

## Медициналық деректер негізінде жүрек-қантамыр аурулары қаупін болжауда машиналық оқыту әдістерін қолдану

ЖАРИЯЛАНДЫ  
29.05.2026

СІЛТЕМЕ  
<https://bilimger.kz/189016/>

УДК 004.85:616.1

### **Мейрамбеков Мухамедали Мейрамбекулы**

магистрант, «Автоматтандыру және басқару» кафедрасы

Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті

(Алматы қ., Қазақстан)

Ғылыми жетекші: **Жусупбеков Сарсенбек Сейтбекұлы**, профессор, «Автоматтандыру және басқару» кафедрасы

Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті

(Алматы қ., Қазақстан)

Аңдатпа: Жұмыста медициналық деректер негізінде жүрек-қантамыр аурулары қаупін болжауда машиналық оқыту әдістерін қолдану мүмкіндіктері қарастырылады. Мақсаты – пациенттердің клиникалық, зертханалық және мінез-құлықтық көрсеткіштерін кешенді талдай отырып, тәуекел деңгейін автоматтандырылған түрде бағалауға арналған модельдің әдістемелік және техникалық негізін айқындау. Жұмыста дәстүрлі тәуекелді бағалау тәсілдерінің шектеулері талданып, медициналық деректердің құрылымы мен оларды алдын ала өңдеу кезеңдері сипатталады; логистикалық регрессия, Random forest, Support vector machine және XGBoost алгоритмдерінің мүмкіндіктері салыстырмалы тұрғыда қарастырылады. Бағдарламалық іске асыру үшін Python тілі мен оның арнайы кітапханаларын (Pandas, NumPy, Scikit-learn, XGBoost) пайдаланудың орындылығы негізделеді. Күтілетін нәтиже – пациентті қауіптің жоғары немесе төмен/орташа тобына жатқызатын бинарлық классификация моделі, оның сапасы Accuracy, Precision, Recall, F1-score және ROC-AUC метрикалары бойынша бағаланады.

Зерттеу медициналық шешім қабылдауды қолдау жүйелерін әзірлеуде қолданбалы маңызға ие.

Кілт сөздер: жүрек-қантамыр аурулары, машиналық оқыту, тәуекелді стратификациялау, жекелендірілген медицина, логистикалық регрессия, Random forest, Support vector machine, XGBoost, медициналық деректер, болжамды модель.

### Кіріспе

Дүниежүзілік денсаулық сақтау ұйымының (ДДҰ) деректері бойынша жүрек-қантамыр аурулары (ЖҚА) әлемдегі өлім-жітімнің негізгі себептерінің бірі болып табылады және жыл сайын шамамен 17,9 миллион адамның өмірін қияды, бұл жалпы өлім-жітімнің 31 %-ын құрайды [1]. Қазақстан Республикасында да аталған патологиялар аурушаңдық пен өлім-жітім құрылымында жетекші орын алады және отандық денсаулық сақтау жүйесіне айтарлықтай экономикалық жүктеме түсіреді. Осы жағдайда тәуекел топтарын ерте кезеңде анықтайтын, дәрігерге ақпараттық қолдау көрсететін интеллектуалды шешімдерді әзірлеу маңызды ғылыми-практикалық міндетке айналады.

Жүрек-қантамыр аурулары қаупін бағалаудың классикалық тәсілдері (Framingham, SCORE, ASCVD) пациенттердің жалпы топтарына негізделген шектеулі параметрлерді қолданады және популяциялық деңгейде қалыптастырылған [2, 5]. Осы себепті олар жекелендірілген болжамды жеткілікті дәлдікпен ұсына алмайды. Сонымен қатар заманауи медициналық деректердің көлемі мен әртүрлілігі тез өсіп келеді: зертханалық көрсеткіштер, физиологиялық параметрлер, өмір салтына қатысты мәліметтер және анамнез деректері күрделі көпфакторлы құрылым құрайды. Мұндай деректерді талдау үшін машиналық оқыту әдістері тиімді құрал болып табылады, өйткені олар үлкен көлемдегі ақпарат арасынан жасырын заңдылықтарды дербес анықтап, болжамдық дәлдікті едәуір арттыруға мүмкіндік береді [3, 4].

Зерттеудің мақсаты – медициналық деректер негізінде жүрек-қантамыр аурулары қаупін болжауда машиналық оқыту әдістерін қолдану мүмкіндіктерін зерттеу және жекелендірілген тәуекелді стратификациялауға арналған модельді әзірлеудің әдістемелік және техникалық негізін айқындау.

Қойылған мақсатқа жету үшін келесі міндеттер шешіледі:

- жүрек-қантамыр аурулары қаупін бағалаудың дәстүрлі әдістерінің ерекшеліктері мен шектеулерін талдау;
- тәуекелді болжау үшін қажетті медициналық деректердің құрылымы мен сипаттамасын анықтау;
- деректерді алдын ала өңдеу кезеңдерінің реттілігін негіздеу;
- логистикалық регрессия, Random forest, Support vector machine және XGBoost

алгоритмдерінің қолданылу мүмкіндіктерін салыстырмалы талдау;

– модельді бағдарламалық іске асыруға арналған құралдар мен сапаны бағалау метрикаларын анықтау.

Зерттеудің ғылыми жаңалығы пациенттің демографиялық, клиникалық, зертханалық және мінез-құлықтық көрсеткіштерін кешенді талдау арқылы жүрек-қантaмыр аурулары қаупін жекелендірілген түрде бағалайтын машиналық оқыту моделінің сәулетін ұсынуда және оны медициналық шешім қабылдауды қолдау жүйесінің құрамдас бөлігі ретінде позициялауда.

Материалдар мен әдістер

Дәстүрлі әдістердің шектеулері және машиналық оқытудың рөлі

Жүрек-қантaмыр ауруларының алдын алу үшін олардың тәуекелін ерте кезеңде бағалау маңызды міндет болып табылады. Framingham, SCORE және ASCVD сияқты дәстүрлі шкалалар бірнеше шектеулі көрсеткішке (жас, жыныс, қан қысымы, холестерин деңгейі, темекі шегу) сүйенеді [2, 5]. Олардың негізгі кемшілігі – популяциялық сипатта болуы, яғни нақты пациенттің ерекшеліктерін толық ескере алмауы. Сондай-ақ дәстүрлі модельдер бұрын қалыптасқан статистикалық тәуелділіктерге негізделген, бұл деректер құрылымы өзгерген жағдайда олардың сезімталдығын төмендетеді.

Машиналық оқыту әдістері осы кемшіліктерді жеңуге мүмкіндік береді. Олар үлкен көлемдегі деректерден заңдылықтарды дербес анықтап, көрсеткіштер арасындағы сызықтық емес тәуелділіктерді ескереді және әрбір пациент үшін жекелендірілген болжам жасайды [3, 4, 6]. Аталған әдістердің медицинада қолданылуы – жекелендірілген медицина (personalized medicine) тұжырымдамасының негізгі құралдарының бірі.

Медициналық деректердің құрылымы мен сипаттамасы

Жүрек-қантaмыр аурулары қаупін болжау міндетінде қолданылатын медициналық деректер бірнеше топқа бөлінеді: демографиялық (пациенттің жасы, жынысы), клиникалық (артериялық қан қысымы, жүрек соғу жиілігі, дене салмағы индексі, қосалқы аурулардың болуы), зертханалық (жалпы холестерин, төмен (LDL) және жоғары (HDL) тығыздықтағы липопротеидтер, триглицеридтер, қандағы глюкоза мөлшері) және мінез-құлықтық (темекі шегу, алкоголь тұтыну, физикалық белсенділік деңгейі, тамақтану режимі).

Деректер кестелік форматта ұсынылады: әрбір жол жеке пациентке, ал әрбір баған белгілі бір медициналық сипаттамаға сәйкес келеді. Нысаналы айнымалы (target) ретінде пациенттің қауіп тобына жату фактісі бинарлық белгі түрінде алынады: «жоғары қауіп тобы» немесе «төмен/орташа қауіп тобы». Дене салмағы индексі (BMI) дене

салмағы мен бой ұзындығының арақатынасы бойынша есептеледі:  $BMI = m / h^2$ , мұнда  $m$  – дене салмағы (кг),  $h$  – бой (м). BMI мәнінің 25-тен жоғары болуы жүрек-қантамыр аурулары қаупінің артуымен байланысты.

Медициналық деректердің негізгі ерекшеліктері – көпөлшемділік, гетерогенділік, бос мәндердің болуы, өлшеу қателіктері және әртүрлі форматтағы (сандық, категориялық) белгілердің кездесуі. Аталған ерекшеліктер деректерді машиналық оқыту алгоритмдеріне тікелей беруге мүмкіндік бермейді және алдын ала өңдеу кезеңін міндетті етеді.

Деректерді алдын ала өңдеу кезеңдері

Машиналық оқыту моделінің сапасы көп жағдайда деректердің сапасына тікелей тәуелді. Деректерді алдын ала өңдеу мынадай негізгі кезеңдерден тұрады:

– деректерді тазалау – қайталанатын жазбаларды, қате енгізілген мәндерді және шектен тыс ауытқуларды (outliers) анықтау мен жою;

– бос мәндерді өңдеу – толық емес жазбаларды орташа мәнмен, медианамен, ең жиі кездесетін мәнмен немесе KNN-imputer сияқты арнайы алгоритмдер арқылы толықтыру;

– категориялық белгілерді кодтау – «жыныс», «темекі шегу» сияқты белгілерді сандық форматқа түрлендіру (one-hot encoding немесе label encoding);

– деректерді масштабтау – әртүрлі ауқымдағы белгілерді стандарттау (Z-score) немесе Min-Max қалыпқа келтіру арқылы салыстырмалы форматқа келтіру:  $z = (x - \mu) / \sigma$ , мұнда  $\mu$  – орташа мән,  $\sigma$  – стандарттық ауытқу;

– маңызды белгілерді таңдау – корреляциялық талдау, Recursive Feature Elimination (RFE) немесе ансамбльдік әдістерге негізделген маңыздылықты бағалау арқылы ең ақпараттық параметрлерді іріктеу;

– деректерді бөлу – жиынтықты оқыту (training), валидациялық (validation) және тестілік (test) бөліктерге 70/15/15 немесе 80/20 пропорциясымен бөлу.

Сонымен қатар медициналық деректерде жиі кездесетін кластардың дисбалансын (қауіп тобындағы пациенттер санының салыстырмалы аздығын) жою үшін SMOTE сияқты арнайы әдістер қолданылуы мүмкін [7].

Қарастырылатын машиналық оқыту алгоритмдері

Жүрек-қантамыр аурулары қаупін болжау міндеті машиналық оқыту тұрғысынан бинарлық классификация есебі ретінде қойылады және бақыланатын оқыту (supervised learning) парадигмасында шешіледі. Зерттеу шеңберінде төрт алгоритм қарастырылады.

Логистикалық регрессия – бинарлық классификация үшін қолданылатын классикалық статистикалық әдіс. Алгоритм пациенттің берілген класқа жату ықтималдығын

логистикалық (сигмоидтық) функция арқылы есептейді:  $P(y = 1) = 1 / (1 + \exp(-(b_0 + b_1x_1 + \dots + b_nx_n)))$ ). Артықшылықтары – қарапайымдылығы, есептеу жылдамдығы, түсіндірмелілігі және әрбір белгінің модельге қосатын үлесін бағалау мүмкіндігі. Бұл әдіс көбіне базалық (baseline) модель ретінде қолданылады [5].

Random forest – көптеген шешім ағаштарының болжамдарын біріктіруге негізделген ансамбльдік әдіс. Алгоритм әр ағашты деректердің кездейсоқ ішкі жиынында (bootstrap sample) және белгілердің кездейсоқ ішкі жиынында оқытады, бұл үлгінің тұрақтылығын арттырып, шамадан тыс бейімделу (overfitting) тәуекелін айтарлықтай азайтады. Random forest белгілердің маңыздылығын (feature importance) сандық түрде бағалауға мүмкіндік береді – бұл қасиет медициналық интерпретация үшін аса құнды [6, 7].

Support vector machine (SVM) – кластар арасындағы максималды шекараны (margin) құрайтын оңтайлы гипержазықтықты табуға бағытталған әдіс. Сызықтық емес ядро функцияларын (RBF, полиномдық, sigmoid) қолдану арқылы күрделі құрылымды деректерде де жоғары дәлдікке қол жеткізуге болады. SVM әсіресе шағын және орта көлемді деректер жиынтығында тиімді жұмыс істейді, дегенмен оның есептеу күрделілігі деректер көлемінің артуымен өседі.

XGBoost (Extreme Gradient Boosting) – градиенттік бустинг қағидатына негізделген заманауи алгоритм. Ол шешім ағаштарын кезең-кезеңімен құрастырады, мұнда әрбір жаңа ағаш алдыңғыларының қателіктерін түзетуге бағытталған. XGBoost үлкен көлемдегі деректермен жұмыс істеуде, белгілер арасындағы сызықтық емес тәуелділіктерді анықтауда және жоғары болжамдық дәлдікке жетуде ең тиімді әдістердің бірі болып саналады [4, 8]. Қосымша артықшылықтарына регуляризацияның кірістірілген механизмдері, бос мәндермен жұмыс істеу қабілеті және параллельді есептеулерді қолдау жатады.

Аталған төрт алгоритмнің салыстырмалы талдауы зерттеу шеңберінде әр модельдің медициналық деректерге сай келу дәрежесін объективті бағалауға және практикалық қолдану үшін ең тиімді шешімді негіздеп таңдауға мүмкіндік береді.

Модельдің сапасын бағалау метрикалары мен бағдарламалық іске асыру құралдары

Бинарлық классификация моделінің сапасын объективті бағалау үшін стандартты метрикалар жиыны қолданылады: Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$ , Precision =  $TP / (TP + FP)$ , Recall =  $TP / (TP + FN)$ , F1-score (Precision мен Recall арасындағы гармоникалық орта) және ROC-AUC (қисық астындағы аудан). Медициналық міндеттерде Recall ерекше маңызға ие, өйткені ол жоғары қауіп тобындағы пациенттерді жіберіп алу тәуекелін сипаттайды.

Бағдарламалық іске асыру Python тілінде, Jupyter Notebook немесе Google Colab ортасында жүзеге асырылады. Деректерді өңдеу үшін Pandas және NumPy

кітапханалары; машиналық оқыту алгоритмдерін құру үшін Scikit-learn (логистикалық регрессия, Random forest, SVM) және XGBoost; нәтижелерді визуализациялау үшін Matplotlib және Seaborn кітапханалары пайдаланылады. Аталған құралдар жиынтығы деректерді өңдеуден бастап нәтижелерді ұсынуға дейінгі барлық кезеңдерді бір ортада жүзеге асыруға мүмкіндік береді.

#### Талқылау

Жүрек-қантaмыр аурулары қаупін болжауда машиналық оқыту әдістерін қолдану дәстүрлі тәсілдермен салыстырғанда бірқатар принципті артықшылықтарды ұсынады. Біріншіден, машиналық оқыту көрсеткіштер арасындағы сызықтық емес және жоғары ретті өзара әрекеттерді анықтай алады, бұл жекелендірілген болжам үшін аса маңызды. Екіншіден, мұндай модельдер деректер құрылымы өзгерген жағдайда қайта оқытылуы мүмкін, бұл олардың клиникалық тәжірибеге бейімделгіштігін арттырады. Үшіншіден, олар көп ғана көрсеткіштермен қатар уақыттық қатарларды (мысалы, артериялық қысым динамикасын) талдауға да мүмкіндік береді [4, 6, 9].

Алынған теориялық нәтижелер әдебиеттегі бағыттарға сай. Liu және әріптестерінің электрондық денсаулық сақтау жазбалары негізіндегі жүйелі шолуы машиналық оқыту модельдері классикалық регрессиялық шкалалардан жоғары AUC көрсеткіштерін көрсететінін, бірақ модель сапасы деректер сапасы мен таңдалған алгоритмге айтарлықтай тәуелді екенін көрсетті [4]. Cai және әріптестерінің мета-талдауы Random forest пен градиенттік бустинг алгоритмдерінің медициналық деректерде ең тұрақты нәтиже беретінін атап өтеді [3]. Сондай-ақ Weng және әріптестерінің зерттеуінде машиналық оқытудың рутиналық клиникалық деректерде дәстүрлі тәуекел шкалаларынан 7,6 %-ға дейінгі қосымша оқиғаны дұрыс анықтайтыны көрсетілген [10]. Бұл нәтижелер ұсынылып отырған тәсілдің перспективалылығын растайды.

Күтілетін нәтижелер мен практикалық маңыздылық. Зерттеу шеңберінде нақты медициналық деректер жиынтығында модельді апробациядан өткізу нәтижесінде Accuracy > 85 %, ROC-AUC > 0,90 көрсеткіштеріне қол жеткізілуі мүмкін деп болжанады, бұл әдебиеттегі ұқсас жұмыстармен үйлеседі [3, 4, 8, 10]. Әзірленетін модель пациенттерді тәуекел деңгейіне қарай уақытылы стратификациялауға, профилактикалық бағдарламаларды дербес жоспарлауға және денсаулық сақтау ресурстарын тиімді бөлуге мүмкіндік беретін медициналық шешім қабылдауды қолдау жүйесінің (Clinical Decision Support System, CDSS) құрамдас бөлігі ретінде қолданылуы мүмкін. Зерттеудің шектеуі - клиникалық қолдану үшін модельдің интерпретацияланушылығын қамтамасыз ету қажеттілігі; бұл бағытта SHAP-талдау сияқты түсіндірілетін жасанды интеллект әдістерін қолдану перспективалы болып табылады [7].

#### Қорытынды

Зерттеуде медициналық деректер негізінде жүрек-қантамыр аурулары қаупін болжауда машиналық оқыту әдістерін қолдану мүмкіндіктері талданды. Дәстүрлі тәуекелді бағалау тәсілдерінің жекелендірілген болжам жасауда жеткіліксіз болатыны көрсетіліп, осы кемшілікті машиналық оқыту әдістерінің көмегімен жоюдың перспективалылығы дәлелденді.

Жұмыста медициналық деректердің құрылымы, оларды алдын ала өңдеу кезеңдері, сондай-ақ логистикалық регрессия, Random forest, Support vector machine және XGBoost алгоритмдерінің мүмкіндіктері қарастырылды. Әрбір алгоритмнің күшті және әлсіз жақтары айқындалып, олардың салыстырмалы талдауының қажеттілігі негізделді. Бағдарламалық іске асыру үшін Python тілі мен оның арнайы кітапханаларын (Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib, Seaborn) пайдаланудың орындылығы көрсетілді. Модель сапасын бағалау Accuracy, Precision, Recall, F1-score және ROC-AUC метрикалары бойынша жүзеге асырылатыны дәлелденді.

Жұмыстың ғылыми жаңалығы – пациенттің демографиялық, клиникалық, зертханалық және мінез-құлықтық көрсеткіштерін кешенді талдау арқылы тәуекел деңгейін жекелендірілген түрде бағалайтын модельдің әдістемелік сәулетін ұсыну. Практикалық маңыздылығы – ұсынылған тәсіл медициналық шешім қабылдауды қолдау жүйелерінде қолданылып, дәрігерге пациенттерді уақтылы стратификациялауға және профилактикалық бағдарламаларды дербес жоспарлауға мүмкіндік береді. Болашақ зерттеу бағыты – нақты медициналық деректер жиынтығында модельді практикалық іске асыру, оның сапасын стандартты метрикалар бойынша бағалау, түсіндірілетін жасанды интеллект әдістерін (SHAP, LIME) қолдану және клиникалық тәжірибеге енгізу мүмкіндіктерін зерттеу. Жұмыс «Цифрлы Қазақстан» басымдықтарына сай келеді және ел денсаулық сақтау жүйесін цифрландыруға өз үлесін қосады.

#### Әдебиеттер тізімі

[1] World Health Organization. Cardiovascular diseases (CVDs): Fact sheet. – Geneva: WHO, 2024. – URL: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).

[2] Visseren F.L.J., Mach F., Smulders Y.M., et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice // European Heart Journal. – 2021. – Vol. 42, No. 34. – P. 3227-3337. DOI: 10.1093/eurheartj/ehab484.

[3] Cai Y., Cai Y.-Q., Tang L.-Y., et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review // BMC Medicine. – 2024. – Vol. 22, No. 1. – Art. 56. DOI: 10.1186/s12916-024-03273-7.

[4] Liu T., Krentz A., Lu L., Curcin V. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-

analysis // European Heart Journal – Digital Health. – 2025. – Vol. 6. – P. 7–22. DOI: 10.1093/ehjdh/ztae080.

[5] Arnett D.K., Blumenthal R.S., Albert M.A., et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease // Circulation. – 2019. – Vol. 140, No. 11. – P. e596–e646. DOI: 10.1161/CIR.0000000000000678.

[6] Mao Y., Jimma B.L., Mihretie T.B. Machine learning algorithms for heart disease diagnosis: A systematic review // Current Problems in Cardiology. – 2025. – Vol. 50, No. 8. – Art. 103082. DOI: 10.1016/j.cpcardiol.2025.103082.

[7] Salih A.M., Boscolo Galazzo I., Rauseo E., et al. A review of evaluation approaches for explainable AI with applications in cardiology // Artificial Intelligence Review. – 2024. – Vol. 57, No. 9. – Art. 240. DOI: 10.1007/s10462-024-10852-w.

[8] Teshale A.B., Htun H.L., Vered M., Owen A.J., Freak-Poli R. A Systematic Review of Artificial Intelligence Models for Time-to-Event Outcome Applied in Cardiovascular Disease Risk Prediction // Journal of Medical Systems. – 2024. – Vol. 48, No. 1. – Art. 68. DOI: 10.1007/s10916-024-02087-7.

[9] Krittanawong C., Virk H.U.H., Bangalore S., et al. Machine learning prediction in cardiovascular diseases: a meta-analysis // Scientific Reports. – 2020. – Vol. 10. – Art. 16057. DOI: 10.1038/s41598-020-72685-1.

[10] Weng S.F., Reps J., Kai J., Garibaldi J.M., Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? // PLoS ONE. – 2017. – Vol. 12, No. 4. – e0174944. DOI: 10.1371/journal.pone.0174944.

Мейрамбеков Мухамедали, магистрант, Алматинский университет энергетики и связи им. Г. Даукеева, Алматы, Казахстан, m.meirambekov@aues.kz

Жусупбеков Сарсенбек Сеитбекович, профессор, Алматинский университет энергетики и связи им. Г. Даукеева, Алматы, Казахстан, s.zhusupbekov@aues.kz

#### ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ РИСКА СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ НА ОСНОВЕ МЕДИЦИНСКИХ ДАННЫХ

Аннотация. В работе рассматриваются возможности применения методов машинного обучения для прогнозирования риска сердечно-сосудистых заболеваний на основе медицинских данных. Цель исследования – обосновать методологическую и техническую базу для разработки модели автоматизированной оценки уровня риска на основании клинических, лабораторных и поведенческих показателей пациента. Проанализированы ограничения традиционных подходов к оценке риска, описаны структура медицинских данных и этапы их предварительной обработки, рассмотрены возможности алгоритмов логистической регрессии, Random forest, Support vector machine и XGBoost. Обоснована

целесообразность использования языка Python и его специализированных библиотек (Pandas, NumPy, Scikit-learn, XGBoost) для программной реализации. Ожидаемый результат – модель бинарной классификации, относящая пациента к группе высокого или низкого/среднего риска; качество модели оценивается метриками Accuracy, Precision, Recall, F1-score и ROC-AUC. Исследование имеет прикладное значение для разработки систем поддержки принятия медицинских решений.

Ключевые слова. сердечно-сосудистые заболевания, машинное обучение, стратификация риска, персонализированная медицина, логистическая регрессия, Random forest, Support vector machine, XGBoost, медицинские данные, предиктивная модель.

Meirambekov Mukhamedali, master degree student, Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev, Almaty, Kazakhstan, m.meirambekov@aes.kz

Zhusupbekov Sarsenbek Seitbekovich, professor, Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev, Almaty, Kazakhstan, s.zhusupbekov@aes.kz

#### APPLICATION OF MACHINE LEARNING METHODS FOR PREDICTING CARDIOVASCULAR DISEASE RISK BASED ON MEDICAL DATA

Annotation. This paper considers the application of machine learning methods for predicting cardiovascular disease risk based on medical data. The aim of the study is to substantiate the methodological and technical foundation for developing a model of automated risk-level assessment based on the patient's clinical, laboratory, and behavioural indicators. The limitations of traditional risk-assessment approaches are analysed, the structure of medical data and the stages of their preprocessing are described, and the capabilities of logistic regression, Random forest, Support vector machine, and XGBoost algorithms are reviewed. The use of Python and its specialised libraries (Pandas, NumPy, Scikit-learn, XGBoost) for software implementation is justified. The expected result is a binary classification model that assigns a patient to a high-risk or low/medium-risk group; model quality is evaluated by Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. The study is of applied importance for the development of clinical decision support systems.

Keywords. cardiovascular diseases, machine learning, risk stratification, personalized medicine, logistic regression, Random forest, Support vector machine, XGBoost, medical data, predictive model.

**ҚМ АА** Күәлік нөмірі: **KZ45VPY00102718** — ҚР Мәдениет және Ақпарат министрлігі

© 2026 **Bilimger.kz** Ақпараттық-танымдық білім порталы. Барлық мазмұн авторлық құқықпен қорғалған.