

ӘОЖ 519.872.1
ҒТАХР 28.23.25

<https://doi.org/10.55452/1998-6688-2026-23-1-37-51>

¹ **Черикбаева Л.Ш.,**

PhD, ORCID ID: 0000-0001-8948-4205,
e-mail: cherikbayeva.lyailya@gmail.com

² **Бериков В.Б.,**

т.ғ.д., профессор, ORCID ID: 0000-0002-5207-9764,
e-mail: berikov@math.nsc.ru

^{1*} **Меліс З.М.,**

докторант, ORCID ID: 0000-0003-4244-8121,
e-mail: melis.zarina98@gmail.com

¹ **Елеусинов А.И.,**

докторант, ORCID ID: 0000-0002-0425-6527,
e-mail: armankaznu@mail.ru

¹ **Адилжанова С.А.,**

PhD, ORCID ID: 0000-0003-1768-064,
e-mail: asaltanat81@gmail.com

¹ **Атаниязова А.С.,**

докторант, ORCID ID: 0000-0003-1122-6614,
e-mail: aisulu.ataniyazova@gmail.com

³ **Дайырбаева Э.Н.,**

докторант, ORCID ID: 0000-0002-4255-5456,
e-mail: nurbekkyzyelmira@gmail.com

¹ Өл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

² Новосібір мемлекеттік университеті, Новосібір қ., Ресей

³ Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті,
Алматы қ., Қазақстан

ТРАНСФОРМЕРЛЕР НЕГІЗІНДЕГІ АНСАМБЛЬ КӨМЕГІМЕН 3D КТ КЕСКІНДЕРІНДЕ ИНСУЛЬТТИ СЕГМЕНТАЦИЯЛАУ

Аңдатпа

Ишемиялық инсульт өлім-жітім мен мүгедектіктің басты себептерінің бірі саналады. Бас миының КТ-суреттеріндегі зақымдалған аймақтарды дәл сегментациялау дер кезінде диагноз қою және клиникалық шешімдер қабылдауда маңызды. Бұл мақалада SE-UNETR және Swin UNETR трансформерлік модельдерін салмақталған дауыс беру арқылы біріктіруге негізделген ансамбльдік әдіс ұсынылады. Бағалау үшін Дайс коэффициенті қолданылды, бұл метрика зақымдану аймақтарының болжамды нәтижелері мен эталондық белгілерінің қиылысу дәрежесін өлшейді. Жалғыз модельдерді пайдаланудан айырмашылығы, ансамбльдік нейрондық желілік тәсілдер бірнеше архитектураның болжамдарын үйлестіре біріктіру арқылы сегментацияның анағұрлым жоғары сенімділігі мен дәлдігін қамтамасыз етеді. Ресей Ғылым академиясының Сібір бөлімшесінің халықаралық томография орталығы ұсынған жедел ишемиялық инсульт диагнозы қойылған 98 пациенттің үшөлшемді КТ-суреттері пайдаланылды. Нәтижелер ұсынылған ансамбльдің жеке модельдерден жоғары өнімділік көрсеткенін дәлелдеді. Dice коэффициентінің орташа мәні 0.7983 құрап, ишемиялық ошақтарды сегментациялаудағы әдістің жоғары тиімділігін көрсетті. Талдау барысында ансамбльдік әдіс бас миының КТ-суреттеріндегі зақым шекараларын дәлірек анықтап, сегментация қателерін

азайтты. Ұсынылған әдіс инсультпен қатар, медициналық бейнелерді нақтылы талдау қажет болатын басқа да патологияларды автоматтандырылған түрде анықтау жүйелерінде қолданыс таба алады.

Тірек сөздер: UNETR, Swin Transformer, компьютерлік томография, ишемиялық инсульт, тереңдетілген оқыту, сегментация, модельдер ансамблі.

Кіріспе

Ишемиялық инсульт әлем бойынша өлім-жітім мен мүгедектіктің негізгі себептерінің бірі болып қала беретін ең ауыр аурулардың бірі. Жыл сайын миллиондаған адам инсульттің құрбанына айналып, бұл әртүрлі елдердегі денсаулық сақтау жүйесі үшін елеулі мәселе тудырады. Науқастың болжамын айқындайтын негізгі факторлардың бірі диагностика жылдамдығы мен дәлдігі. Инсультті дер кезінде және сапалы диагностикалау тиімді емдеуді бастап, қайтымсыз салдар мен еңбекке қабілеттіліктің жоғалу қаупін едәуір төмендетуге мүмкіндік береді. Соңғы жылдары медициналық кескіндерге негізделген диагностика әдістерін жетілдіру үшін магниттік-резонанстық томография (МРТ) мен компьютерлік томографияны (КТ) терең оқыту [1] сияқты заманауи технологияларды қолдануға айрықша назар аударылуда.

Медициналық визуализация инсультті диагностикалауда шешуші рөл атқарады, ал мисуреттерін сегментациялау – науқастың жағдайын бағалау үдерісіндегі маңызды кезең. Сегментацияның дәстүрлі әдістері, мысалы, шекті мән бойынша фильтрация [2] немесе белсенді контурларды пайдалану [3], ишемиялық инсульт ошақтарын дәл әрі жылдам бөліп көрсету мүмкіндігінде шектеулі, әсіресе 3D кескіндерде.

Соңғы жылдары медициналық кескіндерді сегментациялау саласындағы айтарлықтай жетістік 3D конволюциялық нейрондық желілерді (3D CNN) [4] қолдану, олар томографиялық деректерден кеңістіктік ақпаратты дәлірек алуға мүмкіндік береді. Қазіргі заманғы терең оқыту модельдері, соның ішінде CNN, кең ауқымды міндеттерде қолданыс табуда. Маңызды міндеттердің бірі – 3D медициналық кескіндермен (МРТ және КТ) тиімді жұмыс істей алатын архитектураларды әзірлеу, өйткені кеңістіктік тәуелділіктерді алу дәл диагноз қою үшін аса маңызды. Ең танымал әрі тиімді архитектуралардың бірі – 3D U-Net. 2021 жылдан бері жүргізілген зерттеулерде бұл модель толық конволюциялық желілерге (Fully Convolutional Networks, FCN) және V-Net модельдеріне қарағанда, әсіресе ишемиялық инсультті қоса алғанда күрделі патологияларды сегментациялауда, анағұрлым жоғары дәлдік көрсететіні анықталды [5–10]. Ишемиялық инсультті медициналық кескіндер негізінде диагностикалау дәлдігі мен тұрақтылығын арттыру үшін бірнеше модельдің болжамдарын біріктірудің маңыздылығы айқындалды [11–12].

[13] мақала авторларының еңбегінде контрастсыз КТ-кескіндерде жедел ишемиялық инсульт ошақтарын сегментациялау үшін 3D конволюциялық нейрондық желілерді қолдану қарастырылған. Авторлар медициналық деректерді автоматтандырылған өңдеуде терең оқытудың тиімділігін көрсетіп, диагностика дәлдігін арттыру үшін үшөлшемді талдаудың маңыздылығын ерекше атап өтеді. Бұл тәсіл инсультті сегментациялау міндеттерінде 3D CNN қолданудың болашағын растап, ансамбльдік әдістерді одан әрі дамытуға негіз бола алады.

Бірақатар зерттеулерге сәйкес, 3D U-Net архитектурасы әртүрлі масштабтарда кеңістіктік белгілерді алу қабілетінің арқасында кеңінен қолданылады және медициналық кескіндерді сегментациялау үшін ең тиімді модельдердің бірі болады. Дегенмен, ядро өлшемінің тұрақты болуына байланысты кеңістіктік контекстіні шектеулі қамту, бейнелер аймақтары арасындағы алыс тәуелділіктерді анықтауда оның тиімділігін төмендетуі мүмкін. Бұл әсіресе ишемиялық инсультті сегментациялауда өзекті, өйткені патологиялық ошақ шекаралары бұлдыр немесе анық емес болуы ықтимал. Бұл шектеуді еңсеру мақсатында соңғы жылдары бастапқыда табиғи тілді өңдеу міндеттеріне арналған трансформерлік архитектураларға деген қызығушылық артып келеді. Осы негізде U-тәрізді құрылым мен трансформерлік энкодерді біріктіретін гибриді архитектуралар пайда болды. Мұндай тәсілдің ең танымал өкілдерінің бірі – UNETR

(U-Net with Transformers) моделі, [14] мақала авторлары ұсынған, мұнда энкодер ретінде Vision Transformer (ViT) қолданылады. Бұл тәсіл кескіндерді алдын ала агрегаттауды қажет етпей кеңістіктік белгілерді алуға мүмкіндік береді және сегментация дәлдігін арттырады. Осы идеяның дамуы ретінде Swin UNETR моделі ұсынылды. Ол Swin Transformer негізінде құрастырылған және терезелік өзіндік назар мен белгілердің иерархиялық құрылымын пайдаланады. Swin UNETR ми құрылымдарын сегментациялауда жоғары дәлдік көрсетіп, соңғы зерттеулерге сәйкес [15–17], инсультті автоматтандырылған диагностикалау міндеттерінде қолдануға айтарлықтай әлеуетке ие.

Бұл зерттеудің мақсаты – үшөлшемді КТ-суреттерде ишемиялық инсультті сегментациялауға арналған жоғары дәлдікті әрі тұрақты нейрожелілік модельді әзірлеу. Трансформерлік модельдер ансамблін (SE-UNETR және Swin UNETR) пайдалану жеке модельдерге қарағанда 3D-КТ кескіндерінде ишемиялық инсульт ошақтарын сегментациялауда анағұрлым жоғары дәлдік пен тұрақтылықты қамтамасыз етеді. Осы мақсатқа жету үшін жұмыста ми КТ-кескіндерінде ишемиялық ошақтарды автоматтандырылған 3D-сегментациялауға арналған ансамбльдік тәсіл ұсынылады. Ансамбль екі трансформерлік архитектурадан құралады: SE-UNETR моделі – бұл UNETR-дің модификацияланған нұсқасы, оған Squeeze-and-Excitation блоктары енгізілген; және Swin UNETR архитектурасы, ол терезелік өзіндік назар мен белгілердің иерархиялық құрылымына негізделген. Болжамдарды біріктіру үшін салмақталған дауыс беру механизмі қолданылды, ол модельдердің сенімділік деңгейін ескеріп, соңғы нәтижені қалыптастыруға мүмкіндік береді. Ұсынылған тәсіл ишемиялық инсульт диагнозы расталған 98 науқастың нақты медициналық деректерінде сыналды.

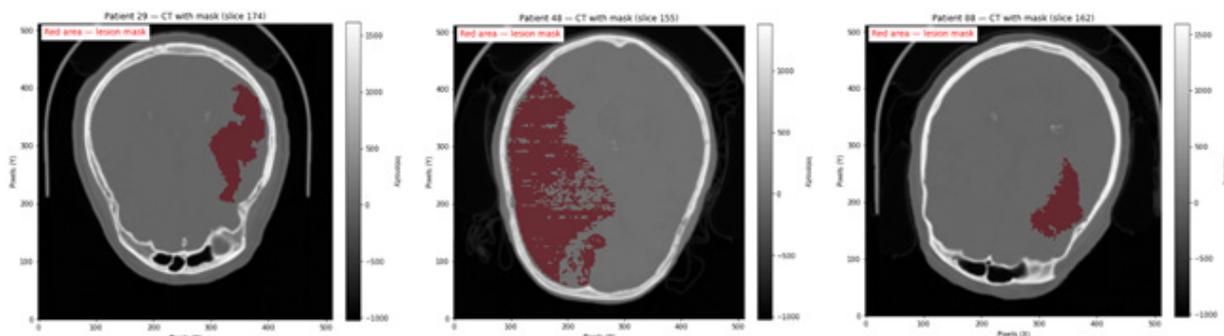
Ұсынылып отырған әдістің жаңашылдығы инсультті контрастсыз КТ-кескіндерде 3D-сегментациялау міндеті үшін трансформерлік модельдер ансамблін жүйелі түрде қолдану мен бағалауда жатыр. Қолданыстағы жұмыстар көбіне жекелеген архитектураларға немесе олардың модификацияларына бағытталса, бұл зерттеу әртүрлі назар механизмдеріне негізделген модельдерді біріктіру сегментация дәлдігін айтарлықтай арттыра алатынын көрсетеді. Ұсынылған зерттеу жұмысы алғаш рет инсультті КТ-кескіндерді сегментациялау міндетінде трансформерлік модельдерді ансамбльдеу тиімділігінің салыстырмалы талдауын жүргізеді. Және осы клиникалық деректер жиынтығында SE-UNETR мен Swin UNETR модельдерінің ансамбліне негізделген тәсілдің тиімділігі бұрын-соңды зерттелмеген. Нақты клиникалық деректер негізінде жүргізілген эксперименттер ұсынылған ансамбльдің тиімділігін растап, жұмыстың түпнұсқалығын айқындайды және медициналық визуализация саласында сенімді шешімдер әзірлеу үшін оның практикалық маңыздылығын атап көрсетеді.

Материалдар мен әдістер

Осы зерттеу үшін жедел ишемиялық инсульт диагнозы қойылған 98 науқастың бас миының КТ-суреттер жиынтығы пайдаланылды. Деректер Philips Ingenuity CT компьютерлік томографиясында алынып, Ресей Ғылым академиясының Сібір бөлімшесінің халықаралық томография орталығы ұсынған [13], бұл олардың жоғары сапасы мен дәлдігін қамтамасыз етеді. Барлық деректер жиынтығы медициналық бейнелеу және нейроғылым саласында стандарт болып табылатын NIfTI [18] форматында сақталған. Әрбір КТ-сурет нақты пациенттің ерекшеліктеріне байланысты 306-дан 505-ке дейінгі срездер сериясынан тұрады. Әр срездің өлшемі 512x512 пиксельді құрайды, бұл талдау үшін жеткілікті деңгейде нақтылықты қамтамасыз етеді. Әрбір срездің қалыңдығы 0,5 мм-ге тең, бұл қазіргі заманғы КТ-сканерлер үшін стандартты параметр болады және сегментация мен талдаудың қажетті дәлдігін қамтамасыз етеді. Сегментацияны радиология және радиациялық терапия саласында 9–13 жыл тәжірибесі бар, PhD дәрежесіне ие екі маман 3D Slicer бағдарламалық жасақтамасын пайдалана отырып жүргізді [19]. Сегментация барысында ишемиялық инсультпен зақымдалған ми аймақтары бөлініп көрсетілді, бұл деректерді диагностика мен сегментация алгоритмдерін оқыту және тестілеу үшін қолдануға мүмкіндік береді. Деректер жиынтығы деп аксиалдық срездердің бірізділігінен тұратын

үшөлшемді КТ-объем түсіндіріледі. Сегментация 3D-объем құрамындағы әр срез бойынша орындалады және нәтижесінде бейнедегі анатомиялық құрылымдар мен патологиялық ошақтарға сәйкес келетін көлемдік маска қалыптасады. Зерттеу барысында жалпы 29 988–49 490 срез өңделді, бұл сегментация моделін құру мен бағалау үшін ауқымды әрі алуан түрлі оқу және тест жиынтығын қамтамасыз етеді.

Деректер құрылымын және белгілеу сапасын көрнекі түрде көрсету мақсатында 1-суретте үш түрлі пациенттің аксиалдық КТ-срезы көрсетілген. Қызыл аймақ тәжірибелі радиологтар жасаған бастапқы сегментацияға сәйкес зақымдану маскасын білдіреді. Мұндай визуализация патологиялық ошақтың орналасуын, сегментация сапасын және суреттің деталізация деңгейін бір мезгілде бағалауға мүмкіндік береді, бұл кейінгі модельді оқыту мен тестілеу үшін маңызды.



Сурет 1 – Үш түрлі пациенттің КТ-кескіндері зақымдану аймақтарымен

Инсультті диагностикалау үшін КТ кескіндерімен жұмыс барысында деректердің сапасын жақсартуға және ары қарайғы талдауға қажетті белгілерді бөліп алуға көмектесетін бірқатар алдын ала өңдеу қадамдарын орындау маңызды. Алдын ала өңдеудің негізгі кезеңдері мыналарды қамтиды:

1. Миды бинаризациялау (НВ-ВЕТ): Алғашқы кезеңде НВ-ВЕТ (Hunt and Bender Brain Extraction Tool) [20] алгоритмі арқылы бас миы аймағы ерекшеленеді. Бұл алгоритм артық құрылымдарды (мысалы, бас сүйек) алып тастап, тек ми бөлігін қалдырады. Алгоритм табалдырықтық сегментацияны қолданады:

$$B(x, y, z) = \begin{cases} 1, & \text{егер } I(x, y, z) \geq T \\ 0, & \text{әйтпесе} \end{cases} \quad (1)$$

мұндағы $I(x, y, z)$ – (x, y, z) нүктесіндегі пиксельдің қарқындылығы, T – кескінді талдау барысында анықталатын табалдырық мәні. Табалдырық мәні T кескіндегі қарқындылықтардың статистикалық сипаттамалары негізінде есептеледі:

$$T = \mu - k \cdot \sigma \quad (2)$$

мұндағы μ – ми аймағы ішіндегі қарқындылықтың орташа мәні, σ – осы мәндердің стандартты ауытқуы, k – эмпирикалық коэффициент.

2. Нөлдік емес аймаққа дейін қиып алу:

Нөлдік мәндері бар және пайдалы ақпарат бермейтін кескін бөліктері жойылады. Барлық остер бойынша нөлдік емес пиксельдердің минималды және максималды координаттары анықталып, кескін сол шекаралар бойынша қиылады.

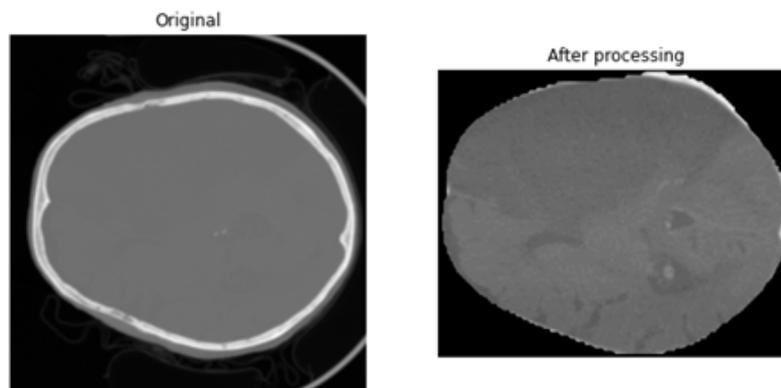
3. Интенсивтілікті нормализациялау: Сканерлеу параметрлеріндегі айырмашылықтарды жою үшін интенсивтілікті минимакс әдісімен $[0, 1]$ диапазонына нормализациялау жүргізіледі [21]:

$$I_{\text{norm}}(x, y, z) = \frac{I(x, y, z) - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \quad (3)$$

мұндағы I_{max} және I_{min} – кескіндегі интенсивтіліктің ең үлкен және ең кіші мәндері.

4. Контрасты теңестіру: Ми құрылымдарының ұсақ немесе әлсіз көрінетін аймақтарын талдау кезінде пайдалы болатын контрастылықты арттыру үшін гистограмма теңестіру әдісі қолданылады.

2-суретте алдын ала өңдеуден кейінгі деректер жиынтығынан алынған бір мысал көрсетілген. Мұнда деректер сапасын арттыруға және әрі қарайғы талдауға қажетті негізгі сипаттамаларды бөліп көрсетуге бағытталған кескінді өңдеу нәтижелерін көруге болады.



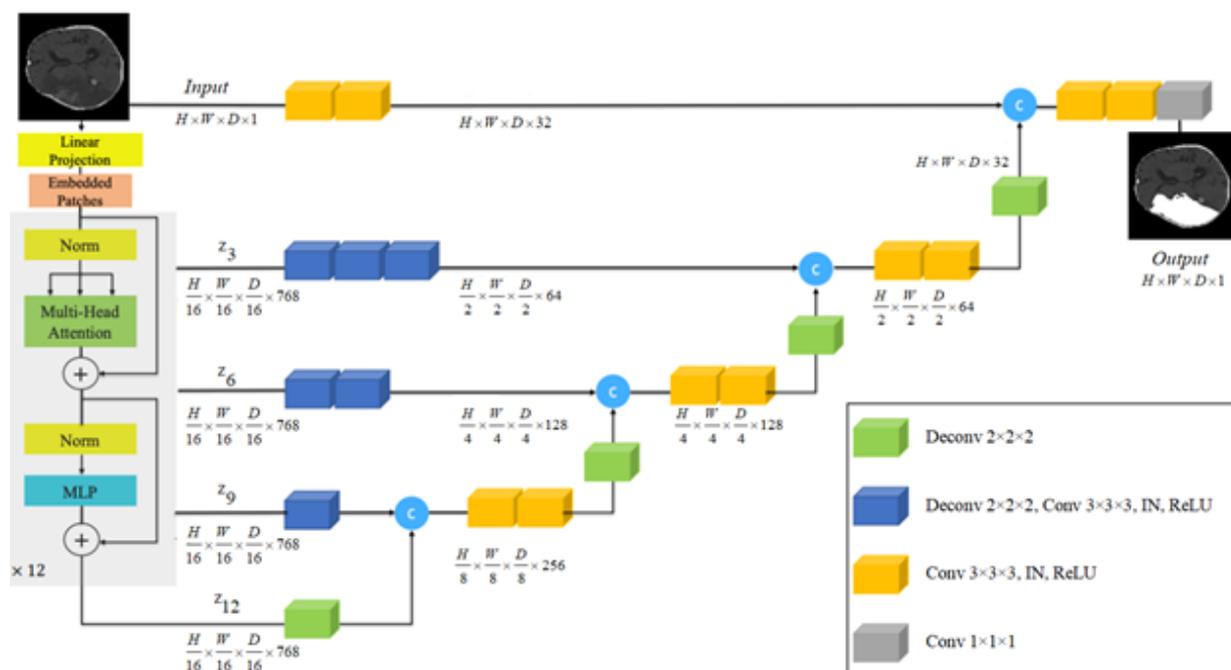
Сурет 2 – Алдын ала өңдеуге дейінгі және кейінгі КТ-кескіндер

Осы зерттеуде ми КТ-кескіндерінде ишемиялық инсульт ошақтарын автоматтандырылған 3D-сегментациялауға арналған ансамбльдік модель архитектурасы ұсынылады. Ұсынылған ансамбль екі трансформерлік модельден – SE-UNETR және Swin UNETR [15] тұрады. Ансамбльдің алғашқы моделі ретінде SE-UNETR пайдаланылады, бұл UNETR трансформер архитектурасының осы зерттеу аясында ми КТ-кескіндерін 3D-сегментациялау үшін әзірленген модификацияланған нұсқасы. Модельдің айрықша ерекшелігі декодтау бөліміне SE-блокларды [22] енгізу, бұл арналарды бейімделген қайта нормализациялауды қамтамасыз етіп, патологиялық аймақтарды дәлірек бөліп көрсетуге мүмкіндік береді. Әзірленген модификация SE-UNETR деп аталды.

UNETR [14] U-Net пен Vision Transformer (ViT) артықшылықтарын біріктіріп, кескіннің жергілікті де, ғаламдық та белгілерін бір уақытта пайдалануға мүмкіндік береді. UNETR-дің негізгі құрылымы 3-суретте көрсетілген. Модель кірісіне үшөлшемді КТ-кескін ($H \times W \times D \times 1$) беріледі, мұндағы H , W , D сәйкесінше көлемнің биіктігі, ені және тереңдігі. Алғашқы кезеңде кіріс кескіні көлемі $16 \times 16 \times 16$ воксель болатын патчтарға бөлініп, олардың әрқайсысы өлшемділігі 768 белгілер кеңістігіне сызықтық түрде проекцияланады. Бұл ViT-энкодерге кіріс деректерін қалыптастырады.

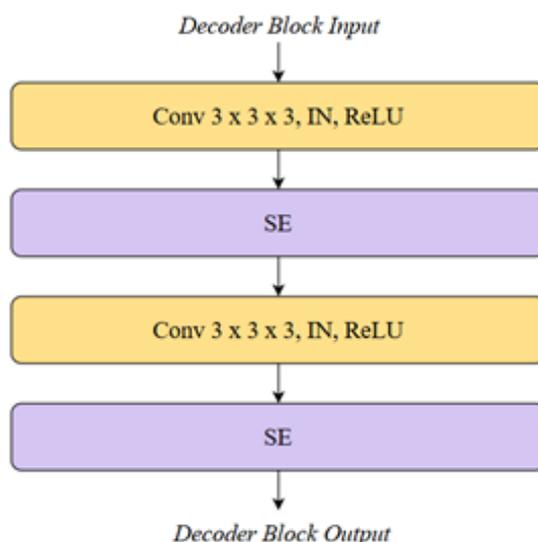
Энкодер 12 трансформер блогынан тұрады, олардың әрқайсысында көпбасты өзіндік назар (Multi-Head Attention) механизмі, нормализация (LayerNorm) және толық байланысқан қабаттар (MLP) бар. Әр үшінші трансформер блогынан кейін (яғни z_3 , z_6 , z_9 , z_{12} қабаттарынан кейін) белгілер алынады, кейін олар декодер бөлігін құру үшін пайдаланылады. Бұл классикалық U-Net-тегі сияқты skip-қосылымдарды қамтамасыз етеді және ерте қабаттардан алынған кеңістіктік ақпаратты сақтауға мүмкіндік береді.

SE-UNETR-дің базалық UNETR-ден басты айырмашылығы – декодердегі әрбір свертка қабатынан кейін SE-блокларды енгізуінде. Осындай бір декодер блогының егжей-тегжейлі сызбасы 4-суретте көрсетілген.



Сурет 3 – UNETR базалық моделінің архитектурасы

Әрбір декодер блогы дәйекті орналасқан екі 3D-свертка қабатынан (ядролары 3×3×3) тұрады, олардың артынан Instance Normalization (IN) нормализациясы және ReLU активация функциясы қолданылады. Ең ақпаратты белгілерге селективтілікті арттыру үшін әрбір свертка қабатынан кейін арналарды бейімделген қайта нормализациялауды орындайтын Squeeze-and-Excitation (SE) блогы [22] енгізіледі. Мұндай блоктар модельге инсульттің әлсіз білінетін, көмескі немесе ұсақ ошақтары сияқты ерекшеліктеріне бейімделе отырып, ең маңызды белгілер арналарға назар аударуға мүмкіндік береді.



Сурет 4 – Декодер бөлігіне SE блоктары енгізілген ұсынылған SE-UNETR моделінің архитектурасы

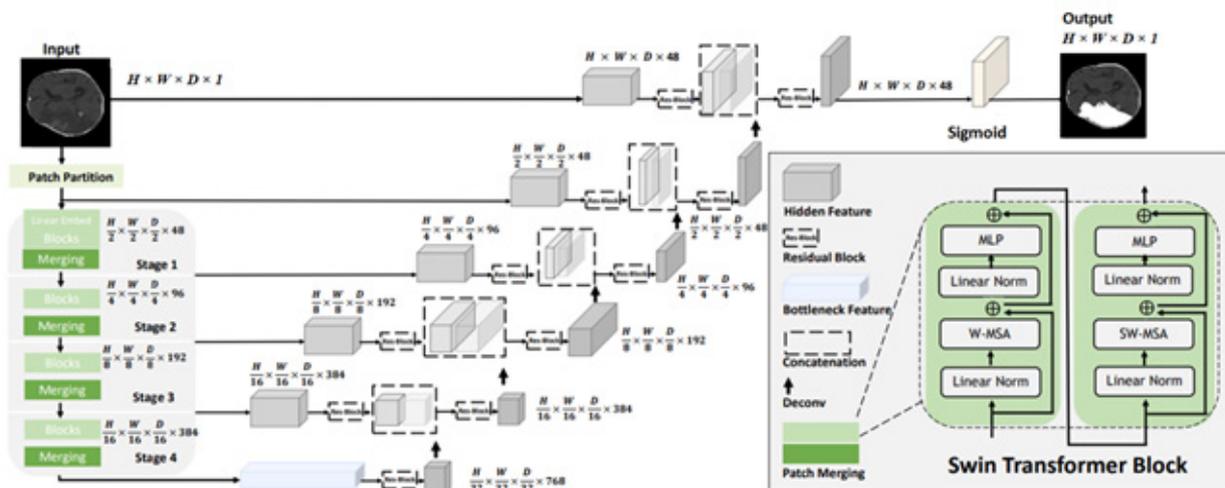
SE-блоктардан кейін шығыс тензоры декодердің келесі деңгейіне беріледі, мұнда транспонирленген свертка операциясы орындалып, энкодерден алынған сәйкес белгілермен бірік-

тіріледі. Соңғы кезеңде модель арналар санын 1-ге дейін азайту үшін $1 \times 1 \times 1$ свертка қабатын қолданады, бұл бинарлы сегментация маскасына (ошақ / ошақ емес) сәйкес келеді. Нәтижелік болжам картасы бастапқы кескінмен бірдей өлшемге ие: $(H \times W \times D \times 1)$.

Ансамбльдің екінші әдісі ретінде үшөлшемді медициналық сегментацияға арналған Swin-UNETR [15] архитектурасы қолданылды. Ол Swin Transformer мен U-Net-тәрізді декодердің артықшылықтарын біріктіреді. Классикалық UNETR ғаламдық назарға (ViT) негізделсе, Swin-UNETR локализацияланған терезелік назар механизмдерін (Window-based Multi-Head Self-Attention, W-MSA) қолданады. Бұл жергілікті де, ғаламдық та тәуелділіктерді тиімді ескеруге мүмкіндік береді, әрі есептеу жүктемесін азайтады. Модель кіріс ретінде өлшемі $(H \times W \times D \times 1)$ болатын үшөлшемді деректерді қабылдайды, олар бір арналы MPT көлеміне сәйкес келеді. Алғашқы кезеңде кескін өзара қиылыспайтын 3D патчтарға бөлінеді, олар өлшемділігі 48 болатын белгілер кеңістігіне проекцияланады. Кейін патчтар төрт кезеңді энкодерде тізбекті түрде өңделеді, мұнда блоктардың тереңдігі (2, 2, 2, 2), ал назар басшыларының саны сәйкесінше (3, 6, 12, 24) болатын Swin Transformer – блоктары қолданылады. Кезеңдер арасында Patch Merging операциясы қолданылады, ол кеңістіктік рұқсаттаманы азайтып, арналар санын арттыра отырып, иерархиялық көрініс қалыптастырады. Әр кезеңнің шығыстары сақталып, U-Net-тәрізді skip-косылымдар арқылы декодерге беріледі.

Декодер транспонирленген свертка блоктарын қамтиды, олар кеңістіктік рұқсаттаманы қалпына келтіріп, объектілердің локализациясын нақтылайды. Соңғы қабат белгілерді шығыс картасына $(H \times W \times D \times 1)$ түрлендіреді. Шығыста сигмоида қолданылады, бұл нәтижені воксельдің мақсатты класқа тиістілік ықтималдығы ретінде интерпретациялауға мүмкіндік береді.

5-суретте Swin-UNETR архитектурасының құрылымы көрсетілген: ол төрт кезеңді энкодерден (Swin Transformer негізінде), U-Net-тәрізді декодерден (skip-косылымдармен) және бинарлы сегментацияға арналған шығыс қабатынан тұрады. Модель DiceLoss функциясын $\text{sigmoid}=\text{True}$ параметрімен пайдалана отырып оқытылды, бұл бинарлы сегментацияда теңгерімсіз деректермен тиімді жұмыс істеуге мүмкіндік береді.



Сурет 5 – Swin-UNETR архитектурасы

Ансамбльді қолдану әрбір жеке модельдің мықты жақтарын біріктіріп, соңғы нәтижені тұрақты етуге мүмкіндік береді. Бұл клиникалық контексте ерекше маңызды, себебі болжамдардың сенімділігі мен тұрақтылығы сыни мәнге ие. Жеке модельге қарағанда ансамбль бірнеше тәуелсіз үлгінің шешімдерін келісу арқылы қате сегментация ықтималдығын төмендетеді.

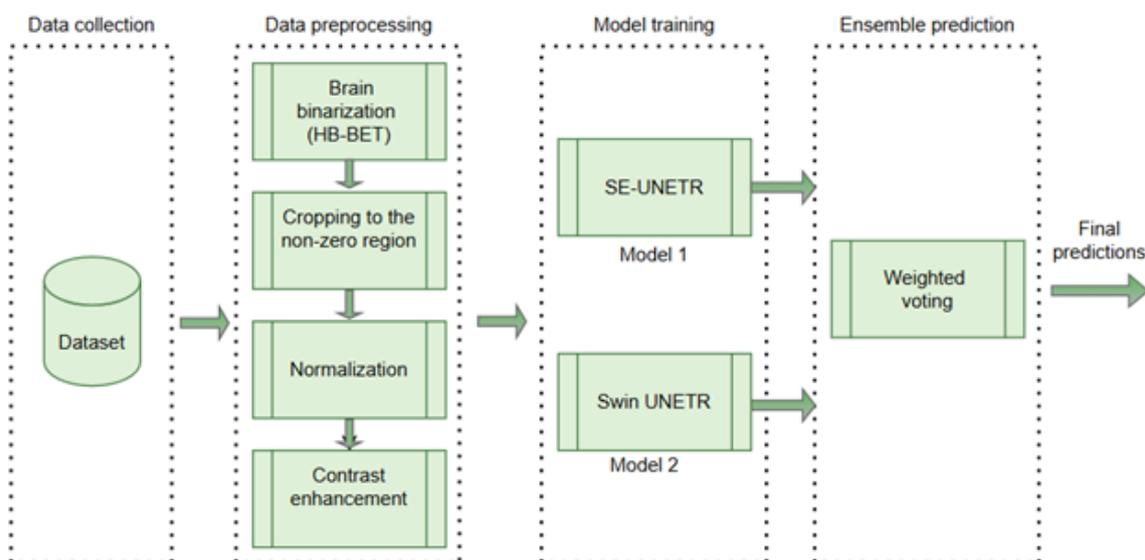
Ансамбль салмақталған дауыс беру әдісі арқылы құрылады, мұнда соңғы болжам барлық модельдердің болжамдары негізінде, олардың сенімділік коэффициенттерін ескере отырып есептеледі. Бұл жағдайда сенімділік коэффициенттері ретінде Dice коэффициенттері қол-

данылады, олар әрбір жеке модельдің валидациялық деректер жиынтығындағы дәлдігін көрсетеді. Ансамбльдің P_{ens} болжамы келесі формула бойынша есептеледі:

$$P_{ens} = \sum_{i=1}^N \left(\frac{D_i}{\sum_{j=1}^N D_j} \right) \cdot P_i \quad (4)$$

мұндағы N – ансамбльдегі модельдер саны (бұл жағдайда $N = 2$), P_i – i -ші модельдің болжамы, D_i – i -ші модель үшін валидациялық деректер жиынтығында есептелген Dice коэффициенті.

Осы зерттеуде ишемиялық инсультті сегментациялау жалпы процесі төрт бірізді кезең түрінде ұсынылған: деректерді жинау, алдын ала өңдеу, модельдерді оқыту және болжамдарды ансамбльдеу. 6-суретте ұсынылған тәсілдің сызбасы көрсетілген. Бұл сызба ансамбльдік модельді құрудың толық циклін, бастапқы бейнелерді дайындаудан бастап соңғы болжамға дейінгі процесті, бейнелейді және процесті көрнекі түрде түсінуге мүмкіндік береді.



Сурет 6 – Медициналық бейнелерді сегментациялауға арналған модельдер ансамблін құру процесі

Нәтижелер мен талқылау

Оқыту ұсынылған трансформерге негізделген модельдер ансамблі SE-UNETR және Swin UNETR көмегімен жүргізілді. Әрбір модель Adam оптимизаторы [23] арқылы, оқу жылдамдығы $1e-4$ параметрімен оқытылды. Барлық желі параметрлері (SE-UNETR және Swin UNETR) PyTorch кітапханасында әдепкі түрде инициализацияланды. Деректер жиыны 80:20 қатынасында оқыту және тест жиынтығына бөлінді. Оқытуды оңтайландыру үшін шығын функциясы ретінде бинарлық кроссэнтропия, ал сегментация сапасын бағалау үшін Дайс коэффициенті қолданылды.

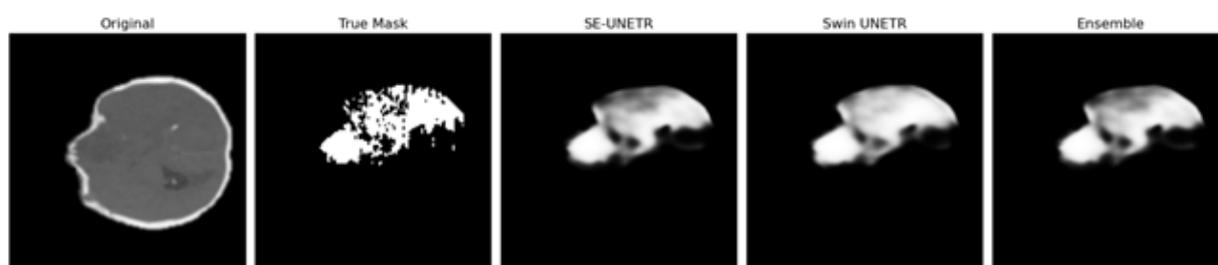
Модельдер Python 3.11.13 тілінде жүзеге асырылды, PyTorch 2.6.0+cu124 фреймворкі мен MONAI 1.5.0 кітапханасы пайдаланылды. Бейнелерді алдын ала өңдеу және онымен жұмыс істеу үшін NumPy 2.0.2 және NiBabel 5.3.2, ал визуализация үшін Matplotlib 3.10.0 қолданылды. Оқыту процесі Google Colab Pro+ ортасында, CUDA 12.4 қолдайтын NVIDIA A100 графикалық процессорында орындалды [24].

Ұсынылған модельдер ансамблінің тиімділігін бағалау нәтижелері 1-кестеде келтірілген. Ең жоғары Dice коэффициенті SE-UNETR және Swin UNETR ансамблін қолдану кезінде 0.7983 мәніне жетті. Бұл әртүрлі архитектуралардың болжамдарын біріктіру сегментацияның дәлдігін арттыратынын көрсетеді.

Кесте 1 – Салыстырмалы талдау

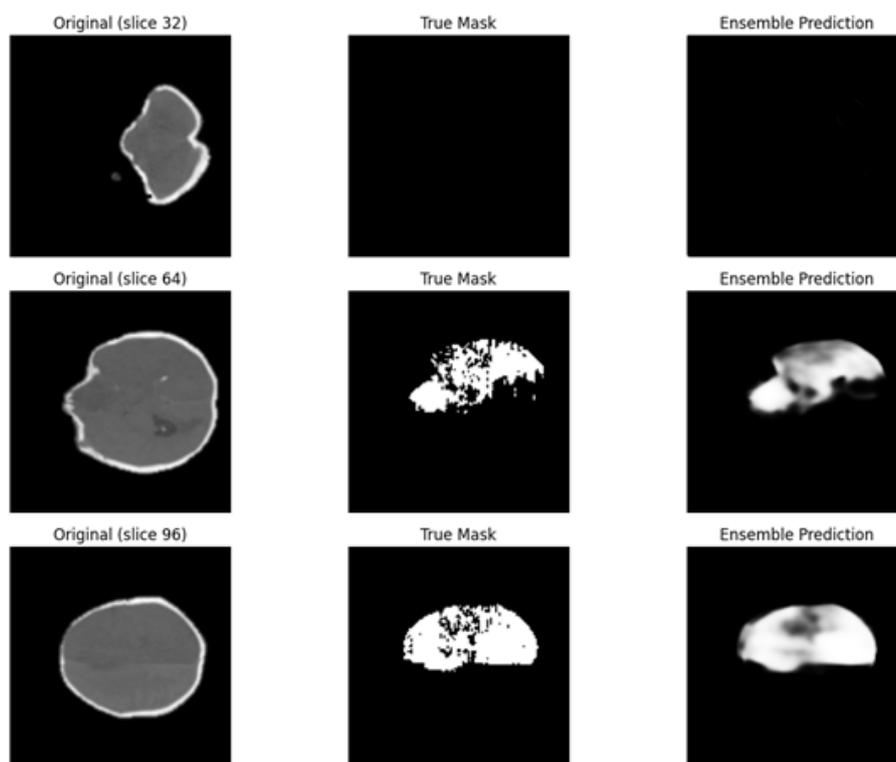
Модель	Sensitivity (%)	Specificity (%)	Precision (%)	Dice
Se-unetr	87.30	99.99	94.31	0.7835
Swin unetr	85.40	99.96	86.58	0.7667
Ensemble	89.75	99.99	94.91	0.7983

7-суретте ишемиялық инсульт ошағын сегментациялау нәтижелерінің визуалды салыстыруы берілген. Сегментация SE-UNETR, Swin UNETR модельдері және олардың ансамблі арқылы орындалды. Бірінші бағанда бастапқы КТ-бейне, екінші бағанда қолмен белгіленген шынайы маска көрсетілген, ал келесі бағандарда екі жеке модельдің және ансамбльдің соңғы болжамдары берілген. Ансамбльдік болжам патологиялық аймақты анағұрлым толық әрі анық қамтиды, бұл оны жеке модельдермен салыстырғанда тиімдірек етеді.



Сурет 7 – КТ-бейнесіндегі ишемиялық ошақты сегментациялау нәтижелерінің визуалды салыстыруы

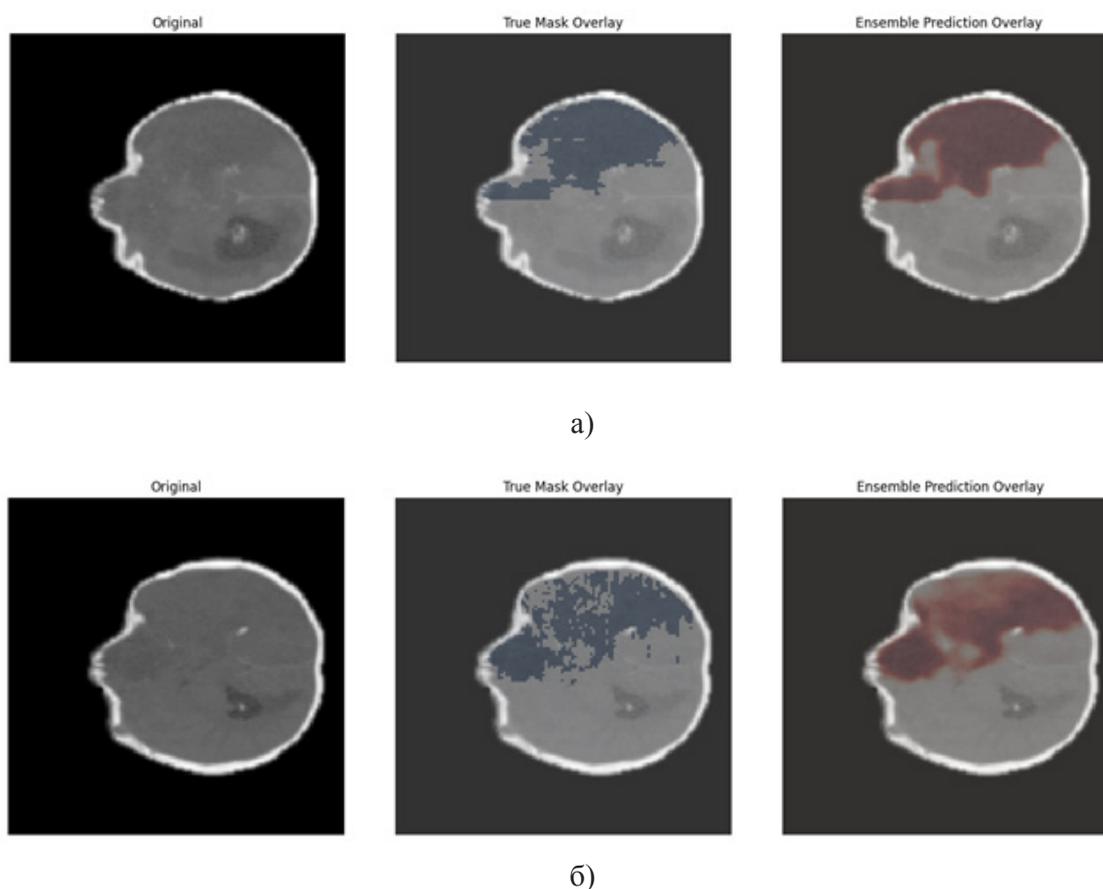
8-суретте эталондық сегментация маскалары мен SE-UNETR және Swin UNETR модельдерінің ансамблі арқылы алынған нәтижелердің салыстыруы көрсетілген.



Сурет 8 – Ансамбль болжамдарының үш түрлі 3D КТ-срездерде салыстырылуы

Талдау үшін ми КТ-сканерлеуінің үш сипаттамалық срезі тандалды: бастапқы (32-срез), ортаңғы (64-срез) және соңғы (96-срез). Бұл модельдің болжамын әртүрлі кезеңдерінде тұрақтылығын бағалауға мүмкіндік береді. Бірінші бағанда мидың анатомиялық құрылымын көрсететін бастапқы КТ-срездер берілген. Екінші бағанда ишемиялық зақымдану аймақтарына сәйкес келетін, сарапшылар белгілеген шынайы маскалар ұсынылған. Үшінші бағанда ансамбльдік модель арқылы алынған болжамдар көрсетілген. Көріп отырғанымыздай, модель зақымданған аймақтарды тиімді анықтап, инсульт ошағының формасы мен орналасуын дәл қайталайды. Ансамбльдік модельдің шу мен артефактілерге төзімділігін, сондай-ақ жалпылау қабілетін көрсетеді. Бұл клиникалық деректерді өңдеуде сегментацияның жоғары сапасы мен ансамбльдік тәсілдің сенімділігін дәлелдейді.

9-суретте екі түрлі пациенттің ишемиялық инсультін сегментациялау нәтижелерінің көрнекі визуализациясы көрсетілген. Әрбір қатар бір КТ-срезге сәйкес келеді және үш бағаннан тұрады: бірінші баған бастапқы КТ-бейне, екінші баған сарапшылар белгілеген шынайы маска (көк түспен), үшінші баған ансамбльдік модельдің болжамы (қызыл түспен).



Сурет 9 – Ишемиялық инсультті сегментациялау мысалдары (а, б)

Бұл мысалдар модельдің ишемиялық ошақтардың әртүрлі орналасуы мен формалары жағдайында тұрақты жұмыс істейтінін көрсетеді. Бірінші жағдайда модель шынайы белгілерге ұқсас ауқымды зақымдалу аймағын дәл анықтайды. Екінші жағдайда болжанған және шынайы маскалардың сапалы қабаттасуы байқалады, әсіресе орталық және маңдай бөлігінде. Бұл визуализациялар ансамбльдік модельдің әртүрлі пациенттерде ишемиялық ошақтарды дәл анықтай алатынын, олардың формасы мен орналасуын эталондық белгілерге сәйкес қайта шығаратынын растайды. Мұндай нәтиже оның клиникалық әлеуетін және КТ-бейнелерді автоматтандырылған талдау тәжірибесінде қолданылу мүмкіндігін айқындайды.

Бақылау тәжірибесі ретінде [13] зерттеу жұмысында ұсынылған нәтижелермен салыстыру жүргізілді. Ол зерттеуде ишемиялық зақымдарды сегментациялау үшін 3D конволюциялық нейрондық желі қолданылған. Дұрыс салыстыруды қамтамасыз ету үшін дәл осы зерттеуде пайдаланылған КТ-мәліметтер жиынтығы пайдаланылды. Алынған нәтижелер ұсынылған ансамбльдік тәсілдің [13] әдісіне қарағанда патологиялық ошақтарды дәлірек сегментациялайтынын көрсетті.

Қосымша түрде бір архитектурадағы үш 3D U-Net модельдерінен құралған баламалы ансамбль іске асырылып, тестілеуден өткізілді. Негізгі айырмашылық олардың салмақтарының әртүрлі кездейсоқ инициализациясында болды, бұл оқыту процесінің стохастикалық сипаты есебінен болжамдардың әртүрлілігін қамтамасыз етті. Мұндай ансамбльдер оқыту кезінде туындайтын кездейсоқ флуктуациялардың әсерін азайтып, шамадан тыс үйрену қаупін төмендетеді. Бұл медициналық кескіндермен жұмыс істегенде өте маңызды, өйткені олар жоғары вариативтілікке және деректер көлемінің шектеулігіне ие. Дегенмен, бұл ансамбль SE-UNETR мен Swin UNETR модельдерін қамтитын трансформерлік ансамбльге қарағанда төмен нәтиже көрсетті. Жеке 3D U-Net модельдері үшін Dice коэффициенттері 0.7132, 0.7145 және 0.7018 құрады, ал олардың ансамблі үшін Dice 0.7253 болды. Бұл ұсынылған трансформерлік ансамбльдің нәтижесінен (0.7983) төмен. Маңыздысы, екі ансамбль де ишемиялық инсульт расталған 98 нақты пациенттің бірдей КТ-мәліметтерінде оқытылды. Мұндай тәсіл салыстырудың объективтілігін қамтамасыз етіп, әртүрлі архитектуралардың тиімділігін дәлірек бағалауға мүмкіндік береді.

Жинақталған нәтижелер 2-кестеде келтірілген. Онда ишемиялық инсультті сегментациялау үшін әртүрлі нейрондық желілер архитектурасының салыстырмалы талдауы берілген. Атап айтқанда, ұсынылған SE-UNETR + Swin UNETR ансамблі (салмақталған дауыс беруімен), баламалы 3D U-Net ансамблі, сондай-ақ бұрын жарияланған шешімдер: SE-Res 3D U-Net [13], METrans трансформерлік модель [25], және модификацияланған 3D U-Net [26] нәтижелері қамтылған. Кестеден көріп тұрғандай, ұсынылған тәсіл Dice коэффициентінің ең жоғары мәнін көрсетіп, КТ-бейнелердегі ишемиялық ошақтарды автоматты сегментациялау кезінде жоғары дәлдігі мен тиімділігін дәлелдейді. Дайс коэффициентінің мәні айтарлықтай артқаны жасалған модельдің сегментацияны дәлірек жүргізіп, патологиялық аймақтарды жақсырақ анықтай алатынын көрсетеді. Алынған нәтижелер медициналық бейнелерді талдау міндеттерінде терең оқытуға негізделген ансамбльдік архитектураларды қолданудың жоғары тиімділігін дәлелдейді.

Кесте 2 – Сегментация нәтижелерінің салыстырмалы талдауы

Дереккөз	Әдіс	Dice Coefficient	Sensitivity (%)	Specificity (%)	Precision (%)
Ұсынылған әдіс	Ensemble UNETR + Swin UNETR (weighted voting)	0.798	89.75	99.99	94.91
Тестіленген балама әдіс (осы зерттеу жұмысында)	Ensemble 3D U-Net (weighted voting)	0.725	61.59	99.72	74.37
Dobshik, 2023 [13]	SE-Res 3D U-Net	0.628	69.9	99.7	61.9
Wang, 2022 [25]	METrans (Transformer-based model)	0.670	64.0	-	72.0
Omarov, 2022 [26]	Modified 3D U-Net	0.580	60.0	-	68.0

Қорытынды

Осы зерттеуде бас мианың КТ-суреттерінде ишемиялық инсульт ошақтарын автоматтандырылған түрде сегментациялау үшін SE-UNETR және Swin UNETR модельдерін біріктіретін трансформерлік ансамбль ұсынылды. Әдіс нақты клиникалық деректерде жоғары дәлдікті көрсетті, Дайс коэффициентінің мәні 0.7983-ке жетті.

Болжамдарды визуалды талдау ансамбльдің жеке модельдерге тән қателіктерді азайтып, патологиялық аймақтардың шекараларын дәлірек қалпына келтіретінін растады. Алынған нәтижелер ұсынылған тәсілдің инсультті және басқа да патологиялық өзгерістерді нақты локализациялауды талап ететін ауруларды автоматтандырылған диагностикалау міндеттерінде қолдану әлеуетін көрсетеді.

Алдағы уақытта әдісті дамыту күрделі ансамбльдеу стратегияларын енгізу және модельдердің интерпретациялануына баса назар аудару арқылы жүзеге асырылмақ. Бұл болжамдарды клиникалық мамандар үшін айқынырақ әрі түсініктірек етіп, жасанды интеллект жүйелеріне деген сенімділікті арттыруға және олардың медициналық тәжірибеде қолданылуын кеңейтуге мүмкіндік береді.

Ұсынылған нәтижелер медициналық сегментация міндеттерінде трансформерлік архитектуралар мен ансамбльдік стратегияларды қолданудың тиімділігін дәлелдейді. Жасалған тәсіл инсульт диагностикасын автоматтандыруға және медициналық көмектің сапасын арттыруға бағытталған клиникалық қолданысқа жарамды шешімдерді әзірлеудің негізі бола алады.

Қаржыландыру туралы ақпарат: Бұл зерттеуді Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландырды (Грант № AP26195405).

ӘДЕБИЕТТЕР

- 1 Bakator, M., & Radosav, D. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction*, 2(3), 47 (2018). <https://doi.org/10.3390/mti2030047>
- 2 Zhu, S., Xia, X., Zhang, Q., & Belloulata, K. An image segmentation algorithm in image processing based on threshold segmentation. *Proceedings of the Third International IEEE Conference on Signal-Image Technologies and Internet-Based Systems, Shanghai, China*, 673–678 (2007). <https://doi.org/10.1109/SITIS.2007.116>
- 3 Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., & Zheng, Y. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11632–11640 (2019).
- 4 Khan, M.Z., Gajendran, M.K., Lee, Y., & Khan, M.A. Deep neural architectures for medical image semantic segmentation: Review. *IEEE Access*, 9, 83002–83024 (2021). <https://doi.org/10.1109/ACCESS.2021.3086530>.
- 5 Tursynova, A., & Omarov, B. 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. *Proceedings of the 16th International Conference on Electronics Computer and Computation (ICECCO), Kazakhstan*, 1–4 (2021). <https://doi.org/10.1109/ICECCO53203.2021.9663825>.
- 6 Khan, W.R., Madni, T.M., Janjua, U.I., Javed, U., Khan, M.A., Alhaisoni, M., Tariq, U., & Cha, J.-H. A hybrid attention-based residual U-Net for semantic segmentation of brain tumor. *Computers, Materials & Continua*, 76(1), 647–664 (2023). <https://doi.org/10.32604/cmc.2023.039188>.
- 7 Zhang, B., Qiu, S., & Liang, T. Dual attention-based 3D U-Net liver segmentation algorithm on CT images. *Bioengineering*, 11, 737 (2024). <https://doi.org/10.3390/bioengineering11070737>.
- 8 Siddique, N., Paheding, S., Elkin, C.P., & Devabhaktuni, V. U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9, 82031–82057 (2021). <https://doi.org/10.1109/ACCESS.2021.3086020>.
- 9 Pinheiro, G.R., Voltoline, R., Bento, M., & Rittner, L. V-Net and U-Net for ischemic stroke lesion segmentation in a small dataset of perfusion data. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes 2018)*. Cham: Springer, 301–309 (2019).

- 10 Atika, L., Nurmaini, S., Partan, R.U., & Sukandi, E. Image segmentation for mitral regurgitation with convolutional neural network based on U-Net, ResNet, V-Net, FractalNet, and SegNet: A preliminary study. *Big Data and Cognitive Computing*, 6(4), 141 (2022). <https://doi.org/10.3390/bdcc6040141>.
- 11 Amirgaliyev, Y.N., Buribayev, Z.A., Melis, Z.M., & Ataniyazova, A.S. On one approach to recognizing fuzzy images of faces based on an ensemble. *Proceedings of the 25th International Conference on Circuits, Systems, Communications and Computers (CSCC)* (2021). <https://doi.org/10.1109/CSCC53858.2021.00011>.
- 12 Berikov, V.B., & Cherikbayeva, L.S. Searching for optimal classifier using a combination of cluster ensemble and kernel method. *CEUR Workshop Proceedings*, 2098, 45–60 (2018).
- 13 Dobshik, A.V., Verbitskiy, S.K., Pestunov, I.A., Sherman, K.M., Sinyavskiy, Y.N., Tulupov, A.A., & Berikov, V.B. Acute ischemic stroke lesion segmentation in non-contrast CT images using 3D convolutional neural networks. *Computer Optics*, 47(5), 770–777 (2023). <https://doi.org/10.18287/2412-6179-CO-1233>.
- 14 Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H., & Xu, D. UNETR: Transformers for 3D medical image segmentation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2022). <https://doi.org/10.1109/WACV51458.2022.00109>.
- 15 Hatamizadeh, A., Yang, D., Roth, H.R., & Xu, D. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. *arXiv preprint arXiv:2201.01266* (2022). <https://arxiv.org/abs/2201.01266>.
- 16 Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., & Xu, D. Self-supervised pre-training of Swin transformers for 3D medical image analysis. *arXiv preprint arXiv:2111.14791* (2022).
- 17 Zhao, W., Li, Y., Lin, S., et al. 3D medical image segmentation using hybrid Swin transformers: A comparative study. *Medical Imaging with Deep Learning (MIDL)* (2023).
- 18 Sriramakrishnan, P., Kalaiselvi, T., Padmapriya, S.T., Shanthi, N., Ramkumar, S., & Kalaichelvi, N. An medical image file formats and digital image conversion. *International Journal of Engineering and Advanced Technology*, 9(1S3), 74–78 (2019).
- 19 Wang, Y., Wang, H., Shen, K., Chang, J., & Cui, J. Brain CT image segmentation based on 3D slicer. *Journal of Complexity in Health Sciences*, 3(1), 34–42 (2020). <https://doi.org/10.21595/chs.2020.21263>.
- 20 Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., & Biller, A. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *NeuroImage*, 129, 460–469 (2016). <https://doi.org/10.1016/j.neuroimage.2016.01.024>.
- 21 Lima, F.T., & Souza, V.M.A. A large comparison of normalization methods on time series. *Big Data Research*, 34, 100407 (2023). <https://doi.org/10.1016/j.bdr.2023.100407>.
- 22 Jin, X., Xie, Y., Wei, X.-S., Zhao, B.-R., & Chen, Z.-M. Delving deep into spatial pooling for squeeze and excitation networks. *Pattern Recognition*, 121, 108159 (2021). <https://doi.org/10.1016/j.patcog.2021.108159>.
- 23 Ogundokun, R.O., Maskeliunas, R., Misra, S., & Damaševičius, R. Improved CNN based on batch normalization and Adam optimizer. In: *International Conference on Computational Science and Its Applications*. Cham: Springer, 593–604 (2022). https://doi.org/10.1007/978-3-031-10545-6_43.
- 24 Choquette, J., Gandhi, W., Giroux, O., Stam, N., & Krashinsky, R. NVIDIA A100 Tensor Core GPU: Performance and innovation. *IEEE Micro*, 41(2), 29–35 (2021). <https://doi.org/10.1109/MM.2021.3053039>.
- 25 Wang, J., Wang, S., & Liang, W. METrans: Multi-encoder transformer for ischemic stroke segmentation. *Electronics Letters*, 58, 340–342 (2022). <https://doi.org/10.1049/ell2.12444>.
- 26 Omarov, M., Ibragimov, M., Kaldybekov, K., & Aytmahanov, M. Modified 3D U-Net for brain stroke lesion segmentation on computed tomography images. *Computers*, 11, 23 (2022). <https://doi.org/10.3390/computers11020023>.

¹ **Черикбаева Л.Ш.,**

PhD, ORCID ID: 0000-0001-8948-4205,
e-mail: cherikbayeva.lyailya@gmail.com

² **Бериков В.Б.,**

д.т.н., профессор, ORCID ID: 0000-0002-5207-9764,
e-mail: berikov@math.nsc.ru

^{1*} **Мелис З.М.,**

докторант, ORCID ID: 0000-0003-4244-8121,
e-mail: melis.zarina98@gmail.com

¹ **Елеусинов А.И.,**

докторант, ORCID ID: 0000-0002-0425-6527,
e-mail: armankaznu@mail.ru

¹ **Адилжанова С.А.,**

PhD, ORCID ID: 0000-0003-1768-064,
e-mail: asaltanat81@gmail.com

¹ **Атаниязова А.С.,**

докторант, ORCID ID: 0000-0003-1122-6614,
e-mail: aisulu.ataniyazova@gmail.com

³ **Дайырбаева Э.Н.,**

докторант, ORCID ID: 0000-0002-4255-5456,
e-mail: nurbekkyzyelmira@gmail.com

¹ КазНУ им. аль-Фараби, г. Алматы, Казахстан

² Новосибирский государственный университет, г. Новосибирск, Россия

³ КазНИТУ им. К.И. Сатпаева, г. Алматы, Казахстан

АНСАМБЛЕВАЯ МОДЕЛЬ НА ОСНОВЕ ТРАНСФОРМЕРОВ ДЛЯ СЕГМЕНТАЦИИ ИШЕМИЧЕСКОГО ИНСУЛЬТА НА 3D КТ

Аннотация

Ишемический инсульт является одной из основных причин смертности и инвалидности. Точная сегментация пораженных областей на КТ-снимках головного мозга имеет решающее значение для своевременной диагностики и принятия клинических решений. В данной работе предлагается ансамблевая методика, основанная на объединении моделей трансформеров SE-UNETR и Swin UNETR с помощью взвешенного голосования. Для оценки качества использовался коэффициент Дайса, метрика, измеряющая степень пересечения предсказанных областей поражения с эталонной разметкой. В отличие от использования одиночных моделей, ансамблевые нейросетевые подходы обеспечивают более высокую надежность и точность сегментации за счет согласованного объединения предсказаний нескольких архитектур. Были использованы трехмерные КТ-снимки 98 пациентов с острым ишемическим инсультом, предоставленные Международным центром томографии Сибирского отделения Российской академии наук. Результаты подтвердили, что предложенный ансамбль демонстрирует более высокую производительность по сравнению с отдельными моделями. Среднее значение коэффициента Дайса составило 0.7983, что свидетельствует о высокой эффективности метода при сегментации ишемических очагов. Анализ показал, что ансамблевая методика позволяет более точно определять границы поражений на КТ-снимках головного мозга и снижает ошибки сегментации. Предложенный подход может быть применен не только для инсульта, но и для других патологий, требующих точного анализа медицинских изображений в автоматизированных системах диагностики.

Ключевые слова: UNETR, Swin Transformer, компьютерная томография, ишемический инсульт, глубокое обучение, сегментация, ансамбль моделей.

¹Cherikbayeva L.Ch.,

PhD, ORCID ID: 0000-0001-8948-4205,
e-mail: cherikbayeva.lyailya@gmail.com

²Berikov V.B.,

Dr. Tech. Sc., Professor, ORCID ID: 0000-0002-5207-9764,
e-mail: berikov@math.nsc.ru

^{1*}Melis Z.M.,

PhD student, ORCID ID: 0000-0003-4244-8121,
e-mail: melis.zarina98@gmail.com

¹Yeleussinov A.I.,

PhD student, ORCID ID: 0000-0002-0425-6527,
e-mail: armankaznu@mail.ru

¹Adilzhanova S.A.,

PhD, ORCID ID: 0000-0003-1768-064,
e-mail: asaltanat81@gmail.com

¹Ataniyazova A.S.,

PhD student, ORCID ID: 0000-0003-1122-6614,
e-mail: aisulu.ataniyazova@gmail.com

³Daiyrbayeva E.N.,

PhD student, ORCID ID: 0000-0002-4255-5456,
e-mail: nurbekkyzyelmira@gmail.com

¹Al Farabi Kazakh National University, Almaty, Kazakhstan

²Novosibirsk State University, Novosibirsk, Russia

³Satbayev University, Almaty, Kazakhstan

TRANSFORMER BASED ENSEMBLE FOR ISCHEMIC STROKE SEGMENTATION ON 3D CT SCANS

Abstract

Ischemic stroke is one of the leading causes of mortality and disability. Accurate segmentation of damaged regions in brain CT images is critical for timely diagnosis and clinical decision-making. In this study, an ensemble approach is proposed, combining SE-UNETR and Swin UNETR transformer models via weighted voting. The Dice coefficient was used for evaluation, measuring the overlap between predicted lesion regions and reference annotations. Unlike single-model approaches, ensemble neural network methods provide higher reliability and segmentation accuracy by integrating predictions from multiple architectures. Three-dimensional CT scans of 98 patients with acute ischemic stroke, provided by the International Tomography Center of the Siberian Branch of the Russian Academy of Sciences, were used. The results demonstrated that the proposed ensemble outperforms individual models. The average Dice coefficient was 0.7983, indicating the high effectiveness of the method in segmenting ischemic lesions. Analysis showed that the ensemble approach more accurately delineates lesion boundaries in brain CT images and reduces segmentation errors. The proposed method can be applied not only to stroke but also to other pathologies requiring precise medical image analysis in automated diagnostic systems.

Keywords: UNETR, Swin Transformer, computed tomography (CT), ischemic stroke, deep learning, segmentation, model ensemble.

Received: September 18, 2025; accepted: October 18, 2025.