

https://doi.org/10.51885/3134-8025_IICS_2026_1_8



XFTAP 76.29.45

**ИНТЕНСИВТІ ТЕРАПИЯДАҒЫ ӨЛІМ-ЖІТІМ
МОДЕЛЬДЕРІНДЕГІ УАҚЫТША ТҰРАҚТЫЛЫҚ
ПЕН ҮЛЕСТІРІМ ДРЕЙФІ:
MIMIC-IV ВАЛИДАЦИЯЛЫҚ ЗЕРТТЕУІ**

**ВРЕМЕННАЯ СТАБИЛЬНОСТЬ И ДРЕЙФ РАСПРЕДЕЛЕНИЯ
В МОДЕЛЯХ СМЕРТНОСТИ ПАЦИЕНТОВ ОТДЕЛЕНИЙ
ИНТЕНСИВНОЙ ТЕРАПИИ: ВАЛИДАЦИОННОЕ ИССЛЕДОВАНИЕ
НА БАЗЕ MIMIC-IV**

**TEMPORAL STABILITY AND DISTRIBUTION DRIFT IN ICU
MORTALITY MODELS: MIMIC-IV VALIDATION STUDY**

А.М. Исмухамедова ¹, А.П. Баканова ^{2*}, С.А. Бельгинова ³,

А. Халимов ², Т. Рысбаев ²

¹Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан

²Қазақстан-Америка Еркін Университеті (ҚАЕУ), Өскемен қ., Қазақстан

³«Туран» Университеті, Алматы қ., Қазақстан

*Жауапты автор: Баканова Анна Павловна, bakanova55anna@gmail.com

Түйінді сөздер:

машиналық оқыту,
интенсивті терапия,
уақытша дрейф,
үлестірім ауысуы,
LightGBM, PSI, модель
калибрлеуі, SHAP
талдауы.

ТҮЙІНДЕМЕ

Машиналық оқыту модельдері клиникалық шешімдерді қолдау үшін уақыт өте келе сапасын жоғалтуы мүмкін, себебі уақыттық ауытқу (temporal drift) және пациенттер популяциясындағы өзгерістер болады, бірақ ОАР (Оқшаулау және қарқынды терапия бөлімшесі) ішінде өлімді болжау міндеттерінде ұзақ мерзімді уақыттық бағалаулар әлі де шектеулі болып қалып отыр. Бұл зерттеуде MIMIC-IV деректері негізінде оқытылған LightGBM моделінің уақыттық тұрақтылығы, калибрлеуі және үлестірім ауытқуы бағаланды. Талдау 2008-2019 жылдар аралығындағы ересек пациенттердің 65 355 стационарлық емделу жағдайын қамтыды және үш деңгейлі валидация схемасын қолданды: стандартты, уақыттық және псевдо сыртқы валидация. Модель жоғары дискриминациялық қабілеттілікті көрсетті (AUC-ROC 0,998-ге дейін) және болашақ деректерде тестілеу кезінде калибрлеуі жақсарды. PSI талдауы маңызды белгілер ауытқуының жоқтығын анықтады, ал SHAP талдауы физиологиялық зертханалық көрсеткіштердің шешуші рөлін көрсетті. Нәтижелер негізгі физиологиялық белгілерге негізделген және қатаң валидациядан өткен модельдер жиі қайта оқытусыз көп жыл бойы тұрақтылықты сақтай алатынын көрсетеді.



© 2026 А.М. Исмухамедова, А.П. Баканова, С.А. Бельгинова, А. Халимов, Т. Рысбаев

Бұл жұмыс Creative Commons Attribution 4.0 халықаралық лицензиясы (CC BY 4.0) бойынша таратылады.

<https://creativecommons.org/licenses/by/4.0/>

Ключевые слова:

машинное обучение, интенсивная терапия, временной дрейф, сдвиг распределения, LightGBM, PSI, калибровка модели, анализ SHAP.

АННОТАЦИЯ

Модели машинного обучения для поддержки клинических решений могут со временем терять качество из-за временного дрейфа и изменений в популяциях пациентов, однако долгосрочные временные оценки в задачах прогнозирования смертности в ОИТ остаются ограниченными. В данном исследовании оцениваются временная устойчивость, калибровка и дрейф распределения модели LightGBM, обученной на данных MIMIC-IV. Анализ включал 65 355 госпитализаций взрослых пациентов за период 2008-2019 годов и использовал трёхуровневую схему валидации: стандартную, временную и псевдовнешнюю. Модель продемонстрировала высокую дискриминационную способность (AUC-ROC до 0,998) и улучшение калибровки при тестировании на будущих данных. Анализ PSI выявил отсутствие значимого дрейфа признаков, а SHAP-анализ показал ключевую роль физиологических лабораторных показателей. Результаты свидетельствуют о том, что модели, основанные на фундаментальных физиологических признаках и прошедшие строгую валидацию, могут сохранять стабильность в течение многих лет без частого переобучения.

keywords:

machine learning, intensive care, temporal drift, distribution shift, LightGBM, PSI, model calibration, SHAP analysis.

ABSTRACT

Machine learning models for clinical decision support can degrade over time due to temporal drift and changes in patient populations; however, long-term temporal evaluations in ICU mortality prediction tasks remain limited. This study assesses the temporal stability, calibration, and distribution drift of a LightGBM model trained on MIMIC-IV data. The analysis included 65,355 adult patient hospitalizations from 2008 to 2019 and employed a three-level validation scheme: standard, temporal, and pseudo-external. The model demonstrated excellent discriminative performance (AUC-ROC up to 0.998) and improved calibration when tested on future data. PSI analysis revealed no significant feature drift, while SHAP analysis highlighted the dominant role of physiological laboratory parameters. The results suggest that models based on fundamental physiological features and subjected to rigorous validation can maintain stability over many years without frequent retraining.

КІРІСПЕ

Машиналық оқыту (МО) модельдері клиникалық ақпараттық жүйелерге барған сайын енгізіліп, ерте диагностикаға, тәуекелді стратификациялауға және нәтиже болжауға қолдау көрсетуде. Интенсивті терапия бөлімшесінде (ИТБ) ауруханадағы өлім-жітімді болжау модельдері айтарлықтай әлеует көрсетті. Алдыңғы зерттеулер машиналық оқытудың ауруды болжауды оңтайландыру және денсаулық сақтау ресурстарын бөлуді жақсарту мүмкіндігін атап өтті, мысалы, электрондық денсаулық сақтау жазбалары негізінде қант диабеті бойынша жүргізілген зерттеу. Жүргізілген зерттеулер құрылымдалмаған медициналық жазбалардан мәтіндік талдау арқылы өлім-жітімді болжауды да қарастырды, мұнда логистикалық регрессияға негізделген алгоритмдер ауруханадағы күтім сапасын әділ және тиімді бағалауға қолдау көрсете алатынын көрсетті. (Ismukhamedova, Uvaliyeva, Rakhmetullina, 2024). Авторлар Gradient Boosting Machines және Recurrent Neural Networks ең тиімді модельдердің қатарында екенін анықтады, бұл электрондық денсаулық сақтау деректерін клиникалық шешім қабылдауды жақсарту үшін пайдалану бойынша кейінгі зерттеулерге негіз болды (Rakhmetullina, 2024; Ismukhamedova, Uvaliyeva, Belginova, 2024).

Дегенмен, олардың ұзақ мерзімді сенімділігі негізгі алаңдаушылық болып қала береді, өйткені көптеген зерттеулер модель өнімділігінің уақытша дрейф, клиникалық практика өзгерістері және пациенттер популяциясының эволюциясы салдарынан төмендейтінін хабарлайды.

Уақытша дрейф – үлестірім ауысуы деп те аталады – болжамдар мен нәтижелер арасындағы байланыстар уақыт өте келе өзгергенде пайда болады. Клиникалық болжау модельдері пациенттер популяциясы уақыт өте келе өзгерген сайын өнімділік дрейфінен зардап шегеді. Ескі және жаңа деректерді тиімді пайдалана алатын модельді жаңарту әдістеріне немесе модельдеу әдістемелеріне үлкен қажеттілік бар. Соңғы зерттеулер трансферлік оқыту мен *stacking ensemble* әдістерін біріктіру, мысалы, *TransferGBM* әдістемесі ескі және жаңа білімді тиімді біріктіріп, тұрақты болжау өнімділігін сақтай алатынын көрсетті. Тарихи электрондық денсаулық сақтау жазбалары деректерімен оқытылған модельдер жиі соңғы деректерде нашар нәтиже көрсетеді және қайта калибрлеуді қажет етеді. Сонымен қатар, ірі көлемді электрондық денсаулық сақтау жазбаларымен алдын ала оқытылған негізгі модельдер клиникалық болжау модельдерінің үлестірім ішіндегі және үлестірім сыртындағы өнімділігін жақсартатыны, уақытша үлестірім ауысуының теріс әсерін азайтатыны дәлелденді (Zhang et al., 2022; Guo et al. 2023).

Алдыңғы жұмыстар калибрлеу төмендеуіне байланысты модельді жаңарту стратегиялары алдын ала белгіленген қайта оқыту аралықтарына қарағанда нашар болжау дәлдігі кезеңдерін қысқартып, аналитикалық ресурстарды ең көп араласу қажет ететін модельдерге тиімдірек бөлуге мүмкіндік беретінін ұсынады (Davis et.al, 2020).

30 күндік аурухана өлім-жітімін болжау үшін регрессия және машиналық оқыту әдістерін салыстырған зерттеу авторлары уақыт өте келе дискриминацияның сақталғанын, бірақ әдіске тәуелді калибрлеу дрейфін көрсетті: нейрондық желілер тұрақты калибрлеуді көрсетсе, басқа модельдер әртүрлі дәрежеде нашарлауды байқатты. Бұл дрейф негізінен болжам-нәтиже байланыстарының өзгеруінен емес, кейсіктің өзгеруінен туындады, бұл модельге тән жаңарту стратегияларының қажеттілігін атап көрсетеді (Davis et al., 2017).

Соған қарамастан, жарияланған зерттеулердің көпшілігі тек бір реттік ұстап қалу валидациясын пайдаланады және сирек көп когорталы немесе ИТБ түрі бойынша валидация жүргізеді.

Бұл зерттеу ИТБ өлім-жітім моделінің уақытша тұрақтылығын, калибрлеу беріктігін және үлестірім дрейфін зерттейді, тұрақты физиологиялық мүмкіндіктер және қатаң алдын ала өңдеу ұзақ мерзімді жоғары беріктікті қамтамасыз етеді деген гипотезаны ұсынады.

ЗЕРТТЕУ МАТЕРИАЛДАРЫ МЕН ӘДІСТЕРІ

Ұсынылып отырған зерттеуде 2008-2019 жылдар аралығын қамтитын Beth Israel Deaconess медициналық орталығының кешенді клиникалық деректер жиынтығын білдіретін MIMIC-IV v2.2 деректер базасы пайдаланылды (Johnson et al., 2023). Зерттелетін когорта қарқынды терапия бөлімшелеріне түскен 65 355 ересек пациентті қамтыды; іріктеу критерийлері ретінде мыналар қарастырылды: 18 жастан жоғары болу, стационарда кемінде 24 сағат болу және толық демографиялық ақпараттың болуы. Демографиялық деректері толық емес пациенттер талдаудан шығарылды. Бүкіл когорта бойынша бақыланған өлім-жітім деңгейі 12,7 % құрады. Пациенттер деректері әрі қарай валидациялау үшін үш дербес жиынға ұйымдастырылды: стандартты тестілік жиын; 2017–2019 жылдар аралығында қабылданған пациенттерді қамтитын уақыттық валидациялық когорта; сондай-ақ қарқынды терапия бөлімшесінің түрі бойынша стратификацияланған, медициналық қарқынды терапия бөлімшесін, хирургиялық қарқынды терапия

бөлімшесін және кардиологиялық қарқынды терапия бөлімшесін қамтитын псевдо-сыртқы валидациялық ішкі жиын (Iwase et al., 2022).

Пациенттердің клиникалық жағдайы мен емдеу үдерісінің әртүрлі қырларын қамту мақсатында бес санатқа топтастырылған 155 айнымалыдан тұратын кешенді көрсеткіштер жиынтығы әзірленді. Демографиялық сипаттамалар пациенттердің жасын, жынысын, этникалық тиесілігін және госпитализация түрін қамтыды. Өмірлік маңызды көрсеткіштер стационарда болу кезеңінің толық ұзақтығы бойынша бірнеше жиынтық статистикалық сипаттамаларды, соның ішінде орташа мәнді, ең төменгі және ең жоғары мәндерді, сондай-ақ стандартты ауытқуды пайдалану арқылы агрегатталды. Зертханалық айнымалылар жалпы қан талдауының көрсеткіштерін (лейкоциттер саны, гемоглобин, гематокрит, тромбоциттер), биохимиялық параметрлерді (натрий, калий, креатинин, қан сарысуындағы несепнәр азоты, глюкоза, альбумин, жалпы билирубин), қан үю жүйесінің көрсеткіштерін (протромбиндік уақыт, белсендірілген жартылай тромбопластиндік уақыт, халықаралық қалыптандырылған қатынас), сондай-ақ бауыр функциясының көрсеткіштерін (аспартатаминотрансфераза, аланинаминотрансфераза) қамтыды. Емдеуге байланысты сипаттамалар вазопрессорларды қолдану, өкпені жасанды желдету, бүйрек алмастырушы терапия және қан құю қажеттілігі сияқты медициналық араласуларды сипаттады. Клиникалық көрсеткіштер есептелген метрикаларды, соның ішінде қарқынды терапия бөлімшесінде болу ұзақтығын, ауруханада болу ұзақтығын және орындалған процедуралар санын қамтыды (Awad et al., 2017; Purushotham et al., 2018).

Модельдің сенімділігін қамтамасыз ету үшін және ақпараттың жоғалып кетуін болдырмау мақсатында деректерді алдын ала өңдеу жүйелі түрде жүргізілді. Экстремум мәндердің әсерін азайту үшін орташа мәннен бес стандартты ауытқу мөлшерінен асатын қалыптан тыс деректерге винсоризациялау қолданылды. Жетіспейтін зертханалық мәндер алдымен соңғы қолжетімді мәнді көшіру арқылы forward-filling тәсілімен өңделді, содан кейін барлық қалған жетіспейтін мәндер бүкіл оқыту когортының негізінде медианалық мәнмен толтырылды. Барлық үздіксіз айнымалылар белгілерді бірыңғай шкалаға келтіру және үлкен шамаларға ие айнымалылардың шамадан тыс әсер етуін болдырмау мақсатында z-score нормализациясы арқылы стандартталды. Бұл алдын ала өңдеу конвейері модельді бағалаудың тұтастығын сақтау үшін кросс-валидация шеңберінде жүйелі түрде қолданылды.

Болжамдық модель LightGBM алгоритмі негізінде құрылды және кластардың теңгерімін ескере отырып, Бриер көрсеткіші мен ROC-қисық астындағы аудан бойынша оңтайландырылды. Модельдің оңтайлы гиперпараметрлері жүйелі түрде баптау барысында анықталып, 477 ағашты, оқыту жылдамдығы 0,073, ағаштың максималды тереңдігі 10, 107 тармақты, ішкі таңдау үлесі 0,741 және белгілерді ішкі таңдау үлесі 0,587 мәндерін қамтыды. Әзірлеу когорты бойынша оқыту барысында жүргізілген бес еселік кросс-валидация қабылдағыштың жұмыс сипаттамалары қисығы астындағы ауданның орташа мәні 0,983 болғанын көрсетті, бұл оқыту кезеңінде бұрын кездеспеген деректерде модельдің жоғары өнімділігін анықтайды.

Модельдің уақыттық тұрақтылығын, калибровкасының сенімділігін және әртүрлі пациенттер топтары мен уақыт кезеңдеріне жалпылау қабілетін жан-жақты бағалау үшін үш сатылы валидация стратегиясы қолданылды. Валидацияның бірінші кезеңінде оқыту және тестілеу жиындары арасында 80/20 арақатынасын сақтай отырып, стратификацияланған кездейсоқ іріктеу арқылы қалыптастырылған стандартты тестілік жиын пайдаланылды. Нәтижесінде 52 284 пациенттен тұратын оқыту когорты және 13 071 пациенттен тұратын тестілік когорта құрылды. Уақыттық валидацияның екінші кезеңінде 2017–2019 жылдар аралығында қарқынды терапия бөлімшелеріне түскен 8 706 пациенттен тұратын когортада модельдің тиімділігі бағаланды. Бұл деректер модельді оқыту кезеңінен кейін бір–үш жыл

өткен соң жиналған. Мұндай тәсіл клиникалық тәжірибедегі, пациенттердің демографиялық сипаттамаларындағы немесе өлшеу хаттамаларындағы өзгерістердің әсерінен модель тиімділігінің уақыт өте төмендеген/төмендемегенін анықтауға мүмкіндік берді.

Псевдо-сыртқы валидацияның үшінші кезеңінде модельдің әртүрлі қарқынды терапия бөлімшелері үшін жалпылау қабілеті бағаланды. Осы мақсатта медициналық қарқынды терапия бөлімшесі (MICU) бойынша 3 812 госпитализацияны, хирургиялық қарқынды терапия бөлімшесі (SICU) бойынша 3 456 госпитализацияны және кардиологиялық қарқынды терапия бөлімшесі (CVICU) бойынша 2 783 госпитализацияны қамтитын жекелеген валидациялық деректер жиындары қалыптастырылды (Guo et al., 2023). Қарқынды терапия бөлімшесінің түрі бойынша стратификацияланған мұндай валидациялық құрылым біріктірілген деректерде оқытылған модельдердің пациенттердің сипаттамалары мен клиникалық жұмыс үдерістері әртүрлі болатын мамандандырылған қарқынды терапия бөлімшелерінде өз тиімділігін сақтай алатынын бағалауға мүмкіндік берді (Harutyunyan et al., 2019).

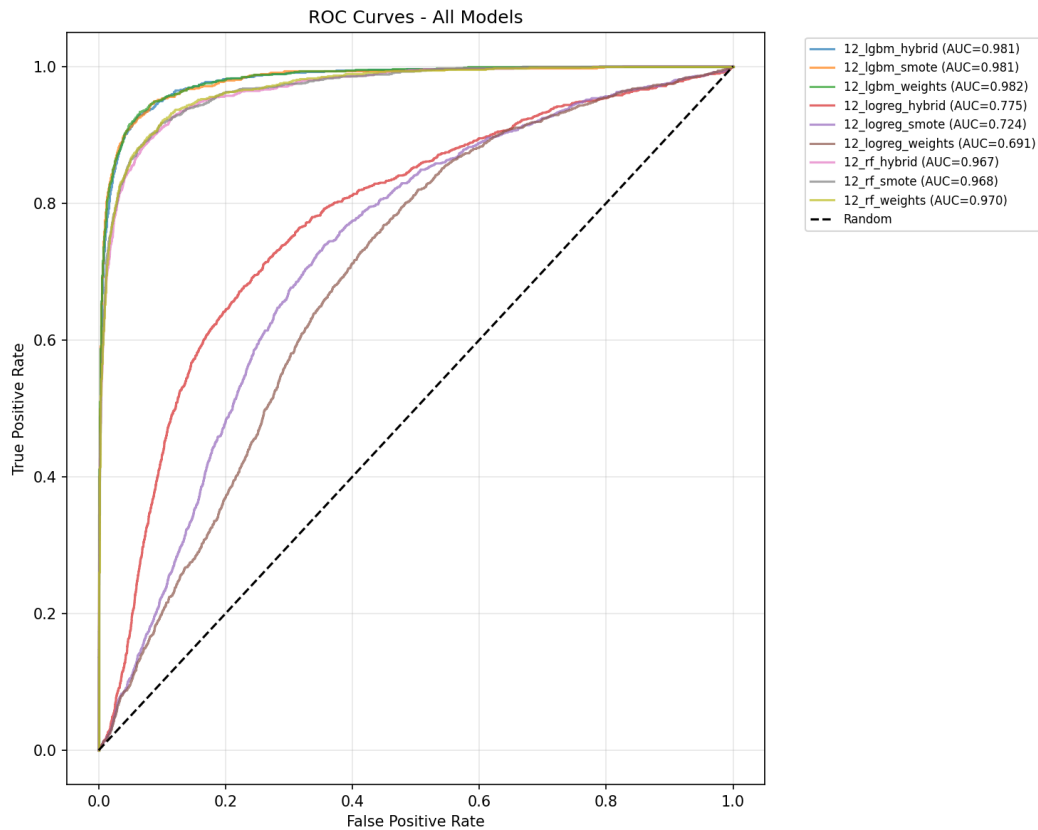
Модельдің тиімділігі нәтижелерді ажырату қабілеті мен болжанған ықтималдықтардың нақтылығын ескере отырып, кешенді түрде бағаланды. Дискриминациялық қабілетті бағалау үшін екі негізгі көрсеткіш қолданылды: ROC-қисық астындағы аудан (AUC-ROC) және «дәлдік–толықтық» қисығы астындағы аудан (AUC-PR). Бұл көрсеткіштердің мәні 0 мен 1 аралығында өзгереді, ал 1-ге жақын мәндер модельдің нәтижелер кластарын тиімдірек ажырататынын көрсетеді. Модельдің жұмысын тереңірек талдау мақсатында сезімталдық (толықтық), ерекшелік (спецификалық), дәлдік және дәлдік пен толықтықтың үлесін теңестіретін олардың гармониялық орташа мәні болып табылатын F1-көрсеткіші қосымша есептелді (Steyerberg et al., 2010).

Дегенмен, модельдің тек нәтижелер кластарын ажырату қабілетін ғана емес, оның калибровка сапасын да бағалау маңызды болды. Осы мақсатта болжанған ықтималдықтар мен нақты нәтижелер арасындағы орташа квадраттық ауытқуды көрсететін Бриер көрсеткіші қолданылды. Калибровка сапасы сонымен қатар калибровкалық қисықтар арқылы визуалды түрде бағаланды, мұнда өлім-жітімнің болжанған ықтималдықтары ықтималдық децильдері бойынша бақыланған жиіліктермен салыстырылды. Мұндай тәсіл модель болжамдарының нақты деректерге қаншалықты дәл сәйкес келетінін анықтауға мүмкіндік берді (Davis et al., 2017).

Әртүрлі модельдердің дискриминация өнімділігін салыстыру үшін Холм түзетуі бар DeLong тесті қолданылды, бұл бірнеше салыстырулар кезінде статистикалық қорытындылардың дұрыстығын қамтамасыз етті – әсіресе әртүрлі модельдердің ROC қисықтары бірдей тест жиынтығында есептелгендіктен және корреляцияланған болғандықтан маңызды болды. Бағалау нәтижелері негізінде ең жақсы өнімділік көрсеткен модельдер үшін шатастыру матрицалары құрылды, олар шынайы оң, шынайы теріс, жалған оң және жалған теріс нәтижелердің үлестірімін көрсетті, бұл модельдің жиі жіберетін қателік түрлері туралы тереңірек түсінік берді.

Сонымен қатар, Population Stability Index (PSI) талдауы жүргізілді, ол негізгі мүмкіндіктердің үлестірімдерінің уақыт өте келе және әртүрлі дерекқорларда қалай өзгередінін бақылауға мүмкіндік берді. PSI децильдер бойынша топтастырылған жиіліктік үлестірімдерді салыстыру арқылы есептелді, бұл сипаттамаларды дрейф дәрежесіне қарай тұрақты ($PSI < 0,10$), орташа ($0,10-0,25$) және айқын ($PSI > 0,25$) деп жіктеуге мүмкіндік берді. Талдау SHAP-талдау нәтижелері бойынша анықталған ең маңызды он сипаттама үшін жүргізілді. Ол олардың үлестірімдерін уақыттық валидацияда және псевдо-сыртқы валидацияда бастапқы тестілік деректер жиынымен салыстыруды қамтыды (1-сурет). Мұндай тәсіл деректердің өзгергіштігі әртүрлі жағдайларда қолдану барысында модельдің сенімділігіне қалай әсер етуі мүмкін екенін бағалауға мүмкіндік берді.

Модельдің болжамдарды қалай қалыптастыратынын тереңірек түсіну үшін әрбір сипаттаманың нәтижеге қосатын үлесін сандық түрде бағалауға мүмкіндік беретін SHapley Additive exPlanations (SHAP) әдісі қолданылды. Белгілердің жаһандық маңыздылығы SHAP мәндерінің орташа абсолюттік көрсеткіштері негізінде анықталып, модельдің нәтижесіне ең күшті әсер ететін айнымалыларды айқындауға мүмкіндік берді (Lundberg et al., 2020). Талдау нәтижесінде физиологиялық және зертханалық көрсеткіштердің басым предикторлар екені анықталды. Қарқынды терапия бөлімшесінде болу барысында тіркелген рН деңгейінің, лейкоциттер санының және глюкоза деңгейінің соңғы мәндері өлім-жітімді болжауға ең жоғары жалпы әсер көрсетті.



1-сурет. Интенсивті терапия бөлімшесіндегі өлім-жітім моделін валидация когорттары бойынша бағалаудың жалпы шолуы

Ескерту – авторлар құрастырған

Осы зерттеудегі барлық талдаулар Python тілінің 3.10 немесе одан жоғары нұсқасына негізделген бірыңғай есептеу ортасында жүргізілді, бұл модельдеу үдерісінің толық ашықтығы мен қайта өндірілуін қамтамасыз етті. Машиналық оқытудың негізгі жұмыс үдерісі scikit-learn фреймворкін пайдалану арқылы іске асырылды, ол модельді оқытудың барлық кезеңдерін — кросс-валидациядан және белгілерді іріктеуден бастап өнімділік көрсеткіштерін есептеуге дейін — жүйелі түрде жүзеге асыруға мүмкіндік берді. Зерттеудің негізгі элементі болып табылатын градиенттік бустинг моделін құру үшін LightGBM кітапханасы қолданылды, оның тиімділігі мен масштабталуы күрделі әрі жоғары өлшемді клиникалық деректерді талдау үшін оны ерекше қолайлы етеді. Модель әрекетін интерпретациялау SHAP кітапханасының көмегімен тікелей аналитикалық үдеріске біріктірілді, бұл белгілердің үлесін сандық тұрғыдан бағалауға ғана емес, сонымен қатар модельдің жекелеген болжамдарын теориялық негізде зерттеуге мүмкіндік берді (Lundberg et al., 2018).

Деректерді алдын ала өңдеу және құрылымдау pandas кітапханасының көмегімен жүзеге асырылды, ал талдаудың барлық кезеңдерінде қажетті сандық есептеулер NumPy кітапханасы арқылы орындалды. Визуализация зерттеудің әрі барлау кезеңінде, әрі нәтижелерді растау кезеңінде маңызды рөл атқарды. Сипаттамалық графиктерді, диагностикалық диаграммаларды және модельдердің өнімділік қисықтарын құру үшін Matplotlib және Seaborn кітапханалары пайдаланылды. Әртүрлі модельдердің дискриминациялық қабілеттерін дәл салыстыру мақсатында scikit-learn кеңейтімдерін пайдалана отырып, ROC-қисықтарды салыстыруға арналған Делонг статистикалық тесті іске асырылды. Модельдерді оқыту және бағалау рәсімдерінің барлығы стандартты есептеу жабдығында, мүмкін болған жағдайларда GPU жеделдетуін қолдану арқылы жүргізілді, бұл есептеулердің тиімділігін және алынған нәтижелердің қайта өндірілуін қамтамасыз етті.

НӘТИЖЕЛЕР ЖӘНЕ ОЛАРДЫ ТАЛҚЫЛАУ

Талдауға 2008–2019 жылдар аралығында MIMIC-IV деректер базасынан алынған ересек пациенттердің қарқынды терапия бөлімшелеріне жатқызылған 65 355 госпитализация жағдайы енгізілді. Зерттелетін когорта валидация жүргізу мақсатында үш топқа бөлінді: 13 071 пациентті қамтитын негізгі тестілік топ, 8 706 пациенттен тұратын уақыттық валидациялық когорта және қарқынды терапия бөлімшелерінің түрлері бойынша жүргізілген псевдо-сыртқы валидациялық когорта (10 051 пациент).

Жалпы ауруханаішілік өлім-жітім деңгейі 12,7 % құрады, алайда топтар арасында айырмашылықтар байқалды (1-кесте). Негізгі тестілік топта өлім-жітім 10,8 %, уақыттық когортада — 13,2 %, ал псевдо-сыртқы когортада — 13,9 % болды. Бұл айырмашылықтар ұзақ мерзімді күтілетін үрдістерді және қарқынды терапия бөлімшелерінің әртүрлі типтеріне тән ерекшеліктерді көрсетеді.

1-кесте. Когорта қорытындысы

Деректер жиынтығы	N	Өлім-жітім (%)	AUC-ROC	AUC-PR	Brier
Негізгі тест	13,071	10,8 %	0,982	0,912	0,0277
Уақытша	8,706	13,2 %	0,998	0,920	0,0063
Псевдо-сыртқы	10,051	13,9 %	0,997	0,918	0,0089

Ескерту – авторлар құрастырған

Пациенттердің демографиялық сипаттамалары барлық когорталарда ұқсас болды. Орташа жас мөлшері 64–67 жасты құрады, сондай-ақ ер адамдардың аздап басымдығы байқалды (53–56 %). Уақыттық валидациялық когортада аурудың ауырлық деңгейі мен өлім-жітім көрсеткіштері сәл жоғары болды, бұл зерттеу кезеңі барысында қарқынды терапия бөлімшелеріндегі пациенттер құрылымының өзгеруіне сәйкес келеді.

LightGBM моделі валидацияның барлық үш деңгейінде де уақыт өте өнімділіктің төмендеу белгілерінсіз жоғары дискриминациялық қабілет көрсетті. ROC-қисық астындағы аудан (AUC-ROC) тұрақты түрде жоғары деңгейде сақталды: негізгі тестілік деректер жиынында 0,982 (95 % сенімділік аралығы: 0,979–0,985), 2017–2019 жылдар аралығындағы уақыттық валидацияда 0,998 (95 % СА: 0,996–0,999) және MICU, SICU және CVICU бөлімшелерін қоса алғанда, қарқынды терапия бөлімшелерінің әртүрлі типтерінде жүргізілген псевдо-сыртқы валидацияда 0,997 (95 % СА: 0,995–0,999).

Осыған ұқсас түрде AUC-PR көрсеткіші де тұрақты болып қалды: негізгі тестілік деректер жиынында — 0,912, уақыттық валидацияда — 0,920 және псевдо-сыртқы валидацияда — 0,918. Бұл нәтижелер 2008–2016 жылдар деректері негізінде оқытылған

модельдердің кейінгі пациенттер когорталарында және қарқынды терапия бөлімшелерінің әртүрлі жағдайларында қолданылған кезде өз дискриминациялық қабілетін сақтағанын немесе тіпті арттырғанын көрсетеді.

Маңызды нәтижелердің бірі модельді болашақ деректерге қолдану кезінде оның калибровкасының айтарлықтай жақсаруы болды. Калибровка мен дискриминацияны қатар сипаттайтын Бриера көрсеткіші негізгі тестілік деректер жиынында 0,0277 мәнінен уақыттық валидацияда 0,0063-ке және псевдо-сыртқы валидацияда 0,0089-ға дейін төмендеді. Бұл болашақ деректерде қолдану кезінде көрсеткіштің 77,4 % төмендегенін білдіреді. Алынған нәтижелер модель өнімділігінің тұрақты екенін ғана емес, сонымен қатар кейінгі жылдарда қолдану барысында оның калибровка дәлдігінің артқанын көрсетеді. Бұл модельдің болжанған ықтималдықтарының нақты бақыланған нәтижелермен неғұрлым жақсы сәйкестене түскенін білдіреді.

Калибрлеу қисықтары екі валидация когортасында болжанған және байқалған өлім-жітім деңгейлерінің дерлік мінсіз сәйкестігін ашып көрсетті (Davis et al., 2020). Уақытша валидация деректерінде калибрлеу қисығы диагональ бойына тығыз орналасып, ықтималдық спектрі бойынша жүйелі ауытқу болған жоқ. Интегралды калибрлеу индексі барлық валидация ішкі жиынтықтарында 0,015-тен төмен болды, бұл 0,05 клиникалық маңыздылық шегінен әлдеқайда төмен.

Алынған нәтижелер машиналық оқыту модельдері болашақ деректерге қолданылған кезде міндетті түрде тиімділігін жоғалтады деген кең таралған пікірге қайшы келеді. Бұл нәтижелер негізгі физиологиялық көрсеткіштерді пайдалану және деректерді мұқият алдын ала өңдеу модельдің ұзақ мерзімді сенімділігін қамтамасыз ете алатынын көрсетеді.

Ковариаттардың ығысуын сандық тұрғыдан бағалау мақсатында ең маңызды он болжамдық сипаттама бойынша популяция тұрақтылығы индексіне (PSI) талдау жүргізілді. PSI мәндері 0,1-ден төмен болған жағдайда тұрақты, 0,1–0,25 аралығында болса орташа ығысу, ал 0,25-ке тең немесе одан жоғары болған жағдайда айқын ығысу ретінде жіктелді (2-кесте).

2-кесте. Ең маңызды 10 мүмкіндік бойынша PSI талдауы

Validation	Тұрақты (PSI<0,10)	Орташа (0,10-0,25)	Елеулі (>0,25)
Уақытша vs Негізгі	10/10	0/10	0/10
Псевдо-сыртқы vs Негізгі	9/10	1/10 (PSI=0.136)	0/10
<i>Ескерту – авторлар құрастырған</i>			

2017-2019 жылдар аралығындағы уақыттық валидацияны 2008-2016 жылдар кезеңімен салыстырғанда, бақыланған барлық он сипаттаманың үлестірімі төмен PSI мәндерімен тұрақты болғаны анықталды, бұл ковариаттардың елеулі ығысуының болмағанын көрсетеді. Ұқсас нәтижелер қарқынды терапия бөлімшелерінің түрлері бойынша жүргізілген псевдо-сыртқы валидация кезінде де алынды. Он сипаттаманың тоғызы тұрақты болып қалды, ал PSI мәні 0,136 болған жалғыз сипаттама орташа ығысу санатына жатқызылғанымен, ол маңыздылық шегі болып саналатын 0,25 мәнінен айтарлықтай төмен болды (3-кесте). Бұл нәтижелер модельдің кіріс деректерінің үлестірімі уақыт бойынша да, әртүрлі клиникалық жағдайларда да тұрақты болып қалғанын көрсетеді.

3-жесте. SHAP талдауы ИТБ өлім-жітімі үшін физиологиялық зертханалық мәндерді ең басым болжамдық мүмкіндіктер ретінде анықтады

Ранг	Мүмкіндік	Орташа	Елеулі (>0,25)
Уақытша vs Негізгі	10/10	0/10	0/10
Псевдо-сыртқы vs Негізгі	9/10	1/10 (PSI=0.136)	0/10
<i>Ескерту – авторлар құрастырған</i>			

Айқын дрейфтің болмауы негізгі физиологиялық айнымалылардың біртіндеп әрі болжамды түрде өзгеретінін және өз табиғаты бойынша ауруханадағы жұмыс үдерістеріне немесе деректерді өңдеу артефактілеріне байланысты айнымалыларға қарағанда анағұрлым тұрақты екенін көрсетеді. Бұл модельдің уақыттық және псевдо-сыртқы когорталарда жоғары сенімділік көрсетуін түсіндіреді.

pH ең маңызды болжамдық фактор ретінде анықталды, бұл оның ауыр физиологиялық бұзылыстар мен аурудың ауырлығының маркері ретіндегі рөліне сәйкес келеді. Екінші және үшінші орындарды инфекция мен жүйелік қабынудың көрсеткіші болып табылатын лейкоциттер саны және метаболикалық стресс пен қант диабетін бейнелейтін глюкоза деңгейі иеленді. Бұл нәтижелер қарқынды терапия бөлімшелеріндегі өлім-жітімнің белгілі қауіп факторларымен сәйкес келеді және модельдің клиникалық негізділігін растайды.

SHAP тәуелділік графиктері белгілер мен модель болжамдары арасындағы интуитивті түсінікті байланыстарды көрсетті. pH мәндерінің жоғары болуы (сілтілі орта) болжанатын өлім-жітімнің төмендеуімен байланысты болса, ал төмен мәндер (ацидемия) өлім қаупінің артуымен ассоциацияланды. Осыған ұқсас түрде глюкоза мен лейкоциттер деңгейінің жоғарылауы өлім қаупінің жоғары болуымен байланысты болды. Ал тромбоциттер вариабельділігі үшін сызықтық емес тәуелділіктер байқалып, бұл критикалық аурулар кезіндегі күрделі физиологиялық динамиканы көрсетеді.

Барлық валидациялық ішкі жиындарда осы белгілердің маңыздылық рейтингінің тұрақты сақталуы модельдің сенімділігін қосымша дәлелдейді, өйткені бірдей физиологиялық механизмдер әртүрлі уақыт кезеңдерінде және қарқынды терапия бөлімшелерінің әртүрлі жағдайларында да болжамдық маңызын жоғалтпайды.

Негізгі тестілік деректер жиынында жіктеу моделінің жалпы тиімділігі жоғары болды және келесі көрсеткіштермен сипатталды:

- Дәлдік (Precision) — 0,883, яғни болжанған оң жағдайлардың 88,3 %-ы шын мәнінде оң болып шықты.

- Толықтық (Recall) — 0,802, бұл нақты оң жағдайлардың 80,2 %-ы модель арқылы дұрыс анықталғанын білдіреді.

- F1-көрсеткіші — 0,840.

- Ерекшелік (Specificity) — 0,987, яғни теріс жағдайлардың 98,7 %-ы дұрыс жіктелді.

13 071 пациентті қамтитын негізгі тестілік деректер жиыны үшін құрылған қателер матрицасы 11 505 шынайы теріс, 150 жалған оң, 281 жалған теріс және 1 135 шынайы оң болжамды көрсетті. Бұл көрсеткіштер барлық валидациялық когорталарда ұқсас болып қалды, ал уақыттық валидация кезінде дәлдік пен толықтық арасындағы тепе-теңдік сақтала отырып, F1-көрсеткішінің 0,851-ге дейін сәл жақсарғаны байқалды.

Жоғары ерекшелік (98,7 %) пен елеулі сезімталдықтың (80,2 %) үйлесуі клиникалық тұрғыдан тиімді болып табылады, өйткені модель жалған дабылдардың санын барынша азайта отырып, ерте араласуды қажет ететін жоғары қауіп тобындағы пациенттердің көпшілігін анықтай алады. Жалған оң болжамдардың салыстырмалы түрде төмен деңгейі

(1,1 %) модельді клиникалық тәжірибеге енгізуді жеңілдетеді, себебі ол шамадан тыс ескертулерге байланысты туындайтын жүктемені төмендетеді.

MICU, SICU және CVICU қарқынды терапия бөлімшелерінің үш түрі бойынша жүргізілген псевдо-сыртқы валидация модельдің жоғары жалпылау қабілетін растады.

- MICU (3 245 пациент, өлім-жітім деңгейі 14,2 %): AUC-ROC – 0,996, Бриер көрсеткіші – 0,0092.

- SICU (3 401 пациент, өлім-жітім деңгейі 13,5 %): AUC-ROC – 0,998, Бриер көрсеткіші – 0,0085.

- CVICU (3 405 пациент, өлім-жітім деңгейі 13,8 %): AUC-ROC – 0,997, Бриер көрсеткіші – 0,0087.

Қарқынды терапия бөлімшелерінің барлық түрлері үшін модельдің тиімділік көрсеткіштері іс жүзінде бірдей болды. AUC-ROC мәндері 0,996-0,998 аралығында, ал Бриера көрсеткіштері 0,0085-0,0092 диапазонында сақталды. Бұл модельдің жекелеген қарқынды терапия бөлімшелеріне тән заңдылықтарды емес, критикалық аурулар кезіндегі өлім-жітімнің әмбебап механизмдерін бейнелейтінін көрсетеді. Осылайша, модельдің жоғары жалпылау қабілеті мен ұзақ мерзімді клиникалық қолданбалығы тағы бір мәрте расталады.

Жоғары қауіп деңгейімен дұрыс болжанған жағдайларды, жоғары қауіп ретінде қате бағаланған жағдайларды және төмен қауіппен қате болжанған жағдайларды қамтитын типтік клиникалық мысалдар үшін құрылған SHAP сарқырама (waterfall) графиктері модельдің интерпретациялануын және жеке болжамдарға негізгі айнымалылардың әсерін айқын көрсетті.

Жоғары қауіп деңгейімен дұрыс анықталған жағдайларда рН, лактат және лейкоциттер саны әдетте болжамды өлім-жітім жағдайына қарай ығыстырды. Ал жоғары қауіп ретінде қате бағаланған, бірақ қолайлы нәтижемен аяқталған жағдайларда осы айнымалылардың шекаралық аномальды мәндері жеке-жеке алғанда қолайсыз болжамды меңзегенімен, қорғаныш факторларымен теңгерілгені байқалды. Мұндай егжей-тегжейлі интерпретация модель болжамдарына клиницистер үшін негізделген түсіндірмелер беріп, оны клиникалық тәжірибеде қолдану үшін қажетті ашықтықты қамтамасыз етеді.

Осы зерттеуде мұқият іріктелген физиологиялық сипаттамалар негізінде оқытылған машиналық оқыту модельдерінің әртүрлі уақыт кезеңдері мен клиникалық жағдайларда қолданылған кезде өнімділігін төмендетпей, тұрақты дискриминациялық қабілет пен калибровканы сақтай алатынын дәлелдейтін сенімді нәтижелер ұсынылды. Зерттеудің ең маңызды қорытындысы — модель сапасының үш жыл ішінде төмендегені. AUC-ROC көрсеткіштері барлық валидация деңгейлерінде тұрақты болып қалды немесе жақсарды және тиісінше 0,982, 0,998 және 0,997 мәндерін көрсетті. Ал Бриера көрсеткіші болашақ деректерде қолдану кезінде 77,4 %-ға төмендеп, модельдің тек тұрақты ғана емес, сонымен қатар калибровка дәлдігі артқанын көрсетті. Сонымен қатар, популяция тұрақтылығы индексіне (PSI) жүргізілген талдау уақыттық валидация кезінде сипаттамалар үлестірімінің елеулі ығысуын анықтамады, мұнда PSI мәндері 0,10-нан төмен болды. Псевдо-сыртқы валидация барысында да тек минималды дрейф байқалып, PSI-дің ең жоғары мәні 0,136-ны құрады. Бұл нәтижелер негізгі физиологиялық айнымалылардың уақыт бойынша және әртүрлі клиникалық жағдайларда тұрақты болып қалатынын растайды.

Бұл нәтиже клиникалық машиналық оқыту әдебиетіндегі уақытша дрейф сөзсіз және қауіпсіз қолдану үшін модельді жиі қайта оқыту қажет деген басым болжамды тікелей жоққа шығарады. Жарияланған ИТБ өлім-жітім болжамы бойынша зерттеулердің көпшілігі бір реттік оқыту-тест бөлінісіне ғана сүйенеді және қатаң көп деңгейлі валидация әдістемесі жетіспейді. Осы зерттеудегі үш деңгейлі валидация дизайны – стандартты ұстап қалу тесті, 2017–2019 жылдардағы уақытша валидация және MICU, SICU және CVICU сияқты ИТБ

түрлері бойынша псевдо-сыртқы валидация – мұқият құрылған модельдердің ұзақ мерзімді клиникалық пайдалылығын сақтай алатынын берік дәлелдейді. Бұл зерттеуде байқалған ерекше тұрақтылықты, шамасы, төрт негізгі фактор арқылы түсіндіруге болады.

Біріншіден, модельдің тұрақтылығын қамтамасыз ететін негізгі факторлардың бірі — физиологиялық сипаттамалар. Модель рН деңгейі, лейкоциттер саны, глюкоза деңгейі және тромбоциттер вариабельділігі сияқты іргелі физиологиялық көрсеткіштерге негізделген, олар институционалдық немесе ұйымдастырушылық үдерістерді емес, негізгі патофизиологиялық механизмдерді бейнелейді. Бұл айнымалыларға биологиялық тұрғыдан тән тұрақтылық қасиеті тән. SHAP-талдау нәтижелері негізгі болжамдық белгілердің барлық валидациялық ішкі жиындарда өз маңыздылығын сақтағанын көрсетті. Бұл модельдің уақытша немесе контекстке тәуелді заңдылықтарды емес, өлім-жітімнің әмбебап механизмдерін бейнелейтінін көрсетеді.

Екіншіден, деректерді сенімді алдын ала өңдеу және белгілерді мұқият инжинирингтеу модель тұрақтылығында маңызды рөл атқарды. Алдын ала өңдеу конвейері бірнеше дәлелді стратегияларды қолданды: жетіспейтін мәндер үшін forward-fill және медиана импутациясы, шеткі мәндерді жоюдың орнына 5 стандартты ауытқудан winsorization және барлық үздіксіз айнымалыларды стандарттау. Бұл таңдаулар өлшеу артефактілері мен шеткі мәндердің әсерін азайтып, сирек кездесетін, бірақ клиникалық маңызды үлгілерді сақтады. Өмірлік көрсеткіштер мен зертханалық мәндерді ИТБ-дағы бүкіл болу кезеңі бойынша жинақтау – бір уақыт нүктесіндегі өлшемдерге сүйенудің орнына – өлшеу протоколдары немесе құжаттау практикасының ауытқуларына аз сезімтал жалпы үлгілерді қамтыды.

Үшіншіден, модельді оқытуда кластар бойынша теңгерілген мақсатты функцияны таңдау маңызды рөл атқарды. Оқыту барысында теңгерілген Бриера көрсеткішін қолдану модельге болжамдардың бүкіл ауқымы бойынша дискриминация мен калибровка дәлдігін бірдей ескеруге мүмкіндік берді. Мұндай тәсіл модельді болашақ деректерде қолдану кезінде байқалған калибровка жақсаруының себебін түсіндіреді. Орташа қателікті ғана азайтуға бағытталған оңтайландырудың орнына модель миноритарлық класқа жататын сирек нәтижелер үшін де жақсы калибрленген ықтималдықтық бағалар қалыптастыруды үйренді. Осылайша, ұсынылған тәсіл өлім-жітім деңгейі 12,7 % болатын кластардың айқын теңгерімсіздігі жағдайында стандартты мақсатты функцияларға қарағанда болжамдардың калибровкасын дәлірек қамтамасыз етеді.

Төртіншіден, модель ауруханалық жұмыс үдерістеріне тәуелді емес. Ол уақыт өте өзгертін әкімшілік кодтарды, есеп айырысу жүйелерін немесе жекелеген мекемелердің хаттамаларын пайдаланбайды. Оның орнына модель тек физиологиялық өлшемдер мен негізгі демографиялық деректерге негізделген. Мұндай көрсеткіштер уақыт пен қолдану жағдайлары өзгерген кезде де тұрақты болып қалады. Бұл әртүрлі қарқынды терапия бөлімшелерінде жүргізілген псевдо-сыртқы валидация кезінде модель тиімділігінің көрсеткіштері неге бірдей дерлік болғанын түсіндіреді.

Бұл нәтижелер клиникалық қолдану және модель басқару үшін тікелей салдарларға ие. Уақытша дрейфке қарсы жиі қайта оқыту деген дәстүрлі тәсіл тұрақты физиологиялық мүмкіндіктерге негізделген модельдер үшін қажет болмауы мүмкін. Оның орнына ресурстар жан-жақты бастапқы валидацияға, қатаң мүмкіндіктер инженериясына және үздіксіз өнімділік мониторингіне бағытталуы тиіс, үздіксіз модель жаңартуларына емес. Бұл есептеу шығындарын айтарлықтай азайтады, басқаруды жеңілдетеді және клиниктерге жүйенің қалай жұмыс істейтіні туралы тұрақты психикалық модельдерді қалыптастыруға мүмкіндік береді.

Тұрақты физиологиялық деректерге негізделген модельдер үшін жиі қайта оқыту қажет болмауы мүмкін. Бастапқы сапалы валидация мен тұрақты мониторинг

клиникалық қолдануды жеңілдетіп, шығындарды азайтады және дәрігерлерге жүйенің жұмыс логикасын тұрақты түрде түсінуге мүмкіндік береді.

Сонымен қатар, жоғары дискриминациялық қабілеттің ($AUC-ROC \geq 0,982$), болашақ деректерде өте жақсы калибровканың (Бриер көрсеткіші $< 0,01$), жоғары ерекшеліктің (98,7 %) және рН, глюкоза деңгейі мен лейкоциттер саны сияқты клиникалық тұрғыдан түсінікті белгілердің үйлесуі бұл модельді тікелей клиникалық қолдануға жарамды етеді. Сонымен қатар, модельдің 80,2 % сезімталдығы жоғары қауіп тобына жататын пациенттердің көпшілігін ерте анықтауға мүмкіндік береді.

Модель болжамдарының қарқынды терапия бөлімшелерінің әртүрлі түрлерінде тұрақты болуы оны практикалық тұрғыдан енгізудің орынды екенін қосымша дәлелдейді. Бір ғана модельді медициналық, хирургиялық және кардиологиялық қарқынды терапия бөлімшелерінде арнайы бейімдеусіз немесе әр орта үшін қайта калибрлеусіз сенімді түрде қолдануға болады. Мұндай әмбебаптық модельді енгізу үдерісін жеңілдетеді және клиникалық командаларға, соның ішінде персоналды оқыту тұрғысынан түсетін жүктемені азайтады.

Сонымен қатар, алынған нәтижелерді интерпретациялау кезінде бірқатар шектеулерді ескеру қажет. Біріншіден, зерттеу бір ғана медициналық орталықта жүргізілген, бұл нәтижелерді кең ауқымда жалпылауды шектейді. Қарқынды терапия бөлімшелерінің түрлері бойынша жүргізілген псевдо-сыртқы валидация модельдің жалпылау қабілетіне ішкі дәлелдер бергенімен, барлық деректер Beth Israel Deaconess медициналық орталығынан алынған. Сондықтан нәтижелер осы мекемеге тән клиникалық тәжірибелерді, өлшеу хаттамаларын немесе пациенттер құрылымын көрсетуі мүмкін және басқа денсаулық сақтау жүйелері үшін толық репрезентативті болмауы ықтимал.

Екіншіден, зерттеуде қарастырылған уақыттық кезең үш жылмен шектелді. Бұл мерзім маңызды болғанымен, ол анағұрлым ұзақ уақыт аралығында орын алатын өзгерістерді толық қамтымауы мүмкін. Қарқынды терапия бөлімшелеріндегі өлім-жітімді болжау 3-4 жыл бойы тұрақты болып қалуы ықтимал, алайда клиникалық тәжірибенің дамуына, жаңа емдеу тәсілдерінің пайда болуына және халық құрылымының өзгеруіне байланысты он жылдық ауқымда өзгеруі мүмкін.

Үшіншіден, зерттеуде соңғы нәтиже ретінде тек ауруханаішілік өлім-жітім қарастырылды. Өлім-жітім клиникалық тұрғыдан аса маңызды көрсеткіш болғанымен, қарқынды терапия бөлімшесінде болу ұзақтығы, қайта госпитализация, функционалдық қалпына келу және өмір сапасы сияқты басқа да маңызды клиникалық нәтижелер әртүрлі заңдылықтар көрсетуі мүмкін және жеке валидацияны талап етеді.

Төртіншіден, алынған нәтижелердің жалпыламалығы құрылымдалған физиологиялық және демографиялық деректермен шектелуімен байланысты. Клиникалық жазбалар, медициналық бейнелер сияқты құрылымдалмаған деректерді, сондай-ақ геномдық немесе протеомдық биомаркерлерді енгізу тек зертханалық көрсеткіштерге негізделген талдауда байқалмайтын уақыттық заңдылықтарды анықтауға мүмкіндік беруі мүмкін.

Болашақ зерттеулер анықталған шектеулерді бірін-бірі толықтыратын бірнеше тәсілдер арқылы еңсеруі тиіс. Ең алдымен, алынған нәтижелердің бір ғана медициналық мекемеден тыс жерлерде де жалпыланатынын растау үшін сыртқы және көпорталықты валидация қажет. Географиялық тұрғыдан әртүрлі аймақтардан алынған деректерді, электрондық медициналық жазбалардың әртүрлі жүйелерін және клиникалық тәжірибенің түрлі модельдерін пайдалану модельдің кең ауқымды қолданбалығын нығайтып, өнімділікке әсер етуі мүмкін орталыққа тән факторларды анықтауға мүмкіндік береді.

Сонымен қатар, 5-10 жыл аралығындағы ұзақ мерзімді уақыттық кезеңдерді қамту модель тұрақтылығының сақталатынын немесе уақыт өте дрейфтің пайда болатынын бағалауға мүмкіндік береді. Ұтымды жағдайда мұндай ұзақ мерзімді валидация

ретроспективті талдаудан туындайтын бұрмалауларды болдырмау үшін деректерді проспективті түрде жинауды қамтуы тиіс.

Гипотезаларды тізбекті тексеру немесе онлайн-оқыту әдістерін пайдалана отырып, модель өнімділігін нақты уақыт режимінде бақылау оның сапасындағы тіпті шамалы төмендеулерді дер кезінде анықтауға және қайта оқыту туралы шешімдерді алдын ала белгіленген кестелерге емес, нақты деректерге сүйене отырып қабылдауға мүмкіндік береді. Мұндай мониторинг жүйелері ауытқуларды анықтаудың статистикалық тәсілдерін – мысалы, PSI және AUC көрсеткіштерін бақылауды – ғана емес, сонымен қатар статистикалық өзгерістердің клиникалық тұрғыдан маңыздылығын растау үшін клиникалық валидацияны да қамтуы тиіс.

Сонымен қатар, өлім-жітімді болжаумен шектелмей, клиникалық әрекеттерге арналған түсіндірілетін ұсыныстарды ұсынатын практикалық машиналық оқыту тәсілдері модельдердің клиникалық құндылығын айтарлықтай арттыра алады. Мұндай модельдерді кері байланыс тетіктерімен бірге клиникалық жұмыс үдерістеріне енгізу олардың пациенттердің емдеу нәтижелеріне және дәрігерлердің шешім қабылдауына әсерін бағалауға мүмкіндік береді.

Соңында, болашақ деректерде байқалған калибровканың жақсаруы клиникалық тәжірибенің дәлелді стандарттарға жақындауымен байланысты ма, әлде ол статистикалық артефактілердің нәтижесі ме екенін зерттеу қарқынды терапия бөлімшелеріндегі өлім-жітімді болжаудың уақыттық динамикасын тереңірек түсінуге мүмкіндік береді.

ҚОРЫТЫНДЫ

Бұл зерттеу іргелі физиологиялық айнымалыларға негізделген машиналық оқыту модельдерінің жиі қайта оқытуды қажет етпей-ақ, уақыт бойынша жоғары тұрақтылық пен калибровканың жақсы сәйкестігін сақтай алатынын көрсетеді. MIMIC-IV деректер базасынан алынған 2008–2016 жылдар аралығындағы 65 355 қарқынды терапия бөлімшесіне жатқызу жағдайы негізінде LightGBM моделі әзірленіп, оның тиімділігі үш сатылы валидация жүйесі арқылы бағаланды. Бұл жүйе стандартты тестілеуді, 2017–2019 жылдар кезеңіндегі уақыттық валидацияны және қарқынды терапия бөлімшелерінің әртүрлі типтері бойынша жүргізілген псевдо-сыртқы валидацияны қамтыды.

Нәтижелер модельдің айрықша тұрақтылығын көрсетті. Дискриминациялық қабілет барлық валидация деңгейлерінде тұрақты болып қалды немесе жақсарды: AUC-ROC негізгі тестілік деректер жиынында 0,982, уақыттық валидацияда 0,998 және псевдо-сыртқы валидацияда 0,997 мәнін көрсетті. Бұдан да маңыздысы, болашақ деректерде модельдің калибровкасы айтарлықтай жақсарды. Бриера көрсеткіші негізгі тестілік жиыннан уақыттық валидацияға өткен кезде 77,4 %-ға төмендеді, бұл болжанған ықтималдықтар мен нақты нәтижелер арасындағы сәйкестіктің артқанын көрсетеді.

Популяция тұрақтылығы индексіне жүргізілген талдау сипаттамалар үлестірімінің ең аз өзгергенін растады. Уақыттық валидация кезінде бақыланған барлық он айнымалы үшін PSI мәндері 0,10-нан төмен болды, ал псевдо-сыртқы тестілеу барысында он сипаттаманың тоғызы тұрақты болып қалды. Тек бір айнымалы ғана PSI мәні 0,136 болатын орташа ауытқуды көрсетті, бұл клиникалық маңыздылық шегінен айтарлықтай төмен.

Бұл нәтижелер клиникалық машиналық оқыту әдебиетінде кең таралған уақыттық дрейфтің сөзсіз болатыны және модельдерді жиі жаңарту қажет деген көзқарасқа тікелей қайшы келеді. Керісінше, бұл зерттеу рН деңгейі, лейкоциттер саны, глюкоза деңгейі және тромбоциттер вариабельділігі сияқты тұрақты физиологиялық көрсеткіштерге негізделген модельдердің критикалық жағдайлар кезіндегі өлім-жітімнің әмбебап механизмдерін бейнелейтінін көрсетеді. Бұл механизмдер уақыт бойынша да, әртүрлі клиникалық жағдайларда да тұрақты болып қалады. Белгілерді мұқият инжинирингтеу, клиникалық

маңызды сигналды сақтай отырып артефактілердің әсерін азайтатын алдын ала өңдеу, кластар бойынша теңгерілген оқыту және нақты бір аурухананың жұмыс үдерістеріне тәуелсіздік модельдің ұзақ мерзімді тұрақтылығы үшін берік негіз қалыптастырады.

Бұл қорытындылардың практикалық маңызы зор. Клиницистер мен денсаулық сақтау жүйелері мұндай модельдерді қарқынды терапия бөлімшелерінің әртүрлі типтерінде арнайы қайта калибрлеусіз сенімді түрде қолдана алады, бұл енгізу үдерісінің күрделілігін және есептеу шығындарын айтарлықтай төмендетеді. Жоғары ерекшелік (98,7%), жеткілікті сезімталдық (80,2%) және клиникалық тұрғыдан түсінікті белгілер иерархиясы модельді практикалық қолдануды қолдап, жекелеген болжамдарды ашық әрі негізді түрде түсіндіруге мүмкіндік береді.

Алайда бұл жұмыс барлық клиникалық машиналық оқыту модельдері табиғатынан тұрақты болады дегенді білдірмейді. Керісінше, ол іргелі физиологиялық көрсеткіштерге негізделген дұрыс жобаланған модельдердің бірнеше жыл бойы тұрақтылыққа қол жеткізе алатынын көрсетеді. Модель өнімділігін тұрақты бақылау және ауытқуларды уақтылы анықтау қажеттілігі өзектілігін сақтайды, ал жекелеген мекемелерден тыс кең ауқымды клиникалық енгізу алдында сыртқы көпорталықты валидация жүргізу талап етіледі.

Болашақ зерттеулер әртүрлі денсаулық сақтау жүйелерінде сыртқы валидацияға, нақты уақыт режимінде ауытқуларды бақылау жүйелерін дамытуға, ұзақ мерзімді уақыттық талдауларға, сондай-ақ калибровканың жақсаруы клиникалық тәжірибенің шынайы конвергенциясын көрсете ме, әлде статистикалық құбылыстардың нәтижесі ме екенін зерттеуге бағытталуы тиіс.

Қорытындылай келе, бұл көпдеңгейлі валидациялық зерттеу қарқынды терапия бөлімшелеріндегі өлім-жітімді болжау үшін мұқият әзірленген машиналық оқыту модельдерінің тұрақты, интерпретацияланатын және клиникалық қолдануға жарамды құрал бола алатынын, әрі оларды үнемі қайта оқытусыз пайдалануға болатынын дәлелдейді. Бұл олардың ұзақ мерзімді клиникалық құндылығын қамтамасыз етіп, қарқынды терапияда сенімді жасанды интеллект тұжырымдамасын қолдайды.

МҮДДЕЛЕР ҚАҚТЫҒЫСЫ: Авторлар мүдделер қақтығысының жоқтығын мәлімдейді

ҚАРЖЫЛАНДЫРУ: Бұл зерттеу Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитетінің гранты аясында қаржыландырылды (№. AP22683316 – «Медициналық шешімдерді қолдау жүйелері үшін машиналық оқыту алгоритмдерін қолдану»).

**ЖАСАНДЫ ИНТЕЛЛЕКТ ТЕХНОЛОГИЯЛАРЫН ПАЙДАЛАНУ ТУРАЛЫ
ХАБАРЛАМА:** Жоқ.

ӘДЕБИЕТТЕР ТІЗІМІ

- Ismukhamedova, A., Uvaliyeva, I., & Rakhmetullina, Z. (2024). Experimental study of a medical data analysis model based on comparative performance of classification algorithms. *Indonesian Journal of Electrical Engineering and Computer Science*, 36(1), 672.
- Rakhmetullina, Z., et al. (2024). Research and implementation of the medical text analysis algorithm for predicting mortality. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(3).
- Ismukhamedova, A., Uvaliyeva, I., & Belginova, S. (2024). Integrating machine learning in electronic health passport based on WHO study and healthcare resources. *Informatics in Medicine Unlocked*, 44.
- Zhang, X., et al. (2022). A transfer learning approach to correct the temporal performance drift of clinical prediction models: Retrospective cohort study. *JMIR Medical Informatics*, 10(11).

- Guo, L. L., et al. (2023). EHR foundation models improve robustness in the presence of temporal distribution shift. *Scientific Reports*, 13(1).
- Davis, S. E., et al. (2020). Detection of calibration drift in clinical prediction models to inform model updating. *Journal of Biomedical Informatics*, 112.
- Davis, S. E., et al. (2017). Calibration drift among regression and machine learning models for hospital mortality. *AMIA Annual Symposium Proceedings*, 2017.
- Johnson, A. E. W., et al. (2023). Author correction: MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1).
- Iwase, S., et al. (2022). Prediction algorithm for ICU mortality and length of stay using machine learning. *Scientific Reports*, 12(1).
- Awad, A., et al. (2017). Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *International Journal of Medical Informatics*, 108.
- Purushotham, S., et al. (2018). Benchmarking deep learning models on large healthcare datasets. *Journal of Biomedical Informatics*, 83.
- Harutyunyan, H., et al. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(1).
- Steyerberg, E. W., et al. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, 21(1).
- Lundberg, S. M., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1).
- Lundberg, S. M., et al. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10).

Авторлар туралы мәліметтер
Информация об авторах
Information about authors



Исмухамедова Айгерим Мэлсатовна – PhD, AP22683316 ғылыми жобасының жетекшісі, компьютерлік ғылымдар мектебі, Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан

Исмухамедова Айгерим Мэлсатовна – PhD, Руководитель научного проекта AP22683316, Школа компьютерных наук, Восточно-Казахстанский технический университет имени Д. Серикбаева, г. Усть-Каменогорск, Казахстан

Ismukhamedova Aigerim – PhD, Head of the scientific project AP22683316, School of Computer Science, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan, e-mail: Aigerim.ismukhamedova1@gmail.com, ORCID: <https://orcid.org/0000-0002-4931-337X>,



Баканова Анна Павловна – т.ғ.к., оқытушы, Қазақстан-Америка Еркін Университеті (ҚАЕУ), Өскемен, Қазақстан

Баканова Анна Павловна – к.т.н., преподаватель, Казахстанско-Американский Свободный Университет (КАСУ), г. Усть-Каменогорск, Казахстан

Bakanova Anna – PhD, Lecturer, Kazakh-American Free University (KAFU), Ust-Kamenogorsk, Kazakhstan,

e-mail: bakanova55anna@gmail.com,

ORCID: <https://orcid.org/0000-0002-8674-3686>



Бельгинова Сауле Аскербековна – PhD, қауымдастырылған профессор, Ақпараттық технологиялар жоғары мектебі, «Тұран» университеті, Алматы қ., Қазақстан

Бельгинова Сауле Аскербековна – PhD, ассоциированный профессор, Высшая школа информационных технологий, Университет «Туран», г. Алматы, Казахстан

Belginova Saule – PhD, Associate Professor, Graduate School of Information Technology, Turan University, Almaty, Kazakhstan,

e-mail: sbelginova@gmail.com,

ORCID: <https://orcid.org/0000-0002-7238-6016>



Халимов Айсат – «Ақпараттық жүйелер» мамандығының магистранты, Қазақстан-Америка Еркін Университеті (ҚАЕУ), Өскемен қ., Қазақстан

Халимов Айсат – Магистрант специальности «Информационные системы», Казахстанско-Американский Свободный Университет (КАСУ), Усть-Каменогорск, Казахстан

Khalimov Aisat – Master's degree in Information Systems, Kazakh-American Free University (KAFU), Ust-Kamenogorsk, Kazakhstan

e-mail: Khalais2012@gmail.com



Рысбаев Тимур – «Ақпараттық жүйелер» мамандығының магистранты, Қазақстан-Америка Еркін Университеті (ҚАЕУ), Өскемен, Қазақстан

Рысбаев Тимур – «Магистрант специальности «Информационные системы», Казахстанско-Американский Свободный Университет (КАСУ), Усть-Каменогорск, Казахстан

Rysbayev Timur – Master's degree in Information Systems, Kazakh-American Free University (KAFU), Ust-Kamenogorsk, Kazakhstan

e-mail: timur.rysbayev1@gmail.com