



DOI 10.51885/1561-4212_2025_3_250

MPHTI 28.23.35

МЕДИЦИНАЛЫҚ ДИАГНОСТИКА МІНДЕТТЕРІНДЕГІ ГЛЮКОМЕТРЛЕР МЕН ГЕМАТОЛОГИЯЛЫҚ АНАЛИЗАТОРЛАРДЫҢ ДЕРЕКТЕРІН ТАЛДАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІ

МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ДАННЫХ ГЛЮКОМЕТРОВ И ГЕМАТОЛОГИЧЕСКИХ АНАЛИЗАТОРОВ В ЗАДАЧАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

MACHINE LEARNING MODELS FOR ANALYZING DATA FROM GLUCOMETERS AND HEMATOLOGY ANALYZERS IN MEDICAL DIAGNOSTICS TASKS

З.Т. Хасенова ¹, И.М. Увалиева ^{1*}, Р.У. Мукашева ¹,

А.Ш. Букунова ¹, Б. Кәрімқызы ²

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

²«А-Медикал» ЖШС, Алматы қ., Қазақстан

*Жауапты автор: Увалиева Индира Махмұтовна, e-mail: iuvalieva@ektu.kz

Түйінді сөздер:

Машиналық оқыту; қант
диабеті; глюкометр;
гематологиялық
анализатор;
биохимиялық көрсеткіш;
интеллектуалды
диагностика; градиентті
күшейту; медициналық
информатика; болжамды
аналитика

ТҮЙІНДЕМЕ

Қант диабеті мен анемиямен сырқаттанушылықтың жаһандық өсуі жағдайында осы жағдайларды ерте диагностикалау және үздіксіз бақылау үшін заманауи, дәл және масштабталатын құралдарды әзірлеу қажеттілігі артып келеді. Осы зерттеудің мақсаты қант диабеті мен анемияның даму қаупін автоматтандырылған жіктеу және бағалау үшін машиналық оқыту алгоритмдерімен заманауи экспресс-глюкометрлерден, гематологиялық және биохимиялық анализаторлардан алынған деректерді интеграциялау тәсілін әзірлеу болып табылады. Мақалада глюкоза, гликатталған гемоглобин (HbA1c), ферритин, гематокрит, MCV, MCH, MCHC және басқа биомаркерлердің көрсеткіштерін қоса алғанда, медициналық құрылғылардан жиналған деректер қолданылады. Деректер стандартталған, өңделген және машиналық оқыту үлгілерін оқыту және сынау үшін пайдаланылды: DecisionTree, RandomForest, AdaBoost, ExtraTrees, GradientBoosting, сондай-ақ логистикалық регрессия, SVM және XGBoost. Сонымен қатар, feature importance және SHAP сияқты модельдерді түсіндіру әдістері қолданылды. Нәтижелер гиперпараметрлерді алдын ала конфигурациялаумен GradientBoosting моделін пайдалану кезінде ең жақсы көрсеткіштерге (0.92 дәлдігі, F1-0.91 көрсеткіші) қол жеткізілгенін көрсетті. Модель сонымен қатар жоғары төзімділік пен интерпретацияны көрсетті, бұл клиникалық тәжірибеде қолдану үшін өте маңызды. Мұндай жүйені енгізу диагностиканың тиімділігін едәуір арттырады, медициналық персоналға жүктемені азайтады және пациенттерді бақылаудың жеке тәсілін қамтамасыз етеді. Зерттеу зертханалық диагностика міндеттерінде AI қолданудың жоғары перспективасын растайды және медициналық құрылғыларды



интеллектуалды аналитикалық платформалармен одан әрі интеграциялау қажеттілігін көрсетеді.

Ключевые слова:

Машинное обучение;
сахарный диабет;
глюкометр;
гематологический
анализатор;
биохимический
показатель;
интеллектуальная
диагностика;
градиентный бустинг;
медицинская
информатика;
предиктивная аналитика

АННОТАЦИЯ

В условиях глобального роста заболеваемости сахарным диабетом и анемиями возрастает необходимость разработки современных, точных и масштабируемых инструментов для ранней диагностики и непрерывного мониторинга этих состояний. Целью настоящего исследования является разработка подхода к интеграции данных, получаемых с современных экспресс-глюкометров, гематологических и биохимических анализаторов, с алгоритмами машинного обучения для автоматизированной классификации и оценки риска развития диабета и анемий. В статье использованы данные, собранные с медицинских приборов, включая показатели глюкозы, гликированного гемоглобина (HbA1c), ферритина, гематокрита, MCV, MCH, MCHC и других биомаркеров. Данные были стандартизированы, обработаны и использованы для обучения и тестирования моделей машинного обучения: DecisionTree, RandomForest, AdaBoost, ExtraTrees, GradientBoosting, а также логистической регрессии, SVM и XGBoost. Дополнительно применялись методы интерпретации моделей, такие как feature importance и SHAP. Результаты показали, что наилучшие метрики (точность 0.92, F1-метрика 0.91) достигнуты при использовании модели GradientBoosting с предварительной настройкой гиперпараметров. Модель также показала высокую устойчивость и интерпретируемость, что критично для применения в клинической практике. Внедрение такой системы может значительно повысить эффективность диагностики, снизить нагрузку на медицинский персонал и обеспечить персонализированный подход к наблюдению пациентов. Исследование подтверждает высокую перспективность применения ИИ в задачах лабораторной диагностики и подчеркивает необходимость дальнейшей интеграции медицинских устройств с интеллектуальными аналитическими платформами.

keywords:

Machine learning; diabetes mellitus; glucose meters; hematology analyzers; biochemical indicators; intelligent diagnostics; gradient boosting; medical informatics; predictive analytics

ABSTRACT

With the global increase in the incidence of diabetes and anemia, there is an increasing need to develop modern, accurate and scalable tools for early diagnosis and continuous monitoring of these conditions. The purpose of this study is to develop an approach to integrating data obtained from modern express glucose meters, hematological and biochemical analyzers with machine learning algorithms for automated classification and assessment of the risk of diabetes and anemia. The article uses data collected from medical devices, including glucose, glycated hemoglobin (HbA1c), ferritin, hematocrit, MCV, MCH, MCHC and other biomarkers. The data was standardized, processed, and used for training and testing machine learning models: DecisionTree, randomForest, AdaBoost, ExtraTrees, GradientBoosting, as well as logistic regression, SVM, and XGBoost. Additionally, model interpretation methods such as feature importance and SHAP were used. The results showed that the best metrics (accuracy 0.92, F1-metric 0.91) were achieved using the GradientBoosting model with hyperparameter presetting. The model also showed high stability and interpretability, which is critical for use in clinical practice. The implementation of such a system can significantly improve the efficiency of diagnostics, reduce the burden on medical staff and provide a personalized



approach to patient monitoring. The study confirms the high prospects of using AI in laboratory diagnostic tasks and highlights the need for further integration of medical devices with intelligent analytical platforms.

КІРІСПЕ

Қант диабеті мен анемия жаһандық ауқымда ең көп таралған созылмалы аурулардың қатарына жатады (Farooqi, M., Tahir, Y., & Rehan, B. (2022)). Дүниежүзілік денсаулық сақтау ұйымының деректеріне сәйкес, диабеттің таралуы тұрақты өсу үрдісін көрсетуде, әсіресе табысы төмен және орташа елдерде Liu J. және басқ. (2022) атап өтті. Анемия, ең алдымен теміртапшылық анемиясы, әлем бойынша 1,6 миллиардтан астам адамға зардап тигізеді және жиі анықталмай қалады, бұл әсіресе әйелдерге, балаларға және қарт адамдарға тән құбылыс деп Iolascon A. және басқ. (2024) атып өткен еді. Екі патологиялық жағдай да өмір сапасына, кәсіби белсенділікке, қатар жүретін аурулардың ағымына және жалпы өлім көрсеткіштеріне айтарлықтай теріс ықпал етеді.

Қиындық тудыратын маңызды факторлардың бірі – қант диабетінің бастапқы кезеңдерінде симптомсыз өтуі, ал анемия белгілері науқастар тарапынан жиі жай ғана шаршау ретінде қабылданып, медициналық көмекке дер кезінде жүгіну ықтималдығын Crasto W. және басқ. (2021) төмендетеді деген. Осыған байланысты неғұрлым дәл, ауқымды және жедел диагностика әдістерін әзірлеу мен енгізу қажеттілігі артып отыр. Заманауи медициналық технологияларды деректерді интеллектуалды талдау әдістерімен біріктіру аталған ауруларды ерте анықтау мен жекелеңдірілген мониторинг жүргізудің жаңа мүмкіндіктерін ашады, бұл денсаулық сақтау жүйесінің цифрлық трансформациясы жағдайында ерекше маңызға ие.

Соңғы онжылдықтарда, негізінен II типті қант диабетінің таралуы айтарлықтай өсті, оған аз қимылдайтын өмір салты, семіздік, теңгерімсіз тамақтану және демографиялық қартаю себеп болып отыр деп Pittas A. G. және басқ. (2023) атап өткен еді. Болжамдарға сәйкес, 2045 жылға қарай диабетпен ауыратындар саны 700 миллионнан асады деп AlFalasi S.M. және басқ. (2024) атап жазған еді. Ал анемия, өз кезегінде, ең кең таралған гематологиялық патологиялардың бірі болып қала береді және дербес ауру ретінде де, созылмалы патологиялық процестердің клиникалық маркері ретінде де көрінуі мүмкін деп өз зерттеулерінде Brittenham G. M. және басқ. (2023) қорытынды жасаған еді. Бұл аурулардың жасырын ағымы, кеш диагностикасы және ұзақ мерзімді терапия қажеттілігі ерекше қауіп тудырады. Мұндай жағдай денсаулық сақтау жүйесіне қосымша жүктеме түсіріп, халықтың еңбекке қабілеттілігін төмендетеді және экономикалық шығындарды едәуір арттырады. Сондықтан ауруларды ерте анықтау, алдын алу және сенімді деректерге негізделген үздіксіз мониторинг «XXI ғасыр індетімен» күрестің негізгі бағыттары болып табылады деп Neumann G. & Kawaoka Y. (2023) атап өткен.

Қант диабеті мен анемияны тиімді басқару дер кезінде қойылған диагнозсыз және науқастың жағдайын динамикалық бақылаусыз мүмкін емес. Бұл екі ауру да ерте кезеңдерде симптомсыз өтуі мүмкін, алайда дәл осы уақытта патологиялық процестер басталады: диабет кезінде – қан тамырлары мен ағзалардың зақымдануы, ал анемия кезінде – тіндік гипоксия. Мұндай бұзылыстарды ерте анықтау ауыр асқынулардың дамуын болдырмауға мүмкіндік береді, оның ішінде ретинопатияны, нефропатияны, жүрек жеткіліксіздігін және когнитивтік бұзылыстарды. Зертханаға мерзімді баруға және талдау нәтижелерін күтуге негізделген дәстүрлі диагностикалық әдістер әрқашан бақылаудың жеткілікті жеделдігі мен үздіксіздігін қамтамасыз ете бермейді. Ал осы уақытта, заманауи медициналық құрылғылар-экспресс-глюкометрлер, гематологиялық және биохимиялық анализаторлар-нақты уақыт режимінде негізгі көрсеткіштерді алуға мүмкіндік береді деп Kubihal S. (2021) атап өткен. Оларды машиналық оқыту алгоритмдерімен біріктіру тәуекел



факторларын мониторингілеудің және автоматтандырылған анықтаудың интеллектуалды жүйелерін қалыптастыру перспективаларын ашады, бұл тәуекелі жоғары топтағы пациенттер үшін, сондай-ақ дәрігерлік көмекке қолжетімділігі шектеулі жағдайларда ерекше маңызға ие.

Қант диабеті мен анемияны диагностикалаудың классикалық әдістері, соның ішінде зертханалық қан анализі, клиникалық тексерулер және медициналық қорытындылар бірқатар шектеулерге ие. Біріншіден, олар пациенттің физикалық қатысуын талап етеді, оны әлеуметтік, экономикалық немесе географиялық факторлар қиындатуы мүмкін. Екіншіден, деректерді түсіндіру көбінесе дәрігердің субъективті тәжірибесіне байланысты, бұл нәтижелердің қайталануын азайтады. Сонымен қатар, дәстүрлі тәсілдер қажетті жеделдікті қамтамасыз етпейді: биоматериалды алған сәттен бастап нәтижелерге қол жеткізуге дейін бірнеше күн қажет. Ақырында, мұндай әдістер тексеріс арасындағы уақыт аралығында жағдайдың динамикасын бағалауға мүмкіндік бермейді, бұл тұрақты бақылауды қажет ететін созылмалы аурулар үшін өте маңызды. Медицинаның цифрлық трансформациясы жағдайында классикалық клиникалық практикадан тыс диагностика мен мониторингке неғұрлым икемді, автоматтандырылған және дербестендірілген тәсілдерді енгізу қажеттілігі өзекті болып отыр.

Медициналық технологиялар мен машиналық оқыту алгоритмдерін дамыту созылмалы ауруларды диагностикалау және бақылау үшін түбегейлі жаңа мүмкіндіктер ашады. Қазіргі заманғы глюкометрлер, гематологиялық және биохимиялық анализаторлар нақты уақыт режимінде өңделетін деректердің айтарлықтай көлемін жасайды деп Jain P. және басқ. (2024) атап өткен. Машиналық оқыту дәстүрлі клиникалық талдауда жоқ жасырын заңдылықтарды анықтауға, аурулардың дамуын болжауға, жағдайларды жіктеуге және клиникалық ұсыныстарды қалыптастыруға мүмкіндік береді. Медициналық құрылғыларды ML алгоритмдерімен біріктіру денсаулық сақтаудың бастапқы деңгейінде де, пациенттердің өзін-өзі бақылауында да қолданылатын интеллектуалды медициналық шешімдерді қолдау жүйелерінің (CDSS) жұмыс істеуінің алғышарттарын жасайды. Мұндай жүйелер гликемия, HbA1c, ферритин, гематокрит және басқа биомаркерлердің көрсеткіштерін автоматты түрде түсіндіруге, тәуекелдерді анықтауға және пациентке немесе емдеуші дәрігерге хабарлама жасауға қабілетті. Бұл әсіресе телемедицина, асқынулардың алдын алу және жеке медицинаға көшу контекстінде маңызды.

Зерттеудің мақсаты қант диабеті мен анемияны диагностикалаудың тиімділігін арттыру үшін заманауи экспресс-глюкометрлерден, гематологиялық және биохимиялық анализаторлардан алынған деректерді машиналық оқыту алгоритмдерімен біріктіру тәсілін әзірлеу және негіздеу болып табылады.

Бұл зерттеу патологиялық жағдайларды дәл жіктеу, жасырын заңдылықтарды анықтау, тәуекел-стратификацияны жүргізу және клиникалық шешім қабылдауды интеллектуалды қолдауды қамтамасыз ету мақсатында объективті зертханалық деректерге негізделген автоматтандырылған жүйені әзірлеуге бағытталған. Жұмыста машиналық оқытудың әртүрлі модельдерінің сапасын кешенді бағалауға, алгоритмдер мен олардың гиперпараметрлерін оңтайландыруға, сондай-ақ нақты клиникалық тәжірибе жағдайында әзірленген шешімдердің практикалық қолданылуын негіздеуге ерекше назар аударылады.

ЗЕРТТЕУ МАТЕРИАЛДАРЫ МЕН ӨДІСТЕРІ

Медициналық технологияларды машиналық оқыту алгоритмдерімен (ML) интеграциялау соңғы жылдары дербестендірілген және цифрлық медицинаны дамытудың негізгі бағыттарының біріне айналды. Ғылыми әдебиеттер созылмалы ауруларды, соның ішінде қант диабеті мен анемияны диагностикалау мен болжауда ML қолдану



мүмкіндіктерін белсенді түрде қарастыруда. Бұл олардың жоғары таралуына, ерте сатыдағы жасырын ағымға және пациенттердің жағдайын үнемі бақылау қажеттілігіне байланысты.

2 типті қант диабеті ML қолдану контекстінде ең көп зерттелген аурулардың бірі болып табылады (Deberneh H. M., & Kim I. (2021)). Pima Indians Diabetes деректерін қолдануға негізделген бірқатар жұмыстар логистикалық регрессия, кездейсоқ орман, SVM және градиентті күшейту әдістері сияқты алгоритмдерді қолданылды деп Mousa A. және басқ. (2023) атап өткен. Зерттеулердің көпшілігі клиникалық белгілердің шектеулі жиынтығына негізделген қант диабеті қаупін болжау кезінде шамамен 85-90% дәлдікке жететін ансамбльдік тәсілдердің (мысалы, XGBoost, Random Forest) жоғары тиімділігін растайды: глюкоза деңгейі, жас, дене салмағының индексі және қан қысымы. Сонымен қатар, нақты глюкометрлер немесе биохимиялық анализаторлардың деректеріне сүйенетін, нақты уақыт режимінде жүргізілген зерттеулер әлі де аз, бұл өзекті зерттеу аясын қалыптастырады.

Анемияны диагностикалауда ML қолдану аз дәрежеде жиі кездеседі, бірақ бұл бағыт белсенді даму тенденциясын көрсетеді. Жіктеу әдістері, соның ішінде Decision Tree, KNN және нейрондық желілер жалпы қан анализінің параметрлеріне негізделген анемия түрлерін (мысалы, микроциттік және макроциттік) ажыратуға мүмкіндік береді – Hb, MCV, MCH, RDW және т.б. деп Ramzan M. және басқ. (2024) атап өткен. Жекелеген зерттеулер нейрондық желілік модельдердің тиімділігін, әсіресе балалар мен жүкті әйелдердегі темір тапшылығы анемиясын болжауда көрсетеді. Бұл саладағы шектеуші фактор белгіленген клиникалық деректер жиынтығының жеткіліксіз болуы болып қала береді, бұл модельдерді оқытуды қиындатады және олардың кеңінен қабылдануын тежейді.

Медициналық мәселелерде ML модельдерінің интерпретациясы ерекше маңызды. Шешімдерді түсіндіру үшін SHAP (Geukes Forpen R.J. және басқ. (2025)), Permutation Importance және LIME (Mitra U. және басқ. (2025)) сияқты тәсілдер кеңінен қолданылады. Бұл дәрігерге модель нәтижесіне қандай белгілердің негізгі әсер еткенін түсінуге мүмкіндік береді, бұл әсіресе диабет пен анемияны диагностикалауда өте маңызды, мұнда алгоритм қорытындыларының клиникалық интерпретациясы аса қажет болып табылады.

Осылайша, қолданыстағы зерттеулер созылмалы ауруларды диагностикалауда машиналық оқытуды пайдаланудың жоғары перспективасын көрсетеді. Сонымен бірге, зертханалық құрылғыларды, үздіксіз деректер жинау технологияларын және интеллектуалды талдауды бірыңғай цифрлық платформаға біріктіретін кешенді жүйелерді құру қажеттілігі сақталуда. Дәл осы міндет осы зерттеудің негізінде жатыр.

Жұмыс аясында қант диабеті мен анемияны диагностикалау үшін заманауи экспресс-глюкометрлер, сондай-ақ гематологиялық және биохимиялық анализаторлар қолданылды. Экспресс-глюкометрлер-бұл үйде де, клиникалық тәжірибеде де қолданылатын қандағы глюкозаны жылдам өлшеуге арналған портативті құрылғылар. Олардың жұмыс принципі көбінесе электрохимиялық әдіске негізделген, онда глюкоза ферментпен (мысалы, глюкоза оксидазасы) әрекеттесіп, қант концентрациясына пропорционалды электр сигналын тудырады. Заманауи модельдер (мысалы, AT Care, Contour Plus, Accu-Chek Guide) автоматты кодтау, өлшеу тарихын жүргізу, Bluetooth немесе USB деректерін беру, мобильді қосымшалармен және еске салу жүйесімен синхрондау мүмкіндіктерімен жабдықталған деп Cournoyer A. және басқ. (2022) атап өткен.

AT Care Multi сияқты кейбір құрылғылар глюкоза деңгейін ғана емес, сонымен қатар метаболикалық синдромы бар науқастар үшін өте маңызды липидті профиль көрсеткіштерін (холестерин, триглицеридтер) анықтауға мүмкіндік беретін мультифункционалдылыққа ие. Осылайша, заманауи экспресс-глюкометрлер қан көрсеткіштерін дәл және ыңғайлы бақылауға арналған жоғары технологиялық құралдар болып табылады (Daves M.



және басқ. (2025)). Қолданыстағы шектеулерге қарамастан, олар диабет кезіндегі өзін-өзі бақылаудың негізгі құралы, сондай-ақ амбулаториялық және жедел жәрдем жағдайында скринингтің тиімді құралы болып қала береді.

«А-Медикал» ЖШС компаниясының заманауи экспресс-глюкометрлері, гематологиялық және биохимиялық анализаторлары сурет 1-де көрсетілген.



1-сурет. «А-Медикал» ЖШС компаниясының заманауи экспресс-глюкометрлері, гематологиялық және биохимиялық анализаторлары

Ескерту: авторлар «А-Медикал» ЖШС компаниясының деректері (2025) негізінде құрастырған

Гематологиялық және биохимиялық анализаторлар заманауи зертханалық диагностикада орталық орын алады, олар физиологиялық және патологиялық жағдайлардың кең ауқымын жоғары дәлдікпен және қайталану мүмкіндігімен бағалауды қамтамасыз етеді. Бұл автоматтандырылған жүйелер қанның жасушалық құрамының көрсеткіштерін, биохимиялық параметрлерді және микроэлементтердің құрамын қамтитын сенімді нәтижелерді қысқа мерзімде алуға мүмкіндік береді, бұл диагноз қоюдың және кейіннен терапияны бақылаудың негізін қалайды.

Гематологиялық анализаторлар (мысалы, ABX Micros ES 60, ABX Pentra XL 80, Yumizen H500 OT) жалпы қан талдауын (ЖҚТ) орындауға арналған. Олар эритроциттердің, лейкоциттердің және тромбоциттердің мөлшерін, гемоглобин концентрациясын (Hb), гематокрит деңгейін (HCT) анықтайды, сондай-ақ әртүрлі анемия түрлерін дифференциалды диагностикалау үшін қажетті эритроцитарлық индекстерді (MCV, MCH, MCHC) есептейді. Сонымен қатар, заманауи құрылғылар лейкоциттерді бес популяция бойынша жіктеуге мүмкіндік беретін 5-DIFF-талдауды қолдайды деп Gómez, J. G. және басқ. (2025) деп атап өткен. Ажыратушы сипаттамалары – жоғары өнімділік (сағатына 120 сынамаға дейін), кіріктірілген ішкі сапаны бақылау жүйелері және зертханалық ақпараттық жүйелермен (LIS) интеграциялау мүмкіндігі, бұл стандарттауды арттырады және қателіктер қаупін азайтады.



Биохимиялық анализаторлар (мысалы, ABX Pentra 400) қан сарысуындағы немесе плазмасындағы метаболиттер мен ферменттерді сандық анықтау үшін қолданылады. Ең көп сұранысқа ие көрсеткіштер – глюкоза, креатинин, ферритин, билирубин деңгейі, ферменттердің белсенділігі (АЛТ, АСТ), жалпы ақуыз бен альбуминнің концентрациясы, сондай-ақ электролиттер (Na^+ , K^+ , Cl^-). Заманауи жүйелер сынамааны дайындауды, ауытқуларды тіркеуді және өзін-өзі калибрлеуді қоса алғанда, 60–100 түрлі сынақты автоматты түрде орындай алады, бұл оператордың қолмен араласу қажеттілігін азайтады және нәтижелердің сенімділігін арттырады.

Жоғары дәлдігі, қайталану мүмкіндігі және жұмыс жылдамдығы арқасында гематологиялық және биохимиялық анализаторлар анемияларды, қант диабетін, бүйрек жеткіліксіздігін, инфекциялық ауруларды және басқа да патологиялық жағдайларды диагностикалау мен бақылауда таптырмайтын құрал болып табылады. Оларды машиналық оқыту алгоритмдерімен біріктіру зертханалық деректердің ақпараттылығын және клиникалық шешімдердің дәлдігін арттыруға, әсіресе мультифакторлы көрсеткіштерді талдау кезінде жаңа мүмкіндіктер ашады.

Осы зерттеу аясында медициналық және зертханалық жүйелерді жеткізуге маманданған «Aksel & A» ЖШС компаниясының жабдықтары қолданылды, бұл алынған деректердің жоғары сапасын және оларды кейіннен интеллектуалды талдау модульдеріне интеграциялау мүмкіндігін қамтамасыз етті.

Медицинадағы жасанды интеллекттің ең белсенді дамып келе жатқан бағыттарының бірі – бұл автоматтандырылған диагностика міндеттері үшін қолданылатын машиналық оқыту (ML). Random Forest, Gradient Boosting, Support Vector Machines және нейрондық желілер сияқты жіктеу алгоритмдері таңбаланған клиникалық деректер (анамнез, зертханалық көрсеткіштер, медициналық кескіндер) негізінде оқытыла отырып, мыналарға мүмкіндік береді (Gómez, J. G. және басқ. (2025)):

- глюкоза деңгейі, HbA1c, дене салмағының индексі, қысым және пациенттің жасы негізінде қант диабетін диагностикалау;
- ЖҚТ (Hb, HCT, MCV, MCHC) деректері бойынша анемияларды анықтау және оларды жіктеу (микроцитарлық, нормоцитарлық, макроцитарлық);
- КТ, МРТ, рентгенография және УДЗ деректері бойынша сверткалық нейрондық желілер (CNN) көмегімен қатерлі ісіктерді тану;
- рекуррентті нейрондық желілерді (RNN, LSTM) пайдалана отырып, ЭКГ немесе эхокардиография негізінде кардиологиялық патологияларды анықтау.

Осылайша, машиналық оқыту диагностикалау мен болжаудың дәлдігін арттыра отырып, медицинаға трансформациялаушы әсерін тигізуде. Болашақта, оның рөлі, әсіресе медициналық құрылғылармен және зертханалық жүйелермен интеграциялануында арта түседі. Алайда, оны тәжірибеде енгізу үшін технологиялық мүмкіндіктер мен клиникалық қауіпсіздік талаптары арасындағы тепе-теңдікті ескеру қажет.

Қант диабеті мен анемияны диагностикалау процестерінде заманауи экспресс-глюкометрлерді, гематологиялық және биохимиялық анализаторларды ML-модельдерімен сәтті интеграциялау үшін бірқатар міндеттерді шешу қажет:

- деректерді жинауды стандарттау;
- алдын ала өңдеу және сүзу процедураларын әзірлеу және енгізу;
- ML модельдерін оқыту және белгілерді іріктеу;
- алгоритм шешімдерін интерпретациялау;
- валидация және клиникалық тестілеу;
- тәуелсіз іріктемелердегі метрикалар (accuracy, F1 score, recall, ROC-AUC) бойынша дәлдікті бағалау;
- нәтижелерді қолданыстағы диагностикалық стандарттармен салыстыру.



Осы зерттеу аясында автоматтандырылған диагностика міндетін жүзеге асыру үшін әртүрлі диагностикалық құрылғылардан алынған зертханалық және биометриялық көрсеткіштерді қамтитын жиынтық деректер жиынтығы жиналды. Негізгі дереккөздер кесте 1-де көрсетілген.

1-кесте. Зерттеудің негізгі дереккөздері

Дереккөз түрі	Медициналық құрылғылардың атауы	Мақсаты
Экспресс-глюкометрлер	AT Care; AT Care Multi	Капиллярлық қандағы глюкозаның, ал бірқатар модельдерде – холестерин мен триглицеридтердің де мандерін береді
Гематологиялық анализаторлар	ABX Micros ES 60; Pentra XL 80; Yumizen H500	Жалпы қан талдауының көрсеткіштерін алу үшін қолданылады
Биохимиялық анализаторлар	ABX Pentra 400; Easy Reader+	Ферритин, креатинин, жалпы ақуыз, глюкоза, сондай-ақ басқа да биохимиялық маркерлердің деңгейлерін анықтауға мүмкіндік береді
Иммунохимиялық жүйелер	eCL8000	Гормоналды және иммундық профильді бағалау үшін қолданылады (қажет болған жағдайда, мысалы, ілеспе жағдайларды диагностикалауда)
Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)		

Деректер жиынтығындағы әрбір жазба пациенттің жеке профилін білдіреді және 2-кестеде көрсетілген көрсеткіштерді қамтиды.

2-кесте. Деректер құрылымы

Көрсеткіштер атауы	Мәндер диапазоны
Жынысы	F - female (әйел) M - male (еркек)
Жасы	Жас топтары: <40, 40 - 49, 50 - 59, 60 >
Адам дене салмағының бойына сәйкестігін бағалау (Дене салмағының индексі, ДСИ)	Салмақ жетіспеушілігі : < 18,5 Қалыпты салмақ: 18,5 – 24,9 Артық салмақ: 25,0 – 29,9 Семіздіктің I дәрежесі: 30,0 – 34,9 Семіздіктің II дәрежесі: 35,0 – 39,9 Семіздіктің III дәрежесі: ≥ 40,0
Симптомдар	1 - бар 0 - жоқ
Глюкоза	Глюкозаның қалыпты деңгейі: <5,6 Диабеттің даму қаупі жоғары: ≥5,6 және ≤6,0 Аш қарынға гликемияның (преддиабет): ≥6,1 және ≤ 6,9 Қант диабеті (диагностика критерийлері бойынша): ≥7,0
Гемоглобин (Hb)	< 5,7 %
Гликозилденген гемоглобин (HbA1c)	Қалыпты деңгей, диабеттің даму қаупі төмен: < 5,5 % Қауіптің жоғарылауы (шекаралық жағдай): ≥ 5,5 % және ≤ 5,9 % Преддиабет, гликемияның бұзылуы: ≥ 6,0 % және ≤ 6,4 % 2 типті қант диабеті (диагностикалық деңгей): ≥ 6,5 %
Гематокрит (HCT)	Ерлер: 40-50 % Әйелдер: 36-46 %
Бір эритроциттің орташа көлемі MCV	Қалыпты: 80-100 fL Микроцитарлық анемия: MCV < 80 fL; Макроцитарлық анемия: MCV > 100 fL; Нормоцитарлық анемия: Қалыпты шегінде



2-кестенің соңы

Көрсеткіштер атауы	Мәндер диапазоны
Бір эритроциттегі гемоглобиннің орташа мөлшері (MCH)	Қалыпты: 27-33 pg Гипохромды анемия: Төмендеген MCH; Макроцитарлық анемия: Жоғарылаған MCH
Эритроциттер массасындағы гемоглобиннің орташа концентрациясы (MCHC)	Қалыпты: 32-36 г/дл Гипохромия: Төмендеген MCHC; Тұқым қуалайтын сфероцитоз: Жоғарылаған MCHC
Қанның рН-ы	Төмен: < 7.3 Жоғары немесе қалыпты рН: > 7.3
Несептегі глюкоза	Қалыпты: глюкоза табылмады; Глюкозурия: глюкоза табылды
Ферритин – организмдегі темір қорын қамтамасыз ететін ақуыз	Ерлер: 30-400 Әйелдер: 15-150 Балалар: 7-140
Тромбоциттер – қан ұю процестеріне (гемостаз) және тамырлардың жазылуына қатысатын қан пластинкалары	Қалыпты шегінде: 150-400 Тромбоцитопения (қан кету қаупі): < 150 Тромбоцитоз (қабыну, ісіктер кезінде болуы мүмкін): > 400
Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)	

Клиникалық шешімдер қабылдауды қолдау міндеттері үшін CNN моделі негізінде Grid Search әдісін қолдану алгоритмін эксперименттік зерттеу Keras және TensorFlow сияқты қажетті кітапханаларды импорттаудан басталады. Бұл кітапханалар нейрондық желілердің күрделі архитектураларын және оңтайландыру әдістерін құруға мүмкіндік береді.

Барлық қажетті кітапханалар жүктелгеннен кейін, деректер жүктеліп, соңғы деректер фреймі жасалады. Деректер фреймін жасаумен қатар, бос орындарды жою, деректерді теңестіру, деректерді біріктіру және кездейсоқ араластыру жүргізіледі, алдыңғы реттілікті жою үшін индекстер қалпына келтіріледі. Нәтижесінде:

- Диабетсіз пациенттер саны: 552336;
- Диабеті бар пациенттер саны: 15969;
- Қорытынды теңдестірілген DataFrame өлшемі: 13 бағанда 31938 жол.

Машиналық оқыту модельдерін оқыту үшін деректерді дайындау бірнеше алдын ала өңдеу кезеңдерін қамтыды. Бастапқы кезеңде бастапқы деректер жиынтығынан Diabetes мақсатты айнымалысы алынып тасталды, оның мәндері Y белгілердің бөлек векторында сақталды. Модельдеуде пайдаланылатын белгілер үшін категориялық айнымалылар One-Hot Encoding әдісі арқылы сандық форматқа түрлендірілді, бұл оларды одан әрі талдауға жарамды бинарлық индикаторлар түрінде көрсетуге мүмкіндік берді. Сандық белгілер бастапқы деректер фреймінен тікелей алынды. Осы операциялардың нәтижесі X белгілер матрицасы болды, ол сандық және кодталған категориялық деректерді біріктірді, бұл машиналық оқыту алгоритмдерін оқыту үшін ақпаратты толық ұсынуды қамтамасыз етті.

Қорытынды белгілер матрицасын қалыптастыру сандық және категориялық атрибуттарды конкатенациялаумен қатар жүрді, бұл зерттелетін нысандардың кешенді сипаттамасын жасауға мүмкіндік берді. X және Y матрицаларының өлшемдерін соңғы тексеру олардың дұрыстығын растады: жиынтық 31 938 дананы және 38 белгіні қамтыды, бұл модельдеудің кейінгі кезеңіне деректердің сәтті дайындалғанын көрсетеді. Әрі қарай, іріктеме оқу және тест іріктемелеріне бөлінді.

Модельдерді оқыту үшін кластарды теңестірудің екі тәсілі сыналды. Бірінші тәсіл кластардың теңгерімсіздігін түзетпей, бастапқы таралудағы деректерді пайдалануды



қарастырды. Екінші тәсіл Tomek Links әдісін қолдануды қамтыды, ол кластардың қиылысатын нысандарын жоюға және олардың қабаттасу дәрежесін төмендетуге бағытталған. Мұндай әдіс теңгерімсіз деректермен жұмыс істеу кезінде шуды азайтуға және жіктеу сапасын арттыруға мүмкіндік береді. Теңестіру орындалғаннан кейін деректер қайтадан оқыту және тест іріктемелеріне бөлініп, әрі қарай әртүрлі жіктеу алгоритмдерін қолдана отырып кросс-валидация жүргізілді.

Зерттеу аясында заманауи экспресс-глюкометрлерден, гематологиялық және биохимиялық анализаторлардан алынған деректер қант диабеті мен анемияны диагностикалау міндеттерін шешу үшін машиналық оқыту алгоритмдерімен өңделді. Модельдеу кезеңінде қолданылған негізгі алгоритмдер 3-кестеде көрсетілген.

3-кесте. Зерттеудің негізгі алгоритмдері

Машиналық оқыту алгоритмі	Мақсаты	Артықшылықтары
Decision Tree Classifier	Модельді ағаш түрінде құрады, мұнда әрбір түйін бір белгі бойынша шарт болып табылады, ал тармақтар – мүмкін болатын мәндер. Жапсырма түйіндері кластарға сәйкес келеді (мысалы, «анемия», «қалыпты», «диабет»).	Интерпретациялау мүмкіндігі; Жылдам оқыту; Белгілерді нормалауды қажет етпейді.
AdaBoost Classifier	Өлсіз модельдердің (әдетте бір-екі тереңдіктегі ағаштар) тізбегін құрады, олардың әрқайсысы алдыңғы қателерге назар аударады. Қорытынды шешім – барлық модельдердің өлшенген дауыс беруі.	Бір ағашпен салыстырғанда дәлдікті арттырады; Салыстырмалы түрде таза, теңдестірілген деректерде жақсы жұмыс істейді.
Random Forest Classifier	Әртүрлі кіші іріктемелердегі деректердің және белгілердің кездейсоқ кіші жиындарындағы көптеген шешім ағаштарын құрады. Қорытынды шешім – барлық ағаштардың дауыс беруі.	Жоғары дәлдік және қайта оқытуға төзімділік; Белгілердің маңыздылығын автоматты түрде бағалау; Теңгерімсіз деректермен жұмыс істейді.
Extra Trees Classifier	Көбірек кездейсоқтық енгізеді: белгілерді кездейсоқ таңдау ғана емес, сонымен қатар бөлу шектерін кездейсоқ анықтау. Бұл ағаштар арасындағы корреляцияны төмендетеді және жалпылау қабілетін арттырады.	Random Forest-қа қарағанда жылдамырақ; Үлкен деректер жиынтықтарымен жақсы жұмыс істейді; Қайта оқытуға жоғары төзімділік.
Gradient Boosting Classifier	Ағаштардың тізбегін оқытады, олардың әрқайсысы жоғалту функциясын (мысалы, log-loss) азайта отырып, алдыңғы қателерді түзетеді. AdaBoost-тан айырмашылығы, қателер бойынша градиенттік түсуді қолданады.	Өте жоғары дәлдік; Икемділік: кез келген жоғалту функциясы бойынша оңтайландыруға болады; Әртүрлі деректер түрлерімен (сандық, категориялық) жұмыс істейді.

Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)

Медициналық міндеттерде машиналық оқыту алгоритмін таңдау талдау мақсаттарымен де, деректердің сипаттамаларымен де анықталады. Базалық жіктеу үшін Decision Tree (Шешім ағашы) (қарапайымдылық және интерпретациялау мүмкіндігі) және Random Forest (Кездейсоқ орман) (тұрақтылық, жалпылау қабілеті) қолданылады. Жоғары дәлдік пен күрделі заңдылықтарды анықтауға басымдық беру Gradient Boosting-ті пайдалануды талап етеді, ал үлкен көлемдегі деректермен жұмыс істегенде, ресурстардың



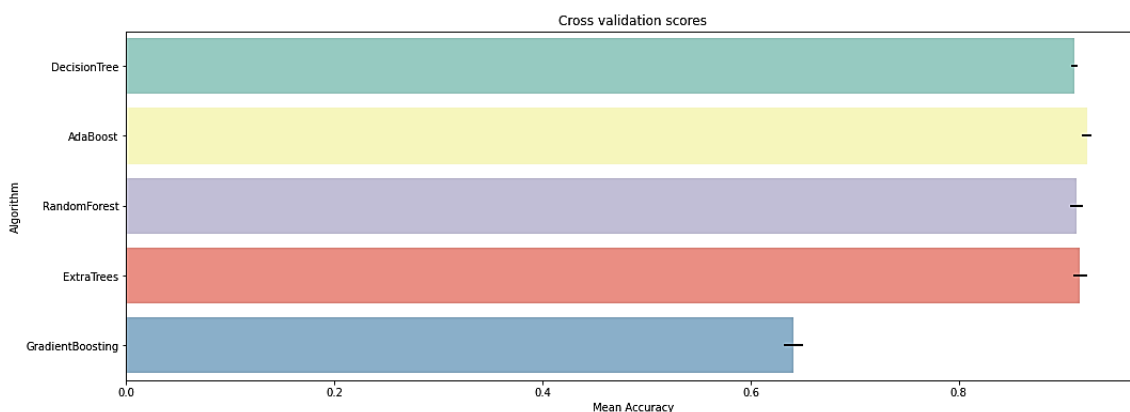
орташа шығынымен жоғары өнімділікті қамтамасыз ететін Extra Trees тиімді болып табылады. Шағын және «таза» іріктемелер үшін теңдестірілген нұсқа – AdaBoost.

Зерттеуде сонымен қатар классикалық алгоритмдер (логистикалық регрессия – интерпретацияланатын базалық модель; SVM – шағын құрылымдалған іріктемелердегі жоғары дәлдік; XGBoost – икемділік және болжау қабілеті) және нейрондық желілік тәсілдер (MLP – кестелік деректер үшін, CNN – кескіндер үшін, LSTM – уақыттық қатарлар үшін, соның ішінде глюкоза мониторингі және ЭКГ) тексерілді. Модельді таңдау деректер түріне және қажетті интерпретациялау деңгейіне байланысты болды, бұл клиникалық тәжірибеде ерекше маңызды деп Yin Y. & Bingi Y. (2023) атап өткен.

Алгоритмдердің сапасы k-fold кросс-валидация және стандартты метрикалар: Accuracy, Precision, Recall, F1-score және AUC-ROC көмегімен бағаланды. Ашықтықты арттыру үшін белгілерді интерпретациялау әдістері: ансамбльдік алгоритмдердегі (Random Forest, XGBoost) feature importance және жекелеген бақылаулар деңгейінде белгілердің үлесін сандық бағалауға мүмкіндік беретін SHAP қолданылды. Бұл тәсілдер модельдердің сапасын және олардың клиникалық қолданылуын арттыруды қамтамасыз етті, себебі дәрігер қандай параметрлер нәтижеге үлкен әсер еткенін түсінеді.

НӘТИЖЕЛЕР ЖӘНЕ ОЛАРДЫ ТАЛҚЫЛАУ

DecisionTree, AdaBoost, RandomForest, ExtraTrees және GradientBoosting алгоритмдері үшін кросс-валидация нәтижелерінің графигі ұсынылған, онда дәлдіктің орташа мәндері мен стандартты ауытқулары 2-суретте көрсетілген.



2-сурет. Кросс-валидация нәтижелерінің графигі

Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)

Кросс-валидация нәтижелеріне сәйкес, 4-кестеде ұсынылған, ағымдағы деректер жиынтығы жағдайында қай модельдердің жіктеу міндетін жақсы орындайтынын бағалауға болады.

4-кесте. Алгоритмдердің кросс-валидация нәтижелері

	CrossValMeans	CrossValerrors	Algorithm
0	0.911332	0.002996	DecisionTree
1	0.923008	0.004382	AdaBoost
2	0.913211	0.005768	RandomForest
3	0.916611	0.006340	ExtraTrees
4	0.640987	0.00891	GradientBoosting

Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)



Кросс-валидация нәтижелеріне (5-кесте, 2-сурет) сүйене отырып, келесідей қорытындылар жасауға болады:

- AdaBoost ең төменгі стандартты ауытқу (0.004382) кезінде ең жоғары орташа дәлдікті (0.923008) көрсетті, бұл модельдің тұрақтылығын растайды.
- ExtraTrees және RandomForest те орташа стандартты ауытқу мәндерінде (0.006340 және 0.005768) жоғары дәлдікті (0.916611 және 0.913211) көрсетті, бұл олардың сенімділігін және жақсы жалпылау қасиеттерін растайды.
- DecisionTree минималды стандартты ауытқу (0.002996) кезінде 0.911332 орташа дәлдікке жетті, бұл тұрақтылықты көрсеткенімен, ансамбльдік әдістерден қалыс қалды.
- GradientBoosting, керісінше, төмен орташа дәлдікті (0.640987) және жоғары стандартты ауытқуды (0.00891) көрсетті, бұл қайта оқытуды немесе деректерге жеткіліксіз бейімделуді көрсетуі мүмкін.

Ансамбльдік алгоритмдерді оңтайландыру үшін EstimatorSelectionHelper класы қолданылды, ол дәлдік (ассурасы) метрикасы бойынша бағалаумен гиперпараметрлерді толық іздеуді (Grid Search) ұйымдастырады. Алынған нәтижелер grid_searches сөздігінде сақталды, бұл оңтайлы модельдерді таңдауды автоматтандыруды, алгоритмдер мен олардың конфигурацияларын қатар салыстыру мүмкіндігін, сондай-ақ олардың өнімділігі туралы тұтас түсінік қалыптастыруды қамтамасыз етті.

5-кесте. GridSearchCV жұмысының нәтижесі

Estimator	Min Score	Mean Score	Max Score	Std Score	Лучшая конфигурация
Gradient Boosting Classifier	0.635	0.794	0.916	0.009	max_depth=9, n_estimators=500
Random Forest Classifier	0.661	0.668	0.678	0.006	max_depth=9, n_estimators=500
Extra Trees Classifier	0.640	0.645	0.654	0.006	max_depth=9, n_estimators=500
<i>Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)</i>					

Жүргізілген эксперименттер негізінде, осы деректер жиынтығы үшін классификатор параметрлерінің ең тиімді конфигурациялары туралы қорытынды жасауға болады. Атап айтқанда, GradientBoostingClassifier max_depth = 9, n_estimators = 500 және max_features = 'auto' параметрлерімен орташа және максималды дәлдік бойынша ең жақсы көрсеткіштерді көрсетті, бұл оны одан әрі қолдану және тереңдетілген тестілеу үшін перспективалы нұсқа ретінде қарастыруға мүмкіндік береді.

Әрбір модель үшін гиперпараметрлердің әртүрлі комбинацияларымен тестілеу жүргізілді, сондай-ақ оқу және тест деректер жиынтықтарында ассурасы, F1-score, recall және precision қоса алғанда, қосымша сапа метрикалары есептелді. Айта кету керек, классификациялық есептердің алдындағы жиынтық нәтижелерде көрсетілген тест іріктемесінің дәлдігі (Test Accuracy) және басқа метрикалардың мәндері түпкілікті болып табылмайды. Дұрыс көрсеткіштер модельдердің нақты тиімділігін толық көрсететін классификация есептерінде және қателер матрицаларында ұсынылған.

Мәселен, GradientBoosting алгоритмі үшін тест деректер жиынтығында 0.92 дәлдікке (Test Accuracy) қол жеткізілді, ал оң класс үшін (True) F1-score мәні 0.91 құрады, бұл precision және recall көрсеткіштерінің теңгерімділігін көрсетеді. Бұл модель үшін қателер матрицасы оң класс үшін 116 бірінші текті қателікті (False Positive) және 669 екінші текті қателікті (False Negative) көрсетті. Ұқсас талдау барлық қалған классификаторлар үшін де жүргізілді, бұл олардың тиімділігін объективті салыстыруға мүмкіндік берді.



Жинақталған нәтижелер GradientBoosting негізгі метрикалар бойынша ең жоғары мәндерді көрсететінін дәлелдейді. Сонымен қатар, модельді іс жүзінде енгізу кезінде қателердің әртүрлі түрлері арасындағы ымыраны ескеру қажет. Атап айтқанда, клиникалық міндетке байланысты False Positive немесе False Negative қателерді азайтуға басымдық берілуі мүмкін. Қосымша, алгоритмнің есептеу ресурстарына қойылатын талаптарын және жұмыс істеу жылдамдығын ескеру қажет, бұл нақты уақыт жүйелеріне интеграциялау үшін өте маңызды. Барлық тестіленген классификаторлар үшін метрикалардың қорытынды мәндері, соның ішінде Train Accuracy, Test Accuracy, Precision, Recall және F1-score көрсеткіштері кесте 6-да келтірілген.

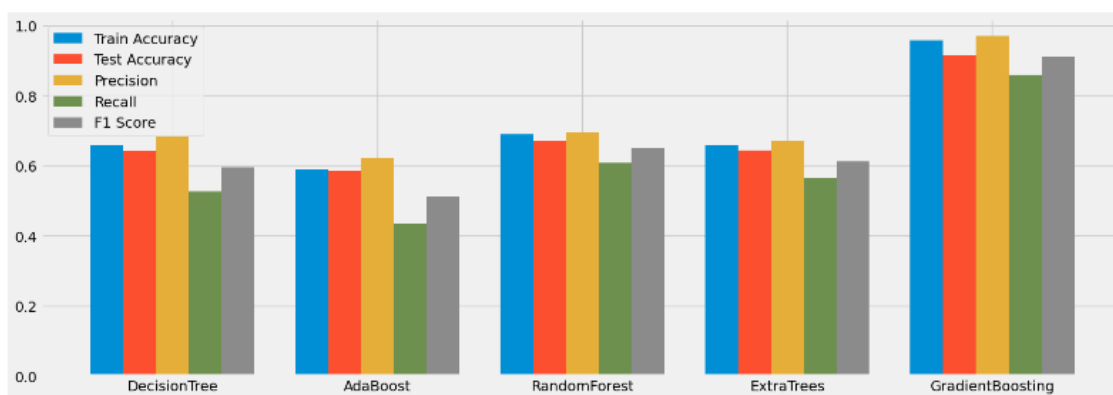
3 кесте. Әртүрлі машиналық оқыту классификаторларын бағалау нәтижелері

Model	Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
DecisionTree	0,658347	0,641202	0,684453	0,526458	0,595148
AdaBoost	0,590222	0,585055	0,623501	0,433333	0,511308
RandomForest	0,688853	0,671467	0,696948	0,608958	0,649989
ExtraTrees	0,658481	0,643394	0,671887	0,563125	0,612717
GradientBoosting	0,958356	0,916301	0,970132	0,859375	0,911401

Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)

GradientBoostingClassifier алгоритмі тестіленген модельдер арасында жіктеу міндетін шешу үшін ең тиімді болып табылатындығын көрсете отырып, барлық негізгі метрикалар бойынша ең жақсы нәтижелерді көрсетті. Бұл алгоритм дәлдік, толықтық және F1-score көрсеткіштері арасындағы оңтайлы теңгерімді қамтамасыз етті, бұл оның жоғары тұрақтылығын және күрделі, мультифакторлы деректерді барабар өңдеу қабілетін растайды.

3-суретте бес классификаторды: DecisionTree, AdaBoost, RandomForest, ExtraTrees және GradientBoosting салыстырмалы бағалау нәтижелері ұсынылған. Әрбір алгоритм үшін келесі көрсеткіштер есептелді және талданды: оқыту деректер жиынтығындағы дәлдік (Train Accuracy), тест деректер жиынтығындағы дәлдік (Test Accuracy), сондай-ақ Precision, Recall метрикалары және F1-score интегралдық көрсеткіші. Салыстырмалы талдау жеке модельдермен (мысалы, DecisionTree) салыстырғанда неғұрлым жоғары тұрақтылық пен болжау қабілетін көрсететін ансамбльдік әдістердің артықшылықтарын анықтауға мүмкіндік береді.



3-сурет. Таңдалған машиналық оқыту алгоритмдері үшін
әртүрлі сапа метрикаларын салыстыру графигі

Ескерту: Автордың деректері негізінде жасалған (Увалиева И.М., 2025)



GradientBoostingClassifier алгоритмі тестіленген модельдер арасында сенімді түрде ерекшеленіп, сапаны бағалаудың барлық бес метрикасы бойынша ең жоғары мәндерді көрсетті. Бұл оның тиімді жалпылауға және жаңа іріктемелерде дұрыс болжауға қабілеттілігін растайды, бұл оны қойылған жіктеу міндетін шешу үшін ең қолайлы нұсқа етеді.

DecisionTree және AdaBoost модельдері салыстырмалы нәтижелер көрсетті, бұл ретте AdaBoost біршама әлсіздеу болды. RandomForest және ExtraTrees алгоритмдері де бір-біріне жақын көрсеткіштерді көрсетті, алайда олардың жіктеу сапасы GradientBoosting нәтижелерінен төмен болды. Эксперимент заманауи глюкометрлерден, гематологиялық және биохимиялық анализаторлардан алынған деректерді талдау кезінде машиналық оқытудың ансамбльдік әдістерінің жоғары тиімділігін растады.

Кросс-валидация кезеңінде AdaBoost ең жақсы нәтижелерді көрсетті, ол төмен стандартты ауытқумен (0.0044) орташа 0.923 дәлдікке жетті, бұл оның әртүрлі кіші іріктемелермен жұмыс істеу кезіндегі тұрақтылығын көрсетеді. Алайда, Grid Search көмегімен гиперпараметрлерді оңтайландырғаннан кейін GradientBoosting көпбасшылыққа шығып, тест іріктемесінде 0.92 дәлдікті және 0.91 F1-score мәнін көрсетті.

Қателер матрицаларын талдау GradientBoosting-тің бірінші және екінші текті қателердің минималды мөлшерін жіберетінін анықтады, бұл жалған оң және жалған теріс нәтижелердің салдары өте ауыр болуы мүмкін медициналық міндеттерде критикалық маңызға ие. Осылайша, оңтайландырылған GradientBoosting моделі дәлдік, толықтық және жіктеу тұрақтылығының ең жақсы үйлесімін көрсетеді, бұл оны іс жүзінде енгізу үшін ең перспективалы етеді.

Алынған нәтижелер ансамбльдік әдістердің, атап айтқанда градиентті бустингтің гематологиялық және биохимиялық деректер негізінде ауруларды диагностикалаудағы жоғары тиімділігін көрсететін заманауи зерттеулермен сәйкес келеді. Әдебиетте GradientBoosting-тің шуларға және лақтырылып тастауларға (выброс) төзімділігі, тіпті шектеулі іріктемелерде тиімді жұмыс істеу қабілеті, сондай-ақ оңтайлы нәтижелерге қол жеткізу үшін гиперпараметрлерді мұқият баптау қажеттілігі атап өтіледі. Бұл заңдылықтар осы зерттеу аясында да расталды, бұл белгілі тәсілдердің қойылған міндет үшін қайталану мүмкіндігін және қолданылуын көрсетеді.

Әзірленген модельді іс жүзінде енгізу перспективалары экспресс-анализаторлардан алынған деректерді машиналық оқыту алгоритмдерімен интеграциялау мүмкіндігімен байланысты. Бұл қант диабеті мен анемияларды диагностикалаудың дәлдігін арттыруға, скрининг пен мониторинг процесін жеделдетуге, сондай-ақ медициналық персоналға түсетін жүктемені азайтуға мүмкіндік береді. Мұндай шешімдерді медициналық ақпараттық жүйелерге және мобильді қосымшаларға енгізу жекелендірілген медицина және пациенттердің жағдайын қашықтықтан бақылау қағидаттарын іске асыруға мүмкіндік береді. Модельдің күнделікті зертханалық көрсеткіштерді өңдеудегі жоғары дәлдігі оны ресурстар мен уақыт көбінесе шектеулі болатын бастапқы медициналық-санитарлық көмек буыны үшін ерекше құнды етеді.

Сонымен қатар, зерттеудің бірқатар шектеулері бар. Біріншіден, іріктеме көлемі мен әртүрлілігі бойынша шектеулі, бұл нәтижелерді әртүрлі этникалық, жас және клиникалық профилі бар кеңірек популяцияларға экстраполяциялауды қиындатады. Екіншіден, деректер глюкометрлер мен анализаторлардың нақты модельдерін пайдалану арқылы алынды, бұл нәтижелердің басқа өндірушілердің құрылғыларына ауыстырылуын төмендетеді. Үшіншіден, талдау офлайн режимінде орындалды, және алгоритмдерді нақты уақыт жүйелеріне интеграциялау мүмкіндіктері әлі зерттелмеген, дегенмен дәл осы аспект практикалық қолдану үшін негізгі болып табылады.

Сонымен қатар, зерттеудің басты назары жіктеу дәлдігіне аударылды, ал клиникалық тәжірибеде басымдыққа ие болатын нақты мақсаттар, мысалы, жүкті әйелдердегі немесе



созылмалы аурулары бар пациенттердегі анемияны диагностикалау кезінде жалған теріс нәтижелердің санын азайту болуы мүмкін.

ҚОРЫТЫНДЫ

Медициналық жабдықтардан алынатын деректерді машиналық оқыту алгоритмдерімен интеграциялау қант диабеті мен анемияны автоматтандырылған, жылдам және дәл диагностикалау үшін жаңа мүмкіндіктер ашады. Мұндай тәсіл ауруларды ерте сатыларда анықтаудың тиімділігін арттырып қана қоймай, емдеудің жекелендірілген стратегияларын қалыптастыруға және пациенттердің жағдайын клиникалық жағдайларда да, үйде де неғұрлым сапалы бақылауды қамтамасыз етуге мүмкіндік береді.

Жүргізілген зерттеу заманауи экспресс-глюкометрлердің, гематологиялық және биохимиялық анализаторлардың деректерін машиналық оқыту алгоритмдерімен біріктіру диагностика дәлдігін айтарлықтай арттыратынын көрсетті. Ең жақсы нәтижелерді GradientBoosting моделі көрсетті, ол гиперпараметрлерді оңтайландырылғаннан кейін 0.92 дәлдікке және 0.91 F1 score мәніне жетті, бұл қалған алгоритмдерден барлық негізгі метрикалар бойынша басым болды.

Әзірленген жүйе патологиялық жағдайларды ерте кезеңдерде анықтауға, клиникалық шешімдерді қолдау құралы ретінде қызмет етуге, сондай-ақ телемедицина және бастапқы медициналық-санитарлық көмек буыны шеңберінде қолданылуға қабілетті. Бұл медициналық ресурстардың тапшылығы және созылмалы ауруларды үздіксіз бақылау қажеттілігі жағдайында ерекше маңызды.

Сонымен қатар, зерттеудің бірқатар шектеулері бар: іріктеменің салыстырмалы түрде шағын және біртекті емес құрамы, анализаторлардың нақты модельдеріне тәуелділік, сондай-ақ жүйені нақты уақыт режимінде тестілеудің болмауы. Алдағы уақытта деректер базасын кеңейту, клиникалық ақпараттық жүйелермен интеграциялау және шешімдерді жабдықтардың әртүрлі түрлеріне бейімдеу қажет.

Осылайша, жұмыс нәтижелері медициналық құрылғылар мен жасанды интеллект алгоритмдерінің мүмкіндіктерін біріктіретін интеллектуалды диагностикалық жүйелерді құрудың жоғары әлеуетін көрсетеді, бұл медициналық көмектің сапасы мен қолжетімділігін арттыруға жол ашады.

МҮДДЕЛЕР ҚАЙШЫЛЫҒЫ: Авторлар мүдделер қақтығысының жоқ екенін мәлімдейді.

ҚАРЖЫЛАНДЫРУ: Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитеті гранттық қаржыландыру шеңберінде қаржыландырылады (грант No AP19679525 «Электрондық денсаулық паспортына арналған клиникалық-гематологиялық синдромдарды диагностикалаудың бағдарламалық кешені»).

АЛҒЫС БІЛДІРУ: Авторлар экспресс-глюкометрлерді, гематологиялық және биохимиялық анализаторларды таңдауда және пайдалануда көрсеткен көмегі үшін «А-Медикал» ЖШС техникалық мамандарына алғыс білдіреді.

ӘДЕБИЕТТЕР ТІЗІМІ

- Farooqi, M., Tahir, Y., & Rehan, B. (2022). Anemia in Patients with Type 2 Diabetes: A Common but Neglected Clinical Finding. *Acta Bio-medica: Atenei Parmensis*, 93(1), e2022044-e2022044. <https://doi.org/10.23750/abm.v93i1.11204>
- Liu, J., Bai, R., Chai, Z., Cooper, M. E., Zimmet, P. Z., & Zhang, L. (2022). Low-and middle-income countries demonstrate rapid growth of type 2 diabetes: an analysis based on Global Burden of Disease 1990–2019 data. *Diabetologia*, 65(8), 1339-1352. <https://doi.org/10.1007/s00125-022-05713-6>



- Iolascon, A., Andolfo, I., Russo, R., Sanchez, M., Busti, F., Swinkels, D., ... & from EHA-SWG Red Cell and Iron. (2024). Recommendations for diagnosis, treatment, and prevention of iron deficiency and iron deficiency anemia. *Hemasphere*, 8(7), e108. <https://doi.org/10.1002/hem3.108>
- Crasto, W., Patel, V., Davies, M. J., & Khunti, K. (2021). Prevention of microvascular complications of diabetes. *Endocrinology and metabolism clinics of North America*, 50(3), 431-455. <https://doi.org/10.1016/j.ecl.2021.05.005>
- Pittas, A. G., Kawahara, T., Jorde, R., Dawson-Hughes, B., Vickery, E. M., Angellotti, E., ... & Balk, E. M. (2023). Vitamin D and risk for type 2 diabetes in people with prediabetes: a systematic review and meta-analysis of individual participant data from 3 randomized clinical trials. *Annals of internal medicine*, 176(3), 355-363. <https://doi.org/10.7326/M22-301>
- AlFalasi, S. M., Abdouli, K. A., & Aldashti, N. A. (2024). Association of anemia and diabetic retinopathy among patients with type 2 diabetes mellitus: retrospective cross-sectional study. *Cureus*, 16(8). <https://doi.org/10.7759/cureus.67995>
- Brittenham, G. M., Moir-Meyer, G., Abuga, K. M., Datta-Mitra, A., Cerami, C., Green, R., ... & Atkinson, S. H. (2023). Biology of anemia: a public health perspective. *The Journal of nutrition*, 153, S7-S28. <https://doi.org/10.1016/j.tjnut.2023.07.018>
- Neumann, G., & Kawaoka, Y. (2023). Which Virus Will Cause the Next Pandemic? *Viruses*, 15(1), 199. <https://doi.org/10.3390/v15010199>
- Kubihal, S., Goyal, A., Gupta, Y., & Khadgawat, R. (2021). Glucose measurement in body fluids: a ready reckoner for clinicians. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15(1), 45-53. <https://doi.org/10.1016/j.dsx.2020.11.021>
- Jain, P., Joshi, A. M., Mohanty, S. P., & Cenkeramaddi, L. R. (2024). Non-invasive glucose measurement technologies: Recent advancements and future challenges. *IEEE Access*, 12, 61907-61936. <https://doi.org/10.1109/ACCESS.2024.3389819>
- Deberneh, H. M., & Kim, I. (2021). Prediction of type 2 diabetes based on machine learning algorithm. *International journal of environmental research and public health*, 18(6), 3317. <https://doi.org/10.3390/ijerph18063317>
- Mousa, A., Mustafa, W., & Marqas, R. B. (2023). A comparative study of diabetes detection using the Pima Indian diabetes database. *Methods*, 7, 8. <https://doi.org/10.26682/sjuod.2023.26.2.24>
- Ramzan, M., Sheng, J., Saeed, M. U., Wang, B., & Duraihem, F. Z. (2024). Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms. *Visual Computing for Industry, Biomedicine, and Art*, 7(1), 18. <https://doi.org/10.1186/s42492-024-00169-4>
- Geukes Foppen R.J., Gioia V., Gupta S., Johnson C.L., Giantsidis J., Papademetris M. (2025). Methodology for Safe and Secure AI in Diabetes Management. *Journal of Diabetes Science and Technology*, 19(3):620-627. <https://doi.org/10.1177/19322968241304434>
- Mitra, U., Sarkar, P., Mondal, J., & Kundu, J. (2025, February). Enhancing Interpretability in Diabetics Prediction: A Comparative Study of SHAP, LIME and Permutation Feature Importance. In *2025 AI-Driven Smart Healthcare for Society 5.0* (pp. 1-6). IEEE. <https://doi.org/10.1109/IEEECONF64992.2025.10962890>
- Cournoyer, A., Deschamps, A., Bau-Gaudreault, L., Dubreuil, P., & Benoit-Biancamano, M. O. (2022). Analytical validation of a portable human Accu-Chek glucometer in honeybee hemolymph. *Journal of Veterinary Diagnostic Investigation*, 34(5), 789-795. <https://doi.org/10.1177/104063872211172>
- Jain, P., Joshi, A. M., Mohanty, S. P., & Cenkeramaddi, L. R. (2024). Non-invasive glucose measurement technologies: Recent advancements and future challenges. *IEEE Access*, 12, 61907-61936. <https://doi.org/10.1109/ACCESS.2024.3389819>



- Daves, M., Roccaforte, V., Lombardi, F., Panella, R., Pastori, S., Spreafico, M., ... & Piccin, A. (2024). Modern hematology analyzers: beyond the simple blood cells count (with focus on the red blood cells). *Journal of Laboratory and Precision Medicine*, 9. <https://doi.org/10.21037/jlpm-23-32>
- Gómez, J. G., Parra Urueta, C., Álvarez, D. S., Hernández Riaño, V., & Ramirez-Gonzalez, G. (2025). Anemia Classification System Using Machine Learning. *Informatics*, 12(1), 19. <https://doi.org/10.3390/informatics12010019>
- Yin, Y., & Bingi, Y. (2023). Using machine learning to classify human fetal health and analyze feature importance. *BioMedInformatics*, 3(2), 280-298. <https://doi.org/10.3390/biomedinformatics3020019>

Авторлар туралы мәліметтер
Информация об авторах
Information about authors



Хасенова Зарина Толеубековна – PhD, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан
Хасенова Зарина Толеубековна – PhD, Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан

Khasenova Zarina Toleubekovna – PhD, East Kazakhstan Technical University named after D. Serikbaev, Ust-Kamenogorsk, Kazakhstan, e-mail: zthasenova@mail.ru, ORCID: <https://orcid.org/0009-0001-1417-7793>



Увалиева Индира Махматовна – PhD, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан
Увалиева Индира Махматовна – PhD, Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан

Uvaliyeva Indira Makhmutovna – PhD, East Kazakhstan Technical University named after D. Serikbaev, Ust-Kamenogorsk, Kazakhstan, e-mail: iuvaliyeva@ektu.kz, ORCID: <https://orcid.org/0000-0002-2117-5390>



Мукашева Роза Урумкановна – техника ғылымдарының кандидаты, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан

Мукашева Роза Урумкановна – кандидат технических наук, Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан

Mukasheva Roza Urumkanovna – Candidate of Technical Sciences, East Kazakhstan Technical University named after D. Serikbaev, Ust-Kamenogorsk, Kazakhstan, e-mail: mukashevaru@mai.ru, ORCID: <https://orcid.org/0000-0002-4722-4576>



Букунова Альмира Шошановна – медицина ғылымдарының кандидаты, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан

Букунова Альмира Шошановна – кандидат медицинских наук, Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан

Bukunova Almira Shoshanovna – Candidate of Medical Sciences, East Kazakhstan Technical University named after D. Serikbaev, Ust-Kamenogorsk, Kazakhstan,

e-mail: ABukunova@edu.ektu.kz,

ORCID: <https://orcid.org/0009-0000-8892-4248>



Кәрімқызы Бекзат – экономика ғылымдарының магистрі, «А-Медикал» ЖШС, Алматы, Қазақстан

Каримкызы Бекзат – магистр экономических наук, ТОО «А-Медикал», г. Алматы, Казахстан

Karimkyzy Bekzat – Master of Economic Sciences, LLP «A-Medical», Almaty, Kazakhstan,

e-mail: bekzastana@gmail.com,

ORCID: <https://orcid.org/0009-0004-4696-7927>