	PM domain	Supporting references	Rank	% of 40 articles
PM1	Stakeholder	[1], [4], [5], [3], [16], [29], [21], [22], [19], [49], [31], [38], [43], [44], [39], [33], [32]	1	42.5%
PM2	Cost	[35], [5], [23], [21], [22], [19], [49], [37], [34], [33], [32]	2	27.5%
PM3	Integration	[3], [16], [23], [34], [25], [52], [30], [47]	3	20.0%
PM4	Resource	[1], [35], [3], [40], [41], [34], [33], [52]	3	20.0%
PM5	Procurement	[27], [30], [33], [32]	5	10.0%
PM6	Time	[33], [32]	6	5.0%
PM7	Risk	[32]	7	2.5%
Note: PM domains are deductive categories. Categories are not mutually exclusive; frequencies indicate recurrence, not causal strength. Source: compiled by the author based on existing data.				

The PM-domain findings show that ESG implementation is most often connected to Stakeholder, Cost, Integration and Resource domains. These domains are not equally represented. The reviewed literature is more concerned with whether ESG implementation is feasible than with how it is controlled once underway. ESG therefore appears more often through coordination, budgeting and organisational alignment than through detailed project-control routines.

An example from Nigerian construction helps clarify this gap. Prevention through design is linked to risk control, procurement, project cost and project time [33]. It also depends on coordination with stakeholders. This shows that safety-related ESG concerns become project-management practice when they are built into design decisions, assigned responsibilities and construction-phase activities.

Procurement, time, and risk appear less often in the coding results. That does not imply these domains lack importance. Instead, direct evidence at the project level tends to cluster contexts, construction safety being one example. ESG is more commonly tied to general managerial concerns than to concrete project tools like schedules, risk registers, or site monitoring [34].

Table 6 – Project type in the reviewed literature

Code	Project type	References	Rank	%
PT1	Organisational-level context / no clearly bounded project	[48], [35], [4], [36], [3], [16], [23], [29], [21], [22], [19], [49], [40], [51], [17], [24], [45], [18], [27], [41], [37], [43], [34], [44], [25], [42], [39], [52], [46], [50], [30], [47]	1	80.0%
PT2	Building / construction	[31], [38], [39], [32]	2	10.0%
PT3	Energy	[5]	3	2.5%
PT4	Green building	[2]	3	2.5%
PT5	Infrastructure	[1]	3	2.5%
PT6	Public sector	[26]	3	2.5%
Note: PT are deductive categories. Categories are not mutually exclusive; frequencies indicate recurrence, not causal strength. Source: compiled by the author based on existing data.				

The project-type findings strengthen this interpretation. Most reviewed articles examine ESG implementation at organisational or sectoral level rather than within clearly bounded projects. Project-specific evidence exists, particularly in construction, green building and infrastructure contexts, but it is less dominant. The project-management relevance of the evidence base is therefore mixed: the literature identifies the conditions that influence ESG implementation more clearly than the mechanisms through which ESG is embedded in project delivery.

This imbalance is also visible in the PM-domain findings. Stakeholder, cost, integration and resource concerns appear more frequently than time, risk and procurement. The pattern suggests that ESG is still discussed mainly through feasibility, coordination and organisational alignment, rather than through detailed project-control routines. The low visibility of schedules, risk registers, procurement criteria and site-level monitoring should not be read as evidence that these tools are unimportant. Instead, it indicates that the existing literature provides limited direct evidence on how ESG commitments are converted into project-specific artefacts and decisions.

The predominance of organisational-level studies is therefore not only a limitation of this review; it is also a substantive finding. In emerging-country contexts, ESG appears to be more developed as a strategic and institutional agenda than as a consistently embedded project-delivery practice. Future empirical research should examine how organisational commitments are translated into responsibilities, monitoring arrangements and decision points across specific project phases.

Conclusion

This study examined the drivers and barriers to implementing ESG principles in organisational and project-related management practices in emerging countries through a systematic literature review. The findings show that ESG implementation is a multi-dimensional and institutionally conditioned process rather than the result of isolated factors. The main drivers include supportive regulation, stakeholder pressure and involvement, governance arrangements, training, leadership commitment and implementation capability. The main barriers include high upfront costs, restricted finance, fragmented regulation, insufficient competence, low awareness, cultural resistance and uneven organisational readiness.

The study contributes to the literature by synthesising fragmented evidence across ESG, sustainability, circular economy, supply-chain management, construction, reporting and organisational implementation studies. It also contributes theoretically by applying institutional theory to explain why regulatory, normative and legitimacy pressures may encourage ESG adoption but remain insufficient without governance capacity, resources, skills and coordination mechanisms.

Practically, the findings suggest that ESG implementation should not be treated as a late-stage compliance activity. Project managers and organisations need to integrate ESG into feasibility assessment, budgeting, procurement planning, stakeholder engagement, monitoring, resource allocation and implementation routines. For policymakers, the findings highlight the importance of coherent regulation, policy support and enabling institutional environments.

The study has several limitations. Its findings depend on the selected Scopus-based evidence base rather than primary field data. Some dimensions, especially implementation approaches and level-of-analysis patterns, required interpretive synthesis. The coding was conducted conservatively, but it was not independently checked by a second coder. In addition, much of the reviewed evidence focuses on organisational, sectoral or supply-chain contexts rather than clearly bounded projects. This limits direct conclusions about project-specific artefacts such as schedule control, site-level risk management or project team dynamics, but it also reveals an important research gap: ESG implementation in emerging countries is still studied more at strategic and organisational levels than at project-delivery level.

Future research should examine the social dimension of ESG more directly, especially worker welfare, health and safety, community impact, stakeholder participation and social equity. More project-level empirical studies are also needed to explain how ESG principles are translated into procurement criteria, schedules, reporting systems, monitoring routines and site-level decision-making across different emerging-country contexts.

REFERENCES

- 1 Dziubaniuk, O., Ivanova-Gongne, M., and Berdysheva, E. Challenges of network interaction in managing sustainable development projects in developing countries: Case of an international consulting company. *Critical Perspectives on International Business*, 18 (4), 546–573 (2022). <https://doi.org/10.1108/cpoib-08-2020-0115>
- 2 Darko, A., Chan, A.P.C., Yang, Y., Shan, M., He, B.-J., and Gou, Z. Influences of barriers, drivers, and promotion strategies on green building technologies adoption in developing countries: The Ghanaian case. *Journal of Cleaner Production*, 200, 687–703 (2018). <https://doi.org/10.1016/j.jclepro.2018.07.318>
- 3 Tran, N.H., and Nguyen, T.T.H. Examining the drivers of ESG adoption: Empirical findings from a developing market. *Corporate and Business Strategy Review*, 4 (4), 168–176 (2023). <https://doi.org/10.22495/cbsrv4i4art15>
- 4 Hirata, R., and Ialnazov, D.S. Corporate environmental behaviour toward sustainable development: Motivations, benefits, and challenges of ISO 14001 adoption at Japanese subsidiaries in Thailand. *Business Strategy and Development*, 7 (3), e70015 (2024). <https://doi.org/10.1002/bsd2.70015>

- 5 Jesus, G.M.K., Jugend, D., Paes, L.A.B., Siqueira, R.M., and Leandrin, M.A. Barriers to the adoption of the circular economy in the Brazilian sugarcane ethanol sector. *Clean Technologies and Environmental Policy*, 25 (2), 381–395 (2023). <https://doi.org/10.1007/s10098-021-02129-5>
- 6 Daugaard, D. Emerging new themes in environmental, social and governance investing: A systematic literature review. *Accounting and Finance*, 60 (2), 1501–1530 (2020). <https://doi.org/10.1111/acfi.12479>
- 7 Govindan, K., and Hasanagic, M. A systematic review on drivers, barriers, and practices towards circular economy: A supply chain perspective. *International Journal of Production Research*, 56 (1–2), 278–311 (2018). <https://doi.org/10.1080/00207543.2017.1402141>
- 8 Seuring, S., and Müller, M. From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, 16 (15), 1699–1710 (2008). <https://doi.org/10.1016/j.jclepro.2008.04.020>
- 9 Aarseth, W., Ahola, T., Aaltonen, K., Økland, A., and Andersen, B. Project sustainability strategies: A systematic literature review. *International Journal of Project Management*, 35 (6), 1071–1083 (2017). <https://doi.org/10.1016/j.ijproman.2016.11.006>
- 10 Osei-Kyei, R., Narbaev, T., Xiaohua, J., Komac, U., Akomea-Frimpong, I., and Castelblanco, G. Critical review of the drivers and barriers for adopting net zero carbon procurement for construction projects. *Sustainable Futures*, 8, 100284 (2024). <https://doi.org/10.1016/j.sft.2024.100284>
- 11 Tranfield, D., Denyer, D., & Smart, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222 (2003). <https://doi.org/10.1111/1467-8551.00375>
- 12 Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., & Moher, D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. 2021. <https://doi.org/10.1136/bmj.n71>
- 13 Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1 (1), 377–386 (2020). https://doi.org/10.1162/qss_a_00019
- 14 Narbaev, T., Hazir, Ö., and Agi, M. A review of the use of game theory in project management. *Journal of Management in Engineering*, 38 (6), 03122002 (2022). [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0001092](https://doi.org/10.1061/(ASCE)ME.1943-5479.0001092)
- 15 Thomas, J., & Harden, A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, Article 45 (2008). <https://doi.org/10.1186/1471-2288-8-45>
- 16 Amoah, P., and Eweje, G. Organisational drivers and sustainability implementation in the mining industry: A holistic theoretical framework. *Business Strategy and the Environment*, 32 (8), 5602–5614 (2023). <https://doi.org/10.1002/bse.3438>
- 17 Mahroof, K., Omar, A., Vann Yaroson, E., Tenebe, S.A., Rana, N.P., Sivarajah, U., and Weerakkody, V. Evaluating the intention to use Industry 5.0 (I5.0) drones for cleaner production in sustainable food supply chains: An emerging economy context. *Supply Chain Management*, 29 (3), 468–496 (2024). <https://doi.org/10.1108/SCM-01-2023-0045>
- 18 Yusuf, Y.Y., Olaberinjo, A.E., Papadopoulos, T., Gunasekaran, A., Subramanian, N., and Sharifi, H. Returnable transport packaging in developing countries: Drivers, barriers and business performance. *Production Planning and Control*, 28 (6–8), 629–658 (2017). <https://doi.org/10.1080/09537287.2017.1294271>
- 19 Piyathanavong, V., Garza-Reyes, J.A., Kumar, V., Maldonado-Guzmán, G., and Mangla, S.K. The adoption of operational environmental sustainability approaches in the Thai manufacturing sector. *Journal of Cleaner Production*, 220, 507–528 (2019). <https://doi.org/10.1016/j.jclepro.2019.02.093>
- 20 Akhmedov, E., Mukashev, Y., & Akhmedov, A. (2025). Barriers to renewable energy transition in a developing oil-exporting country: The case of Kazakhstan. *Journal of Asian Energy Studies*, 9, 207–221. <https://doi.org/10.24112/jaes.090012>
- 21 Agyemang, M., Kusi-Sarpong, S., Khan, S.A., Mani, V., Rehman, S.T., and Kusi-Sarpong, H. Drivers and barriers to circular economy implementation: An explorative study in Pakistan's automobile industry. *Management Decision*, 57 (4), 971–994 (2019). <https://doi.org/10.1108/MD-11-2018-1178>
- 22 Mzembe, A.N., Melissen, F., and Novakovic, Y. Greening the hospitality industry in the developing world: Analysis of the drivers and barriers. *Business Ethics*, 28 (3), 335–348 (2019). <https://doi.org/10.1111/beer.12217>
- 23 Baig, S.A., Abrar, M., Batool, A., Hashim, M., & Shabbir, R. Barriers to the adoption of sustainable supply chain management practices: Moderating role of firm size. *Cogent Business and Management*, 7(1), Article 1841525 (2020). <https://doi.org/10.1080/23311975.2020.1841525>

- 24 Costa, F., Freccassetti, S., Rossini, M., and Portioli-Staudacher, A. Industry 4.0 digital technologies enhancing sustainability: Applications and barriers from the agricultural industry in an emerging economy. *Journal of Cleaner Production*, 408, 137208 (2023). <https://doi.org/10.1016/j.jclepro.2023.137208>
- 25 Lucas, M.M., Moreno-Luna, L., Roets, A.O., and Al-Jaberi, S. Technological, organisational and environmental drivers of sustainability in hotels. *South African Journal of Business Management*, 55 (1), a4815 (2024). <https://doi.org/10.4102/sajbm.v55i1.4815>
- 26 Aragão, C.G., and Jabbour, C.J.C. Green training for sustainable procurement? Insights from the Brazilian public sector. *Industrial and Commercial Training*, 49 (1), 48–54 (2017). <https://doi.org/10.1108/ICT-07-2016-0043>
- 27 Esfahbodi, A., Zhang, Y., and Watson, G. Sustainable supply chain management in emerging economies: Trade-offs between environmental and cost performance. *International Journal of Production Economics*, 181, 350–366 (2016). <https://doi.org/10.1016/j.ijpe.2016.02.013>
- 28 DiMaggio, P.J., & Powell, W.W. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48 (2), 147–160 (1983). <https://doi.org/10.2307/2095101>
- 29 Huq, F.A., and Stevenson, M. Implementing socially sustainable practices in challenging institutional contexts: Building theory from seven developing country supplier cases. *Journal of Business Ethics*, 161 (2), 415–442 (2020). <https://doi.org/10.1007/s10551-018-3951-x>
- 30 Kusi-Sarpong, S., Sarkis, J., and Wang, X. Assessing green supply chain practices in the Ghanaian mining industry: A framework and evaluation. *International Journal of Production Economics*, 181, 325–341 (2016). <https://doi.org/10.1016/j.ijpe.2016.04.002>
- 31 Elsayed, W. Social sustainability in housing as an entry point to achieving quality of urban life in Egypt. *Sustainable Futures*, 10, 101045 (2025). <https://doi.org/10.1016/j.sftr.2025.101045>
- 32 Kineber, A.F., Mostafa, S., Ali, A.H., Mohamed, S., and Daoud, A.O. Breaking barriers: Enhancing construction and demolition waste management in Egyptian residential projects. *Clean Technologies and Environmental Policy*, 27 (7), 3169–3188 (2025). <https://doi.org/10.1007/s10098-024-02999-5>
- 33 Umeokafor, N., Windapo, A.O., Manu, P., Diugwu, I., and Haroglu, H. Critical barriers to prevention through design in construction in developing countries: A qualitative inquiry. *Engineering, Construction and Architectural Management*, 30 (7), 3014–3042 (2023). <https://doi.org/10.1108/ECAM-04-2021-0304>
- 34 Aliu, A.A., Ariff, N.R.M., Said, S.A.A.S., Ametefe, D.S., and John, D. Barriers to digital transformation in facilities management: Insights from a developing country. *Journal of Construction in Developing Countries*, 30 (2), 137–168 (2025). <https://doi.org/10.21315/jcdc.2025.30.2.6>
- 35 Kirsten, M., Mthethwa, S., Sitsha, T., & Niemann, W. Advancing sustainable supply chain management through digitalisation: Insights from the South African automotive industry. *Journal of Transport and Supply Chain Management*, 19, Article a1221 (2025). <https://doi.org/10.4102/jtscm.v19i0.1221>
- 36 Jabber, M.A., Islam, M.T., Hossain, T., & Sultana, R. Unveiling the power of enablers in enacting sustainable supply chain management practices. *Cleaner Logistics and Supply Chain*, 13, Article 100190 (2024). <https://doi.org/10.1016/j.clscn.2024.100190>
- 37 Živković, L., & Štrbac, D. Business leadership in the adoption of eco-innovation in manufacturing: Evidence from firm-level microdata. *Business Ethics and Leadership*, 9 (4), 32–44 (2025). [https://doi.org/10.61093/bel.9\(4\).32-44.2025](https://doi.org/10.61093/bel.9(4).32-44.2025)
- 38 Williams, J., Fugar, F., Adinyira, E., & Danso, H. A framework towards health and safety knowledge transfer from the construction industry to the community in developing countries: A case study of Ghana. *International Journal of Construction Management*, 24 (7), 771–783 (2024). <https://doi.org/10.1080/15623599.2023.2214984>
- 39 Chabán, O., Flores-Hernández, A., López-Pérez, M. E., & Ruz-Mendoza, M. A. The manager's vision of CSR in an emerging economy: From implementation to market impact. *Administrative Sciences*, 14 (3), Article 44 (2024). <https://doi.org/10.3390/admsci14030044>
- 40 Dey, S.K., Dey, R.S.K., Hoang, S.D., & Garza-Reyes, J.A. Unveiling circular economy pathways among small and medium hotels through a hybrid analytical framework. *Business Strategy and the Environment*, 34 (8), 9545–9565 (2025). <https://doi.org/10.1002/bse.70085>
- 41 Menéndez, M.D.L.M., & Lavopa, A. Adoption of twin transition technologies in developing countries: A bivariate analysis. *Journal of Evolutionary Economics*, 35(4), 759–799 (2025). <https://doi.org/10.1007/s00191-025-00910-z>

- 42 Kuzey, C., & Uyar, A. Determinants of sustainability reporting and its impact on firm value: Evidence from the emerging market of Turkey. *Journal of Cleaner Production*, 143, 27–39 (2017). <https://doi.org/10.1016/j.jclepro.2016.12.153>
- 43 Osei, M.B., & Ofori, D. Cultural drivers of sustainability performance in emerging economies: Understanding the effect of supply chain learning. *Journal of Business Logistics*, 46 (4), Article e70036 (2025). <https://doi.org/10.1111/jbl.70036>
- 44 Alharethi, T., Abdelhakim, A., & Mohammed, A. Drivers and barriers towards circular economy in rural tourism destinations: A case study of Tunis Village, Egypt. *Tourism and Hospitality*, 5 (3), 639–656 (2024). <https://doi.org/10.3390/tourhosp5030038>
- 45 Hmamed, H., Cherrafi, A., Benghabrit, A., Tiwari, S., & Sharma, P. The adoption of I4.0 technologies for a sustainable and circular supply chain: An industry-based SEM analysis from the textile sector. *Business Strategy and the Environment*, 33 (4), 2949–2968 (2024). <https://doi.org/10.1002/bse.3645>
- 46 Nartey, E. Determinants of carbon management accounting adoption in Ghanaian firms. *Meditari Accountancy Research*, 26 (1), 88–121 (2018). <https://doi.org/10.1108/MEDAR-03-2017-0133>
- 47 Tran, N.H. Impact factors on the adoption of corporate social responsibility: Empirical evidence from an emerging market. *Corporate Governance and Organizational Behavior Review*, 7 (2 Special Issue), 350–359 (2023). <https://doi.org/10.22495/cgobrv7i2sip13>
- 48 Shanta, M.H., Patwary, P.M., Salman, S., Hanifa, S.A., Shil, S., Hasanat, S., Roshid, M.M., & Dhar, B.K. Strategic barriers to circular waste transitions in emerging economies: A fuzzy DEMATEL approach for sustainable industrial development. *Sustainable Futures*, 10, Article 101374 (2025). <https://doi.org/10.1016/j.sft.2025.101374>
- 49 Bernal-Torres, C.A., Torres-Guevara, L.E., Álvarez-Aros, E.L., Barrera-Canto, J.L., & Ojeda-López, R.N. Collaborative work in the implementation of Industry 4.0 technologies and its relationship with business sustainability in an emerging economy. *International Journal of Innovation Studies*, 9 (3), 218–230 (2025). <https://doi.org/10.1016/j.ijis.2025.05.002>
- 50 Dua, J., & Sharma, A. K. Corporate sustainability and capital costs: A panel evidence from BRICS countries. *Indian Journal of Corporate Governance*, 17 (1), 77–101 (2024). <https://doi.org/10.1177/09746862241240368>
- 51 Bittner, N., Bakker, N., & Long, T.B. Circular economy and the hospitality industry: A comparison of the Netherlands and Indonesia. *Journal of Cleaner Production*, 444, Article 141253 (2024). <https://doi.org/10.1016/j.jclepro.2024.141253>
- 52 Rodríguez-Espíndola, O., Cuevas-Romo, A., Chowdhury, S., Díaz-Acevedo, N., Albores, P., Despoudi, S., Malesios, C., & Dey, P. The role of circular economy principles and sustainable-oriented innovation to enhance social, economic and environmental performance: Evidence from Mexican SMEs. *International Journal of Production Economics*, 248, Article 108495 (2022). <https://doi.org/10.1016/j.ijpe.2022.108495>

¹*Исраждинова, Т.С.,

магистрант, ORCID ID: 0009-0005-7203-6058,

*e-mail: t_israzhdinova@kbtu.kz

¹Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

ДАМУШЫ ЕЛДЕРДЕГІ ЖОБАЛАРДЫ БАСҚАРУ ТӘЖІРИБЕСІНЕ ESG ҚАҒИДАТТАРЫН ЕНГІЗУГЕ ЫҚПАЛ ЕТЕТІН ЖӘНЕ КЕДЕРГІ КЕЛТІРЕТІН ФАКТОРЛАР

Аңдатпа

Мақала дамушы елдердегі жобаларды басқару тәжірибесінде Environmental, Social and Governance (ESG) қағидастарын енгізудің негізгі қозғаушы күштері мен кедергілерін қарастырады. Зерттеу Scopus дерекқоры негізінде PRISMA логикасына сәйкес жүргізілген жүйелі әдеби шолуға сүйенеді, ал нәтижелер институционалдық теория тұрғысынан түсіндіріледі. Іздеу барысында 605 жазба анықталып, скрининг пен сәйкестікті бағалаудан кейін 40 рецензияланған зерттеу талдауға енгізілді. Нәтижелер ESG қағидастарын енгізу реттеуші қысымның, ұйымдық дайындықтың, қаржылық мүмкіндіктің, мүдделі тараптарды үйлестірудің және іске асыру әлеуетінің өзара байланысы арқылы қалыптасатынын көрсетеді.

Негізгі драйверлерге қолдаушы реттеу, мүдделі тараптардың қысымы мен қатысуы, оқыту, басшылықтың қолдауы, басқару тетіктері және технологиялық мүмкіндіктер жатады. Негізгі кедергілер ретінде жоғары бастапқы шығындар, қаржыландырудың шектеулі болуы, реттеудің фрагменттелуі, кәсіби құзыреттіліктің жеткіліксіздігі, хабардарлықтың төмендігі, мәдени қарсылық және мүдделі тараптардың әлсіз қатысуы анықталды. Нәтижелер ESG өлшемдері арасында назардың біркелкі бөлінбегенін көрсетеді. Басқару өлшемі ең көп кездеседі – 62,5%. Экологиялық өлшем одан кейін 55,0%. Әлеуметтік өлшем ең аз назарға ие – тек 32,5%. Мақалада ESG енгізу мен жобалық деңгейдегі нақты операциялық тәжірибелер арасындағы тұрақты алшақтық атап өтіледі. Нәтижелер ESG енгізуді институционалдық жағдайлар мен ұйымдық әлеуетке тәуелді процесс ретінде түсінуге үлес қосады, сондай-ақ жоба менеджерлері, ұйымдар және саясаткерлер үшін практикалық ұсынымдар ұсынады.

Түйін сөздер: ESG, дамушы елдер, жобаларды басқару, жүйелі әдеби шолу институционалдық теория, кедергілер, қозғаушы күштер.

¹*Исраждинова, Т.С.,
магистрант, ORCID ID: 0009-0005-7203-6058,
*e-mail: t_israzhdinova@kbtu.kz

¹Казахстанско-Британский технический университет, г. Алматы, Казахстан

ФАКТОРЫ, СПОСОБСТВУЮЩИЕ И ПРЕПЯТСТВУЮЩИЕ ВНЕДРЕНИЮ ESG-ПРИНЦИПОВ В ПРАКТИКЕ УПРАВЛЕНИЯ ПРОЕКТАМИ В РАЗВИВАЮЩИХСЯ СТРАНАХ

Аннотация

В статье рассматриваются движущие силы и барьеры внедрения экологических, социальных принципов и принципов управления (ESG) в практику управления проектами в развивающихся странах. Проведен систематический обзор литературы на основе базы Scopus с использованием логики PRISMA и интерпретацией через институциональную теорию. Поиск выявил 605 записей, из которых после скрининга и оценки приемлемости были включены 40 рецензируемых исследований. Результаты показывают, что внедрение ESG формируется взаимодействием регуляторного давления, организационной готовности, финансовой осуществимости, координации заинтересованных сторон и способности к реализации. Основными драйверами являются поддерживающее регулирование, давление и вовлечение стейкхолдеров, обучение, приверженность руководства, механизмы управления и технологические возможности. Основными барьерами являются высокие первоначальные затраты, ограниченное финансирование, фрагментарное регулирование, недостаточная профессиональная компетентность, низкая осведомленность, культурное сопротивление и слабое взаимодействие с заинтересованными сторонами. Результаты обнаруживают неравномерное распределение внимания между измерениями ESG. Управленческое измерение наиболее заметно – 62,5% проанализированных источников. Экологическое измерение следует за ним с 55,0%. Социальное измерение получает наименьшее внимание – лишь 32,5%. Статья выявляет устойчивый разрыв перевода между принятием ESG и операционной практикой на уровне проектов. Результаты способствуют пониманию внедрения ESG как институционально обусловленного и зависящего от организационных возможностей процесса и содержат выводы для руководителей проектов, организаций и политиков.

Ключевые слова: ESG, барьеры, движущие силы, институциональная теория, развивающиеся страны, систематический обзор литературы, управление проектами.

UDC 658.310.9
IRSTI 06.81.19

<https://doi.org/10.55452/1998-6688-2026-23-2-551-559>

¹***Nurman A.D.,**

HR Officer Central Asia, ORCID ID: 0009-0008-8608-1660,

*e-mail: aidana.nurman@partners.basf.com

²**Tarkhanov P.K.,**

Business Development Manager, ORCID ID: 0009-0001-3243-7405,

e-mail: pavel.tarkhanov@partners.basf.com

¹BASF Central Asia, Almaty, Kazakhstan

²BASF Coatings Turkey, Istanbul, Turkey

TURNING DIVERSITY INTO PERFORMANCE: HOW COHESION DRIVES KAZAKHSTANI PROJECT TEAMS

Abstract

The present paper investigates the potential returns from diversity within project teams in Kazakhstan, a country hardly considered in the body of mainstream organizational studies, as well as through which mechanism this return may be achieved. Using a survey of 103 managers across several industries, we conducted a set of nine hypothesis tests concerning the relationship between cultural diversity and communication, trust, conflict, cohesion, and performance within teams. There exists a significant positive correlation between diversity and performance ($r = .463$, $p < .001$). In contrast to typical relationships found in Western literature, diversity in our sample positively correlates with both trust ($r = .448$) and cohesion ($r = .478$), having no impact whatsoever on conflict levels. All five predictors together account for 63.8% of performance variance, with cohesion being the only significant predictor ($\beta = .673$, $p < .001$), while the direct relationship between diversity and performance loses its significance ($\beta = -.057$, $p > .05$), suggesting full mediation. We suggest that the collectivist nature of Kazakhstan's society and the existence of officially guaranteed ethnic peace and harmony mitigate significantly the adverse effects of the "double-edged sword," and thus cohesion emerges as the main mediator for translating diversity into performance improvement. Of nine hypotheses, five turn out to be supported.

Keywords: cultural diversity, team performance, team cohesion, project management, Kazakhstan, collectivist culture, mediation.

Received May 13, 2026; revised May 28, 2026; accepted June 4, 2026.

Introduction

The issue of managing culturally diverse teams has been a key element of modern project management practices in virtually any knowledge-based industry. The influx of foreign investments into the oil and gas industry, globalization processes in the international banking and IT companies, as well as national-level digitalization strategies have resulted in the emergence of work teams consisting of employees with different cultural, linguistic, and professional backgrounds. However, research on the influence of such diversity on the efficiency of teamwork in Kazakhstan appears to be rather scant. In fact, many empirical insights on diversity-performance relationships originate from North American or Western European contexts [1, 2], and their applicability to the Kazakh situation is not clear at all.

The dominant approach in international research sees cultural diversity as a "double-edged sword," delivering positive results by enhancing understanding and perspectives and creating negative effects through social classification, lack of trust, and increased conflict [1]. Therefore, the overall effect on productivity will be roughly neutral, since the advantages and disadvantages tend

to cancel each other out. It is argued by some researchers that there is no need for these two effects to counteract each other; various moderators including national culture, organizational culture, and group experience may affect the balance between the positive and negative effects [3, 4]. Nevertheless, this idea has remained empirically underexplored until recently.

In terms of addressing the identified research gap, the present study makes a contribution from three different perspectives. Firstly, in theoretical terms, our study demonstrates that the predictions of social categorization theory are context-dependent, do not hold under conditions of collectivism and harmony-based institutions, and consequently questions the universalism of the double-edged sword paradigm. Secondly, from an empirical perspective, our study offers one of the few systematic data analyses pertaining to the performance-diversity relationship within the Central Asian setting, filling a gap in a field where most empirical findings have been generated from Western settings. Thirdly, our study suggests that team cohesion rather than conflict resolution mediates diversity benefits in Kazakhstani teams.

Materials and methods

Theoretical frameworks

There have been two major traditions of theorizing about the phenomenon of diversity that dominate empirical research on diverse teams. First, the information processing tradition asserts that diverse teams outperform homogeneous groups because they share more extensive knowledge bases and apply a wider range of problem solving strategies. [5] This idea seems to be highly consistent with the empirical evidence for the benefits of diversity in situations where creative problem solving skills are needed. Second, according to the social identity and self-categorization theories, people perceive those who seem different from them culturally as members of the out-group, which negatively affects communication, lowers trust, and contributes to conflicts. [6] Finally, the categorization-elaboration model combines the above ideas into one framework, which suggests that diversity promotes information elaboration and social categorization, the balance between which depends on the presence of moderating variables, including team context, task characteristics, and leadership practices. [3]

Team processes as pathways

Processes that help explain how team composition is turned into team outcomes include team communication, trust, conflict, and cohesion [7, 8]. Effective communication allows for the exchange of different knowledge instead of the stockpiling of knowledge. Trust helps in taking the risks required for intercultural cooperation. Conflict related to tasks can be beneficial, while conflict related to relationships always negatively impacts task performance [9]. Cohesion, or attraction and identification of team members with the group, is positively correlated with team performance [10].

In many studies carried out among Western societies, higher levels of diversity are linked with lower levels of trust, weaker cohesion, and higher conflict [11]. Nevertheless, various researchers have warned that these relationships do not hold true universally but are dependent on specific national cultural and institutional circumstances [4]. Modern diversity studies emphasize the importance of looking into issues beyond WEIRD (Western, Educated, Industrialised, Rich, Democratic) societies [12].

Kazakhstan as context

The Kazakhstani society offers an entirely different setting for the discussion on diversity management, because it is made up of over 130 ethnically recognized nations united into one state. Cooperation between members of different nationalities in the workplace was promoted under the influence of Soviet administrative system, while the Assembly of the People of Kazakhstan, together with the corresponding public programs, has established the culture of interethnic harmony at the civic level [13]. Additionally, being collectivist in terms of culture, Kazakhstan places more importance on group cohesion than on individualism, which was usually the focus of previous diversity management researches [14].

The following paper can be complemented by an example from Kazakhstan itself. In their research on intercultural communicative competence among multicultural project teams of Kazakhstani-based multinational enterprises, Atakoziyeva et al. [15] found that positive attitudes to diversity and good communication competencies determine successful performance of the projects.

Research Hypotheses

Based on the literature reviewed above, nine hypotheses were formulated. H3, H4, and H5 follow conventional Western-based predictions from social categorisation theory; they are included precisely to test whether those predictions replicate in Kazakhstan:

H1: Cultural diversity is positively related to team performance.

H2: Cultural diversity is positively related to team communication quality.

H3: Cultural diversity is negatively related to team trust.

H4: Cultural diversity is positively related to team conflict.

H5: Cultural diversity is negatively related to team cohesion.

H6: Communication quality is positively related to team performance.

H7: Trust is positively related to team performance.

H8: Conflict is negatively related to team performance.

H9: Team cohesion is positively related to team performance.

Research design and data collection

A cross-sectional online questionnaire was distributed through Google Forms in March 2026. The sample was selected based on purposive-convenience sampling without any financial remuneration for participating in the study. The inclusion criterion entailed ongoing involvement in a project team consisting of at least three people. Every question on the form was obligatory; therefore, no missing answers emerged.

The tool was offered bilingually – in English and Russian – to capture the reality of a bilingual professional environment in Kazakhstan. In total, 103 valid responses were collected among professionals working in different industries: IT (18.4%), education (18.4%), chemistry and production (15.5%), finance and banking (11.7%), land cadaster and governmental services (15.5%), oil and gas (2.9%), and other areas (17.5%). As per gender, there were more women (54.4%) compared to men (45.6%). In terms of age, respondents were young (64.1% were aged 18–25, 16.5% – aged 26–35, and 15.5% – aged 36–45). As for group size, 3–5 people comprised 35.0%, 6–10 people – 28.2%, 11–20 people – 10.7%, and over 20 people – 26.2%.

One of the limitations inherent to the research is that it used non-probability sampling design; hence, the results will probably be representative of well-educated and urban individuals.

Measures and reliability

Measurement items for all constructs utilized five-point Likert scales. The construct of cultural diversity was measured through five items dealing with age, educational background, native language, work experience, and personal values drawn from existing measures [16, 17]. Communication quality was measured through five items on openness, information exchange, listening, accommodating language usage, and cultural sensitivity [17]. The concept of trust was operationalized through five items including reliability, honesty, psychological safety, respect, and competence [18, 19]. Three items measuring conflict frequency related to disagreements at work, personal problems, and cross-cultural issues (high scores indicated greater conflict frequency) were included [20]. The construct of cohesion was operationalized through five items related to collaboration, goal commitment, team identification, supportiveness, and teamwork [21]. Performance of the work group was measured by seven items assessing timeliness, quality, goals attainment, problem solving, innovation, stakeholder satisfaction, and continuous improvement [22, 23].

The reliability of measurement scales was estimated using Cronbach's alpha coefficient. All six coefficients exceeded .70, which is considered an acceptable level [24], while their minimum value (.837) and maximum value (.951) corresponded to conflict and performance, respectively.

Data analysis

Statistical analysis was carried out in Jamovi 2.6 [25]. We first examined Pearson bivariate correlations to test H1–H9. We then ran simple linear regression (diversity predicting performance only) and multiple regression entering all five predictors simultaneously. We report both unstandardised (B) and standardised (β) coefficients, as well as variance inflation factors (VIF) to assess multicollinearity. Basic regression assumptions were inspected via Q–Q plots and residual plots; no serious violations were detected.

Results and discussion

Descriptive statistics and reliability

Table 1 reports means, standard deviations, reliability coefficients, and bivariate correlations for all six study variables. Respondents perceived their teams as moderately diverse ($M = 3.58$, $SD = 1.11$). Trust was the highest-rated construct ($M = 4.08$), while conflict was the lowest ($M = 2.30$), suggesting generally positive team climates with low interpersonal friction. Cohesion ($M = 3.95$) and performance ($M = 3.90$) were both rated above the scale midpoint.

Table 1 – Descriptive statistics, reliability coefficients, and correlations

Variable	M	SD	α	1	2	3	4	5	6
1. Diversity	3.58	1.11	.878	–					
2. Communication	3.96	0.91	.849	.492*	–				
3. Trust	4.08	0.94	.926	.448*	.771*	–			
4. Conflict	2.30	1.04	.837	.047	–.115	–.098	–		
5. Cohesion	3.95	0.91	.926	.478*	.745*	.770*	–.127	–	
6. Performance	3.90	0.90	.951	.463*	.595*	.660*	.018	.780*	–

Note. $N = 103$. * $p < .001$ (two-tailed).

Bivariate correlations and hypothesis testing

Diversity correlates positively and significantly with performance ($r = .463$, $p < .001$), supporting H1, and with communication quality ($r = .492$, $p < .001$), supporting H2. Contrary to H3 and H5, diversity correlates positively with trust ($r = .448$, $p < .001$) and with cohesion ($r = .478$, $p < .001$). The diversity-conflict correlation is negligible and non-significant ($r = .047$, $p = .635$), failing to support H4.

Among the process-to-performance paths, cohesion shows the strongest bivariate association with performance ($r = .780$), followed by trust ($r = .660$) and communication ($r = .595$), supporting H9, H7, and H6 respectively. Conflict is essentially unrelated to performance ($r = .018$, $p = .858$), so H8 is not supported.

Regression analysis

In simple regression, diversity alone accounts for 21.4% of performance variance ($R^2 = .214$, $B = 0.378$, $SE = .072$, $p < .001$). Adding the four process variables in multiple regression raises the explained variance to 63.8% (Table 2). The only significant predictor in the full model is cohesion ($B = 0.669$, $\beta = .673$, $p < .001$). The coefficient for diversity drops to $B = 0.080$ and becomes non-significant ($p = .181$) – a pattern consistent with full mediation through cohesion.

VIF values range from 1.12 to 3.82, well below the conventional threshold of 10, confirming that multicollinearity is not a serious concern. The non-significant negative coefficient for communication in the full model reflects shared variance among the three correlated positive process variables, not a genuine negative effect.

Table 2 – Multiple regression predicting team performance (DV)

Predictor	B	SE	β	t	p	VIF
Intercept	0.386	0.311	–	1.24	.217	–
Diversity	0.080	0.059	.066	1.35	.181	1.59
Communication	–0.065	0.105	–.066	–0.62	.537	3.81
Trust	0.155	0.104	.161	1.48	.142	3.82
Conflict	0.092	0.054	.106	1.71	.090	1.12
Cohesion	0.669	0.104	.673	6.41	< .001	3.64

Note. N = 103. β = standardised coefficient. VIF = variance inflation factor. $R^2 = .638$, ΔR^2 vs simple regression = .424.

Table 3 – Summary of hypothesis testing

H#	Relationship	Direction predicted	r	p	Outcome
H1	Diversity → Performance	+	.463	< .001	Supported
H2	Diversity → Communication	+	.492	< .001	Supported
H3	Diversity → Trust	–	.448	< .001	Not supported (opposite)
H4	Diversity → Conflict	+	.047	.635	Not supported (n.s.)
H5	Diversity → Cohesion	–	.478	< .001	Not supported (opposite)
H6	Communication → Performance	+	.595	< .001	Supported
H7	Trust → Performance	+	.660	< .001	Supported
H8	Conflict → Performance	–	.018	.858	Not supported (n.s.)
H9	Cohesion → Performance	+	.780	< .001	Supported (strongest)

Note. n.s. = not significant. H3 and H5 show significant relationships in the direction opposite to the hypothesis. Although H1 is supported in bivariate correlation, the multivariate model shows diversity's effect operates primarily through cohesion rather than independently.

Interpreting the positive diversity effects

The association between diversity and effectiveness (H1 supported, $r = .463$), consistent with the information-processing approach [5], suggests that teams with diverse backgrounds and experiences seem to be better positioned to deal with the complex tasks inherent to project work in Kazakhstan's knowledge economy. Moderately large effect size implies statistical and practical significance.

More importantly from the theoretical point of view, it can be stated that there is no statistically significant evidence for the null and reversed hypotheses, which would be consistent with the predictions of the social categorization theory [6] and the dominant paradigm of Western empirical studies on diversity [11]. Such results imply that, for some reason, social categorization mechanisms in Kazakhstan's context are substantially weakened. In turn, there are several possible reasons behind such results.

First of all, the administrative heritage of the Soviet period led to the widespread practice of inter-ethnic cooperation in work-related activities; therefore, the “other” person within a project team does not necessarily activate out-group biases [13]. Second, official discourse of nation-building, according to which ethnic diversity is regarded as a national advantage, creates a favorable institutional framework for non-categorical relations in project teams (Assembly of Peoples of Kazakhstan and the state program “Kazakhstan – Our Common Home”). Finally, a collectivist orientation of Kazakhstan society, which emphasizes importance of cohesive interaction in contrast to individualist societies in which the double-edged sword effect has been discovered [14].

The above analysis is taken to suggest attenuation rather than total absence of the negative blade of the double-edged sword. At the same time, one cannot exclude the need for further testing the hypotheses on larger and more representative samples.

Cohesion as the mediating mechanism

Cohesion emerges as the only variable with a significant effect in the whole model, acting as the mediator in the relationship between diversity and performance ($\beta = .673$, $p < .001$). The introduction of cohesion results in the disappearance of any direct effect diversity may have had, thus supporting a scenario of full mediation. It should be noted that since formal bootstrapping tests for mediation could not be conducted, owing to the small sample of 103 participants – below what is considered acceptable for SEM analysis [26] – convergent evidence is provided by both analyses.

If diversity rarely leads to friction within an organization, cohesion should be promoted to manage the process effectively. This has particular importance for project managers in Kazakhstan, where diverse ethnic backgrounds characterize the working environment.

Practical implications

From a project manager's perspective and that of an HR manager in Kazakhstan, the main conclusion to draw from the data presented above is that cultural diversity should be treated as a resource for the team. It is evident from the findings above that there is no need to spend money on training programs to prevent conflicts arising due to diversity. Instead, efforts should be made to increase cohesion by creating a common goal, reciprocal acknowledgment between people from different cultures, and interdependence among them [12].

Conclusion

The present research aims at examining the existence of two faces of cultural diversity's impact on organizational performance, especially in Kazakhstan, and proves that the negative one is virtually nonexistent. Diversity leads to improved performance, more trust, better cohesion, and good communication without triggering any conflict. In multiple regression analysis, 63.8 percent of performance variance is explained by such predictor variables as cohesion, which plays the most important role and serves as the channel of diversity's influence on performance.

Three contributions can be made by this paper. First, theoretically, the adverse expectations based on social categorization theory seem to be true only for specific environments, such as collectivistic countries with institutions aimed at maintaining harmony. Second, empirically, this is the first quantitative study of diversity-performance relationship in Kazakhstani project teams, enriching a body of literature, which tends to be WEIRD-focused [12]. Finally, practically, the present findings imply a shift from trying to avoid conflict to striving for achieving cohesion.

Several limitations constrain the conclusions. The cross-sectional design precludes causal inference; reverse causality cannot be ruled out. Convenience sampling means the respondent pool over-represents younger, educated, urban professionals; findings may not generalise to more experienced or sector-specific populations. All measures are self-reported, introducing potential common-method variance; future work would benefit from multi-source data and objective performance metrics. The sample size ($N = 103$) precluded structural equation modelling [26]; studies with $N \geq 200$ could formally test the mediation. Comparative studies across Central Asian countries would help establish how much of what we found is specific to Kazakhstan versus the broader region.

REFERENCES

- 1 Stahl, G.K., Maznevski, M.L., Voigt, A., Jonsen, K. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41 (4), 690–709 (2010). <https://doi.org/10.1057/jibs.2009.85>
- 2 Joshi, A., Roh, H. The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, 52 (3), 599–627 (2009). <https://doi.org/10.5465/amj.2009.41331491>
- 3 van Knippenberg, D., De Dreu, C.K.W., Homan, A.C. Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, 89 (6), 1008–1022 (2004). <https://doi.org/10.1037/0021-9010.89.6.1008>

- 4 Kirkman, B.L., Tesluk, P.E., Rosen, B. The impact of demographic heterogeneity and team leader-team member demographic fit on team empowerment and effectiveness. *Group & Organization Management*, 29 (3), 334–368 (2004). <https://doi.org/10.1177/1059601103257412>
- 5 Cox, T.H., Blake, S. Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Executive*, 5 (3), 45–56 (1991). <https://doi.org/10.5465/ame.1991.4274465>
- 6 Tajfel, H., Turner, J.C. The social identity theory of intergroup behavior. *Psychology of intergroup relations* / ed. by S. Worchel, W.G. Austin. Chicago: Nelson-Hall, 1986. P. 7–24. <https://doi.org/10.4324/9781315127767>
- 7 Mesmer-Magnus, J.R., DeChurch, L.A. Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94 (2), 535–546 (2009). <https://doi.org/10.1037/a0013773>
- 8 Williams, K.Y., O'Reilly, C.A. Demography and diversity in organizations: A review of 40 years of research // *Research in Organizational Behavior*, 20, 77–140 (1998). [https://doi.org/10.1016/S0191-3085\(98\)20002-7](https://doi.org/10.1016/S0191-3085(98)20002-7)
- 9 De Dreu, C.K.W., Weingart, L.R. Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, 88 (4), 741–749 (2003). <https://doi.org/10.1037/0021-9010.88.4.741>
- 10 Beal, D.J., Cohen, R.R., Burke, M.J., McLendon, C.L. Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88 (6), 989–1004 (2003). <https://doi.org/10.1037/0021-9010.88.6.989>
- 11 Harrison, D.A., Price, K.H., Bell, M.P. Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of Management Journal*, 41 (1), 96–107 (1998). <https://doi.org/10.2307/256901>
- 12 Jonsen, K., Maznevski, M.L., Schneider, S.C. Gender differences in leadership – believing is seeing: Implications for managing diversity. *Equality, Diversity and Inclusion*, 29 (6), 549–572 (2010). <https://doi.org/10.1108/02610151011067504>
- 13 Dave, B. Kazakhstan: Ethnicity, language and power. London: Routledge, 2007. 238 p. <https://doi.org/10.4324/9780203961575>
- 14 Hofstede, G. Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. 2nd ed. Thousand Oaks: Sage, 2001. 596 p. <https://doi.org/10.1177/0170840603024002082>
- 15 Atakoziyeva, Y.Y., Moldakhmetova, S.K., Khonai, A., Charman, K.P. Using intercultural communicative competence dimensions in managing multicultural team performance. *Herald of the Kazakh-British Technical University*, 1 (68), 169–182 (2024). <https://doi.org/10.55452/1998-6688-2024-21-1-169-182>
- 16 Pelled, L.H., Eisenhardt, K.M., Xin, K.R. Exploring the black box: An analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly*, 44 (1), 1–28 (1999). <https://doi.org/10.2307/2667029>
- 17 O'Reilly, C.A., Roberts, K.H. Task group structure, communication, and effectiveness in three organizations. *Journal of Applied Psychology*, 62 (6), 674–681 (1977). <https://doi.org/10.1037/0021-9010.62.6.674>
- 18 McAllister, D.J. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38 (1), 24–59 (1995.). <https://doi.org/10.2307/256727>
- 19 Costa, A.C., Roe, R.A., Taillieu, T. Trust within teams: The relation with performance effectiveness. *European Journal of Work and Organizational Psychology*, 10 (3), 225–244 (2001.). <https://doi.org/10.1080/13594320143000654>
- 20 Jehn, K.A. A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 40 (2), 256–282 (1995). <https://doi.org/10.2307/2393638>
- 21 Carless, S.A., De Paola, C. The measurement of cohesion in work teams. *Small Group Research*, 31 (1), 71–88 (2000). <https://doi.org/10.1177/104649640003100104>
- 22 Hoegl, M., Gemuenden, H.G. Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science*, 12 (4), 435–449 (2001). <https://doi.org/10.1287/orsc.12.4.435.10635>
- 23 Ancona, D.G., Caldwell, D.F. Demography and design: Predictors of new product team performance. *Organization Science*, 3 (3), 321–341 (1992). <https://doi.org/10.1287/orsc.3.3.321>
- 24 Nunnally, J.C. Psychometric theory. 2nd ed. New York: McGraw-Hill, 1978. 701 p.
- 25 The jamovi project. jamovi (Version 2.6) [Computer software]. 2024. URL: <https://www.jamovi.org>
- 26 Hair J.F., Black W.C., Babin B.J., Anderson R.E. Multivariate data analysis. 8th ed. Andover: Cengage, 2019. 813 p. <https://doi.org/10.1002/9781119409137>

¹*Нұрман А.Д.,

Орталық Азия бойынша қызметкерлер жөніндегі маман,

ORCID ID: 0009-0008-8608-1660,

*e-mail: aidana.nurman@partners.basf.com

²Тарханов П.К.,

бизнесі дамыту жөніндегі менеджер,

ORCID ID: 0009-0001-3243-7405,

e-mail: pavel.tarkhanov@partners.basf.com

¹BASF Central Asia, Алматы қ., Қазақстан²BASF Coatings Turkey, Ыстанбұл қ., Түркия

АЛУАНТҮРЛІЛІКТІ ТИІМДІЛІККЕ АЙНАЛДЫРУ: ҚАЗАҚСТАНДЫҚ ЖОБАЛЫҚ КОМАНДАЛАРДА ЫНТЫМАҚТАСТЫҚ ҚАНДАЙ НӘТИЖЕГЕ ӘКЕЛЕДІ

Аңдатпа

Зерттеу Қазақстандағы жобалық командалардың тиімділігіне мәдени алуантүрліліктің оң ықпал ететін-етпейтінін және бұл әсердің қандай механизмдер арқылы жүзеге асатынын анықтауға бағытталған. Ұйымдастырушылық зерттеулерде сирек қарастырылатын осы мәселені зерттеу мақсатында әртүрлі салалардағы 103 маманнан сауалнама деректері жиналып, мәдени алуантүрліліктің командалық процестерге (коммуникация, сенім, жанжал және ынтымақтастық) және команданың тиімділігіне әсеріне қатысты тоғыз гипотеза тексерілді. Нәтижелер мәдени алуантүрліліктің командалық тиімділікпен оң байланыста екенін растады ($r = .463$, $p < .001$). Батыс елдерінде жүргізілген зерттеулердің нәтижелерінен айырмашылығы, мәдени алуантүрлілік сеніммен ($r = .448$) және ынтымақтастықпен ($r = .478$) оң корреляция көрсетті, ал жанжалмен статистикалық тұрғыдан мәнді байланыс анықталмады. Регрессиялық талдау нәтижесінде ынтымақтастық командалық тиімділіктің жалғыз маңызды болжаушысы болып шықты ($\beta = .673$, $p < .001$), ал толық модель тиімділік дисперсиясының 63,8%-ын түсіндірді. Сонымен қатар, ынтымақтастық айнымалысы модельге енгізілген кезде мәдени алуантүрліліктің тиімділікке тікелей әсері статистикалық мәнін жоғалтты, бұл толық медиацияның бар екенін көрсетеді. Авторлар Қазақстанның ұжымдық мәдениеті мен этносаралық келісімді мемлекеттік деңгейде институционализациялау мәдени алуантүрліліктің «екі жүзді семсер» тұжырымдамасында сипатталатын теріс әсерлерін айтарлықтай әлсіретеді деп болжайды. Зерттеудің практикалық қорытындысы бойынша, мұндай жағдайда ұйымдар үшін жанжалды басқаруға қарағанда, командалардағы ынтымақтастықты нығайтуға инвестиция салу анағұрлым тиімді.

Түйін сөздер: мәдени алуантүрлілік, топтың тиімділігі, ынтымақтастық, жоба менеджменті, Қазақстан, ұжымдық мәдениет, медиация.

¹*Нұрман А.Д.,

специалист по персоналу в Центральной Азии,

ORCID ID: 0009-0008-8608-1660

*e-mail: aidana.nurman@partners.basf.com

²Тарханов П.К.,

менеджер по развитию бизнеса,

ORCID ID: 0009-0001-3243-7405,

e-mail: pavel.tarkhanov@partners.basf.com

¹BASF Central Asia, г. Алматы, Казахстан²BASF Coatings Turkey, г. Стамбул, Турция

ПРЕВРАЩЕНИЕ РАЗНООБРАЗИЯ В РЕЗУЛЬТАТ: КАК СПЛОЧЕННОСТЬ ОПРЕДЕЛЯЕТ ЭФФЕКТИВНОСТЬ КАЗАХСТАНСКИХ ПРОЕКТНЫХ КОМАНД

Аннотация

В данной статье исследуется вопрос о том, способствует ли культурное разнообразие эффективности проектных команд в Казахстане – контексте, редко представленном в международных организационных

исследованиях, – и какой механизм при этом задействован. На основе данных опроса 103 специалистов из нескольких отраслей проверены девять гипотез о связях между разнообразием, командными процессами (коммуникация, доверие, конфликт, сплоченность) и результативностью. Результаты показывают значимую положительную связь культурного разнообразия с эффективностью ($r = .463$, $p < .001$). В отличие от паттернов, характерных для западных выборок, разнообразие положительно коррелирует как с доверием ($r = .448$), так и со сплоченностью ($r = .478$) и не имеет значимой связи с конфликтом. В модели множественной регрессии единственным значимым предиктором оказалась сплоченность ($\beta = .673$, $p < .001$); полная модель объясняет 63,8% дисперсии эффективности. Прямой эффект разнообразия при включении сплоченности в модель становится незначимым, что согласуется с полной медиацией. Мы полагаем, что коллективистская культура Казахстана и институционализированная межэтническая гармония существенно нейтрализуют негативное «лезвие» метафоры двустороннего меча разнообразия. Практический вывод: в данном контексте инвестиции в формирование сплоченности команды, а не в управление конфликтами, являются более эффективным инструментом использования потенциала разнообразия.

Ключевые слова: культурное разнообразие, эффективность команды, сплоченность, управление проектами, Казахстан, коллективистская культура, медиация.

UDC 336.7:316.346.32
IRSTI 06.52.01

<https://doi.org/10.55452/1998-6688-2026-23-2-560-571>

¹**Ahmed Sait S.,**

Assistant Professor, ORCID ID: 0009-0007-4844-9332,

e-mail: sofiamaqbool@ymail.com

^{2*}**Vijesh P.V.,**

PhD, ORCID ID: 0000-0002-1503-0806,

e-mail: vijesh@rajagiri.edu

¹Department of Commerce, Loyola Academy Degree and PG College
(Osmania University), Hyderabad, India

²Rajagiri College of Social Sciences, Kalamassery, Kerala, India

NEXT-GEN SUSTAINABILITY: UNVEILING THE FINANCIAL INTENTIONS, CHALLENGES, AND TRANSFORMATIVE POWER OF GENERATION Z INVESTORS

Abstract

Sustainable finance (SF) has emerged as a pivotal tool in addressing global and societal challenges by integrating economic, social, and governance factors into investment and decision-making processes. In Hyderabad, a rapidly growing urban hub with a significant youth population, Generation Z investors are increasingly influencing the sustainable finance landscape. This study explores the awareness, motivations, and investment behaviours of Gen Z in Hyderabad, with a particular focus on their engagement through social media platforms. Using Structural Equation Modelling (SEM) via SmartPLS, primary data were collected from 152 Generation Z respondents through a structured survey. The study finds that sustainable finance attitude is the strongest predictor of investment behaviour ($\beta = 0.5583$, $p = 0.0092$), followed by technology awareness ($\beta = 0.4125$, $p = 0.0486$), while financial literacy alone is not a statistically significant predictor ($\beta = 0.3087$, $p = 0.1253$). These findings challenge the conventional assumption that financial literacy drives sustainable investment participation, positioning technology accessibility and values alignment as more powerful behavioural enablers. The research highlights the transformative potential of Generation Z in reshaping Hyderabad into a model city for sustainable finance in India, offering actionable insights for policymakers, financial institutions, and educators.

Keywords: Generation Z; Sustainable Finance; ESG Investing; SmartPLS; Technology Awareness; Hyderabad; Green Investments

Received March 11, 2026; accepted June 4, 2026

Introduction

The European Commission defines sustainable finance as the integration of environmental, social, and governance (ESG) concerns into investment decisions, with the goal of making long-term investments in sustainable economic activities and projects [1]. This approach has evolved over decades, with the introduction of responsible investment and, most recently, the broader term sustainable finance [2, 3, 4, 5, 6, 7]. The increasing focus on SF has had a profound impact on financial markets, affecting not just investments but also the regulatory and legislative environment as a whole [8, 9, 10, 11]. Green finance, climate finance, and blue finance are associated terms that describe SF applied to specific environmental or social objectives including biodiversity conservation, climate change mitigation, and sustainable ocean economies [12, 13, 14, 15, 16, 17, 18].

Generation Z, born between the mid-1990s and the 2010s, is increasingly influencing the sustainable finance landscape. Gen Z investors are attracted to businesses that share their values, focus on social impact and sustainability, and respond to environmental changes in financial markets [19, 20, 21, 22, 23]. Sait and Vijesh [24] demonstrate that macroeconomic and industrial indicators such as GDP, IIP, and the Green Exchange Index significantly influence carbon credit dynamics in India, providing a macro-level foundation for examining how Gen Z can leverage these factors to advance sustainable investment practices. Gen Z's familiarity with digital tools and social media has increased their capacity to demand openness from corporations and financial institutions [25, 26].

Despite growing scholarly interest in Gen Z and sustainable finance globally, the majority of existing studies originate from Western economies or adopt pan-Indian approaches that obscure city-level dynamics [27, 28]. Hyderabad presents a theoretically compelling research site for four reasons. First, it is one of India's leading technology hubs, home to HITEC City, Cyberabad, and over 1,500 technology startups [29], creating a Gen Z population that is disproportionately digitally literate. Second, despite this digital exposure, formal financial literacy infrastructure targeted at young urban investors remains underdeveloped, creating a knowledge-action gap. Third, the macroeconomic perspective has received attention in the Indian context, but individual investment choices of young urban residents have not been addressed [24]. Fourth, no published study has investigated how Hyderabad's strong fintech penetration, sizeable Gen Z population, and rapidly changing startup-driven economy affect sustainable investment intentions and behaviours.

This study is distinguished from prior literature by two elements. Although existing studies concentrate on independent variables in isolation, this study uses SEM via SmartPLS to examine SF attitude, financial literacy, and technology awareness as combined predictors of investment behaviour, providing a comprehensive perspective. In contrast to studies that focus on intention rather than action and use non-Indian samples, this study specifically examines investment behaviour among Gen Z in Hyderabad and analyses differences between groups with varying educational backgrounds. The contributions of this research are twofold: first, it offers empirical findings on the sustainable finance behaviour of Gen Z from an Indian technology cluster; second, it integrates the Technology Acceptance Model (TAM), the Stimulus-Organism-Response (S-O-R) framework, and the Theory of Planned Behaviour (TPB) within one structural model.

The literature on Generation Z and sustainable finance spans five intersecting domains. The following synthesis builds a cumulative theoretical argument leading to the specific research gap this study addresses.

Financial Literacy and Sustainable Finance

Financial literacy constitutes the cognitive foundation for informed investment decisions, yet a consistent finding across the literature is that literacy is necessary but insufficient to drive sustainable investment action. Akibun et al. [30] highlighted that despite Gen Z's intensive use of digital innovations, a large number lack the essential financial decision-making capabilities, confirming that digital fluency and financial literacy are distinct competencies. The study established that financial literacy and digital fluency are not interchangeable – possessing one does not ensure the other. Abdullah et al. [31] extended this finding to digital banking, demonstrating that limited financial literacy created major hurdles in efficient financial management even among highly technology-engaged Gen Z users. Together, these studies establish a foundational paradox: Gen Z inhabits a technologically rich environment but remains cognitively under-equipped to extract its full financial value – a tension particularly acute in Hyderabad, where digital exposure is high but structured financial education for young investors remains limited [29].

Sustainable Investment Motivation and Entrepreneurial Orientation

Sustainable investing is increasingly viewed by Generation Z as a reflection of social duty, ethical principles, and personal identity rather than merely a financial endeavour [20]. Aini [32] analysed the motivational drivers influencing Generation Z's inclination toward sustainable business ventures, finding that Gen Z seeks to align economic activities with ethical responsibility, environmental

stewardship, and social impact – framing sustainable entrepreneurship as identity formation in addition to an economic venture. Filippini et al. [27] examined sustainable financial literacy and identified significant knowledge gaps, behavioural obstacles, and institutional barriers that impede the adoption of long-term financial instruments, demonstrating the complex interactions among product accessibility, financial literacy, and trust in financial systems. Judijanto et al. [20] found that Gen Z has a high propensity for socially and ecologically conscious investing, but perceived financial literacy and risk management skills moderate actual investment habits. Ramesh [28] corroborated this in the German context, finding ESG measures crucial in directing investment choices with financial literacy and risk tolerance playing moderating roles – a pattern this study tests in Hyderabad.

Green Investment Accessibility, Structural Barriers, and the Platform Gap

Ravani et al. [33] found that while Gen Z showed high interest in long-term green financial products, regulatory limits, platform usability issues, and low exposure prevented active engagement. Malzara et al. [34] found that high entry barriers and knowledge gaps are the two largest obstacles to wider involvement, while fintech solutions offer significant opportunities to engage Gen Z in sustainable investing through accessible and technologically advanced platforms. Saliha and Dinesh [35] found that socioeconomic conditions, cultural shifts, and digital platforms jointly shape Gen Z's sustainable purchasing patterns – suggesting that interventions cannot adopt a uniform approach across diverse urban populations.

Social Media as Enabler and Misinformation Channel

Sethuraman et al. [36] found that Instagram and TikTok were the most effective channels for spreading sustainable finance awareness among Gen Z, but that false information was a major barrier, as content lacked veracity and authenticity. Ghouse et al. [37] found that digital media and environmental education jointly shape green consumer behaviour, with Gen Z demonstrating a stronger commitment to ethical sourcing and brand transparency than Gen Y.

Long-Term Financial System Alignment

Gutterman [38] emphasised that sustainable entrepreneurship requires structural financial alignment through mixed finance – combining public, private, and philanthropic capital – alongside policy stability and intergenerational capital stewardship, framing individual literacy interventions as necessary but insufficient without enabling systemic conditions.

Research Gap

The foregoing review reveals that Gen Z's interest in sustainable finance is high, but practical participation is constrained by intersecting knowledge gaps, platform accessibility barriers, social media misinformation, and structural market limitations. Critically, no study has simultaneously examined these dynamics in the Hyderabad urban context using a multivariate analytical framework. This gap – a city-specific, Gen Z-focused empirical examination of the drivers and moderators of sustainable finance investment behaviour in an emerging Indian metropolitan market – is what the present study addresses.

Theoretical Framework

This study is grounded in three complementary theoretical frameworks that together explain the pathways from awareness to sustainable investment behaviour.

Technology Acceptance Model (TAM)

The Technology Acceptance Model, developed by Davis [39], posits that a person's intention to use a technology is determined by perceived usefulness and perceived ease of use. Applied to this study, TAM explains how Gen Z's knowledge of and experience with digital financial platforms – including fintech applications, ESG investment portals, and blockchain-based instruments – affects their willingness to invest in sustainable finance. TAM underpins the hypothesis that technology awareness is a more powerful predictor of SF participation than financial literacy, because platform accessibility reduces perceived effort and increases perceived utility.

Stimulus-Organism-Response (S-O-R) Framework

The S-O-R framework [40] explains how external stimuli – sustainable finance information from social media, educational institutions, startups, and government policies – interact with internal cognitive and attitudinal states (the organism) to produce behavioural responses. In this study, the organism represents Gen Z's SF attitude and financial literacy, while the response is actual SF investment behaviour. The S-O-R framework explains why SF attitude, as an internalised organism-level construct, is a stronger predictor of investment behaviour than external knowledge inputs (financial literacy) alone.

Theory of Planned Behaviour (TPB)

Ajzen's [41] TPB posits that behaviour is determined by attitude toward the behaviour, subjective norms, and perceived behavioural control. In this study, TPB explains why Gen Z does not invest in sustainable finance even when aware of it: positive attitude and social norms may be present, but perceived complexity of access (low behavioural control) prevents action. TPB underpins the multiple group analysis testing whether finance students – with higher perceived control – exhibit significantly different SF investment behaviour than non-finance students.

Integration of Frameworks

The three frameworks function as a layered model: S-O-R explains how external stimuli shape Gen Z's internal cognitive and attitudinal states; TAM explains how technological perceptions translate into investment platform engagement; and TPB explains how attitudes, norms, and perceived control jointly determine whether engagement converts into actual investment behaviour. This layered structure provides the theoretical basis for the study's three hypotheses.

Materials and methods

Data Collection

This study focuses on Generation Z individuals aged 18–30, specifically university students and young professionals residing in Hyderabad, India. A stratified sampling technique was employed, with strata defined along two criteria: academic level (undergraduate, graduate, postgraduate, and above/others) and prior investment experience (yes or no). Within each stratum, random selection was applied to minimise bias. Respondents were reached through WhatsApp groups and in-class distribution of the survey link. A proportionate stratified approach was followed, with the final sample of 152 respondents reflecting the natural distribution across strata. Data were collected through a structured Google Forms questionnaire.

The questionnaire assessed six domains using a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree): demographics; awareness of sustainable finance; investment behaviour and financial literacy; influence of social media and external resources; challenges and barriers to participation; and future outlook and perception of the financial industry. Items were adapted from prior research on Gen Z and green finance [34, 37].

Methods

Survey responses were analysed using Structural Equation Modelling (SEM) via SmartPLS (Partial Least Squares). Path coefficient analysis evaluated the strength and direction of relationships between latent constructs. Multiple Group Analysis (MGA) assessed whether model relationships varied across academic background subgroups (finance versus non-finance students). Moderation analysis using SmartPLS bootstrapping examined interaction effects. Regression analysis through path modelling assessed how independent variables influenced sustainable investment intentions. Data preparation was carried out in Microsoft Excel, while all path modelling, bootstrapping, multigroup, and moderation analyses were conducted in SmartPLS.

Results and discussion

Demographic Profile

Table 1 – Age-wise Distribution of Respondents

Age Group	Count	Percentage
18–22	119	78.3%
23–26	19	12.5%
27–30	14	9.2%
Total	152	100%

Source: Primary data, n = 152.

The age distribution confirms a strong representation of early-career Gen Z (78.3% aged 18–22), consistent with the study’s theoretical expectation that SF attitude rather than formal financial literacy will dominate investment behaviour, as this cohort is in an identity-formation phase where values-driven decision-making typically precedes analytical financial evaluation.

Table 2 – Education Level Distribution of Respondents

Education Level	Count	Percentage
Undergraduate	107	70.4%
Postgraduate	16	10.5%
Graduate	13	8.6%
Above and Others	16	10.5%
Total	152	100%

Source: Primary data, n = 152.

The predominance of undergraduates (70.4%) confirms the sample is captured at the knowledge-formation stage of financial literacy development, before significant investment experience has accumulated. This demographic profile strengthens the relevance of the TAM-based hypothesis that perceived platform utility – rather than accumulated financial knowledge – will drive sustainable investment engagement.

Objective 1: Green Bonds, Social Impact Bonds, and ESG-Oriented Funds

H_0 : There is no significant difference between finance and non-finance Gen Z students in their awareness and behavioural response towards green bonds, social impact bonds, and ESG-oriented funds.

H_1 : Awareness of green bonds, social impact bonds, and ESG funds significantly differs in its influence on investment behaviour between finance and non-finance Gen Z students.

Table 3 – Multiple Group Analysis – Finance vs. Non-Finance Students

Relationship	Path Coefficient (Finance)	Path Coefficient (Non-Finance)	p-value (MGA)	Significance
Attitude towards SF → Challenges and Barriers	-0.2301	-0.1954	0.152	Not Significant
Attitude towards SF → Future Outlook	-0.3105	-0.2598	0.112	Not Significant
Attitude towards SF → Influence and Resources	0.2583	0.2109	0.184	Not Significant
Attitude towards SF → Investment Behaviour	-0.6021	-0.5054	0.095	Marginally Significant

Source: Primary data. MGA conducted via SmartPLS, 5,000 bootstrap samples.

The MGA results show no statistically significant differences between finance and non-finance students across most structural paths, confirming that Gen Z students' attitudes toward SF tools are not significantly affected by their academic background in finance. Both groups face similar barriers including limited awareness, restricted access, and lack of formal financial literacy. The only exception is the pathway for investment behaviour ($p = 0.095$, marginally significant), where finance students show slightly greater caution toward conventional investments when sustainable alternatives are considered – consistent with their training raising awareness of financial trade-offs rather than generating enthusiasm for SF instruments per se.

Importantly, the MGA tests whether structural relationships differ between groups – it does not directly test whether awareness of ESG instruments causes investment behaviour. The direct effect of awareness on behaviour is assessed in the regression analysis (Objective 3). On the basis of MGA results, H_0 is accepted: awareness of green bonds, social impact bonds, and ESG funds does not produce significantly different investment decision patterns across Gen Z subgroups defined by financial education background – consistent with Judijanto et al. [20], who similarly found perceived financial literacy moderated but did not independently drive sustainable investment behaviour.

Objective 2: Technology, Digital Platforms, and Sustainable Finance Participation

H_0 : The use of digital platforms, fintech, and blockchain does not significantly impact Gen Z's participation in sustainable finance.

H_1 : The use of digital platforms, fintech, and blockchain significantly enhances Gen Z's participation in sustainable finance.

Table 4 – Moderation Analysis – SmartPLS Bootstrapping

Relationship	Interaction Effect (β)	T-Statistic	P-Value	Significance
Financial Literacy $\times$ SF Attitude $\rightarrow$ Investment Behaviour	0.1524	1.092	0.124	Not Significant
Technology Awareness $\times$ SF Attitude $\rightarrow$ Investment Behaviour	0.2058	1.328	0.097	Marginally Significant

Source: Primary data. SmartPLS bootstrapping, 5,000 subsamples, two-tailed test.

The moderation analysis reveals an asymmetry between financial literacy and technology awareness as moderators of the SF attitude–investment behaviour pathway. The interaction between financial literacy and SF attitude ($\beta = 0.1524$, $p = 0.124$) is not statistically significant, indicating that financial knowledge alone does not amplify the translation of positive SF attitudes into investment action. Technology awareness produces a marginally significant moderating effect ($\beta = 0.2058$, $p = 0.097$), indicating that greater exposure to digital finance tools strengthens the pathway from SF attitude to investment behaviour.

Under the TAM framework, perceived ease of use – operationalised through technology awareness – functions as the critical enabler converting positive SF intentions into behaviour. Financial literacy, by contrast, may increase the cognitive complexity of the investment evaluation process without lowering the perceived effort of platform participation. This finding suggests that platform design interventions – reducing interface complexity, improving mobile accessibility, and embedding decision-support tools – will be more effective at increasing SF participation than standalone financial literacy programmes, consistent with Malzara et al. [34]. H_1 is partially accepted: technology awareness has a marginally significant positive moderating effect on SF participation.

Objective 3: Financial Literacy, Education, and Sustainable Finance Engagement

H_0 : Higher financial literacy levels among Gen Z do not lead to significantly greater engagement with sustainable finance.

H_1 : Higher financial literacy levels among Gen Z lead to significantly greater engagement with sustainable finance.

Table 5 – Regression Analysis – SmartPLS Path Modelling

Predictor	Regression Coefficient (β)	T-Statistic	P-Value	Significance
Financial Literacy $\rightarrow$ SF Investment	0.3087	1.1488	0.1253	Not Significant
Technology Awareness $\rightarrow$ SF Investment	0.4125	1.9823	0.0486	Significant
SF Attitude $\rightarrow$ SF Investment	0.5583	2.6214	0.0092	Highly Significant

Source: Primary data. SmartPLS path modelling, 5,000 bootstrap samples.

SF attitude is the strongest predictor of SF investment behaviour ($\beta = 0.5583$, $p = 0.0092$), confirming that a positive orientation toward sustainable finance is the dominant driver of actual investment participation. Technology awareness is the second significant predictor ($\beta = 0.4125$, $p = 0.0486$), confirming that familiarity with digital finance tools materially increases SF investment participation. Financial literacy is not a statistically significant independent predictor ($\beta = 0.3087$, $p = 0.1253$), suggesting it operates as a necessary but not sufficient condition. H_0 for Objective 3 is accepted.

This finding is consistent with Judijanto et al. [20] and challenges the dominant policy assumption that financial literacy programmes alone drive sustainable investment participation. Under the S-O-R framework, SF attitudes (organism-level responses to sustainability stimuli) predict behaviour more powerfully than cognitive knowledge stocks, because attitudes integrate both emotional and rational evaluations in a way that literacy alone does not.

Measurement Model and Structural Model Results

Table 6 – Measurement Model – Construct Reliability and Validity

Construct	Items (n)	Cronbach's α	CR	AVE	Item Loadings
Financial Literacy (FL)	3	0.055*	0.604	0.349*	FL1=0.761, FL2=0.428*, FL3=0.534*
SF Attitude (ATT)	4	-0.111*	0.537	0.236*	A1=0.355*, A2=0.557, A3=0.626, A4=0.343*
Technology Awareness (TA)	3	-0.255*	0.555	0.309*	TA1=0.667, TA2=0.329*, TA3=0.611
SF Investment Behaviour (IFB)	3	0.159*	0.641	0.376*	IFB1=0.641, IFB2=0.508*, IFB3=0.678
Future Outlook (FO)	2	0.414	0.773	0.631	FO1=0.809, FO2=0.779
Challenges and Barriers (CB)	2	0.340	0.752	0.603	CB1=0.782, CB2=0.771

Source: Primary data, $n = 152$. PLS-SEM composite score approach. Thresholds: $\alpha \geq 0.70$; $CR \geq 0.70$; $AVE \geq 0.50$ [42]. * Items below 0.70 threshold retained as overall model fit is acceptable for exploratory research.

Table 7 – Variance Inflation Factors (VIF) – Collinearity Assessment

Predictor Construct	VIF	Assessment
Financial Literacy (FL)	1.004	No collinearity concern
SF Attitude (ATT)	1.011	No collinearity concern
Technology Awareness (TA)	1.009	No collinearity concern

Note: VIF values below 5.0 indicate no multicollinearity concern [42].

Table 8 – Structural Model – R^2 , Effect Size (f^2), and Predictive Relevance (Q^2)

Endogenous Construct	Predictors	R^2	f^2 (per predictor)	Q^2
SF Investment Behaviour (IFB)	FL, ATT, TA	0.075	FL=0.011 (negligible), ATT=0.009 (negligible), TA=0.054 (small)	0.018
Future Outlook (FO)	ATT, CB	0.072	—	—

Note: R^2 = variance explained; f^2 = effect size (small ≥ 0.02 , medium ≥ 0.15 , large ≥ 0.35); $Q^2 > 0$ indicates predictive relevance. Source: Primary data.

Table 9 – Bootstrapping Results – Path Coefficients and Significance (1,000 Resamples)

Path	β (Bootstrap Mean)	SE	t-Statistic	95% CI (Lower)	95% CI (Upper)	Significance
FL $\rightarrow$ IFB	0.090	0.080	1.127	-0.062	0.239	Not Significant
ATT $\rightarrow$ IFB	0.094	0.087	1.090	-0.082	0.260	Not Significant
TA $\rightarrow$ IFB	0.256	0.096	2.670	0.068	0.435	Significant ($p < 0.05$)

Note: β = standardised path coefficient; SE = standard error; 95% CI from percentile bootstrap. Significance: $|t| > 1.96$. TA $\rightarrow$ IFB is the only significant path. Source: Primary data.

Measurement model results (Table 6) indicate that Future Outlook and Challenges and Barriers achieve acceptable reliability and validity thresholds (Cronbach's $\alpha > 0.30$, CR > 0.70 , AVE > 0.50). However, the CR scores for the four main constructs – Financial Literacy, SF Attitude, Technology Awareness, and SF Investment Behaviour – range from 0.54 to 0.64, which fall below the 0.70 threshold. This is acknowledged as a limitation of the exploratory approach and the use of a five-point Likert scale with a limited item pool. Multicollinearity is not an issue since all VIF scores are approximately 1.0 (Table 7). The structural model explains 7.5% of variance in SF Investment Behaviour ($R^2 = 0.075$; Table 8), comparable to prior research on attitudinal-behavioural linkages in emerging sustainable finance domains [34]. Predictive relevance is confirmed by $Q^2 = 0.018$. Technology Awareness is the sole predictor showing a significant influence ($\beta = 0.256$, $t = 2.670$, $p < 0.05$; Table 9), corroborating TAM-based hypotheses that Gen Z engagement in sustainable finance is mediated by perceived digital platform utility.

Conclusion

This study provides empirical evidence on the relationship between Generation Z and sustainable finance in the specific context of Hyderabad, India. Three key findings emerge. First, sustainable finance attitude is the dominant predictor of SF investment behaviour ($\beta = 0.5583$, $p = 0.0092$), confirming that Gen Z's values-driven orientation toward sustainability is more powerful than financial literacy in driving investment participation. Second, technology awareness significantly enhances SF participation as both a direct predictor ($\beta = 0.4125$, $p = 0.0486$) and a marginal moderator ($\beta = 0.2058$, $p = 0.097$), establishing digital platform familiarity as a critical enabler. Third, financial literacy alone is not a statistically significant predictor ($\beta = 0.3087$, $p = 0.1253$), challenging the conventional assumption that knowledge-based interventions are sufficient to promote sustainable investment among young investors.

Based on these findings, four recommendations are proposed. First, financial institutions and fintech firms should invest in mobile-first, gamified platforms that provide seamless access to ESG-compliant products. Second, financial literacy programmes must integrate practical sustainable investing applications – simulation-based exercises, scenario-based learning, and digital platform usage – rather than rely on classroom instruction alone. Third, governments and regulators should introduce incentives including tax relief and subsidised green bonds to offset perceived risk for younger investors. Fourth, AI-driven advisory tools can help align portfolios with Gen Z's personal values and risk appetite, raising both participation rates and decision-making confidence.

This study has four limitations. First, the cross-sectional design precludes causal inference; longitudinal data are needed to establish whether SF attitudes and technology awareness produce sustained investment behaviour over time. Second, the sample is geographically confined to Hyderabad and skewed toward undergraduate respondents (78.3% aged 18–22), limiting generalisability. Third, self-reported Likert-scale data are subject to social desirability bias. Fourth, several construct reliability metrics fall below conventional thresholds, reflecting the exploratory nature of this research.

Future research should extend this study through longitudinal designs, comparative studies across Indian metropolitan markets, and structural equation models incorporating mediating variables such as trust in financial institutions and peer influence. Studies should also incorporate objective financial data – including actual investment records – to bridge the attitude-behaviour gap documented here.

REFERENCES

- 1 Ho, M.N., Abhayawansa, S., & Adams, C.A. Sustainable finance. *Elgar Encyclopedia of Corporate Communication*, August, 200–205 (2024). <https://doi.org/10.4337/9781802200874.ch32>
- 2 Gutterman, A.S. Sustainable Finance. *SSRN Electronic Journal*. 2024. <https://doi.org/10.2139/ssrn.3804435>
- 3 IFC. SBFN Toolkit: Developing Sustainable Finance Roadmaps. 2023. https://www.sbfnetwork.org/sites/default/files/2025/Mar/SBFN_Sustainable_Finance_Roadmap-Toolkit_Guide.pdf
- 4 KPMG. Sustainable Investing: Fast Forwarding its Evolution. 2020. <https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/02/sustainable-investing.pdf>
- 5 Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S.K. Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 345(2), 1061–1104 (2025). <https://doi.org/10.1007/s10479-021-04410-8>
- 6 Martiny, A., Taglialatela, J., Testa, F., & Iraldo, F. Determinants of environmental social and governance (ESG) performance: A systematic literature review. *Journal of Cleaner Production*, 456, 142213 (2024). <https://doi.org/10.1016/j.jclepro.2024.142213>
- 7 Taneja, S., Bhatnagar, M., & Ozen, E. Greening the economy: Harnessing the power of sustainable finance. *Encyclopedia of Management Practice*, 537–539 (2025). <https://doi.org/10.1016/B978-0-44-313776-1.00150-1>
- 8 Ahmed, R., Chen, X. H., Hoang, Y. H., & Do-Linh, C. Climate change effects and their implications for the financial markets: Evidence from the United Kingdom. *Journal of Environmental Management*, 366, 121782 (2024). <https://doi.org/10.1016/j.jenvman.2024.121782>
- 9 Eren, E., Merten, F., & Verhoeven, N. Pricing of climate risks in financial markets: a summary of the literature (Issue 130) (2022).
- 10 International Monetary Fund. *Global Financial Stability Report*, April 2025: Chapter 1. <https://www.imf.org/en/Publications/GFSR/Issues/2025/04/22/global-financial-stability-report-april-2025>
- 11 Wijethunga, A.W.G.C.N., Rahman, M. M., & Sarker, T. Environmental impact of financial market's development in Australia. *Environmental and Sustainability Indicators*, 23, 100438 (2024). <https://doi.org/10.1016/j.indic.2024.100438>
- 12 Gutterman, A. Sustainable Entrepreneurship. *SSRN Electronic Journal*. 2022. <https://doi.org/10.2139/ssrn.4192086>
- 13 Cooperation, I., Annual, C., & Report, P. *Green Consensus and High Quality Development: CCICED Annual Policy Report 2020*. (2022). <https://doi.org/10.1007/978-981-16-4799-4>
- 14 Faria, J.R., Tindall, G., & Terjesen, S. The Green Tobin's q: theory and evidence. *Energy Economics*, 110 (2022). <https://doi.org/10.1016/j.eneco.2022.106033>
- 15 Financial Stability Board. *The Implications of Climate Change for Financial Stability*. 2020. <https://www.fsb.org/wp-content/uploads/P231120.pdf>
- 16 Ozili, P.K. Green finance research around the world: a review of literature. *International Journal of Green Economics*, 16(1), 1(2022). <https://doi.org/10.1504/ijge.2022.10048432>
- 17 Rahman, M.H., Tanchangya, T., Rahman, J., Aktar, M.A., & Majumder, S.C. Corporate social responsibility and green financing behavior in Bangladesh. *Innovation and Green Development*, 3(3), 100133 (2024). <https://doi.org/10.1016/j.igd.2024.100133>

- 18 Roth, N., Thiele, T., & von Unger, M. Blue Bonds: Financing Resilience of Coastal Ecosystems. IUCN GMPP Technical Report. 2019. URL: <https://www.greenfinanceplatform.org/research/blue-bonds-financing-resilience-coastal-ecosystems>
- 19 Cagarman, K., & Stiftung, B. Gen Z Pioneering Entrepreneurs: Reflections and Policy Considerations. 2023. <https://doi.org/10.11586/2023090>
- 20 Judijanto, L., Utami, E.Y., & Siagian, A.O. Gen Z Investor Behavior: Analysing the influence of Sustainable Investment Practices, Financial Education and Risk Tolerance in Indonesia. *Journal of Management and Business*, 15(1), 37–48 (2024). <https://doi.org/10.31539/jomb.v6i3.8832>
- 21 Lulewicz-Sas, A., Kinowska, H., Zubek, M., & Danilewicz, D. Examining the impacts of ESG on employee engagement: a study of Generation Z. *Central European Management Journal*, 33(4), 634–649 (2025). <https://doi.org/10.1108/CEMJ-08-2024-0242>
- 22 Sengupta, D., Mathews, M., Bridges, L., D’Costa, R., & Bastian, B.L. Sustainability Orientation of Generation Z and Its Role in Their Choice of Employer. *Administrative Sciences*, 14(10), 249 (2024). <https://doi.org/10.3390/admsci14100249>
- 23 Zaka, K., & Al Juboori, M. The Impact of Generation-Z on International Branding Strategy. Linnaeus University Working Paper. 2023.
- 24 Sait, S.A., & Vijesh, P.V. Green Economics in Action: Exploring the Nexus Between Industrial Growth, Market Forces, and Carbon Credit Movements in Emerging Economies – With Reference to India. *Journal of Vasyl Stefanyk Precarpathian National University*, 12(4), 49–65 (2025). <https://doi.org/10.15330/jpnu.12.4.49-65>
- 25 Gómez Sánchez, L., & Tobon, S. Sustainable investing among young generations: balancing ideals and financial realities. *Discover Sustainability*, 6(1) (2025). <https://doi.org/10.1007/s43621-025-01053-8>
- 26 Adel, N. The Impact of Digital Literacy and Technology Adoption on Financial Inclusion. *Heliyon*, 10(24), e40951 (2024). <https://doi.org/10.1016/j.heliyon.2024.e40951>
- 27 Filippini, M., Leippold, M., & Wekhof, T. Sustainable Finance Literacy and the Determinants of Sustainable Investing. *Journal of Banking & Finance*, 163, 107167 (2024). <https://doi.org/10.1016/j.jbankfin.2024.107167>
- 28 Ramesh, A. Attitudes of Gen Z Germans towards Renewable & Sustainable Investments. HEC Liège. 2024.
- 29 United News of India. The city is home to major IT campuses like HITEC City and Cyberabad, as well as over 1,500 technology startups. 2026. URL: <https://www.uniindia.com/news/business-economy/business-telangana-nasscom-gcc-summit-minister/3835537.html>
- 30 Akibun, F., Prayitno, H., Z, R., & Otto, N. Financial Literacy In Gen Z Generation. *Jurnal Ekonomi, Manajemen, Akuntansi Dan Keuangan*, 6(2), 8 (2025). <https://doi.org/10.53697/emak.v6i2.2286>
- 31 Abdullah, A., Basuki, H., Santoso, W.I., Sukresna, I.M., & Indriani, F. Utilization of Digital Banking Services for Generation Z Economic Sustainability. *Research Horizon*, 1(3), 86–93 (2024). <https://doi.org/10.54518/rh.4.4.2024.298>
- 32 Aini, N. The Motivation of Gen Z Towards Creating a Sustainable Business. *JUSTEK: Jurnal Sains dan Teknologi*, 7(4), 380–386 (2024).
- 33 Ravani, K., Verma, H., & Vidani, J. Comparative Study of Investment Preferences Among Millennial and Gen Z in Ahmedabad. *International Journal of Economics, Finance and Business Statistics*, 2(4), 203–228 (2024). <https://doi.org/10.59890/ijefbs.v2i4.261>
- 34 Malzara, V.R.B., Widyastuti, U., & Buchdadi, A.D. Analysis of Gen Z’s Green Investment Intention: The Application of Theory of Planned Behavior. *Jurnal Dinamika Manajemen Dan Bisnis*, 6(2), 63–84 (2023). <https://doi.org/10.21009/jdmb.06.2.5>
- 35 Saliha, & Dinesh, P. A review on the sustainable purchase behavior of Gen Z. *Journal of Harbin Engineering University*. 2023. URL: <https://harbinengineeringjournal.com/index.php/journal/article/view/1806>
- 36 Sethuraman, P., Arasuraja, G., & Rajapriya, M. Social Media’s Effect on Millennials and Generation Z’s Green Purchasing Habits. *International Journal of Professional Business Review*, 8(5), 1–18 (2023). <https://doi.org/10.26668/businessreview/2023.v8i5.1470>
- 37 Ghouse, S., Shekhar, R., & Chaudhary, M. Sustainable choices of Gen Y and Gen Z: exploring green horizons. *Management & Sustainability: An Arab Review*, 4(3), 533–559 (2024). <https://doi.org/10.1108/MSAR-04-2024-0018>

- 38 Gutterman, A.S. Sustainable Finance and Impact Investment: A Guide for Sustainable Entrepreneurs. SSRN Electronic Journal. 2024. <https://doi.org/10.2139/ssrn.4944162>
- 39 Davis, F. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13, 319 (1989). <https://doi.org/10.2307/249008>
- 40 Mehrabian, A., & Russell, J.A. An Approach to Environmental Psychology. MIT Press. 1974.
- 41 Ajzen, I. The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211 (1991). [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- 42 Hair, J.F., Risher, J.J., Sarstedt, M., & Ringle, C.M. When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24 (2019). <https://doi.org/10.1108/EBR-11-2018-0203>
- 43 Dwivedi, Y.K., et al. Setting the future of digital and social media marketing research. International Journal of Information Management, 59, 102168 (2021). <https://doi.org/10.1016/j.ijinfomgt.2020.102168>
- 44 Agarwal, H., & Vaghela, P. Work Values of Gen Z: Bridging the Gap to the Next Generation. National Conference on Innovative Business Management Practices in 21st Century, 1–26 (2018).
- 45 Fuchs, O., Lorenz, E., & Fuchs, L. Generational Differences in Attitudes Towards Work and Career. International Journal of Innovative Research and Advanced Studies, 11(7) (2024).
- 46 Salvadorinho, J., Hines, P., Kumar, M., Ferreira, C., & Teixeira, L. Empowering Generation Z in manufacturing organizations. Journal of Work-Applied Management. 2024. <https://doi.org/10.1108/JWAM-07-2024-0087>

¹**Ахмед Саит С.,**

ассистент-профессор, ORCID ID: 0009-0007-4844-9332,

e-mail: sofiamaqbool@ymail.com

^{2*}**Виджеш П.В.,**

PhD, ORCID ID: 0000-0002-1504-0806,

*e-mail: vijesh@rajagiri.edu

¹Сауда кафедрасы, «Лойола Академия» бакалавриат және магистратура колледжі
(Османия университеті), Хайдарабад қ., Үндістан

²Раджагири әлеуметтік ғылымдар колледжі,
Каламассери қ., Керала штаты, Үндістан

ЖАҢА БУЫННЫҢ ТҰРАҚТЫЛЫҒЫ: Z ҰРПАҒЫ ИНВЕСТИТОРЛАРЫНЫҢ ҚАРЖЫЛЫҚ МАҚСАТТАРЫН, ҚИЫНДЫҚТАРЫН ЖӘНЕ ӨЗГЕРІСТЕР ЖАСАУ ӘЛЕУЕТІН АНЫҚТАУ

Аңдатпа

Тұрақты қаржы жаһандық және қоғамдық мәселелерді шешудің маңызды құралына айналды, өйткені ол инвестициялық және басқарушылық шешімдер қабылдау үдерісіне экономикалық, әлеуметтік және басқарушылық факторларды біріктіреді. Үндістанның жастар саны көп әрі қарқынды дамып келе жатқан қалалық орталығы саналатын Хайдарабадта Z ұрпағы инвесторлары тұрақты қаржы саласының қалыптасуына барған сайын ықпал етуде. Бұл зерттеу Хайдарабадтағы Z ұрпағы тұрақты қаржы туралы хабардарлығын, мотивациясын және инвестициялық мінез-құлқын, әсіресе олардың әлеуметтік медиа платформалары арқылы қатысуын талдауға бағытталған. SmartPLS бағдарламасы арқылы құрылымдық теңдеулерді модельдеу әдісін қолдана отырып, құрылымдалған сауалнама негізінде Z ұрпағына жататын 152 респонденттен бастапқы деректер жиналды. Зерттеу нәтижелері тұрақты қаржыға деген көзқарастың инвестициялық мінез-құлықтың ең күшті болжаушысы екенін көрсетті ($\beta = 0.5583$, $p = 0.0092$), одан кейін технологиялық хабардарлық маңызды орын алады ($\beta = 0.4125$, $p = 0.0486$), ал қаржылық сауаттылықтың өзі статистикалық тұрғыдан маңызды болжаушы болып табылмады ($\beta = 0.3087$, $p = 0.1253$). Бұл нәтижелер қаржылық сауаттылық тұрақты инвестицияларға қатысудың негізгі факторы деген дәстүрлі түсінікке күмән келтіріп, технологиялық қолжетімділік пен құндылықтық сәйкестікті мінез-құлықты қалыптастыратын анағұрлым ықпалды факторлар ретінде көрсетеді. Зерттеу Z ұрпағының Хайдарабадты Үндістандағы тұрақты қаржының үлгілі қаласына айналдырудағы трансформациялық әлеуетін көрсетеді және саясаткерлерге, қаржы институттарына және білім беру ұйымдарына практикалық ұсыныстар ұсынады.

Түйін сөздер: Z ұрпағы, тұрақты қаржы, ESG-инвестициялау, SmartPLS, технологиялық хабардарлық, Хайдарабад, жасыл инвестициялар.

¹Ахмед Саит С.,

ассистент-профессор, ORCID ID: 0009-0007-4844-9332,

e-mail: sofiamaqbool@ymail.com

^{2*}Виджеш П.В.,

PhD, ORCID ID: 0000-0002-1504-0806,

e-mail: vijesh@rajagiri.edu

¹Кафедра коммерции, Колледж бакалавриата
и последипломного образования «Лойола Академия»
(Университет Османия), г. Хайдарабад, Индия

²Колледж социальных наук Раджагири,
г. Каламассери, штат Керала, Индия

УСТОЙЧИВОСТЬ НОВОГО ПОКОЛЕНИЯ: РАСКРЫТИЕ ФИНАНСОВЫХ НАМЕРЕНИЙ, ВЫЗОВОВ И ТРАНСФОРМАЦИОННОГО ПОТЕНЦИАЛА ИНВЕСТИРОВ ПОКОЛЕНИЯ Z

Аннотация

Устойчивые финансы стали важным инструментом решения глобальных и социальных проблем, поскольку они интегрируют экономические, социальные и управленческие факторы в процессы инвестирования и принятия решений. В г. Хайдарабаде, быстрорастущем городском центре с высокой долей молодежи, инвесторы поколения Z оказывают все более заметное влияние на развитие сферы устойчивых финансов. Данное исследование направлено на изучение осведомленности, мотивации и инвестиционного поведения представителей поколения Z в г. Хайдарабаде, с особым акцентом на их вовлеченность через социальные медиаплатформы. С использованием моделирования структурными уравнениями в SmartPLS первичные данные были собраны на основе структурированной анкеты у 152 респондентов, относящихся к поколению Z. Результаты исследования показывают, что отношение к устойчивым финансам является наиболее сильным предиктором инвестиционного поведения ($\beta = 0.5583$, $p = 0.0092$), далее следует технологическая осведомленность ($\beta = 0.4125$, $p = 0.0486$), тогда как финансовая грамотность сама по себе не является статистически значимым предиктором ($\beta = 0.3087$, $p = 0.1253$). Эти результаты ставят под сомнение традиционное предположение о том, что финансовая грамотность является главным фактором участия в устойчивом инвестировании, и показывают, что технологическая доступность и ценностное соответствие выступают более сильными поведенческими факторами. Исследование подчеркивает трансформационный потенциал поколения Z в превращении Хайдарабада в модельный город устойчивых финансов в Индии и предлагает практические выводы для политиков, финансовых институтов и образовательных организаций.

Ключевые слова: поколение Z, устойчивые финансы, ESG-инвестирование, SmartPLS, технологическая осведомленность, Хайдарабад, зеленые инвестиции.

**REQUIREMENTS TO THE DESIGN
OF ARTICLES PUBLISHED IN THE JOURNAL KBTU BULLETIN**

Previously **unpublished material by the author (s)** corresponding to the profile of the journal in Russian, Kazakh or English is accepted for publication. The article is provided in electronic format (in formats .doc, .docx,) only by downloading it through the functionality of the journal's website (<https://vestnik.kbtu.edu.kz/jour/about>), which has detailed instructions for authors on the process of submitting articles, as well as on payment for publication.

Mandatory requirements for registration:

- ♦ The MRNTI index is placed in the upper-left corner. (Times New Roman 14 size)
- ♦ One single interval below the authors' surnames and initials in bold. Detailed names of organizations, cities, countries where the work was performed, e-mail of one of the authors. Specify footnotes to the place of work of the authors.
- ♦ Below, in one interval, the full title of the article is in bold.
- ♦ Below, a brief (150–250 words) abstract (abstract) at one interval. Abstract should outline the general orientation of the article and formulate its main results and be self-sufficient, it should not contain references to literary sources or to any sections and elements of the article to which it relates.
- ♦ Key words related to the submitted article.

The text of the article is typed in a straight font Times New Roman, size 14. The line spacing is single. All fields are 2 cm long.

The article should be divided into sections: Introduction, Materials and Methods, Results and Discussion, Conclusion, Information on funding (if available).

At the end, a "List of References" is given. Literary references in the text are in square brackets [1]. The list of references is drawn up twice: according to the above GOST. The list should fully correspond to the main text of the article, contain all the works mentioned (and only mentioned) in it. Next, the same list of references should be presented in the generally accepted English transliteration (see the sample of the literature design below).

The article undergoes "blind" review, reviewers are appointed by the editorial board of the journal.

At the end of the article please indicate full name, return address, phone numbers, fax, e-mail address.

The author(s) are responsible for the content of the article. The opinion of the Scientific and Editorial Board does not always coincide with the opinion of the author (authors). The Editorial Board reserves the right to publish or reject articles.

The reference to our edition in reprinting is obligatory.

Our address: **59, Tole Bi St., Almaty, 050000, Kazakhstan-British Technical University.**

Website: <https://vestnik.kbtu.edu.kz/jour/pages/view/EditorialC> Editorial Board.

ЖУРНАЛЫНДА ЖАРИЯЛАНАТЫН МАҚАЛАЛАРДЫ РӘСІМДЕУГЕ ҚОЙЫЛАТЫН ТАЛАПТАР

Жариялау үшін **автормен (авторлармен) бұрын жарияланбаған**, журнал бейініне сәйкес келетін орыс, қазақ немесе ағылшын тілдеріндегі материалдар қабылданады. Мақала журнал сайтының (<https://vestnik.kbtu.edu.kz/jour/about>) функционалдығы арқылы жүктеу жолымен тек электронды форматта (.doc, .docx форматтарында) қабылданады, сайтта авторларға мақаланы жүктеу, ұсыну үрдісі туралы, сондай-ақ жарияланымға ақы төлеу туралы толық нұсқаулық бар.

Ресімдеу бойынша міндетті талаптар:

Жоғарғы сол жақ бұрышта FTAMI индексі жазылады. (Times New Roman 14 өлшем)

- ♦ Бір бірлік интервалдан кейін авторлардың тегі мен аты-жөні жартылай қалың қаріппен жазылады. Жұмыс орындалған ұйымдардың, қалалардың, елдердің толық атаулары, авторлардың бірінің e-mail поштасы көрсетіледі. Авторлардың жұмыс орнына сілтеме беріледі.

- ♦ Бір интервалдан кейін мақаланың толық атауы жартылай қалың қаріппен жазылады.

- ♦ Бір интервалдан кейін қысқаша (150–250 сөз) аңдатпа (abstract) ұсынылады. Аңдатпа мақаланың жалпы бағытын көрсетіп, оның негізгі нәтижелерін тұжырымдауы тиіс, аңдатпада әдеби дереккөздерге де, оған қатысты мақаланың кез-келген бөлімдері мен элементтеріне сілтеме жасауға болмайды.

- ♦ Ұсынылған мақалаға қатысты кілт сөздер (тірек сөздер) жазылады.

Мақала мәтіні 14 өлшемді Times New Roman шрифтімен теріледі. Жолдар арасында бірлік интервал сақталады. Шеттер – барлық жағынан 2 см.

Мақала келесі бөлімдерді қамтуы тиіс: Кіріспе, Материалдар және әдістер, Нәтижелер мен талқылау, Қорытынды, Қаржыландыру туралы ақпарат (болған жағдайда көрсетіледі).

Соңында «Әдебиеттер тізімі (References)» жазылады. Мәтіндегі әдеби сілтемелер төртбұрышты жақшада ұсынылады [1]. Әдебиеттер тізімі MEMCT-ке сәйкес екі рет рәсімделеді. Тізім мақаланың негізгі мәтініне толық сәйкес келіп, мақалада аталған (тек аталған) барлық жұмыстарды қамтуы керек. Әрі қарай, дәл осы сілтемелер жалпы қабылданған ағылшын транслитерациясында жазылады.

Мақала «жасырын» рецензиялаудан өтеді, рецензенттерді журналдың редакциялық алқасы тағайындайды.

Мақаланың соңында авторлардың толық аты-жөнін, кері байланыс мекенжайын, телефондар, факс және электрондық пошта мекенжайын (e-mail) жазуыңызды сұраймыз.

Мақаланың мазмұнына автор (авторлар) жауапты. Ғылыми-редакция алқасының пікірі әрдайым автордың (авторлардың) пікірімен сәйкес келе бермейді. Редакция алқасы мақалаларды жариялау немесе қабылдамау құқығын өзіне қалдырады.

Қайта басылым кезінде біздің жарияланымға сілтеме жасау керек.

Мекенжайымыз: **050000, Алматы қ., Төле би көш., 59, Қазақстан-Британ техникалық университеті.**

Сайт: <https://vestnik.kbtu.edu.kz/jour/pages/view/EditorialC> Редакция алқасы.

**ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ,
ПУБЛИКУЕМЫХ В ЖУРНАЛЕ ВЕСТНИК КБТУ**

Для публикации принимается **ранее неопубликованный автором (авторами) материал**, соответствующий профилю журнала, на русском, казахском или английском языках. Статья предоставляется в электронном формате (в форматах .doc, .docx,) только путем ее загрузки через функционал сайта журнала (<https://vestnik.kbtu.edu.kz/jour/about>), в котором имеется подробная инструкция для авторов о процессе подачи статей, а также об оплате публикации.

Обязательные требования по оформлению:

- ♦ В левом верхнем углу ставится индекс МРНТИ (Times New Roman 14 размер).
- ♦ Через один одинарный интервал ниже фамилии и инициалы авторов полужирным шрифтом. Развернутые названия организаций, городов, стран, где выполнена работа, e-mail одного из авторов. Указать сноски на место работы авторов.
- ♦ Ниже через один интервал полное название статьи полужирным шрифтом.
- ♦ Ниже через один интервал краткая (150–250 слов) аннотация (abstract). Аннотация должна обрисовывать общую направленность статьи, формулировать ее основные результаты и быть самостоятельной, в ней недопустимы ссылки ни на литературные источники, ни на какие-либо разделы и элементы статьи, к которой она относится.
- ♦ Ключевые слова (Key words), относящиеся к представляемой статье.

Текст статьи набирается прямым шрифтом Times New Roman, размер 14. Интервал между строками одинарный. Поля все по 2 см.

Статья должна быть разделена на разделы: Введение, Материалы и методы, Результаты и обсуждение, Заключение, Информация о финансировании (при наличии).

В конце дается «Список литературы (References)». Литературные ссылки в тексте в квадратных скобках [1]. Список литературы оформляется дважды: согласно ГОСТу. Список должен полностью соответствовать основному тексту статьи, содержать все упомянутые (и только упомянутые) в нем работы. Далее тот же список литературы представить в общепринятой английской транслитерации.

Статья проходит «слепое» рецензирование, рецензенты назначаются редколлегией журнала.

В конце статьи просим указать Ф.И.О. полностью, обратный адрес, телефоны, факс, адрес электронной почты (e-mail).

Ответственность за содержание статьи несет автор (авторы). Мнение Научно-редакционного совета не всегда совпадает с мнением автора (авторов). Редакционный совет оставляет за собой право публикации или отклонения статей.

Ссылка на наше издание при перепечатке обязательна.

Наш адрес: **050000, г. Алматы, ул. Толе Би, 59, Казахстанско-Британский технический университет.**

Сайт: <https://vestnik.kbtu.edu.kz/jour/pages/view/EditorialC> Редакционная коллегия.

ҚАЗАҚСТАН-БРИТАН ТЕХНИКАЛЫҚ УНИВЕРСИТЕТІНІҢ
ХАБАРШЫСЫ
HERALD
OF THE KAZAKH-BRITISH TECHNICAL UNIVERSITY
ВЕСТНИК
КАЗАХСТАНСКО-БРИТАНСКОГО ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Ответственный за выпуск
Алагузов Рашид Жуманович

Редакторы
Кабылдаева Толганай Кайсаровна,
Скуратова Ирина Михайловна

Компьютерная верстка
Жуйкова Майя Антоновна

Подписано в печать 16.06.2026 г.
Тираж 300 экз. Формат 60х84 1/16.
Бумага тип. Усл.печ.л. 72,0 Уч.-изд.л. 74,8
Заказ №

Издание Казахстанско-Британского технического университета
Издательский центр КБТУ, Алматы, ул. Толе би, 59

