Вестник РУДН. Серия: Инженерные исследования
RUDN Journal of Engineering Research

2025;26(2):168–180

ISSN 2312-8143 (Print); ISSN 2312-8151 (Online)
journals.rudn.ru/engineering-researches

DOI: 10.22363/2312-8143-2025-26-2-168-180
EDN: MBIJVQ

Research article / Научная статья

Machine Learning Methods for Predicting Cardiovascular Diseases: A Comparative Analysis

Aiym B. Temirbayeva✉, **Arshyn Altybay**

Astana IT University, *Astana, Republic of Kazakhstan*
✉ aiymtemirbaeva@gmail.com

Article history

Received: December 11, 2024
Revised: February 14, 2025
Accepted: February 25, 2025

Conflicts of interest

The authors declare that there is no conflict of interest.

Abstract. The study aims to accurately predict the presence of heart disease using machine learning models. The research evaluates and compares the performance of five algorithms — Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and Gradient Boosting — on a dataset containing clinical features of patients. The primary research question is to identify which algorithm demonstrates the best predictive performance for heart disease diagnosis. The study used a dataset of 270 patients with 13 clinical features. The data was preprocessed, and target variables were converted into binary values for classification. The dataset was split into training and test sets in a 70–30 ratio. Five machine learning models were trained and evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Confusion matrices were analyzed to gain additional insights into model performance. Logistic Regression and Random Forest showed the best results among all models, with an accuracy of 86.4 and 80.2%, respectively. The Logistic Regression showed a ROC-AUC score of 0.844, while the Random Forest showed a score of 0.88. The confusion matrices revealed the strengths and weaknesses of each model in terms of forecasting. Logistic Regression and Random Forest were identified as the most reliable models for predicting heart disease in this dataset. Future work will explore hyperparameter tuning and ensemble methods to further enhance model performance, providing valuable insights for early diagnosis and treatment of cardiovascular diseases.

Keywords: Random Forest, support vector machine, gradient boosting, decision tree, Logistic Regression, accuracy

Authors' contribution

Temirbayeva A.B. — writing – original draft, visualization, validation; *Altybay A.* — methodology, investigation, validation.

For citation

Temirbayeva AB, Altybay A. Machine learning methods for predicting cardiovascular diseases: A comparative analysis. *RUDN Journal of Engineering Research.* 2025;26(2):168–180. http://doi.org/10.22363/2312-8143-2025-26-2-168-180

Методы машинного обучения для прогнозирования сердечно-сосудистых заболеваний: сравнительный анализ

А.Б. Темирбаева✉, А. Алтыбай

Астанинский IT-университет, *Астана, Республика Казахстан*

✉ aiymtemirbaeva@gmail.com

История статьи

Поступила в редакцию: 11 декабря 2024 г.
Доработана: 14 февраля 2025 г.
Принята к публикации: 25 февраля 2025 г.

Заявление о конфликте интересов

Авторы заявляют об отсутствии конфликта интересов.

Аннотация. Цель исследования — точное предсказание наличия сердечно-сосудистых заболеваний с помощью моделей машинного обучения. Оценивалась и сравнивалась эффективность пяти алгоритмов: логистической регрессии, машины опорных векторов, дерева решений, случайного леса и градиентного бустинга на наборе данных, содержащем клинические характеристики пациентов. Определялось, какой из алгоритмов демонстрирует наилучшие прогностические характеристики для диагностики заболеваний сердца. Использован набор данных 270 пациентов с 13 клиническими признаками. Данные были предварительно обработаны, а целевые переменные преобразованы в бинарные значения для классификации. Набор данных был разделен на обучающий и тестовый в соотношении 70–30. Пять моделей машинного обучения были обучены и оценены с помощью таких метрик, как точность, recall, precision, F1-score и ROC-AUC. Для получения дополнительной информации о производительности моделей были проанализированы матрицы ошибок. В результате логистическая регрессия и случайный лес показали наилучшие результаты среди всех моделей с точностью 86,4 и 80,2 %, соответственно. Логистическая регрессия продемонстрировала ROC-AUC на уровне 0,844, а случайный лес — 0,88. С помощью матриц путаницы выявлены прогностические достоинства и недостатки каждой модели. Авторами сделаны следующие выводы: логистическая регрессия и случайный лес были определены как наиболее надежные модели для прогнозирования сердечно-сосудистых заболеваний в этом наборе данных. В дальнейшем планируется изучение методов настройки гиперпараметров и ансамбля для повышения эффективности моделей, что позволит получать ценные сведения для ранней диагностики и лечения сердечно-сосудистых заболеваний.

Ключевые слова: случайный лес, машина опорных векторов, градиентное усиление, дерево решений, логистическая регрессия, точность

Вклад авторов

Темирбаева А.Б. — написание исходного текста, визуализация, валидация; *Алтыбай А.* — методология исследования, валидация.

Для цитирования

Temirbayeva A.B., Altybay A. Machine learning methods for predicting cardiovascular diseases: A comparative analysis. // Вестник Российского университета дружбы народов. Серия: Инженерные исследования. 2025. Т. 26. № 2. С. 168–180. http://doi.org/10.22363/2312-8143-2025-26-2-168-180

Introduction

Cardiovascular diseases (CVDs) cast a long shadow around the world, claiming millions of lives every year. CVDs are the leading cause of death worldwide, claiming an estimated 17.9 million lives in 2019 [1]. This translates to one person dying every 34 s because of CVD. In Kazakhstan, the situation is similar, with a high prevalence of CVDs. The prevalence of CVDs among the popu-lation increased from 1845.4 per 100,000 people in 2004 to 2597.5 per 100,000 people by 2017 [2]. These

statistics emphasize the urgent need for improved methods of diagnosis, treatment, and prevention. Fortunately, advances in artificial intelligence (AI) have offered promising solutions that may prove particularly effective. Traditional diagnostic methods for heart disease can be subjective and prone to human error. AI algorithms excel at analyzing complex medical images, such as echocardiograms, potentially leading to more accurate diagnosis. A recent study published in Nature demonstrated an AI framework's ability to effectively classify cardiac diseases using audio signals [3]. AI can be used to analyze an individual's specific medical data, including demographics, lifestyle factors, and genetic information. This personalized approach allows for tailored risk assessments and treatment plans, potentially leading to better patient outcomes. Early detection is crucial for improving the outcome of heart disease. The AI-powered analysis of various data points, including electrocardiograms (ECGs), heart sounds, and electronic health records, has shown promise in identifying patterns and predicting the likelihood of developing heart disease. Studies suggest that machine learning (ML) models, particularly Random Forests (RF), can achieve high accuracy rates (approximately 90%) in heart disease prediction [4]. This research hypothesizes that ML algorithms, when trained on a diverse dataset of patient attributes, can effectively learn patterns associated with heart disease risk and achieve high levels of predictive accuracy. Additionally, it is hypothesized that certain algorithms may outperform others in terms of predictive performance depending on the characteristics of the dataset and the complexity of the underlying relationships. Therefore, the purpose of this study is to investigate the effectiveness of ML algorithms in predicting the risk of heart disease based on various patient attributes. By analyzing a dataset containing features related to patient demographics, medical history, and diagnostic tests, the aim is to develop models that can accurately classify individuals as either having a high risk of heart disease. The objectives of this research were to preprocess and analyze a dataset containing features relevant to heart disease prediction and evaluate multiple ML algorithms for heart disease classification. The performance of the developed models was assessed using key performance indicators, such as accuracy, recall, and ROC AUC, to compare the effectiveness of different ML algorithms in predicting heart disease risk. In conclusion, with heart disease statistics painting a concerning picture, AI-powered methods for analysis and classification offer significant potential. Early detection, improved diagnostics, and personalized medicine facilitated by ML hold the promise of revolutionizing cardiac care, ultimately saving lives, and improving overall heart health.

Background and Related Work. Heart disease, which includes a variety of disorders affecting the heart and blood arteries, continues to be a major global health problem owing to its high prevalence, morbidity, and mortality rates. Early identification and precise risk assessment are critical for the effective prevention and management of cardiovascular disorders. In recent years, ML techniques have received increased attention in the healthcare profession due to their potential to aid in disease prediction and diagnosis. Several researchers have investigated the use of ML algorithms to forecast the risk of heart disease using various datasets and approaches. For instance, [5] aimed to create a novel end-to-end technique for detecting and classifying heart-sound abnormalities that can be applied to a variety of heart-sound diagnosis activities. They created a Multi-feature Decision Fusion Network (MDFNet) composed of two modules: Multi-dimensional Feature Extraction (MFE) and Multi-dimensional Decision Fusion (MDF). This approach was applied to two datasets, which are open-access databases of heart-sound recordings. There were four experiments with an overall accuracy of 94.44% and an F1-score of 86.90% for the binary classification task and 99.30% for the five-classification task. This technology surpassed other cutting-edge methods and showed promising therapeutic applications. CVDs remain a constant threat in areas with low resources and moderate incomes. [6] Other researchers have used deep learning techniques to reveal a model that is enhanced by bispectrum-inspired feature extraction and the Vision Trans-

former (ViT) model's cutting-edge capabilities. This paradigm leads to the binary classification of cardiac sounds as 'normal' or 'abnormal.' Their algorithm relies on data from the PhysioNet Challenge 2022 database, which contains 3163 data points from 942 patients. The model demonstrates an excellent classification process with impressive consistency, particularly when distinguishing between pregnant and non-pregnant individuals' heart sounds. The model described in this work, which employs bispectrum for feature extraction and the ViT model for classification, achieves an accuracy of 0.91 and an AUC of 0.98 in the test set drawn from the PhysioNet Challenge 2016 and 2022 databases. This study [7] applied ML to classify cardiovascular disorders accurately. The authors used Naive Bayes, RF, Decision Tree, and Multilayer Perceptron algorithms to combine predictions using the Bagging process, which is being investigated as a way to improve the accuracy of less accurate algorithms. Bagging, boosting, voting, and stacking are among the ensemble methods employed. The study used UCI's Cleveland heart dataset (CHD) for people with heart disease. This dataset contained 303 occurrences and only 14 attributes. Consequently, the accuracy of the ensemble approach utilizing boosting and bagging was superior to that of the individual classifiers. The initial accuracy of the RF algorithm was 81.53%. Upon incorporating the feature selection, the accuracy increased to a maximum of 90.52%. The Multilayer Perceptron method increased significantly from 78.52 to 96.18%. The findings highlight the importance of critical feature selection in increasing the accuracy of the models. They were able to reduce noise and concentrate on the main risk factors for cardiovascular disease by choosing the most pertinent attributes, which led to more precise forecasts. The study demonstrates that feature selection and ensemble classification algorithms can greatly increase the accuracy of cardiac disease risk prediction. [8] employed the *K*-nearest neighbors (KNN) algorithm to handle classification problems related to coronary heart disease. There are six forms of coronary heart disease, however only two were selected for classification. These include angina pectoris (AP) and acute myocardial infarction (AMI). The CHD dataset was obtained from the National Center for Health Statistics (NCHS) and reviewed by the medical team at the Center for Specialty Cardiology. The dataset for this study consisted of 100 case histories of CHD patients. Following the analysis, the data was separated into training and test data-sets, with the training features containing 80 records (80% of the data) and the test features containing 20 records (20% of the data). Two distinct datasets were created: Boolean values (D1), and values consisting of twos within a certain range (D2). The primary goal is to compare the F1-score diagnostic levels and choose the most appropriate one.

As a result, by selecting a random value of k equal to 5 for D1 data, the k-nearest neighbors algorithm achieves an accuracy of 93%, outperforming some experiments with ML methods, such as RF, Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression. However, the same k value for D2 data resulted in a lower accuracy of 76%. The experimental results showed that for coronary heart disease classification, the D1 dataset produced better F1-score results of 92 and 94% compared to the D2 dataset of 70 and 81%, respectively, using the KNN algorithm. This shows that Boolean attributes can be an effective dataset for categorization. As technology and medical diagnostics become more integrated, data mining and the storage of medical information can improve patient care. Therefore, it is critical to investigate the interrelation of risk factors in patients' medical histories and comprehend their individual contributions to cardiovascular disease prognosis. Another study aimed to analyze numerous components of patient data to accurately forecast heart illnesses [9]. The most important qualities for predicting heart disease were discovered utilizing a correlation-based subset feature selection and a best-fit search approach. Age, sex, smoking, obesity, food, physical activity, stress, type of chest pain, previous chest pain, diastolic blood pressure, diabetes, tro-ponin, ECG, and aim were identified as the most important factors in the diagnosis of

heart disease. Consequently, data were gathered from hospitals, diagnostic centers, and clinics throughout Bangla-desh. The patients were questioned, analytical results were evaluated, and information on essential cha-racteristics was gathered. The dataset consisted of test results from 59 patients and their responses to various questionnaires. The data were separated into two parts: training data (67%) and test data (33%). Various AI approaches (e.g., Logistic Regression, Naive Bayes, K-NN, SVM, decision tree, RF, and MLP) were tested for two types of heart disease datasets (all and chosen features). RF with selected features achieved 90% accuracy, 90.91% precision, 100% recall, 90.91% F1-score, and 89.90% ROC-AUC, the highest among the AI approaches. The proposed method can be utilized to assist in the early diagnosis of heart disease. Audio-based heart disease detection is an exciting research topic that uses the audio signals produced by the heart to detect and diagnose cardiovascular disorders. ML and deep learning (DL) are important techniques for classifying and identifying cardiac disorders using acoustic inputs [3]. Evaluated ML and DL algorithms for detecting cardiac disorders using noisy audio input. This study used two subsets of the Pascal Challenge datasets, which contained real cardiac audio signals from 400 participants. Spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs) were used to process and visualize the signals throughout the study. To improve the model performance, we used data aug-mentation, which involves inserting synthetic noise into heartbeat signals. In addition, a feature en-semble was created to combine various sound feature extraction approaches. Several machine-learning and deep-learning classifiers have been used to diagnose cardiac diseases. The multilayer perceptron model outperformed the other models and earlier experiments with an accuracy of 95.65%. This study revealed how this technology can accurately diagnose cardiac problems using acoustic signals. This study [10] evaluated the performance of seven machine-learning algorithms for heart disease diagnosis using a dataset consisting of 4,238 records and 16 patient characteristics. The algorithms tested were Naive Bayes, decision trees (DT), RF, SVM, artificial neural networks (ANN), KNN, and Logistic Regression (LR). The results showed a significant difference in accuracy between the models. The LR model showed the highest accuracy (85.5%), followed by RF (83.9%) and artificial neural networks (83.7%). The K-nearest neighbors algorithm also performed well, with an accuracy of 83.4%, but its accuracy was slightly lower than that of the RF and ANN models. The decision tree models achieved an accuracy of 79.9%, surpassing Naive Bayes and SVM, which showed lower accuracy values of 78.9% and 70.9%, respectively. These results highlight the potential of ML algorithms for improving the diagnosis of heart disease, with the LR, RF, and ANN models performing the best for accurate predictions. It is evident that the KNN and DT models hold some value in this context, though it should be noted that their performance lags slightly behind that of the most effective algorithms.

1. Implemented Algorithm

This study compares the performance of Logistic Regression SVM decision tree, RF and gradient boosting techniques. Decision trees are tree-like structures used to make predictions. They divide the feature space into discrete regions based on feature tests, with each leaf node representing a class label or goal value. Decision trees are simple and easy to understand. RF is an evolution of this concept, which uses several decision trees to increase prediction accuracy and generalization. It generates an ensemble of decision trees from arbitrary subsets of training data and features. Individual tree findings are combined to yield the final prediction [8]. The RF approach extends the summarizing method by combining summarization and feature randomness to generate an uncorrelated forest of decision trees. The randomized feature approach, also known as the bag-of-features method or the "random subspace method" creates a random subset of features with low correlation between decision trees. This is an important distinction between decision trees and RF. While decision trees

analyze all possible feature partitions, RF only select a subset of these features [11]. In contrast, SVM is an efficient algorithm for both classification and regression. SVM is a sophisticated ML technique that can address linear or nonlinear classification, regression, and even outlier identification problems. SVM can be used for a variety of applications, including text classification, picture classification, spam detection, handwriting identification, gene expression analysis, face and anomaly detection [12]. SVMs are versatile and useful in various applications because they can handle high-dimensional data and nonlinear dependencies. SVM algorithms are particularly effective because they attempt to find the largest separation hyperplane among the many classes present in the target feature. LR is a supervised ML algorithm used in classification problems to predict whether an instance belongs to a specific class or not. LR is a statistical procedure that is used to determine the relationship between two data points. It is used for binary classification with a sigmoidal function that treats the input data as independent variables and returns a probability value between 0 and 1 [13]. Gradient boosting is a powerful boosting approach that combines numerous weak training models into strong training models by training each new model to minimize a loss function, such as the prior model's mean square error or cross-entropy, with gradient descent. At each iteration, the approach calculates the gradient of the loss function with respect to the predictions of the current ensemble, and then trains a new weak model to minimize this gradient. The predictions of the new model are added to the ensemble, and the procedure is repeated until a stopping requirement is met [14].

2. Methodology

2.1. Dataset

This study used a dataset from the University of California Irvine's Machine Learning Repository. It consists of different attributes that are used to predict patients at high risk of heart disease.

The dataset[1] consists of 14 columns and 270 rows. Thirteen of them have integer and decimal data types, whereas only one column is in the string data type. The details of the columns are summarized in Table 1 according to the original data source[2]. After analyzing the dataset, we replaced the values of the "Heart Disease" column with integer values. Thus, the value of "Absence" is 0 and "Presence" is 1. It is necessary to compare the ML methods that focus on this column.

Table 1

Description of dataset

Variable Name	Role	Type	Description
Age	Feature	Integer	Age of the patient
Sex	Feature	Integer	1= male, 0 = female
Chest Pain Type	Feature	Integer	1= typical angina; 2 = atypical angina; 3 = non-anginal pain; 4 = asymptomatic
BP	Feature	Integer	Resting blood pressure
Cholesterol	Feature	Integer	Serum cholestoral
FBS over 120	Feature	Integer	Fasting blood sugar > 120 mg/dl
EKG results	Feature	Integer	0 = normal; 1 = having ST-T wave abnormality; 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria
Max HR	Feature	Integer	Maximum heart rate achieved

[1] Damarla R. Heart Disease Prediction. Predicting probability of heart disease in patients. Kaggle. Available from: https://www.kaggle.com/datasets/rishidamarla/heart-disease-prediction/data (accessed: 12.09.2024).

[2] UCI Machine Learning Repository. (n.d.). Available from: https://archive.ics.uci.edu/dataset/45/heart+disease (accessed: 12.09.2024).

Ending of the Table 1

Variable Name	Role	Type	Description
Exercise Angina	Feature	Integer	Exercise induced angina (1 = yes; 0 = no)
ST Depression	Feature	Decimal	ST depression induced by exercise relative to rest
Slope of ST	Feature	Integer	the slope of the peak exercise ST segment: 1 = upsloping; 2 = flat; 3 = downsloping
Number of Vessels Fluro	Feature	Integer	Number of major vessels (0–3) colored by flourosopy
Thallium	Feature	Integer	3 = normal; 6 = fixed defect; 7 = reversable defect
Heart Disease	Target	String	Presence, absence of heart disease

S o u r c e: by A.B. Temirbayeva

2.2. Algorithms

The following ML algorithms were implemented and tested to predict the risk of heart disease:

1) *Logistic Regression:* LR is the most common modeling strategy for binary outcomes in epidemiology and medicine[3]. LR uses a logistic function to calculate the likelihood that an observation belongs to one of two classes.

2) *Support vector machine (SVM):* The Support Vector Machine algorithm is based on finding the best way to separate hyperplanes between multiple classes. The best hyperplane is often determined by determining the optimal curvature of the hyperplane and maximizing the separation distance between the nearest data points from each class [15].

3) *Decision Tree:* Decision Trees are a set of computational heuristics that solve problems by creating binary splitting rules on data features based on the criterion of maximizing the information acquired from the split [16].

4) *Random Forest:* RF is an ensemble technique that mixes multiple classification trees to generate a prediction based on the majority vote of a single tree. A random subset of the dataset is used to fit each constituent tree with predictors selected at random [17].

5) Gradient Boosting: Gradient boosting is an iterative approach for fitting simple statistical models to data. GB models the data using classifier trees, which are simple statistical models. Iteratively, GB examines the current model's performance, adds another tree to the previous errors, and updates the model by adding the regression tree to the ensemble [18].

Each model was implemented using the scikit-learn library in Python. For each model, hyperparameters were tuned using cross-validation and Grid Search to select the best parameters for training.

2.3. Code

This section consists of the following stages.

1) *Data preparation:* The first step is data cleaning. Raw data often contains noise, errors, or missing values, which can negatively impact the performance of machine-learning models. In our case, we started by loading an initial dataset containing 270 observations and 14 traits, each providing information about the patient's health such as age, sex, cholesterol levels, and other measures. Following data loading, we conducted a check for missing values, which turned out to be non-existent. The column which indicates the presence or absence of cardiovascular disease, was then converted from its original string format ("Absence "Presence") to a numeric format (0 and 1 respectively). This conversion is a crucial step for the ML algorithms to effectively handle the target variable. Figure 1 shows the actual values of the healthy and unhealthy patients.

[3] UCI Machine Learning Repository. (n.d.). Available from: https://archive.ics.uci.edu/dataset/45/heart+disease (accessed: 12.09.2024).

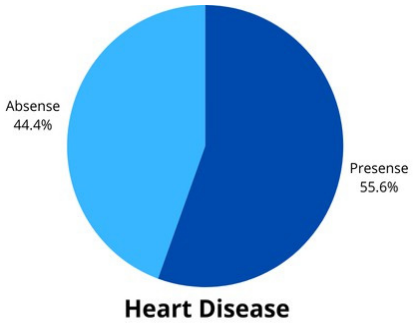

Figure 1. Patients with absense/presense of heart disease
Source: by A.B. Temirbayeva

After these initial steps, we identified the "Heart Disease" column as the target value, the one we aim to predict, and the remaining columns as features that will be used to make those predictions.

2) *Modeling:* After data preparation, the dataset was divided into training and test samples at a ratio of 70:30. The training sample accounted for 70% of the total data and was used to build ML models. The test sample, which accounted for 30% of the data, was used to evaluate the performance of the trained models.

Training set dimension: 189, 13.

Test set dimension: 81, 13.

This separation allowed us to test how well the models could generalize their predictions to new, unseen data. To ensure a comprehensive approach, five diverse ML algorithms were selected and trained to predict the risk of heart disease: Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and Gradient Boosting. Each model was trained using a training data sample.

3) *Evaluation:* Key accuracy measures were used to evaluate the performance of the trained models. The Confusion Matrix (Confusion Matrix) was used for a more detailed analysis of the model performance. The results were compared after these metrics were obtained.

2.4. Key Performance Indicators

The following Key Performance Indicators were used to evaluate the performance of the trained models:

◆ *Accuracy:* Accuracy was evaluated as the number of correct heart disease predictions divided by the total number of datasets. The accuracy comparison was based on the performance of the four classification methods.

◆ *Recall:* The proportion of correctly classified positive observations out of the total number of positive observations. Completeness is essential in medical applications, where missed cases must be minimized.

◆ *F1 Score:* Harmonic mean accuracy and completeness. A high F1 score indicated perfect precision and recall of the proposed model.

◆ The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is an efficiency metric for classification. Utilization of the AUC-ROC metric is a method for evaluating the predictive capacity of the model. The model performs better when the AUC is larger. It can be calculated quantitatively by comparing the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold values.

A Confusion Matrix was used to analyze the performance of the models in more detail. The error matrix allows us to visualize how often the model makes mistakes when classifying each class. This includes the following elements:

◆ *True Positives (TP):* Number of correctly predicted positive cases (cardiac patients correctly classified as sick).

◆ *True Negatives (TN):* Number of correctly predicted negative cases (patients without heart disease correctly classified as healthy).

◆ *False Positives (FP):* Number of false predictions of positive cases (patients without heart disease incorrectly classified as sick).

◆ *False Negatives (FN):* Number of incorrect predictions of negative cases (heart disease patients misclassified as healthy).

3. Results and Discussions

After training the models, key performance indicators were identified for each model. Logistic regression 0.827 and Random Forest 0.79 exhibited the highest accuracy values, indicating their proficiency in accurately classifying most of the samples (Tables 2 and 3).

In particular, logistic regression had the highest ROC AUC 0.844, signifying its superior ability to distinguish between the healthy and diseased classes. The precision, Recall, and F1-Score metrics further support the superiority of Logistic Regression and RF, with both models consistently outperforming the other models. Of note, Logistic Regression achieved a remarkable F1-Score for class 1 (0.81), a key metric in balancing Precision and Recall when identifying patients, underscoring its importance in our analysis.

Gradient boosting (Table 4) had the highest accuracy 0.777 among the three models, followed by the Decision Tree 0.765 and SVM 0.641. The highest ROC AUC 0.75 indicated a superior ability to distinguish between classes compared to the Decision Tree 0.741 and SVM 0.595. With values of Precision 0.78, Recall 0.88, and F1-Score 0.83, it demonstrated a lower performance in predicting healthy patients. The decision tree, while showing comparable results, lagged with Precision 0.78, Recall 0.86 and F1-Score 0.82.

Unfortunately, SVM (Table 5) showed the worst results for class 0, with Precision 0.67, Recall 0.82, and F1-Score 0.73. For class 1, gradient boosting continued to lead with Precision 0.57, Recall 0.38, and F1-Score 0.45, indicating its superior performance in predicting patients with heart disease. As demonstrated in Table 6, the decision tree exhibited lower values for class 1, with Precision 0.75, Recall 0.62 and F1-Score 0.68. SVM showed the worst results for class 1, with Precision 0.57, Recall 0.38, and F1-Score 0.45.

The results of each model were compared across all the key metrics. The graph below summarizes the accuracy scores of each model.

The ROC curve for SVM, but significantly lower than that of Logistic Regression, still demonstrated moderate performance with an AUC of 0.73. The position of the curve further from the top-left corner reflects a lower actual positive rate and a higher false positive rate than the other models. This suggests that SVM, although less effective in distinguishing between patients with and without heart disease, still provides valuable insights.

Table 2

Classification for Logistic Regression

Target	Precision	Recall	F1-Score	Support
0	0.81	0.94	0.87	49
1	0.88	0.66	0.75	32
Accuracy	0.827			
ROC AUC	0.797			

Source: by A.B. Temirbayeva

Table 3

Classification for RF

Target	Precision	Recall	F1-Score	Support
0	0.78	0.92	0.84	49
1	0.83	0.59	0.69	32
Accuracy	0.790			
ROC AUC	0.756			

Source: by A.B. Temirbayeva

Table 4

Classification for gradient boosting

Target	Precision	Recall	F1-Score	Support
0	0.78	0.88	0.83	49
1	0.77	0.62	0.69	32
Accuracy	0.777			
ROC AUC	0.751			

Source: by A.B. Temirbayeva

Table 5

Classification for SVM

Target	Precision	Recall	F1-Score	Support
0	0.67	0.82	0.73	49
1	0.57	0.38	0.45	32
Accuracy	0.641			
ROC AUC	0.595			

Source: by A.B. Temirbayeva

Table 6

Classification for decision tree

Target	Precision	Recall	F1-Score	Support
0	0.78	0.86	0.82	49
1	0.74	0.62	0.68	32
Accuracy	0.7658			
ROC AUC	0.741			

Source: by A.B. Temirbayeva

The ROC curve for the Decision Tree model, slightly below that of the SVM, with an AUC of 0.74, indicates a similar performance. This balanced comparison suggests that the Decision Tree performs similarly but is slightly worse than the SVM.

The ROC curve for the RF model, a model that consistently shows a high ability to distinguish between classes, with an AUC of 0.89, is a testament to its robust performance. The proximity of the curve to the top left corner indicated good performance in terms of sensitivity and specificity. RF, therefore, not only demonstrates a robust performance but also provides a high level of reliability in predicting heart disease. The ROC curve for Gradient Boosting was similar to that of the RF, with an AUC of 0.88. The model, with its high actual positive rate and low false positive rate, is another strong performer in distinguishing between patients with and without heart disease. The graph of this curve is shown in Figure 2. The error matrix for each model was also calculated and used to visualize the classification errors. Logistic Regression and RF, with their high performance and low FP and FN values, demonstrated a practical balance between identifying healthy individuals and those with heart disease. Therefore, these models are reliable tools for real-world applications in predicting heart diseases. Gradient Boosting performs well but has slightly higher FP and FN than Logistic Regression and RF. Decision Tree provides moderate performance, with a higher number of FP and FN, indicating it is less reliable than the top models. While SVM shows the weakest performance, with high FP and FN, it is important to note that it may not be suitable for this specific task. This acknowledgement of the limitations of the models underscores the transparency and honesty of our research. Through a comprehensive analysis of the confusion matrix, Logistic Regression and RF have emerged as the top-performing models for predicting heart disease. This comparative approach underscores the credibility and robustness of the findings (Figures 3, 4.)

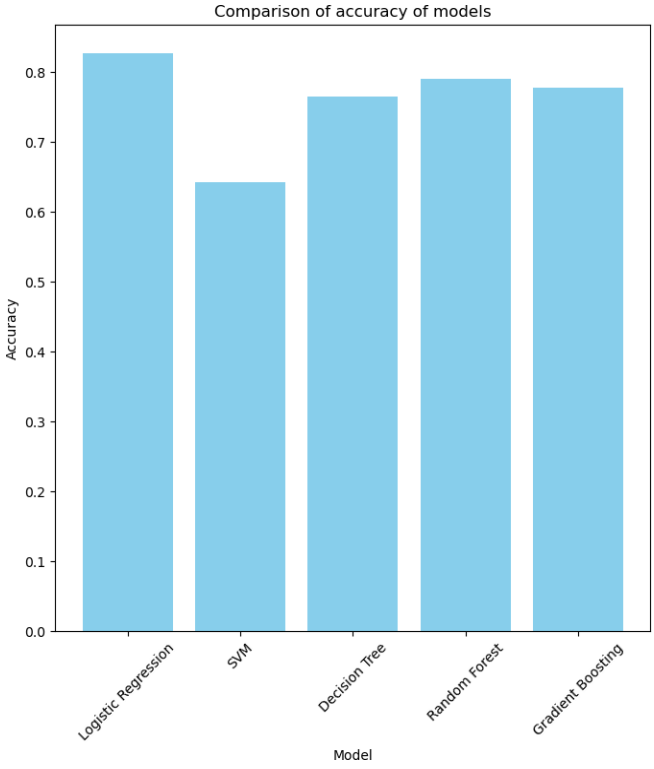

Figure 2. Comparison of accuracy

S o u r c e: by A.B. Temirbayeva

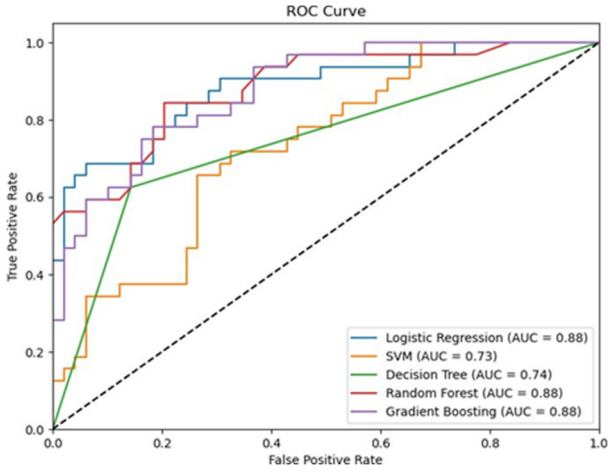

Figure 3. ROC curve for all models

S o u r c e: by A.B. Temirbayeva

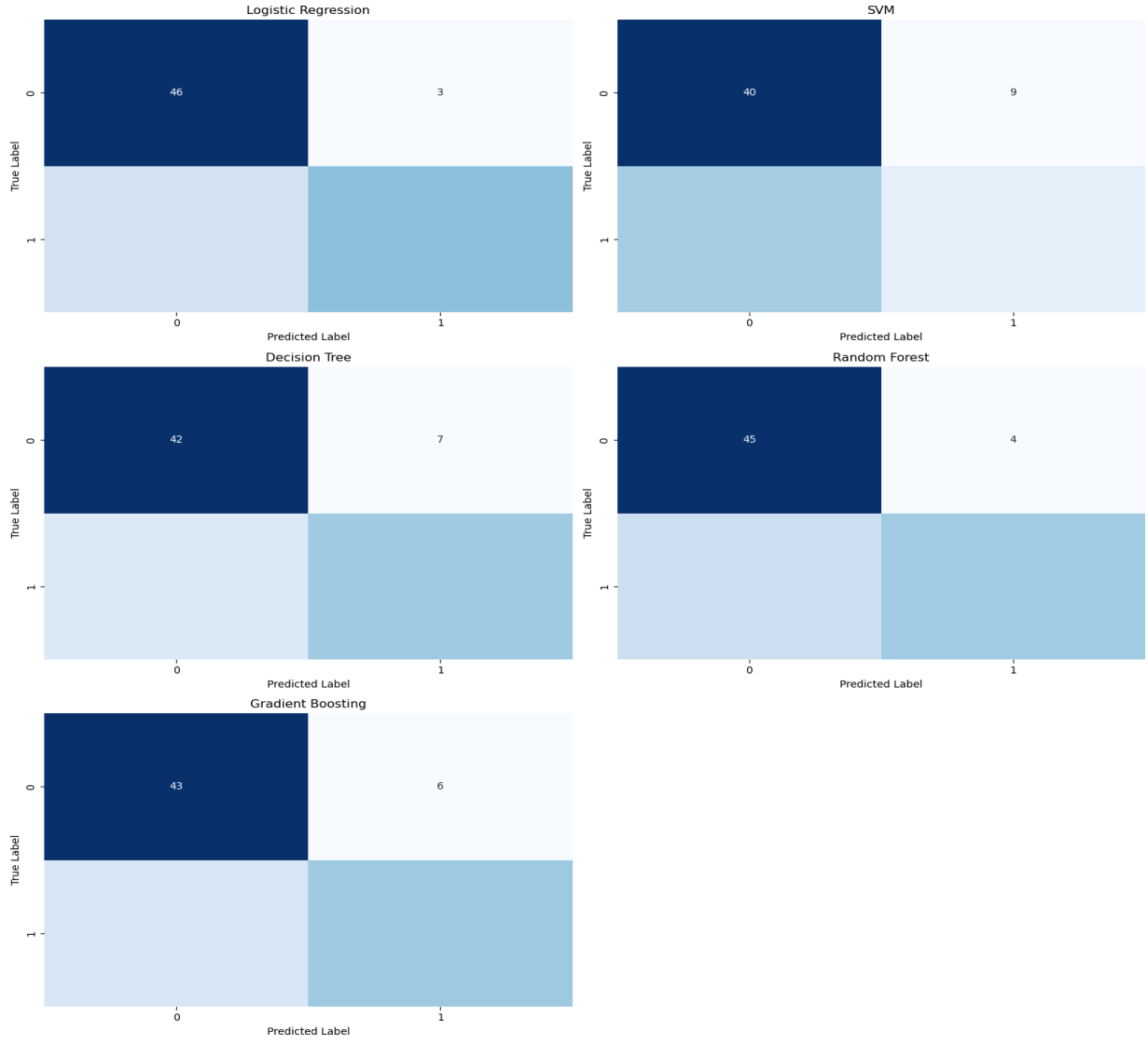

Figure 4. Confusion Matrices for Models

Source: by A.B. Temirbayeva

Conclusion

After careful evaluation of the results presented in the tables above, it becomes clear that the Logistic Regression model is the best choice for this dataset and for predicting CVDs. Its performance, as reflected in the estimates, leaves no room for doubt. The Logistic Regression model demonstrated high accuracy and ROC AUC values, indicating its ability to distinguish between patients with and without heart disease.

◆ *Precision:* shows the proportion of correctly predicted positive cases among all cases predicted as positive. For class 1 (patients with heart disease), the Precision was 0.88, meaning that 89% of the "with heart disease" predictions were correct.

◆ *Recall (Completeness):* shows the proportion of correctly predicted positive cases among all true positive cases. The completeness of class 1 was 0.75, indicating that the model correctly identified 75% of all cardiac patients.

◆ *F1-Score:* is the Harmonic mean of accuracy and completeness. For class 1, the F1-score was 0.79, indicating a balance between accuracy and completeness.

◆ *ROC AUC:* evaluates the classification quality based on the curve of "positive false classification rate" vs. "negative false classification rate". A value of 0.79 means that the model has a good ability to discriminate between classes.

◆ AUC of 0.88, the highest among all the models, is a significant indicator of the superior performance of Logistic Regression in predicting heart disease in this dataset.

Therefore, it is clear that Logistic Regression is the optimal model for this dataset and the task of predicting heart diseases. Its high-performance evaluation metrics and balanced ratio between accuracy and completeness underscore its superiority.

References

1. Mendis S, Graham I, Narula J. Addressing the global burden of cardiovascular diseases; need for scalable and sustainable frameworks. *Global Heart.* 2022;17(1):46. https://doi.org/10.5334/gh.1139 EDN: ALVXJY

2. Mukasheva G, Abenova M, Shaltynov A, Tsigengage O, Mussabekova Z, Bulegenov T, Shalgumbaeva G, Semenova Yu. Incidence and mortality of cardiovascular disease in the Republic of Kazakhstan: 2004–2017. *Iranian Journal of Public Health.* 2022;51(4):821–830. https://doi.org/10.18502/ijph.v51i4.9243 EDN: DHJPUR

3. Abbas S, Ojo S, Hejaili AA, Sampedro GA, Almadhor A, Zaidi M, Kryvinska N. Artificial intelligence framework for heart disease classification from audio signals. *Scientific Reports.* 2024;14(1)3123. https://doi.org/10.1038/s41598-024-53778-7 EDN: UPLLIK

4. Hossain MI, Maruf MH, Khan MAR, Prity FS, Fatema S, Ejaz MS, Khan M. Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran Journal of Computer Science*. 2023;6(4):397–417. https://doi.org/10.1007/s42044-023-00148-7 EDN: IKJGNI

5. Zhang H, Zhang P, Wang Z, Chao L, Chen Y, Li Q. Multi-Feature decision fusion network for heart sound abnormality detection and classification. *IEEE Journal of Biomedical and Health Informatics*. 2024;28(3):1386–1397. https://doi.org/10.1109/jbhi.2023.3307870 EDN: SSTBYM

6. Liu Z, Jiang H, Zhang F, Ouyang W, Li X, Pan X. Heart sound classification based on bispectrum features and Vision Transformer mode. *Alexandria Engineering Journal*. 2023;85:49–59. https://doi.org/10.1016/j.aej.2023.11.035 EDN: EKYJWK

7. Mahajan RA, Balkhande B, Wanjale K, Chitre A, Jadhav TA, Hundekari SN. Enhancing Heart Disease Risk Prediction Accuracy through Ensemble Classification Techniques. *International Journal of Intelligent Systems and Applications in Engineering.* 2023;11(10s):701–713. Available from: https://ijisae.org/index.php/IJISAE/article/view/3325 (accessed: 12.09.2024).

8. Rakhimov M, Akhmadjonov R, Javliev S. Artificial intelligence in Medicine for Chronic disease classification using Machine learning. *2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT).* 2022:1–6 https://doi.org/10.1109/aict55583.2022.10013587

9. Hossain I, Maruf M, Khan MAR, Prity FS, Fatema S, Ejaz MS, Khan M. Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran Journal of Computer Science.* 2023;6(4):397–417. https://doi.org/10.1007/s42044-023-00148-7 EDN: IKJGNI

10. Erdem K, Yildiz MB, Yasin ET, Koklu M. A detailed analysis of detecting heart diseases using artificial intelligence methods. *Intelligent Methods in Engineering Sciences*. 2023;2(4):115–124 https://doi.org/10.58190/imiens.2023.71 EDN: DYZTFY

11. Salman HA, Kalakech A, Steiti A. Random Forest algorithm Overview. *Babylonian journal of machine learning*. 2024;2024:69–79. https://doi.org/10.58496/bjml/2024/007 EDN: HWNARA

12. Wang Q. Support Vector machine algorithm in machine learning. *2022 IEEE International Conference on Artificial Intelligence and Computer Applications* (*ICAICA*). 2022:750–756. https://doi.org/10.1109/icaica54878.2022.9844516

13. Berrendero JR, Bueno-Larraz B, Cuevas A. On functional logistic regression: some conceptual issues. *Test*. 2022;32(1):321–349. https://doi.org/10.1007/s11749-022-00836-9 EDN: XCAHRR

14. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*. 2020;54(3):1937–1967. https://doi.org/10.1007/s10462-020-09896-5

15. Levy JJ, O'Malley AJ. Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. *BMC Medical Research Methodology*. 2020;20(1):171. https://doi.org/10.1186/s12874-020-01046-3

16. Liew BXW, Kovacs FM, Rugamer D, Royuela A. Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain. *European Spine Journal*. 2022;31(8):2082–2091. https://doi.org/10.1007/s00586-022-07188-w EDN: YWKGZQ

17. Becker T, Rousseau A, Geubbelmans M, Burzykowski T, Valkenborg D. Decision trees and random forests. *American Journal of Orthodontics and Dentofacial Orthopedics*. 2023;164(6):894–897. https://doi.org/10.1016/j.ajodo.2023.09.011 EDN: QKTJHR

18. Mahajan RA, Balkhande B, Kirti Wanjale K, Chitre A, Jadhav TA, Hundekar SN. Enhancing Heart Disease Risk Prediction Accuracy through Ensemble Classification Techniques. *International Journal of Intelligent Systems and Applications in Engineering*. 2023;11(10s):701–713. Available from: https://ijisae.org/index.php/IJISAE/article/view/3325/1911 (accessed: 12.09. 2024).

About the authors

Aiym B. Temirbayeva, MS student in Applied Data Analytics, Astana IT University, 55/11 Mangilik El avenue, Business center EXPO, block C1, Astana, 010000, Kazakhstan; ORCID: 0009-0003-6131-2884; e-mail: aiymtemirbaeva@gmail.com

Arshyn Altybay, PhD of Philosophy, Senior Researcher of the Department of Differential Equations, Institute of Mathematics and Mathematical Modeling, 28 Shevchenko St, 050010, Almaty, Republic of Kazakhstan; ORCID: 0000-0003-4939-8876; e-mail: arshyn.altybay@gmail.com

Сведения об авторах

Темирбаева Айым Болатовна, магистрант департамента по прикладной аналитике данных, Астанинский IT-университет, Казахстан, 010000, г. Астана, пр-т Мангилик Ел, 55/11, Бизнес-центр EXPO, блок C1; ORCID: 0009-0003-6131-2884; e-mail: aiymtemirbaeva@gmail.com

Алтыбай Аршын, доктор философии, старший научный сотрудник департамента дифференциальных уравнений, Институт математики и математического моделирования, Республика Казахстан, 050010, г. Алматы, ул. Шевченко, 125; ORCID: 0000-0003-4939-8876; e-mail: arshyn.altybay@gmail.com