
**COMPUTER SCIENCE
КОМПЬЮТЕРЛІК ҒЫЛЫМДАР
КОМПЬЮТЕРНЫЕ НАУКИ**

ӘОЖ 681.327.12
FTAXP 65.011.56

<https://doi.org/10.55452/1998-6688-2025-22-4-10-22>

^{1,2,3}**Еримбетова А.,**

т.ғ.к., қауымдастырылған профессор, PhD, ORCID ID: 0000-0002-2013-1513,

e-mail: aigerian8888@gmail.com

^{2*}**Бержанова У.,**

докторант, ORCID ID: 0009-0000-2467-5721,

*e-mail: berzhanovaulmekenn@gmail.com

^{1,4}**Дайырбаева Ә.,**

оқытушы, ORCID ID: 0000-0002-4255-5456,

e-mail: nurbekkyzy_e@mail.ru

¹**Сәкенов Б.,**

инженер-программист, ORCID ID: 0000-0002-9849-6176,

e-mail: sbakzhan22@gmail.com

⁵**Самбетбаева М.,**

PhD, ORCID ID: 0000-0001-9358-1614,

e-mail: madina_jgtu@mail.ru

¹Ақпараттық және есептеуіш технологиялар институты, Алматы қ., Қазақстан

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

³Еуразиялық технологиялық университеті, Алматы қ., Қазақстан

⁴Қ.И. Сәтбаев атындағы Қазақ ұлттық зерттеу техникалық университеті,
Алматы қ., Қазақстан

⁵Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

**MEDIAPIRE ЖӘНЕ ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІН
ПАЙДАЛАНУ АРҚЫЛЫ НАҚТЫ УАҚЫТ РЕЖИМІНДЕ
ҚАЗАҚ ҒЫМ ТІЛІН ТАНУ ҮШІН МОДЕЛЬ ЖАСАУ**

Аңдатпа

Бұл мақалада MediaPipe платформасын қолдана отырып, қазақ ғым тілін тану жүйесін әзірлеу процесі қарастырылған. Платформа ғым тілі қимылдарын нақты уақытта дәл анықтауға мүмкіндік береді. Негізгі назар ғым қимылдарын тану модельдерін құруға, нейрондық желілерді оқытуға және MediaPipe платформасымен интеграциялауға бөлінген. Маңызды аспектілердің бірі – нейрондық желілердің архитектурасын пайдалану арқылы қимылдарды өңдеудің жоғары дәлдігі мен жылдамдығына қол жеткізу. Жүйе аннотацияланған

қимылдардан алынған үлкен деректер жиынтығы негізінде оқытылған, бұл танудың сапасын айтарлықтай жақсартты. Қазақ ым тілі қимылдарын тану үшін LSTM нейрондық желісі қолданылды, себебі ол уақыттық қатарлар мен деректердің реттіліктерімен тиімді жұмыс істейді. Модель қазақ ым тілінің 30 қимылына оқытылып, ым қимылдарын нақты уақытта мәтінге айналдыруға мүмкіндік береді. Бұл әдіс есту және сөйлеу қабілеті шектеулі адамдармен қарым-қатынасты жеңілдетіп, инклюзивтілікті арттыруға ықпал етеді. Сонымен қатар, пайдаланушымен өзара әрекеттесу үшін нейрондық желіні ым тілі қимылдарын тануға арналған қосымшалармен оңай интеграциялауға мүмкіндік беретін ыңғайлы веб-интерфейс жасалды. Жұмыстың негізгі аспектілерінің бірі – тану дәлдігін жақсарту мақсатында деректерді аннотациялау және өңдеу әдістерін жетілдіру. Жүйені болашақта дамыту ым тілі қимылдарының дерекқорын кеңейту мен веб-қосымшалармен интеграциялауды көздейді. Бұл есту және сөйлеу қабілеті шектеулі адамдар үшін әлеуметтік қамтуды жақсартып, кең ауқымды және қолжетімді платформаны құруға мүмкіндік береді.

Тірек сөздер: қазақ ым тілі, MediaPipe, терең оқыту, жасанды нейрондық желі, нақты уақыт режимі.

Кіріспе

Қазіргі заманғы технологиялар мүмкіндігі шектеулі жандардың өмір сүру сапасын жақсартуға айтарлықтай үлес қосуда. Солардың бірі – ым тілі қимылдарын тану жүйелері, олар есту және сөйлеу қабілеті нашар адамдар үшін аса маңызды рөл атқарады. Мұндай жүйелер ым-ишараларды мәтін немесе сөйлемдерге түрлендіру арқылы қарым-қатынасты жеңілдетіп, инклюзивтілікті арттыруға мүмкіндік береді.

Көптеген зерттеушілер [1–4] CNN (Convolutional Neural Networks) негізінде терең оқыту әдістерін қолдана отырып, өздерінің тәсілдері мен әдістемелерін ұсынған. Бұл жұмыстарда CNN модельдері бейнелерден ерекшеліктерді автоматты түрде бөліп алып, оларды тиімді өңдеу мүмкіндігін көрсетеді. Бірінші және төртінші зерттеулер тек қана CNN модельдеріне сүйенсе, екінші және үшінші зерттеулер бірнеше үлгілердің комбинациясын пайдаланған. Зерттеулерде жақсы нәтижелерге жету көбіне қолданылған деректердің сапасы мен көлеміне байланысты болған.

Ым тілін тану бойынша YOLO (You Only Look Once) объектілерді анықтау моделінің әртүрлі нұсқаларының пайдаланылуы [5–8] зерттеу жұмыстарында ұсынылды. Бұл модельдер ым тілін тануға арналған жүйелерде қолданылып, тез және дәл анықтау үшін дамытылған. Алғашқы жұмыста, YOLO-ның түрлі нұсқаларын қолдану арқылы ым тілін тану үшін арнайы алгоритм ұсынылып, нәтижесінде қолданыстағы модельдерден жоғары көрсеткіштер алынған; содан кейінгі зерттеу жұмысында YOLO-ға негізделген трансформерлік архитектура зерттеліп, ым тілін тану кластарын және шекараларын тікелей болжау мүмкіндігі қарастырылған; үшінші жұмыста Raspberry Pi мен YOLOv3 негізінде автоматты ым тілін тану жүйесі жасалып, кескіндерді өңдеу және тану дәлдігін арттыруға бағытталған; соңғысында, YOLOv5 моделінің графикалық интерфейсі арқылы видеолардың нақты уақытта ым тілін тануын жоғары дәлдікпен жүзеге асыру мүмкіндігі көрсетілген. YOLO моделінің қолданылуы осы төрт жұмыстың нәтижелерін жақсартуда шешуші рөл атқарады. YOLO жүйесінің жылдамдығы мен дәлдігі, сонымен қатар оның онлайн режимде жұмыс істеу қабілеті, ым тілін тану процесін тиімді әрі нақты жасауға мүмкіндік береді. Нәтижесінде, дәлдік пен жылдамдықтың үйлесімі табиғи өзара әрекеттесуді қамтамасыз ететін адам-компьютер қарым-қатынасында жақсы нәтижелерге қол жеткізуге ықпал етеді.

Келесі зерттеушілердің [9–13] еңбектерінде MediaPipe кітапханасы қолданылады, ол қол қимылдарын және ым тілін тану жүйелерінде маңызды рөл атқарады. Бұл кітапхана бейнемазмұнды өңдеу мен қозғалыстарды талдау бойынша тиімді құралдар ұсына отырып, ым тілін тану технологияларының дамуына үлкен үлес қосады. Алдымен бірінші жұмыста MediaPipe кітапханасын пайдаланып, ASL ым тілін тану үшін CNN негізіндегі жүйе ұсынылды, ол 99,95% дәлдік көрсетті. Кейін екінші жұмыста MediaPipe-пен бірге компьютерлік көру әдістерін қолданып, ASL ым тілін дәлірек интерпретациялауға бағытталған жүйе әзірленді, мұнда OpenCV және MediaPipe кітапханаларының мүмкіндіктері пайдаланылды. Үшінші жұмыста MediaPipe негізінде RSL ым тілінің оқшауланған қол қимылдарын тану

жүйесі компоненттік тәсілмен ұсынылды, алайда динамикалық жесттерді тануда қосымша ерекшеліктер мен трансформациялар енгізу қажеттілігі анықталды. Төртінші жұмыста динамикалық қимылдарды тану үшін MediaPipe RNN модельдерімен біріктіріліп, жаңа DSL10 деректер жиынтығы жасалды және модельдің 99%-дан жоғары дәлдікке жеткені көрсетілді. Алайда, төрт жұмыста да динамикалық қимылдар мен статикалық қимылдар арасындағы нақты шекаралар мен деректер жиынтығының шектелуі, сондай-ақ нақты уақыт режиміндегі жүйелердің тиімділігі толыққанды зерттелмеген. MediaPipe кітапханасының қолданылуы ым тілін тану дәлдігін арттыруға ықпал етіп, әртүрлі дене кемістігі бар адамдар мен машиналар арасындағы өзара әрекеттестікті жақсартуға мүмкіндік берді.

Келесі зерттеуде ResNet-50 және VGG-19 негізіндегі ансамбль әдісі, ол қазақ ым тілінің 42 дактилін жіктейді. 57 мыңнан астам бейнелерден тұратын деректер жиыны қалыптастырылып, қол жеткізілген дәлдік 95,7%-ды құраған [14].

Сонымен қатар, қазақ мәтінін KSL глоссина аудару үшін трансформерлерді қолдануда деректердің параллель корпусы әзірленіп, машиналық аударма моделі жоғары дәлдікті көрсетіп, оқытылған [16]. Ал зерттеу [15] деректер сапасын жақсарту үшін аугментация әдістерін қолданудың маңыздылығын атап өткен. Негізгі мәселелерге аннотацияланған деректердің шектеулілігі, кластарды теңестіру қажеттілігі баяндалған.

Ым тілін тану технологиясын жетілдіру бағытындағы зерттеулер [17,18,19] ым қимылдарының динамикасын талдауға негізделген әдістерді ұсынады. Бұл әдістер уақыттық жинақтаушы сипаттамалар мен траекторияларды динамикалық түсіруді қолдану арқылы, сондай-ақ кросс-модальды ақпаратты біріктіретін DFCNet+ моделінің көмегімен жоғары дәлдікке қол жеткізеді. Мұндай тәсілдер ымдардың мағынасын контексте дәлірек анықтауға және түрлі деректер жиынтығында тиімділікті арттыруға мүмкіндік береді.

Бұл жұмыстарда қарастырылған зерттеулер мен әдістер ым тілін тану жүйелерінің тиімділігін арттыруға бағытталғанымен, динамикалық және статикалық қимылдар арасындағы шекараларды анықтау, нақты уақыт режимінде жүйелердің тиімділігін толық зерттеу және LSTM рекуррентті нейрондық желілерін қолдану арқылы қимылдар тізбегін өңдеудің жаңа әдістерін енгізу мәселелері әлі де толық зерттелмеген.

Осы мақалада авторлар бейне және қозғалыстарды дәл талдауға арналған MediaPipe платформасының мүмкіндіктерін қазақ ым тілінің динамикалық қозғалыстарын тану үшін LSTM рекуррентті нейрондық желісімен біріктіретін жаңа әдістемесін ұсынады. Бұл әдіс қимылдар тізбегін тиімді өңдеу және оларды нақты уақыт режимінде тану дәлдігін арттыру үшін қолданылатын шешім ретінде ұсынылады.

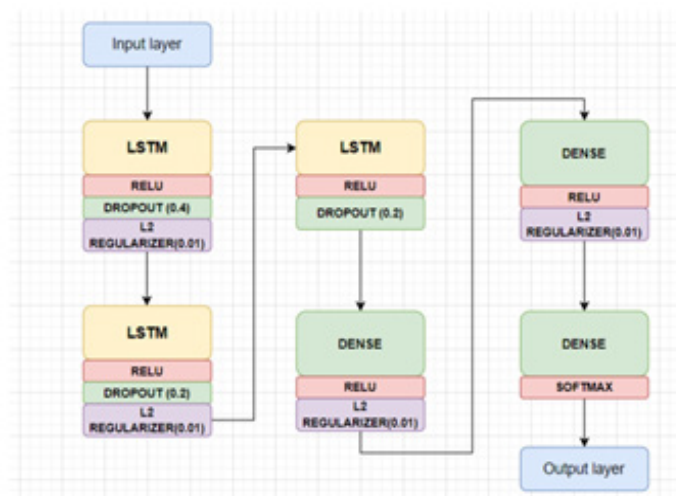
Материалдар мен әдістер

Зерттеу қазақ ым тіліне тән деректерді жинау және өңдеу әдістемесін ұсынады, ол тек статикалық емес, динамикалық қимылдарды да қамтиды. Бұл тәсіл ым тілінің басқа ым жүйелерінен ерекшеленетін бірегей ерекшеліктерін ескере алатын жүйелерді құру үшін маңызды.

Динамикалық қимылдарды тану үшін бірнеше LSTM қабатынан тұратын рекурренттік нейрондық желі құрылды [20]. Модельдің тиімділігіне айтарлықтай әсер ететін нейрондық желілерді құру кезінде қабаттардың саны мен әр қабаттағы нейрондардың саны. Бұл параметрлерді таңдау эксперименттік жолмен жүзеге асырылады, себебі олар оқыту мен тест үлгілеріндегі болжау дәлдігін тікелей анықтайды.

Жүйенің артық оқытылуы (overfitting) жиі кездеседі, егер жасырын қабаттарының саны тым көп болатын күрделі модельдер құрылса. Мұндай жағдайда модель тек оқыту деректерін жаттап алуға бейімделеді де, жаңа деректерді өңдеуде жинақтау қабілеті нашарлайды. Керісінше, қабаттар мен нейрондар саны тым аз болса, модель қажетті ақпаратты үйреніп үлгермей, жеткіліксіз оқытылған күйде қалады. Бұл балансқа қол жеткізу нейрондық желінің функционалдық мүмкіндіктерін оңтайландырудың басты шарты.

Жоғарыда ұсынылған теориялық тұжырымдарды ескере отырып, 1-суретте жасанды нейрондық желінің архитектурасы көрсетілген.



Сурет 1 – Жасанды нейрондық желінің архитектурасы

Бірнеше маңызды құрамдас бөліктерден тұратын нейрондық желі архитектурасы 1-суретте сипатталған. Бұл модельде кіріс қабаты, үш LSTM қабаты, үш тығыз қабат және шығыс қабаты бар.

Жасанды нейрондық желінің кіріс деректері жүйеге әрбір кадрда бөлінген 1662 нүктенің орындарын көрсетеді. Бұл жағдайда өлшемі (30, 1662) болып табылады: мұндағы 30 – тізбектегі кадрлар саны, ал 1662 – MediaPipe кітапханасын қолдану нәтижесінде кадрды өңдеу арқылы алынған нүктелердің координаттарының саны.

Relu (Rectified Linear Unit) белсендіру функциясы градиенттің жоғалуын болдырмайды және кері тарату кезінде оңтайландыруды жеңілдететіндіктен, нейрондық желілердің өнімділігін арттыруға көмектеседі. Алғашқы екі LSTM қабатының әрқайсысы 128 нейроннан тұрады және осы қабаттарда Relu белсендіру функциясы қолданылған.

Желінің артық үйрену (overfitting) мәселесін шешуге көмектескен Dropout және $L2$ реттеуі сияқты оңтайландыру әдістері модельде қолданылды. Dropout әдісі кейбір нейрондарды кездейсоқ өшіру арқылы модельдің сенімділігін арттырды, ал $L2$ реттеуі нейрондардың салмағын шектеу арқылы модельдің күрделілігін төмендетіп, жалпы өнімділік пен үйлесімділікке оң әсер етті.

Dropout әдісі нейрондық желілерді оңтайландыру үшін артық үйренудің алдын алу мақсатында қолданылды. Бұл әдіс оқыту барысында әрбір epoch кезінде нейрондардың белгілі бір санын өшіруді қамтамасыз етеді. Нәтижесінде, нейрондық желі белгілі бір түйіндерге шамадан тыс тәуелді болмайды және барлық нейрондар тең дәрежеде оқытылады. Dropout әдісінде нейронның өшірілу ықтималдығы P $[0, 1]$ шегінде анықталады.

Нейрондық желі түріне байланысты ықтималдықты реттеуге арналған жалпы нұсқаулар:

- ♦ Толық байланысқан (Dense) қабаттар үшін Dropout дәрежесі әдетте 0,2 мен 0,5 аралығында орнатылады. Бұл қабаттар шу мен шамадан тыс оқуға (overfitting) бейім болғандықтан, нейрондардың жартысына дейін өшірілуі мүмкін;

- ♦ CNN (конволюциялық нейрондық желілер) үшін P мәндері әдетте 0,1 мен 0,3 арасында орнатылады. Деректердегі маңызды ерекшеліктерді тиімді тануға қабілетті CNN желілері кеңістік ерекшеліктерін сақтайтын құрылымның арқасында тиімділігін арттырады және шамадан тыс оқудан (overfitting) көп зардап шекпейді. Сол себепті, басқа оңтайландыру әдістері қолданылғаннан кейін көбінесе Dropout қабаты енгізіледі;

♦ Модельдің тиімділігін арттырып, артық үйренуді (overfitting) болдырмауға көмектесу үшін, LSTM және RNN модельдері бойынша кіріс пен негізгі қабаттарында Dropout мәндерін дәл орнатқан маңызды. Модельдің тиімді жұмысын қамтамасыз етіп, нейрондардың тек қажетті бөлігі ғана белсенді болып, жасырын қабаттарда Dropout мәні 0,2-ге дейін азайып, кіріс қабаттарында Dropout мәні 0,4 пен 0,8 арасында болуы ықтимал. Себебі RNN жоғары мәнді Dropout мәліметтердің берілуіне кедергі келтіріп, оқыту процесін тұрақсыз етуі мүмкін.

Жоғарыда аталған ұсыныстарға сүйене отырып, соңғы нейрондық желіде бірінші кіріс қабаты үшін Dropout мәні 0,4 деп белгіленді, ал келесі екі қабатта бұл мән екі есе төмендетілді.

Барлық төрт қабатында нейрондық желінің $L2$ сызықтық реттеуі қолданылды, модельге шектеулер қою оқыту процесінде жүзеге асырылып, бұл оңтайландыру техникасы оның жалпыға бірдей болу қабілетін жақсартуға ықпал етеді. Регуляризация – бұл оқу процесіне жаттығу үлгісінде әсер етпейтін, тест үлгісіндегі шығын функциясының мәнін төмендетуге бағытталған модельде енгізілетін кішігірім өзгеріс.

Белгілі бір міндеттерді шешуге бағытталған регуляризацияның екі негізгі түрі бар. Нақтырақ айтқанда, $L1$ реттеуі (Lasso реттеуі) модельде артық салмақтардың пайда болуын алдын алу үшін қолданылады. Бұл әдісте қателік функциясы барлық бар салмақтардың модульдерінің қосындысын белгілі бір коэффициентке көбейтіп қосу арқылы есептеледі. Формула (1) $L1$ реттеуін қолдана отырып, шығын функциясының мәні келесідей болады:

$$J(\theta) = MSE + \lambda \sum |\theta|_i, \quad (1)$$

мұндағы $J(\theta)$ – шығын функциясы және MSE – бұл орташа квадратталған қателік болса, сонда $\lambda \sum |\theta|_i$ – $L1$ реттеуінің коэффициенті. Қате функциясында қосымша терминнің қосылуы нәтижесінде модель тек кішірек салмақтармен жұмыс істеуге мүмкіндік алады. $L1$ реттеуінің ерекшелігі – бұл кейбір салмақтардың нөлге айналуына әкелуі мүмкін, яғни модельде босатылған салмақтар пайда болады. Ал $L2$ реттеу (жолды реттеу) салмақтар тым үлкен болса, үлгі жазаланады. Әр салмақтың квадраты алынады және модельде барлық салмақтардың қосындылары есептеледі (модуль алынғанымен, таңбалары әртүрлі салмақтардың мәндері жойылмайды). Осыдан кейін, (2) формуласында көрсетілгендей, осыдан кейін RMS қатесіне негізделген жоғалту функциясы төмендегідей болады:

$$J(\theta) = MSE + \sum \theta_i^2, \quad (2)$$

Мұнда егер ұсынылған $J(\theta)$ – қателіктің функциясы болса, ендеше MSE – бұл орташа квадратталған қателік, ал машиналық оқытудың моделінің салмақтары (weights) квадраттарын қосуды – $\sum \theta_i^2$ деп белгілейміз. Бұл $L1$ реттеуінде көрінбейтін, бірақ $L2$ реттеуінде нөлге тең болатын разрядталған салмақтар. Нөлдің орнына маңыздылығы аз салмақтар нөлге жақын шағын мәндерді қабылдайды, бірақ әлі де үлгіге қатысады. Тапсырмада әр қимыл 1662 координаттың 30 кадрының 30 тізбегін білдіретін 30 қимылды ажырату керек болғандықтан, жаттығу үшін салмақтар аз болады және оларды модельден толығымен алып тастаудың қажеті жоқ. Сондықтан $L2$ реттеуі таңдалды.

Нәтижелер мен талқылау

Тәжірибелік компонент. Ұсынылған модель сурдоаудармашыны пайдаланбай, ым тілін қолданатын адамдармен тікелей қарым-қатынас орнатуға арналған. Модельді қоғамдық ұйымдардың қызметіне, мысалы, Халыққа қызмет көрсету орталықтары, банктер, университеттер және басқа да мекемелерге интеграциялау жоспарлануда.

Жобаны әзірлеу барысында Python бағдарламалау тілі және оның бірнеше кітапханалары, атап айтқанда MediaPipe, Tensorflow, Keras, Numpy, Matplotlib, Sklearn және басқа да қажетті құралдар қолданылды. Модельді оқыту кезінде нейрондық желіні оңтайландыру әдістерінің

бірі ретінде жүйелеу (regularization) қолдану ұсынылады. Сонымен қатар, нейрондық желілердің бірнеше архитектурасын тексеріп, олардың ішінен ең тиімдісін таңдау керек. Нейрондық желіні оқыту барысында шығындар функциясы ретінде кросс-энтропия және Adam оптимизаторын пайдалану ұсынылады. Модельдің тиімділігін бағалау үшін Accuracy сияқты негізгі дәлдік көрсеткішін қолдануға болады.

Нейрондық желіні оқыту. Модель компьютеріміздің ресурстарын үнемдеу үшін Kaggle Kernels ортасында оқытылды. Модуль 1-кестеде келтірілген ым-ишараға үйретілуі керек болатын:

Кесте 1 – Оқытуға арналған модельдің қимылдары

Қимылдардың нөмірі	Қимылдардың қазақ тілінде (латын) ұсынылуы	Қимылдардың ағылшын тілінде ұсынылуы	Қимылдардың орыс тілінде ұсынылуы
I.	Сәлеметсіз бе (salemetsiz be)	Hello	Здравствуйте
II.	Мен (men)	I	Я
III.	Үйрену (uyrenu)	Learn	Учить
IV.	Ым (ym)	Gesture	Жест
V.	Бүгін (bugin)	Today	Сегодня
VI.	Ауа райы (aua rayi)	Weather	Погода
VII.	Жақсы (jaksy)	Good	Хорошо
VIII.	Мен сені сүйемін (men seni suyemin)	I love you	Я тебя люблю
IX.	Суық (suuq)	Cold	Холодно
X.	Ыстық (ystyq)	Hot	Жарко
XI.	Рақмет (rakhmet)	Thank you	Спасибо
XII.	Телефон (telefon)	Telephone	Телефон
XIII.	Тыныш (tynysh)	Silent	Тихо
XIV.	Билеу (bileu)	Dance	Танцевать
XV.	Кеуде (keude)	Chest	Грудь
XVI.	Ойнау (oynau)	Play	Играть
XVII.	Ақылды (akyldy)	Clever	Умный
XVIII.	Ұшақ (ushaq)	Plain	Самолет
XIX.	Бет (bet)	Face	Лицо
XX.	Адам (adam)	Person	Человек
XXI.	Әтеш (atesh)	Cock	Петух
XXII.	Такси (taxi)	Taxi	Такси
XXIII.	Дөрекі (doreki)	Rude	Грубый
XXIV.	Ит (it)	Dog	Собака
XXV.	Керемет (keremet)	Perfect	Отлично
XXVI.	Жаман (jaman)	Bad	Плохо
XXVII.	Су (su)	Water	Вода
XXVIII.	Сау болыңыз (sau bolynyz)	Good bye	До свидания
XXIX.	Қатты (katty)	Loud	Громко
XXX.	Үлкен (ulken)	Big	Большой

Kaggle платформасында толық деректер жиынтығы ретінде құрылып, ым тілінің қимылдары бір мұрағатта жинақталды. Қимылдарды өңдеу кезінде олардың толық құрылымын қалыптастыру мақсатында, бір қимылдың барлық тізбектерін бір-біріне қосу қажеттілігі туындады, өйткені әр қимыл үшін жеке каталогтар бар және әрбір каталогта кадрлардан тұратын реттіліктер орналасқан.

Келесі кезеңде sklearn кітапханасын пайдаланып, жалпы деректер жиынтығы 90/10 қатынасында оқу және тест үлгілеріне бөлінді. Зерттеу жұмысымызда негізгі мақсат – классификация тапсырмасын орындау болғандықтан, шығын функциясы ретінде көпкатегориялы кросс-энтропия пайдаланылды, ал оңтайландыру үшін модельде Adam алгоритмі қолдану таңдалды. Оқуды ерте тоқтату әдісі оңтайландыру процесінде қолданылады, бұл әдіс модельдің артық үйренуін тежеп, тиімді оқыту процесін қамтамасыз етуге мүмкіндік береді, бұл шамадан тыс оқытуды болдырмау және градиенттің жоғалуын алдын алу үшін көмектесті. Бұл жағдайда, егер қате көрсеткіші 10 қатарынан epoch бойы өзгермесе (patience = 10), онда оқыту ағымдағы epoch-тан кейін тоқтайды, тіпті оқу үшін әлі де epoch қалған болса да.

Модель 150 epoch бойы жаттықтырылды, ал негізгі оқу метрикасы ретінде көпкатегориялы дәлдік (categorical_accuracy) таңдалды (себебі бұл классификация тапсырмасы). Тест деректерінде нақты уақыт режимінде тестілеу жүргізгенде, LSTM және Dropout + L2 регуляризациясымен RNN моделі 92% жалпы дәлдік көрсеткені 2-кестеде ұсынылған.

Кесте 2 – LSTM және Dropout + L2 регуляризациясы бар RNN моделінің нәтижелері

Модель	Метрикасы	Мәні	Эксперименттік шарттар
RNN with LSTM and Dropout + L2 Regularization	Accuracy	92%	30 кадр әр ишараға, 30 класстан

LSTM және Dropout + L2 регуляризациясы бар RNN моделі жоғары дәлдік пен жылдам өңдеу уақытын көрсетеді, бұл оны YOLOv5 және SVM моделдерімен бәсекеге қабілетті етеді (3-кесте).

Кесте 3 – Ым тілі қимылдарын тану үшін модельдерді және олардың сипаттамаларын салыстыру

Модель	Деректер жиыны	Дәлдігі (%)	Кадрды өңдеу уақыты (мс).	Іске асырудың күрделілігі
YOLOv5	[8]	99,6	-	Жоғары
SVM	[13]	99	-	Жоғары
RNN with LSTM and Dropout + L2 Regularization	Біздің деректер жиынтығы	92	30	Орташа

Сынақ үлгісі 100 кадр тізбегінен тұрды және бір қимылда орта есеппен 3 кадр болды, сондықтан үлгіні теңдестірілген деп санауға болады. Оқу және бастапқы бағалау кезеңін аяқтағаннан кейін модельді нақты уақыт режимінде тексерілді. Бұл сценарий, деректерді белгілеуді жүзеге асырып, оларды бөлек функция арқылы өңдеп, модельге жіберу процесін автоматтандырады, алдымен камерада 30 кадрды оқытын нейрондық желі үшін дайын сынақ сценарийі іске қосылды, бұл модельдің кірісі үшін деректерді белгілеу мақсатында болатын. Тиімді оқыту процесі қамтамасыз етіледі, егер де осылайша жеке бір функция арқылы оқылып, модельге қосылып, дайын файл ашып, ол модельдің оқыту нәтижесінде алынған және бөлек сақталған оңтайланған салмақтары бар болуы қамтамасыз етілсе, бұл модельдің дәлдігін және өнімділігін арттыруға мүмкіндік береді.

Содан соң, сол функцияда кіріс ретінде алынған мәліметтер өңделіп, салмақтармен бірге нейрондық желінің кірісіне жіберіледі, нәтижесінде болжам жасалады. Функциядан алынған болжамдар класс белгісі түрінде қайтарылады, олар көрсетілген қимылды нақты бір кадрда анықтап, оқу процесінде қолданылып, қимылдардың дәл сәйкестігі анықталады.

Бірінші кезекте камера объектісі құрылып, оған (0) параметрлі Video Capture тағайындалады. Бұл әдетте жүйенің қателіктерге қарсы тұрақтылығын арттырады, бұл процесс автоматты түрде орындалуды тоқтатып, кадрды оқу барысында скриптта ақау байқалса, жүйенің дұрыс жұмыс істеуін қамтамасыз ету үшін destroyAll Windows () әдісі іске қосылады, кейін кадр арнайы функция арқылы өңделеді, алдымен қате коды мен кадр түрі бойынша оқылып, камерадан бейне ағыны 0 тағайындалады, өйткені көптеген ноутбуктерде камераның идентификаторы енгізу параметрі ретінде 0 болады.

MediaPipe кітапханасында кірістірілген process әдісін пайдаланып, беттегі және денедегі поза нүктелері анықталады. Алдымен draw_face() әдісі, содан кейін draw_hands(), кейін draw_pose() сияқты әртүрлі бөлек әдістер арқылы өңделіп, бұл әдістер нәтижесінде алынған ақпарат бейне ағынына қайтадан жолданады. Нәтижесінде, алынған нүктелер бейнеде бір-бірімен жалпылай сызықтармен байланысып, олар нақты нүктелер ретінде көрініс алады, осылайша бейнедегі адамның беті мен позасы нақты әрі айқын көрсетіледі.

Бұл дегеніміз, кадрда нүктелерді көрсету міндетті емес сияқты көрінуі мүмкін, себебі бізге тек болжанған қимылды сөз түрінде көру жеткілікті, ал кадрда нүктелерді артық көрсету қажет емес. Алайда тәжірибе көрсеткендей, модель екі қолды әрқашан анықтай алмайды, сондықтан бұл жерде нүктелерді көрсету қажеттілігі туындайды.

Модельді түзету және оңтайландыру арқылы ресурстарды жұмсау артық және болашақта кері әсерін тигізуі мүмкін, осындай жағдайда кадрда екі қолдың нүктелері көрсетілмесе, модель толық емес деректер негізінде қате болжамдар жасауы мүмкін немесе мүлдем болжам жасай алмауы ықтимал.

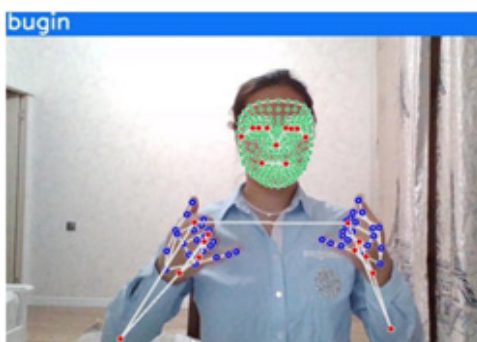
Осы фактіні ескере отырып, барлық нүктелер кадрға орналастырылады. Содан кейін модель алынған кадрларды өңдеуді бастайды. Алдымен скрипт 30 кадр жинауды қамтамасыз етеді.

Кейіннен 30 кадрдан тұратын видео ағысы жинақталып, одан кейін модельдегі predict функциясы шақырылады. Бұл функцияның нәтижесі answ айнымалысына енгізіліп, соңында болжам жасалады. Барлық болжанған сыныптар арасынан ең жоғары ықтимал мәні таңдалады. Егер бұл мән алдын ала белгіленген шектен асып кетсе, ол дұрыс нәтиже ретінде қабылданады.

Бұл модель нақты осы қимыл бейнеленгенін көрсетеді, өйткені ықтималдық мәндерінің ішінен ең жоғарысы argmax() әдісі арқылы таңдалады. Соңында 30 ықтимал мән қайтарылады, бұл жағдайда 30 қимыл = 30 класс құрайды. Массивтің ұзындығы жалпы сыныптардың санына тең болады және ықтималдық мәндерінің толық массиві беріледі. Мұнда модельдің жауабы тек бір нақты санды, сынып белгісін немесе жеке ықтималдық мәнін емес, барлық ықтималдық мәндерін қамтиды, себебі ол көпсыныпты классификация міндетін шешеді. Болжанған ым тілінің қимылы браузерде 2–3-суреттерде көрсетілгендей бейнеленеді:

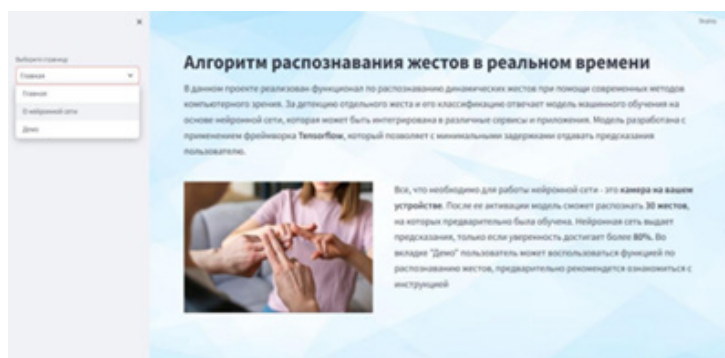


Сурет 2 – «Salemetsiz be!» деген қимылдың браузердегі бейнесі



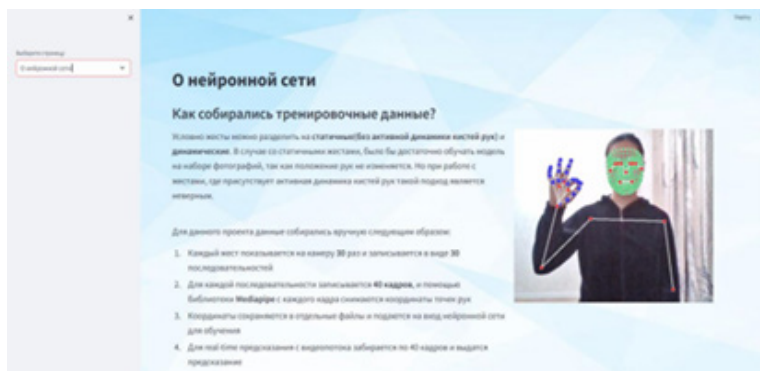
Сурет 3 – «Bugin» деген қимылдың браузердегі бейнесі

Веб-интерфейстің кідіріссіз жұмыс істеуі өте маңызды болды. Сондықтан оның таңдалуының басты себебі – пайдаланушыларға модельдермен нақты уақытта жұмыс істеу мүмкіндігін беруі. Бұл құралдың негізгі артықшылығы – Streamlit ашық бастапқы кодты платформа ретінде Python негізінде машиналық оқыту жобаларына арналған дербес веб-қосымшаларды құруға мүмкіндік береді. Сынақтан өткен соң нейрондық желімен пайдаланушының ыңғайлы өзара әрекеттесуін қамтамасыз ету үшін арнайы қарапайым веб-қосымша интерфейсін құру қажеттілігі туындады. Веб-қосымшаның бөліктерін әзірлеу үшін Streamlit фреймворкі таңдалды. Қосымша үш беттен тұрады: алғашқысы – басталу парақшасы, онда жұмыстың басты мақсаты көрсетілген (4-сурет).



Сурет 4 – Браузердегі «Басталу» парақшасы

Қолданушы «Нейрондық желі туралы» тармағына басқанда, ол екінші бетке өтеді, онда нейрондық желінің архитектурасы және оқу деректерін жинаудың артықшылықтарына байланысты қысқаша ақпарат ұсынылады (5-сурет). Бұл мәліметтер қалқымалы навигация панеліндегі бүйірлік мәзірде орналасқан сол жақ бөлігі арқылы қолжетімді.



Сурет 5 – Браузердегі «Ақпарат алу» парақшасы



Сурет 6 – Браузердегі «Көрсету» парақшасы

Қолданушы қимылдарды көрсетуге және сол терезеде болжамдарды алуға мүмкіндік алады, түймені басқан сәтте камераның терезесі ашылып, «Көрсету» бетінде болжамдардың дәлдігін қамтамасыз ету үшін машиналық оқытудың үлгісін пайдалану бойынша нұсқаулықтар берілген (6-сурет). Кейін, «Көрсету» тармағын басу қолданушы ым тілі қимылдарын тану үшін тікелей үлгіні қолдана алатын негізгі бетке апарды.

Қорытынды

Бұл мақалада камерада 30 қазақ ым тілі қимылын нақты уақытта тануға және оларды мәтінге айналдыруға қабілетті LSTM нейрондық желісі әзірленіп, тестіленді. Сондай-ақ, нейрондық желі мен пайдаланушы арасындағы ыңғайлы өзара әрекеттесу үшін шағын веб-интерфейс жасалды. Жобаның негізгі артықшылықтарының бірі – деректерді жинаудағы стандартты емес, сонымен қатар шаблонсыз тәсіл, бұл бір ым тілі қимылын кадрлар тізбегі ретінде қабылдауға мүмкіндік береді, осылайша динамикалық қимылдарды дұрыс өңдеуге жол ашады. Барлық жинақталған деректер қазақ ым тілінің материалдарына жатады, бірақ болашақта орыс тілі сөздері мен алфавитін ескере отырып, оқу үлгісінің көлемін кеңейту жоспарланған. Қазақстандағы ең басты қарым-қатынас тіліне арналған жалғыз ым тілі қимылы базасының болуы нейрондық желіні жасаудың алғашқы қадамы болады, ол болашақта әртүрлі жүйелер мен қосымшаларға интеграцияланады. Нейрондық желінің болжамдау дәлдігі 80%-дан жоғары, бұл қазірдің өзінде жақсы көрсеткіш, бірақ болашақта архитектураны оңтайландыру арқылы дәлдік шегін 95%-ға дейін арттыру жоспарланған. Сондай-ақ, нейрондық желі моделі дискідегі бос орын тұрғысынан ықшам екенін атап өткен жөн, өйткені тек салмақтар жиынтығы сақталады, олар шығаруға оңай пішімде сақталған нақты сандар.

Осылайша, болжамдардың жоғары дәлдігі, кеңістіктің аздығы және архитектураның өнімді Python сценарийі түріндегі көрінісі нейрондық желіні кейіннен дайын шешімдермен біріктіруге ыңғайлы етеді. Есту және сөйлеу қабілеті нашар адамдарға қызмет көрсету жағдайында әлеуметтік қызметтерді көрсетудегі инклюзивтілік деңгейін арттырады. Нейрондық желіні әртүрлі қосымшалар мен қызметтерге жаппай енгізгеннен кейін, есту және сөйлеу қабілеті нашар адамдар сурдоаудармашы қызметтерін қолданбай-ақ мемлекеттік мекемелерге оңай бара алады.

Қаржыландыру туралы ақпарат. Бұл зерттеу Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (Грант № BR24992875).

ӘДЕБИЕТТЕР

- 1 Sharma, S., Singh, S. Recognition of Indian sign language (ISL) using deep learning model. *Wireless personal communications*, 123 (1), 671–692 (2022). <https://doi.org/10.1007/s11277-021-09152-1>.
- 2 Mukhanov, S. et al. Gesture recognition of the Kazakh alphabet based on machine and deep learning models. *Procedia Computer Science*, 241, 458–463 (2024). <https://doi.org/10.1016/j.procs.2024.08.064>.
- 3 Wu, J. et al. Data glove-based gesture recognition using CNN-BiLSTM model with attention mechanism. *Plos one*, 18 (11), e0294174 (2023). <https://doi.org/10.1371/journal.pone.0294174>.
- 4 Mohamed, T. et al. Intelligent hand gesture recognition system empowered with CNN. 2022 International Conference on Cyber Resilience (ICCR). IEEE, 2022, pp. 1–8. <https://doi.org/10.1109/iccr56254.2022.9995760>.
- 5 Mesbahi, S.C. et al. Hand gesture recognition based on various deep learning YOLO models. *International Journal of Advanced Computer Science and Applications*, 14 (40) (2023). <https://doi.org/10.14569/ijacsa.2023.0140435>.
- 6 Doždor, Z. et al. TY-Net: Transforming YOLO for hand gesture recognition. IEEE access.2023, <https://doi.org/10.1109/access.2023.3341702>.
- 7 Zhang, Z., Wu, B., Jiang, Y. Gesture recognition system based on improved YOLO v3. 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 2022, pp. 1540–1543. <https://doi.org/10.1109/icsp54964.2022.9778394>.
- 8 Ling, L., Tao, J., Wu, G. Research on gesture recognition based on YOLOv5. 2021 33rd Chinese Control and Decision Conference (CCDC). IEEE, 2021, pp. 801–806. <https://doi.org/10.1109/ccdc52312.2021.9602731>.
- 9 Kumar, R., Bajpai, A., Sinha, A. MediaPipe and cnns for real-time asl gesture recognition. arXiv preprint arXiv:2305.05296, 2023. <https://doi.org/10.48550/arxiv.2305.05296>.
- 10 Chong, K.S., Subaramaniam, K., Al-Hadi, I.A.A.Q. Developing a Prototype Hand Gesture Recognition System in Interpreting American Sign. *Interpretation*, 3, 1 (2022). <https://doi.org/10.5954/icarob.2024.gs7-4>.
- 11 Grif, M.G., Kondratenko, Y.K. Recognition of Isolated Gestures of the Russian Sign Language Based on the Component Approach //2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE). IEEE, 2023, pp. 1510–1513. <https://doi.org/10.1109/apie59731.2023.10347694>.
- 12 Samaan, G.H. et al. MediaPipe's landmarks with rnn for dynamic sign language recognition. *Electronics*, 11 (19), 3228 (2022). <https://doi.org/10.3390/electronics11193228>.
- 13 Zholshiyeva, L. et al. A Real-Time Approach to Recognition of Kazakh Sign Language //2022 International Conference on Smart Information Systems and Technologies (SIST). IEEE, 2022, pp. 1–6. <https://doi.org/10.1109/sist54437.2022.9945799>.
- 14 Amirgaliyev, Y., Ataniyazova, A., Buribayev, Z., Zhassuzak, M., Urmashiev, B., & Cherikbayeva, L. Application of neural networks ensemble method for the Kazakh sign language recognition. *Bulletin of Electrical Engineering and Informatics*, 13(5), 3275–3287 (2024). <https://doi.org/10.11591/eei.v13i5.7803>.
- 15 Yerimbetova, A., Sakenov, B., Berzhanova, U., Mukazhanov, N., Daiyrbayeva, E., & Othman, M. Development of A Model of Kazakh Sign Language Recognition Based on Deep Learning Method. In 2024 9th International Conference on Computer Science and Engineering (UBMK) (pp. 822-827). IEEE, 2024, October. <https://doi.org/10.11591/eei.v13i5.780310.1109/UBMK63289.2024.10773578>.
- 16 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>.
- 17 Vijitkunsawat, W., Anunvapong, P., & Chantngarm, P. Human Joint Coordinate Sequencing in Video-Based Thai Finger Spelling Recognition. In 2024 8th International Conference on Information Technology (InCIT) (pp. 681–686). IEEE, 2024, November. <https://doi.org/10.1109/InCIT63192.2024.10810526>.
- 18 Katti, R.K., Sujatha, C., Desai, P., & Shankar, G. Character and word level gesture recognition of Indian Sign language. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1–6). IEEE, 2023, April. <https://doi.org/10.1109/I2CT57861.2023.10126314>.
- 19 Feng, Y., Chen, N., Wu, Y., Jiang, C., Liu, S., & Chen, S. DFCNet+: Cross-modal dynamic feature contrast net for continuous sign language recognition. *Image and Vision Computing*, 151, 105260 (2024). <https://doi.org/10.1016/j.imavis.2024.105260>.
- 20 Aleshin, N.A. Recurrent neural networks. *World science: problems and innovations*, pp. 10–12, 2021.

^{1,2,3}**Еримбетова А.,**

к.т.н., ассоциированный профессор, PhD, ORCID ID: 0000-0002-2013-1513,

e-mail: aigerian8888@gmail.com

^{2*}**Бержанова У.Г.,**

докторант, ORCID ID: 0009-0000-2467-5721,

*e-mail: berzhanovaulmekenn@gmail.com

^{1,4}**Дайырбаева Э.,**

преподаватель, ORCID ID: 0000-0002-4255-5456,

e-mail: nurbekkyzy_e@mail.ru

¹**Сакенов Б.,**

инженер-программист, ORCID ID: 0000-0002-9849-6176,

e-mail: sbakzhan22@gmail.com

⁵**Самбетбаева М.,**

PhD, ORCID ID: 0000-0001-9358-1614

e-mail: madina_jgtu@mail.ru

¹Институт информационных и вычислительных технологий КН МНВО РК,
г. Алматы, Казахстан

²Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

³Евразийский технологический университет, г. Алматы, Казахстан

⁴Казахский национальный исследовательский технический университет им. К.И. Сатпаева,
г. Алматы, Казахстан

⁵Евразийский университет им. Л.Н. Гумилева, г. Астана, Казахстан

СОЗДАНИЕ МОДЕЛИ ДЛЯ РАСПОЗНАВАНИЯ КАЗАХСКОГО ЖЕСТОВОГО ЯЗЫКА В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ С ПОМОЩЬЮ MEDIAPIPE И МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

Аннотация

В данной статье рассматривается процесс разработки системы распознавания казахского жестового языка с использованием платформы MediaPipe. Платформа позволяет эффективно распознавать жесты в реальном времени. Основное внимание уделено созданию моделей для распознавания жестов, обучению нейронных сетей и интеграции с платформой MediaPipe. Один из важных аспектов – достижение высокой точности и скорости обработки жестов с использованием архитектуры нейронных сетей. Система была обучена на большом наборе аннотированных жестов, что значительно улучшило качество распознавания. Для распознавания жестов казахского языка использовалась нейронная сеть типа LSTM, так как она эффективно работает с временными рядами и последовательностями данных. Модель была обучена на 30 жестах казахского жестового языка, что позволяет преобразовывать жесты в текст в реальном времени. Этот подход значительно облегчает коммуникацию с людьми, имеющими проблемы со слухом и речью, и способствует повышению инклюзивности. Кроме того, был разработан удобный веб-интерфейс, который позволяет легко интегрировать нейронную сеть с приложениями для распознавания жестов. Одним из ключевых аспектов работы является улучшение методов аннотирования и обработки данных для повышения точности распознавания. Будущее развитие системы включает расширение базы данных жестов и интеграцию с веб-приложениями. Это позволит улучшить социальную интеграцию людей с нарушениями слуха и речи и создать широкую и доступную платформу.

Ключевые слова: казахский жестовый язык, MediaPipe, глубокое обучение, искусственная нейронная сеть, режим реального времени.

^{1,2,3}**Yerimbetova A.,**

Cand. Tech. Sc., Associate Professor, PhD, ORCID ID: 0000-0002-2013-1513,
e-mail: aigerian8888@gmail.com

^{2*}**Berzhanova U.G.,**

PhD student, ORCID ID: 0009-0000-2467-5721,
e-mail: berzhanovaulmekenn@gmail.com

^{1,4}**Daiyrbayeva E.,**

Senior Lecturer, ORCID ID: 0000-0002-4255-5456,
e-mail: nurbekkyzy_e@mail.ru

¹**Sakenov B.,**

Software engineer, ORCID ID: 0000-0002-9849-6176,
e-mail: sbakzhan22@gmail.com

⁵**Sambetbayeva M.,**

PhD, ORCID ID: 0000-0001-9358-1614,
e-mail: madina_jgtu@mail.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

²Al-Farabi Kazakh National University, Almaty, Kazakhstan

³Eurasian Technological University, Almaty, Kazakhstan

⁴Satbayev University, Almaty, Kazakhstan

⁵L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

DEVELOPMENT OF A MODEL FOR REAL-TIME RECOGNITION OF KAZAKH SIGN LANGUAGE USING MEDIAPIPE AND DEEP LEARNING METHODS

Abstract

This article discusses the process of developing a Kazakh sign language recognition system using the MediaPipe platform. The platform allows for efficient real-time gesture recognition. The main focus is on creating models for gesture recognition, training neural networks, and integrating with the MediaPipe platform. One of the key aspects is achieving high accuracy and speed in gesture processing by using neural network architecture. The system was trained on a large dataset of annotated gestures, which significantly improved the recognition quality. For recognizing Kazakh sign language gestures, an LSTM neural network was used because it effectively works with time series and data sequences. The model was trained on 30 Kazakh sign language gestures, enabling the conversion of gestures into text in real-time. This approach greatly facilitates communication with people who have hearing and speech impairments and contributes to increased inclusivity. Additionally, a user-friendly web interface was developed, allowing easy integration of the neural network with applications for gesture recognition. One of the key aspects of the work is improving data annotation and processing methods to enhance recognition accuracy. The future development of the system includes expanding the sign language gesture database and integration with web applications. This will improve social inclusion for people with hearing and speech impairments and create a broad, accessible platform..

Key words: Kazakh sign language, MediaPipe, deep learning, artificial neural network, in real-time.

Мақаланың редакцияға түскен күні: 13.05.2025