

**Ж.Е. Байғараева^{1,2}, А.К. Болтабоева^{1,2*}, Б.Т. Иманбек¹, М.И. Кожамбердиева^{1,2},
Ж.Қ. Жолдыбаева³**

¹Әл-Фараби атындағы Қазақ ұлттық университеті,
050040, Қазақстан, Алматы, Әл Фараби даңғылы 71

²ТОО «Kazakhstan R&D Solutions»,
050056, Қазақстан, Алматы, Кожедуба көшесі 3

³Абай атындағы Қазақ ұлттық педагогикалық университеті,
050010, Қазақстан, Алматы, Достық даңғылы 13

*e-mail: boltaboyeva_assiya3@live.kaznu.kz

КОРОНАРЛЫҚ ЖҮРЕК АУРУЫН БОЛЖАУДА МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ: UCI HEART DISEASE ДЕРЕКТЕР ЖИЫНЫ НЕГІЗІНДЕ

Аннотация: Коронарлық жүрек ауруы әлемдік өлім мен мүгедектіктің негізгі себептерінің бірі болып қала береді. Уақтылы диагностика асқынуларды азайтып, денсаулық сақтау жүйесіне түсетін жүктемені жеңілдетеді. Дегенмен дәстүрлі әдістер қымбат, инвазивті және қолжетімділігі шектеулі. Соңғы зерттеулер машиналық оқытудың клиникалық тәжірибеде қолданылу мүмкіндігін растайды. Осыған байланысты тек клинико-демографиялық деректерді пайдаланып, визуализациясыз ауруды сенімді болжау мүмкін бе деген сұрақ туындайды.

Бұл жұмыстың мақсаты – мұндай модельдердің дәлдігі мен қолданбалық құндылығын бағалау. Біріктірілген протокол бойынша UCI Heart Disease деректер жиынында ($n = 920$) LightGBM алгоритмі оқытылып, мына нәтижелерге жетті: accuracy = 0.8696, precision = 0.8679, recall = 0.9020, F1-score = 0.8846. Нәтижелер алдыңғы визуализацияға сүйенген зерттеулерді толықтырды.

Зерттеу бірнеше алгоритмді бірдей алдын ала өңдеу мен валидация шарттарында салыстырды, ықтималдықтардың калибрленуін тексерді және SHAP әдісі арқылы интерпретация жасады. Талдау негізгі предикторлардың (мысалы, ST-сегмент депрессиясы) клиникалық біліммен сәйкес екенін көрсетті. Бұл модельді бастапқы скрининг пен қосымша диагностикаға жолдауда қолдануға болатынын дәлелдейді. Жалпы, ашық клиникалық деректерге негізделген калибрленген әрі түсіндірілетін алгоритмдер ресурсы шектеулі жағдайларда пациенттерді маршрутизациялауда құнды құрал бола алады.

Түйін сөздер: коронарлық жүрек ауруы (КЖА); машиналық оқыту; LightGBM; классификация; клинико-демографиялық деректер; тәуекелді стратификация; SHAP интерпретациясы; ROC-AUC; калибровка.

Кіріспе

Коронарлық жүрек ауруы (КЖА) – әлемдегі ең жиі кездесетін жүрек-қантамыр патологиясы және өлімнің басты себебі. Оның таралуына өмір салтының өзгеруі, халықтың қартаюуы және негізгі қауіп факторлары (гипертензия, диабет, дислипидемия, темекі шегу) әсер етеді. КЖА миокард инфарктісі мен жүрек жеткіліксіздігінен бөлек, еңбекке қабілеттілікті төмендетіп, DALY жоғалуына ықпал етеді.

Ерте диагностика шешуші мәнге ие, алайда клиникалық көріністің әркелкілігі мен қолданыстағы әдістердің шектеулілігі қиындық туғызады. «Алтын стандарт» саналатын инвазивті коронарлық ангиография қымбат әрі қауіпті, ал КТ-ангиография, МРТ, СPECT және биомаркерлердің (тропонин, микроРНҚ, hs-CRP) сезімталдығы шектеулі.

Соңғы жылдары машиналық оқыту (SVM, нейрондық желілер, CNN, RNN) КЖА диагностикасында жетістіктер көрсеткенімен, зерттеулер көбіне шағын деректерге сүйеніп, жоғары есептеу қуатын талап етеді. ЖИ модельдерін клиникалық тәжірибеге интеграциялау қажеттілігі артып отыр.

Бұл зерттеу тек клинико-демографиялық деректер негізінде КЖА бар-жоғын және оның ауырлығын болжау мүмкіндігін қарастырды. Бірыңғай алдын ала өңдеу (импутация, кодтау, стандартизация, SMOTE) және бірнеше алгоритм (градиенттік бустинг, ансамбльдер, KNN, SVM, логистикалық регрессия) салыстырылды. Бағалау accuracy, precision, recall, F1, ROC-AUC және SHAP интерпретациясы арқылы жүргізілді. Нәтижелер визуализациясыз, рутинді

деректер негізінде сенімді болжам жасауға болатынын көрсетіп, тәуекелді стратификацияда практикалық құндылыққа ие екенін дәлелдеді.

Материалдар мен әдістер

Зерттеуде UCI Machine Learning Repository-ден алынған классикалық Heart Disease Dataset пайдаланылды – коронарлық жүрек ауруын болжауға арналған ең кең таралған эталондық деректер жиыны [40]. Ол 303 пациенттің демографиялық және клиникалық көрсеткіштерін қамтиды: жас, жыныс, қан қысымы, холестерин, қант деңгейі, ЭКГ нәтижелері, максималды жүрек соғу жиілігі, стенокардия, ST сегменті параметрлері, тамырлар саны және талассемия индикаторы.

Нысаналы айнымалы (num) ҚЖА жоқтығын (0) немесе ауырлық дәрежесін (1-4) көрсетеді, сондықтан тапсырма бинарлы да, көпклассты да қарастырылады. Деректер Cleveland, Hungarian, Switzerland және VA Long Beach когорталарынан жиналған, бұл репрезентативтілікті арттырады.

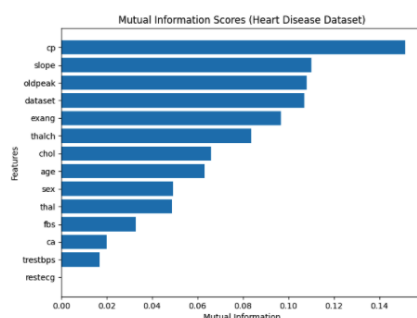
Алдын ала өңдеу міндетті болды: категориалды айнымалылар Label Encoding арқылы сандық форматқа түрлендірілді (мысалы, кеудедегі ауырсыну түрі, ЭКГ нәтижелері, жыныс, талассемия). Кейін көрсеткіштер Z-баға стандартизациясымен қалыпқа келтірілді, бұл барлық белгілердің үлесін теңестіріп, модельді оқытуда бір айнымалының басым болуын болдырмады [41, 42]. Формула келесідей:

$$z = \frac{x - \mu}{\sigma}$$

Мұндағы x – бақыланатын мән, μ – белгінің орташа мәні, σ – стандартты ауытқу. Бұл трансформация айнымалылардың масштабын теңестіріп, олардың модельді үйретудегі үлесін тең қылды.

Алдын ала өңдеуден кейін деректер машиналық оқытуға дайын болды. Белгілердің жүрек ауруын болжаудағы үлесін бағалау үшін өзара ақпарат (Mutual Information, MI) талдауы жүргізілді [43].

1-суретте көрсетілгендей, нысаналы айнымалымен ең жоғары өзара ақпаратқа ие белгілер қатарында стенокардия түрі (*cp*), ST сегментінің еңісі (*slope*), физикалық жүктемемен шақырылған ST депрессиясы (*oldpeak*), сондай-ақ деректер жиынының индикаторы (*dataset*) болды. Бұл көрсеткіштер ең жоғары болжамдық маңызға ие. Ал тыныштықтағы ЭКГ нәтижелері (*restecg*) және тыныштықтағы артериялық қысым деңгейі (*trestbps*) салыстырмалы түрде төмен мәндер көрсетті, бұл олардың болжамдағы үлесінің аздығын аңғартады. Осылайша, бұл талдау ең ақпаратты клинико-демографиялық сипаттамаларды бөліп көрсетуге және клиникалық түсіндіруді жеңілдетуге мүмкіндік берді.



Сурет 1 – Өзара ақпарат көрсеткіштері

Деректер 80/20 қатынасында train/test жиынтықтарына бөлінді [44]. Сынып теңгерімсіздігін жою үшін оқыту жиынына SMOTE қолданылды [45], ол сирек сыныптар үшін синтетикалық үлгілер жасап, модельдердің сезімталдығын арттырды. Oversampling тек оқытуда қолданылып, тест жиыны тәуелсіз күйінде қалды.

Нәтижесінде деректер құрылымдалып, тұрақты әрі клиникалық мәні бар модельдер құруға дайын болды.

IoT device Qimyl

Модельдер бірдей алдын ала өңдеуден кейін үйретіліп, төрт метрика бойынша бағаланды: accuracy, precision, recall және F1 [46]. Бұл көрсеткіштер медицинада маңызды, өйткені жалған теріс пен жалған оң нәтижелердің салдары зор.

Ансамбльдер мен бустинг әдістері сызықтық модельдер мен дара шешім ағаштарынан жоғары нәтиже көрсетті. RandomForest, CatBoost және HistGradientBoosting ең жақсы теңгерімді дәлдікке жетті. Logistic Regression мен Decision Tree төменірек көрсеткіш бергенімен, интерпретация үшін пайдалы болды. KNeighbors precision бойынша жақсы болғанымен, recall әлсіздігін көрсетті (1 кесте).

Кесте 1 – Коронарлық жүрек ауруын болжауда машиналық оқыту модельдерінің салыстырмалы нәтижелері

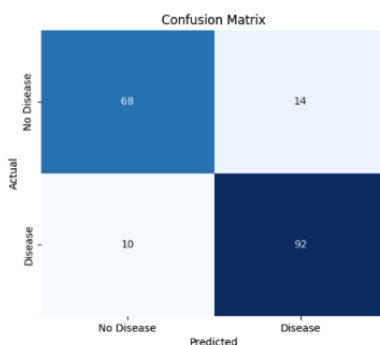
Модели	Accuracy	Precision	Recall	F1 Score
XGBoost+	0.8587	0.8585	0.8922	0.8750
CatBoost	0.8478	0.8491	0.8824	0.8654
KNeighbors+	0.8478	0.8854	0.8333	0.8586
ExtraTrees	0.8478	0.8558	0.88725	0.8641
LGBMClassifier	0.8696	0.8679	0.9020	0.8846
RandomForest	0.8587	0.8952	0.8624	0.8785
Decision Tree	0.8207	0.8710	0.7941	0.8308
HistGradientBoosting	0.8478	0.8426	0.8922	0.8667
Logistic Regression	0.8098	0.8454	0.8039	0.8241
SVM	0.8533	0.8713	0.8627	0.8670

Модельдерді бағалау accuracy, precision, recall және F1-score бойынша жүргізілді. Ең үздік нәтиже LightGBM көрсетті: accuracy = 0.8696, recall = 0.9020, бұл КЖА бар пациенттерді сенімді анықтау қабілетін дәлелдейді [47].

RandomForest және XGBoost та бәсекеге қабілетті болды (accuracy = 0.8587, F1 ≈ 0.87). ExtraTrees recall = 0.8872 және F1 = 0.8641 мәндеріне жетті. SVM теңгерімді көрсеткіштер берді (accuracy = 0.8533, F1 = 0.8670), ал KNeighbors F1 = 0.8586 болса да, recall төмен болды.

Decision Tree (accuracy = 0.8207, F1 = 0.8308) және Logistic Regression (accuracy = 0.8098, F1 = 0.8241) ең әлсіз нәтижелерді көрсетті, оларды тек базалық салыстыру үшін қолдануға болады.

Жалпы LightGBM ең тиімді алгоритм ретінде ерекшеленді, себебі ол күрделі бейсызық тәуелділіктерді таба алады және клиникалық тәжірибеде қолдануға әлеуеті жоғары (2 сурет).

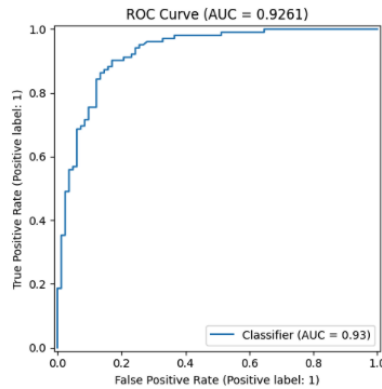


Сурет 2 – LightGBM моделі үшін қателік матрицасы (Confusion Matrix)

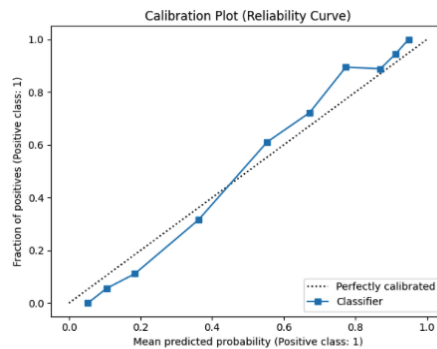
Қателік матрицасы модель сапасын талдауға мүмкіндік береді. LightGBM нәтижелері: 68 TN, 92 TP, 14 FP және 10 FN. Бұл сезімталдық пен ерекшелік арасындағы жақсы тепе-теңдікті көрсетеді. Жалған терістердің аз болуы ауруды өткізіп алмауда маңызды, ал жалған оңдардың орташа деңгейі гипердиагностикаға бейімділіктің төмен екенін дәлелдейді [48].

ROC-қисық классификатордың әртүрлі шектердегі жұмысын бағалайды. LightGBM үшін AUC = 0.93 болып, модельдің КЖА бар және жоқ пациенттерді ажыратудағы жоғары қабілетін көрсетті (3 сурет). Қисық диагональдан едәуір жоғары орналасып, кездейсоқ болжаудан айқын артықшылығын дәлелдеді. AUC > 0.9 медицинада тамаша нәтиже саналады, бұл LightGBM-нің сенімділігін растайды.

Калибровка графигі LightGBM ықтималдықтарының шынайы көрсеткіштерге сәйкес екенін көрсетті: қисық диагональға жақын орналасқан, яғни модель бағалаулары нақты тәуекелді дұрыс бейнелейді [49]. Бұл дәрігерлерге ықтималдықтарды клиникалық шешімдерге тікелей қолдануға мүмкіндік береді (4 сурет).



Сурет 3 – LightGBM моделі үшін ROC-қисығы және AUC мәні



Сурет 4 – LightGBM моделі үшін калибровка графигі

SHAP талдауы LightGBM-нің негізгі предикторлары ретінде артериялық қысым, холестерин, жас және жынысты анықтады. Бұл факторлар дәстүрлі қауіп көрсеткіштерімен сәйкес келеді және модельдің медициналық негізділігін дәлелдейді. SHAP сонымен қатар белгілердің КЖА ықтималдығына қалай әсер ететінін көрсетті, бұл дәрігерлерге нәтижелерді түсіндіруге және пациенттерге қауіп факторларын жеткізуге көмектеседі [50].

Талқылау

Бұл зерттеуде бірыңғай протокол жағдайында бірнеше алгоритм салыстырылды. Бустингтік әдістер базалық модельдерден тұрақты түрде жоғары нәтиже берді. Ең жақсы теңгерім LightGBM көрсетті (accuracy = 0.8696, recall = 0.9020). ROC–AUC = 0.93, ал калибровка қисығы диагональға сәйкес болды. Қателік матрицасында 92 TP, 68 TN, 14 FP және 10 FN тіркелді, бұл модельдің сезімталдық пен ерекшелікті тиімді үйлестіретінін көрсетті.

LightGBM бейсызық тәуелділіктерді жақсы ұстап, Random Forest және XGBoost-тен артық нәтижелер берді. Практикада жалған терістердің аздығы науқастарды өткізіп алмауға, ал жалған оңдардың орташа деңгейі қажетсіз зерттеулерді шектеуге мүмкіндік береді. Модельдің шектері клиникалық мақсатқа қарай реттелуі мүмкін: скрининг үшін сезімталдық, растау үшін ерекшелік.

SHAP талдауы басты факторларды анықтады: oldpeak (ST депрессиясы), exang (жүктемелік стенокардия) және ca (ірі тамырлар саны). Олардың жоғары мәндері КЖА тәуекелін арттырды, бұл кардиологиядағы дәстүрлі түсініктермен сәйкес келеді. Осылайша, LightGBM тек статистикалық дәл ғана емес, клиникалық тұрғыдан да интерпретацияланатын құрал ретінде ерекшеленді.

Қорытынды

Бұл зерттеуде UCI Heart Disease деректер жиынының кеңейтілген нұсқасы негізінде КЖА-ны болжауға арналған машиналық оқытуға бірыңғай әрі ашық схема ұсынылды. Алдын ала өңдеу деректерді үйлестіріп, алгоритмдерді әділ салыстыруға жағдай жасады.

Сынаққа бустинг, ансамбльдік, қашықтыққа негізделген және қарапайым интерпретацияланатын модельдер енгізілді. Нәтижесінде LightGBM ең жоғары нәтижелер көрсетті (accuracy = 0.8696, recall = 0.9020, ROC–AUC = 0.93). Қателік матрицасы модельдің сезімталдық пен ерекшелікті тиімді үйлестіретінін дәлелдеді. Жалған терістердің аздығы КЖА бар науқастарды өткізіп алмауға мүмкіндік берді.

SHAP талдауы басты факторларды (артериялық қысым, холестерин, жас, жыныс, oldpeak, exang) бөліп көрсетті. Бұл медициналық біліммен сәйкес келіп, модельдің клиникалық негізділігін растады.

Шектеулерге ашық деректердің жалпылау қабілеті мен сыртқы валидация қажеттілігі жатады. Болашақ бағыттарға қосымша деректерді (кескіндер, биомаркерлер, генетика) интеграциялау және decision-curve analysis қолдану кіреді.

Жалпы, ұсынылған конвейер КЖА тәуекелін клиникалық тұрғыдан мәнді әрі интерпретацияланатын түрде болжай алатынын көрсетті, бұл оны шешім қабылдауды қолдауда құнды құрал етеді.

Список литературы

1. Reducing the Global Burden of Cardiovascular Disease, Part 1: The Epidemiology and Risk Factors / P. Joseph et al // *Circulation research*. – 2017. – № 121(6). – P. 677-694. <https://doi.org/10.1161/CIRCRESAHA.117.308903>.
2. Clinical, Economical, and Organizational Impact of Chronic Ischemic Cardiovascular Disease in Italy: Evaluation of 2019 Nationwide Hospital Admissions Data / F. Pietrantonio et al // *Int. J. Environ. Res. Public Health*. – 2025. – № 22. – P. 530. <https://doi.org/10.3390/ijerph22040530>.
3. Prevalence and trends of coronary artery disease risk factors and their effect on age of diagnosis in patients with established coronary artery disease: Tehran Heart Center (2005–2015) / K. Hosseini et al // *BMC Cardiovasc Disord* 21. – 2021. – № 477(2021). <https://doi.org/10.1186/s12872-021-02293-y>.
4. Economic Analysis of the CADScor System for Ruling Out Coronary Artery Disease in England / M. Javanbakht et al // *PharmacoEconomics*. – 2022. – Open 6. – P. 123-135. <https://doi.org/10.1007/s41669-021-00297-0>.
5. Roth G. The Global Burden of Cardiovascular Diseases and Risks: A Compass for Global Action / G. Roth, G. Mensah, V. Fuster // *JACC*. – 2020. – № 76(25). – P. 2980-2981. <https://doi.org/10.1016/j.jacc.2020.11.021>.
6. Estimation of the Burden of Ischemic Heart Disease in the Tabasco Population, Mexico, 2013–2021 / J.J. Cárdenas-Anguiano et al // *Int. J. Environ. Res. Public Health*. – 2025. – № 22. – P. 423. <https://doi.org/10.3390/ijerph22030423>.
7. Fadah K. Epidemiology, Pathophysiology, and Management of Coronary Artery Disease in the Elderly / K. Fadah, A. Hechanova, D. Mukherjee // *The International journal of angiology : official publication of the International College of Angiology*. – 2022. – № 31(4). – P. 244-250. <https://doi.org/10.1055/s-0042-1751234>.
8. National prevalence of coronary heart disease and its relationship with human development index: A systematic review / Ke-Fu Zhu et al // *European Journal of Preventive Cardiology*. – 2016. – Vol. 23, Issue 5. – P. 530-543. <https://doi.org/10.1177/2047487315587402>.
9. Prevalence of coronary artery disease and its risk factors in Kerala, South India: a community-based cross-sectional study / M.N. Krishnan et al // *BMC Cardiovasc Disord*. – 2016. – № 16. – P. 12. <https://doi.org/10.1186/s12872-016-0189-3>.
10. Risk Factor Burden and Long-Term Prognosis of Patients With Premature Coronary Artery Disease / M. Zeitouni et al // *Journal of the American Heart Association*. – 2020. – № 9(24). – P. e017712. <https://doi.org/10.1161/JAHA.120.017712>.
11. Stratification of Patients with Coronary Artery Disease by Circulating Cytokines Profile: A Pilot Study / C. Iside et al // *M. J. Clin. Med*. – 2023. – № 12. – P. 6649. <https://doi.org/10.3390/jcm12206649>.
12. Myocardial Ischemia: Differentiating between Epicardial Coronary Artery Atherosclerosis, Microvascular Dysfunction and Vasospasm in the Catheterization Laboratory / G. Monizzi et al // *J. Clin. Med*. – 2024. – № 13. – P. 4172. <https://doi.org/10.3390/jcm13144172>.
13. Genetics of coronary artery disease and myocardial infarction / X. Dai et al // *World journal of cardiology*. – 2016. – № 8(1). – P. 1-23. <https://doi.org/10.4330/wjc.v8.i1.1>.
14. Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries / R. Alizadehsani et al // *Computer methods and programs in biomedicine*. – 2018. – № 162. – P. 119-127. <https://doi.org/10.1016/j.cmpb.2018.05.009>.
15. Symptom Presentation among Women with Suspected Ischemia and No Obstructive Coronary Artery Disease (INOCA) / Y.K. Taha et al // *J. Clin. Med*. – 2023. – № 12. – P. 5836. <https://doi.org/10.3390/jcm12185836>.

16. Comparison of Outcomes in Patients With Acute Coronary Syndrome Presenting With Typical Versus Atypical Symptoms / Y. Hammer et al // *The American journal of cardiology*. – 2019. – № 124(12). – P. 1851-1856. <https://doi.org/10.1016/j.amjcard.2019.09.007>.
17. Silent Myocardial Ischemia: From Pathophysiology to Diagnosis and Treatment / P. Theofilis et al // *Biomedicines*. – 2024. – № 12. – P. 259. <https://doi.org/10.3390/biomedicines12020259>.
18. Nelson AJ. Current approach to the diagnosis of atherosclerotic coronary artery disease: more questions than answers / AJ Nelson, M Ardissino, PJ. Psaltis // *Therapeutic Advances in Chronic Disease*. – 2019. – №10. <https://doi.org/10.1177/2040622319884819>.
19. Diagnostic accuracy of non-invasive cardiac imaging modalities in patients with a history of coronary artery disease: a meta-analysis / RA Jukema et al // *Heart*. – 2025. – № 111. – P. 4-10. <https://doi.org/10.1136/heartjnl-2024-324248>.
20. Speckle Tracking Echocardiography: Early Predictor of Diagnosis and Prognosis in Coronary Artery Disease / M.C. Pastore et al // *BioMed research international*. – 2021. <https://doi.org/10.1155/2021/6685378>.
21. Challenges and Burdens in the Coronary Artery Disease Care Pathway for Patients Undergoing Percutaneous Coronary Intervention: A Contemporary Narrative Review / M. Kodeboina et al // *Int. J. Environ. Res.* – 2023. – № 20. – P. 5633. <https://doi.org/10.3390/ijerph20095633>.
22. Wang L.W. Non-invasive screening for coronary artery disease: current perspectives, patient, public health and ethical considerations in evaluating symptomatic and asymptomatic individuals / L.W. Wang // *Internal medicine journal*. – 2025. – № 55(4). – P. 555-563. <https://doi.org/10.1111/imj.16585>.
23. Advancements in Biomarkers for Early Detection and Risk Stratification of Cardiovascular Diseases-A Literature Review / A. Nazir et al // *Health science reports*. – 2025. – № 8(5). – P. e70878. <https://doi.org/10.1002/hsr2.70878>.
24. Zachariadis C.B. Harnessing Artificial Intelligence for Automated Diagnosis / C.B. Zachariadis, H.C. Leligou // *Information*. – 2024. – № 15. – P. 311. <https://doi.org/10.3390/info15060311>.
25. A new machine learning technique for an accurate diagnosis of coronary artery disease / M. Abdar et al // *Computer methods and programs in biomedicine*. – 2019. – № 179. – P. 104992. <https://doi.org/10.1016/j.cmpb.2019.104992>.
26. Multi-constraints based deep learning model for automated segmentation and diagnosis of coronary artery disease in X-ray angiographic images / M. Algarni et al // *PeerJ Computer Science*. – 2022. – № 8. – P. e993. <https://doi.org/10.7717/peerj-cs.993>.
27. Machine Learning Quantitation of Cardiovascular and Cerebrovascular Disease: A Systematic Review of Clinical Applications / C. Boyd et al // *Diagnostics*. – 2021. – № 11. – P. 551. <https://doi.org/10.3390/diagnostics11030551>.
28. Akella A. Machine Learning Algorithms for Predicting Coronary Artery Disease: Efforts Toward an Open Source Solution / A. Akella, S. Akella // *Future Science OA*. – 2021. – № 7(6). <https://doi.org/10.2144/foa-2020-0206>.
29. A 3D deep learning classifier and its explainability when assessing coronary artery disease / Cheung, Wing Keung et al // – 2023. 10.48550/arXiv.2308.00009.
30. Integration of artificial intelligence into clinical patient management: focus on cardiac imaging / F. Loncaric et al // *Revista Española de Cardiología (English Edition)*. – 2021. – Vol. 74, Issue 1. – P. 72-80. <https://doi.org/10.1016/j.rec.2020.07.003>.
31. Non-invasive Assessment of Coronary Artery Disease: The Role of AI in the Current Status and Future Directions / F. Ezekwueme et al // *Cureus*. – 2025. – № 17(2). – P. e78994. <https://doi.org/10.7759/cureus.78994>.
32. Using artificial intelligence in the development of diagnostic models of coronary artery disease with imaging markers: A scoping review / X. Wang et al // *Front. Cardiovasc.* – 2022. – Med. 9:945451. <https://doi.org/10.3389/fcvm.2022.945451>.
33. Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities / Anichur Rahman et al // [J]. *AIMS Public Health*. – 2024. – № 11(1). – P. 58-109. <https://doi.org/10.3934/publichealth.2024004>.
34. Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases / A. Ogunpola et al // *Diagnostics*. – 2024. – № 14. – P. 144. <https://doi.org/10.3390/diagnostics14020144>.

35. Implementing Machine Learning in Interventional Cardiology: The Benefits Are Worth the Trouble / Ben Ali W. Et al // Front. Cardiovasc. Med. – 2021. – № 8. – P. 711401. <https://doi.org/10.3389/fcvm.2021.711401>.
36. Wong K. Deep learning-based cardiovascular image diagnosis: A promising challenge / K. Wong, G. Fortino, D. Abbott // Future Generation Computer Systems. – 2019. – № 110. – P. 10. 1016/j.future.2019.09.047.
37. Advances in Diagnosis, Therapy, and Prognosis of Coronary Artery Disease Powered by Deep Learning Algorithms / M. Chu et al // JACC: Asia. – 2023. – № 3(1). – P. 1-14. <https://doi.org/10.1016/j.jacasi.2022.12.005>.
38. The Role of Artificial Intelligence in Improving Patient Outcomes and Future of Healthcare Delivery in Cardiology: A Narrative Review of the Literature / D. Gala et al // Healthcare. – 2024. – № 12. – P. 481. <https://doi.org/10.3390/healthcare12040481>.
39. Classification models for assessing coronary artery disease instances using clinical and biometric data: an explainable man-in-the-loop approach / AD. Samaras et al // Sci Rep. – 2023. – № 13. – P. 6668. <https://doi.org/10.1038/s41598-023-33500-9>.
40. Heart Disease [Dataset] / A. Janosi et al // UCI Machine Learning Repository. – 1989. <https://doi.org/10.24432/C52P4X>.
41. Hancock J.T. Survey on categorical data for neural networks / J.T. Hancock, T.M. Khoshgoftaar // J Big Data 7, 28 (2020). <https://doi.org/10.1186/s40537-020-00305-w>.
42. The choice of scaling technique matters for classification performance / B.V. Lucas de Amorim et al // Applied Soft Computing. – 2023. – Vol. 133. – P. 109924, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2022.109924>.
43. Conditional mutual information-based feature selection algorithm for maximal relevance minimal redundancy / X. Gu et al // Appl Intell. – 2022. – № 52. – P. 1436-1447. <https://doi.org/10.1007/s10489-021-02412-4>.
44. Evaluation of clinical prediction models (part 1): from development to external validation / GS. Collins et al // BMJ. – 2024. – № 8. – P. 384. e074819. <https://doi.org/10.1136/bmj-2023-074819>. PMID: 38191193; PMCID: PMC10772854.
45. Demircioğlu A. Applying oversampling before cross-validation will lead to high bias in radiomics / A. Demircioğlu // Sci Rep. – 2024. – № 14. – P. 11563. <https://doi.org/10.1038/s41598-024-62585-z>.
46. On evaluation metrics for medical applications of artificial intelligence / S.A. Hicks et al // Sci Rep. – 2022. – № 12. – P. 5979. <https://doi.org/10.1038/s41598-022-09954-8>.
47. Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework / J. Wang et al // Front. Cardiovasc. Med. – 2024. – № 11. – P. 1360548. <https://doi.org/10.3389/fcvm.2024.1360548>.
48. Evaluation of clinical prediction models (part 1): from development to external validation / G.S. Collins et al // BMJ. – 2024. – № 384. – P. e074819. <https://doi.org/10.1136/bmj-2023-074819>.
49. Predicting COVID-19 mortality with electronic medical records / H. Estiri et al // npj Digit. Med. – 2021. – № 4. – P. 15. <https://doi.org/10.1038/s41746-021-00383-x>.

АЛҒЫС

Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің АР23488586 гранты есебінен қаржыландырылды.

Ж.Е. Байғараева^{1,2}, А.К. Болтабоева^{1,2*}, Б.Т. Иманбек¹, М.И. Кожамбердиева^{1,2}, Ж.Қ. Жодыбаева³

¹Қазақский национальный университет имени аль-Фараби,
050040, Республика Казахстан, г. Алматы, Кпроспект аль-Фараби, 71,

²ТОО «Kazakhstan R&D Solutions»,

050056, Республика Казахстан, г. Алматы, улица Кожедуба 3,

³Қазақский национальный педагогический университет имени Абая,

050010, Республика Казахстан, г. Алматы, проспект Достық 13

*e-mail: boltaboyeva_assiya3@live.kaznu.kz

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ КОРОНАРНОЙ БОЛЕЗНИ СЕРДЦА: НА ОСНОВЕ НАБОРА ДАННЫХ UCI HEART DISEASE

Коронарная болезнь сердца остаётся одной из основных причин смертности и инвалидизации в мире. Своевременная диагностика позволяет снизить частоту осложнений и уменьшить нагрузку на систему здравоохранения. Однако традиционные методы остаются

дорогими, инвазивными и ограниченно доступными. Недавние исследования подтверждают перспективность применения методов машинного обучения в клинической практике. В этой связи возникает вопрос: можно ли достоверно прогнозировать наличие заболевания, используя только клинико-демографические данные, без визуализационных методов?

Целью данной работы было оценить точность и практическую ценность таких моделей. На объединённом протоколе данных UCI Heart Disease ($n = 920$) был обучен алгоритм LightGBM, который показал следующие результаты: $accuracy = 0.8696$, $precision = 0.8679$, $recall = 0.9020$, $F1-score = 0.8846$. Полученные данные дополняют результаты предыдущих исследований, основанных на визуализации.

Исследование включало сравнение нескольких алгоритмов при единых условиях предварительной обработки и валидации, проверку калибровки вероятностей и интерпретацию с помощью метода SHAP. Анализ показал, что ключевые предикторы (например, депрессия ST-сегмента) согласуются с клиническими знаниями. Это подтверждает возможность применения модели для первичного скрининга и направления на дополнительную диагностику. В целом, калиброванные и интерпретируемые алгоритмы на основе открытых клинических данных могут стать ценным инструментом маршрутизации пациентов в условиях ограниченных ресурсов.

Ключевые слова: коронарная болезнь сердца (КБС); машинное обучение; LightGBM; классификация; клинико-демографические данные; стратификация риска; интерпретация SHAP; ROC–AUC; калибровка.

Zh.Y. Baigarayeva^{1,2}, A.K. Boltaboyeva^{1,2*}, B.T. Imanbek¹, M.I. Kozhamberdiyeva^{1,2}, Zh.K. Zholdybayeva³

¹Al-Farabi Kazakh National University,
050040, Kazakhstan, Almaty, Al-Farabi Avenue 71,

²LLP «Kazakhstan R&D Solutions»,
050056, Kazakhstan, Almaty, 3 Kozheduba str.,

³Abai Kazakh National Pedagogical University,
050010, Kazakhstan, Almaty, Dostyk Avenue 13

*e-mail: boltaboyeva_assiya3@live.kaznu.kz

COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS FOR PREDICTING CORONARY HEART DISEASE: EVIDENCE FROM THE UCI HEART DISEASE DATASET

Coronary artery disease remains one of the leading causes of death and disability worldwide. Timely diagnosis can reduce the incidence of complications and ease the burden on healthcare systems. However, traditional methods are often costly, invasive, and limited in accessibility. Recent studies confirm the potential of machine learning for clinical applications. This raises the question: is it possible to reliably predict the presence of disease using only clinical and demographic data, without imaging methods?

The aim of this study was to evaluate the accuracy and practical value of such models. Using the UCI Heart Disease dataset ($n = 920$) under a unified protocol, the LightGBM algorithm was trained and achieved the following results: $accuracy = 0.8696$, $precision = 0.8679$, $recall = 0.9020$, $F1-score = 0.8846$. These findings complement previous research based on imaging approaches.

The study compared multiple algorithms under identical preprocessing and validation conditions, assessed probability calibration, and applied SHAP for interpretability. The analysis revealed that the main predictors (e.g., ST-segment depression) aligned with established clinical knowledge. This confirms that the model can be used for initial screening and referral to additional diagnostics. Overall, calibrated and interpretable algorithms based on open clinical data can serve as a valuable tool for patient routing in resource-limited settings.

Key words: coronary heart disease (CHD); machine learning; LightGBM; classification; clinical and demographic data; risk stratification; SHAP interpretation; ROC–AUC; calibration.

Авторлар туралы ақпарат

Жанель Ермашқызы Байғараева – магистр, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы; e-mail: zhanel.baigarayeva@gmail.com. ORCID: <https://orcid.org/0000-0003-1919-3570>.

Асия Кубланди кизи Болтабоева* – магистр, PhD 3 курс студенті, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы; e-mail: boltaboyeva_assiya3@live.kaznu.kz. ORCID: <https://orcid.org/0000-0002-7279-9910>.

Бағлан Талғатқызы Иманбек – PhD, доцент, профессор-зерттеуші, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы; e-mail: baglan.imanbek@kaznu.edu.kz. ORCID: <https://orcid.org/0000-0001-7249-380X>.

Мергул Иманбековна Кожамбердиева – педагогика ғылымдарының кандидаты, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы; e-mail: kozhamberdiyeva.m@outlook.com.

Жанел Қайратқызы Жолдыбаева – бакалавр 4 курс студенті, Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы; e-mail: zoldybaevazanel@gmail.com.

Информация об авторах

Жанель Ермашқызы Байғараева – магистр, Казахский национальный университет имени аль-Фараби, Алматы; e-mail: zhanel.baigarayeva@gmail.com. ORCID: <https://orcid.org/0000-0003-1919-3570>.

Асия Кубланди кызи Болтабоева* – магистр, PhD 3 курс студенті, Казахский национальный университет имени аль-Фараби, Алматы; e-mail: boltaboyeva_assiya3@live.kaznu.kz. ORCID: <https://orcid.org/0000-0002-7279-9910>.

Бағлан Талғатқызы Иманбек – PhD, доцент, профессор-зерттеуші, Казахский национальный университет имени аль-Фараби, Алматы; e-mail: baglan.imanbek@kaznu.edu.kz. ORCID: <https://orcid.org/0000-0001-7249-380X>.

Мергул Иманбековна Кожамбердиева – педагогика ғылымдарының кандидаты, Казахский национальный университет имени аль-Фараби, Алматы; e-mail: kozhamberdiyeva.m@outlook.com.

Жанел Қайратқызы Жолдыбаева – бакалавр 4 курс студенті, Казахский национальный педагогический университет имени Абая; e-mail: zoldybaevazanel@gmail.com.

Information about the authors

Zhanel Yermashkyzy Baigarayeva – Master's degree holder, Al Farabi Kazakh National University, Almaty; e-mail: zhanel.baigarayeva@gmail.com. ORCID: <https://orcid.org/0000-0003-1919-3570>.

Assiya Kublandi kyzi Boltaboyeva* – Master's degree holder, 3rd-year PhD student, Al Farabi Kazakh National University, Almaty; e-mail: boltaboyeva_assiya3@live.kaznu.kz. ORCID: <https://orcid.org/0000-0002-7279-9910>.

Baglan Talgatkyzy Imanbek – PhD, Associate Professor, Research Professor, Al Farabi Kazakh National University, Almaty; e-mail: baglan.imanbek@kaznu.edu.kz. ORCID: <https://orcid.org/0000-0001-7249-380X>.

Mergul Imanbekovna Kozhamberdiyeva – Candidate of Pedagogical Sciences, Al Farabi Kazakh National University, Almaty; e-mail: kozhamberdiyeva.m@outlook.com.

Zhanel Kairatkyzy Zholdybayeva – 4th year Bachelor's student, Abai Kazakh National Pedagogical University, Almaty; e-mail: zoldybaevazanel@gmail.com.

Редакцияға енуі 18.09.2025

Өңдеуден кейін түсуі 18.10.2025

Жариялауға қабылданды 20.10.2025

[https://doi.org/10.53360/2788-7995-2025-4\(20\)-7](https://doi.org/10.53360/2788-7995-2025-4(20)-7)



MPHTI: 14.01.35, 10.01.01

А.К. Шайханова*, А.К. Токкулиева¹, И.М. Увалиева², К.Е. Мырзаханов³, Б.Т. Смаилова⁴

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
10000, Қазақстан, Астана қ., Сәтбаев к-сі, 2

²Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті,
Қазақстан Республикасы, Өскемен қ., Серікбаев көшесі, 19

³«Ұлттық аударма бюросы» АҚ,
Қазақстан Республикасы, Астана қ.

⁴Шәкәрім Университеті,
Қазақстан Республикасы, Семей қ., Глинки к-сі, 20 А

*e-mail: aigul.shaikhanova@gmail.com

ПЕДАГОГТЕРДЕН БАСТАПҚЫ ДЕРЕКТЕРДІ ЖИНАУ НЕГІЗІНДЕ ЦИФРЛАНДЫРУДЫҢ БІЛІМ БЕРУ ТӘЖІРИБЕСІНЕ ӘСЕРІН БАҒАЛАУ

Аңдатпа: Мақала цифрландырудың білім беру тәжірибесі мен мұғалімдердің кәсіби қызметіне әсерін бағалауға арналған. Зерттеудің мақсаты сандық құзыреттіліктердің қазіргі деңгейін, цифрлық құралдар мен жасанды интеллект (ЖИ) технологияларын пайдалану сипатын, сондай-ақ олардың білім беру үдерісі мен студенттердің нәтижелеріне әсерін анықтау үшін педагогикалық қоғамдастықтан жиналған бастапқы деректерді талдау болды. Зерттеу әдістемесі әр түрлі білім беру ұйымдарының 832 оқытушы-профессорлар құрамының онлайн сауалнамасы арқылы жүргізілген сандық және сапалық сауалнамасын қамтиды. Деректерді жинау респонденттердің демографиялық сипаттамаларын, цифрлық құралдарды (MS Office, Kahoot) және ЖИ (ChatGPT, Gemini) пайдалану жиілігін, сабақта цифрлық технологияларды қолдануды (интерактивті тестілеу, AR), олардың