International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 6, December 2024, pp. 6734~6742
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i6.pp6734-6742

Assessing risk factors for heart disease using machine learning methods

Natalya Maxutova[1], Jamalbek Tussupov[1], Kamilya Kedelbayeva[2], Assemgul Tynykulova[3], Zulfiya Balabayeva[4], Zauresh Yersultanova[5], Zhainagul Khamitova[1], Kamila Zhunussova[6]
[1]Department of Information Systems, Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan
[2]Department of Cardiology, Asfendiyarov Kazakh National Medical University, Almaty, Republic of Kazakhstan
[3]Higher School of Information Technology and Engineering, Astana International University, Astana, Republic of Kazakhstan
[4]Medical Center Hospital of the President's Affairs Administration of the Republic of Kazakhstan, Astana, Kazakhstan
[5]Department of Physics, Mathematics, and Digital Technology, U. Sultangazin Pedagogical Institute Non-Profit Limited Company "Akhmet Baitursynuly Kostanay Regional University", Kostanay, Republic of Kazakhstan
[6]President of Kazakhstan National Association of Nutritionists and Health Coaches, Astana, Republic of Kazakhstan

Article Info

Article history:

Received May 11, 2024
Revised Jul 13, 2024
Accepted Aug 6, 2024

Keywords:

Biochemical indicators
Cardiovascular diseases
Machine learning technologies
Mean squared error
Pathology
Vanilla CNN
XGBoost

ABSTRACT

This paper examines various machine learning methods for assessing risk factors for cardiovascular diseases. To build predictive models, two approaches were used: the extreme gradient boosting (XGBoost) algorithm and a convolutional neural network (CNN). The focus is on analyzing the performance of each model in classification and regression tasks, as well as their ability to identify key biomarkers and risk factors such as cholesterol, ferritin, homocysteine and aspartate aminotransferase (AST) levels. XGBoost parameters have been optimized for working with tabular data, demonstrating high accuracy in risk prediction. The CNN model, despite the initial reduction in error on the training set, showed signs of overfitting when analyzing validation data. Performance evaluation using the metrics of mean squared error (MSE), coefficient of determination (R^2), Akaike information criterion (AIC), and Bayesian information criterion (BIC) revealed significant differences between the models. The study results confirm the effectiveness of XGBoost in analyzing tabular data and summarizing risk factor knowledge, while the CNN model needs further optimization to handle sparse data. The work demonstrates the importance of choosing the right model architecture and training parameters to ensure reliable diagnosis of cardiovascular diseases.

Corresponding Author:

Jamalbek Tussupov
Department of Information Systems, Faculty of Information Technologies, L.N. Gumilyov Eurasian National University
Astana, Republic of Kazakhstan
Email: tussupov@mail.ru

1. INTRODUCTION

In the modern world, cardiovascular diseases (CVD) [1]–[3] remain the leading cause of death and disability among the population throughout the world, which emphasizes the need to develop more effective methods for their prevention and treatment [4]–[6]. Expanding the capabilities of medical diagnostics and prognosis through the introduction of machine learning technologies [7]–[10] opens up new prospects for early diagnosis and assessment of CVD risk [11]. The main objective of this article is to analyze various risk factors for heart disease using machine learning techniques, which can help in developing predictive models

to estimate the likelihood of developing CVD in individual patients. Such models are particularly valuable in clinical practice because they help optimize prevention and intervention strategies aimed at reducing the risk of developing diseases.

Machine learning methods, including both classic algorithms such as logistic regression [12]–[14] and random forests [15]–[17] and more modern approaches such as deep learning, offer significant advantages for processing and analyzing large volumes of medical data. In [18]–[20], these methods are capable of identifying non-obvious relationships and patterns in data that may not be available with traditional statistical approaches. However, despite promising results, the application of machine learning in medicine faces several challenges. These include issues of model interpretability, ensuring sufficient accuracy and reliability of predictions, and the need to adapt models to clinical needs. The article discusses these and other problems and also suggests possible solutions to them.

From a data point of view, the use of a wide range of clinical, laboratory, and demographic indicators allows for the creation of increasingly accurate and tailored risk models. Along with traditional risk factors such as age, gender, medical history, cholesterol levels, and blood pressure, more specific biomarkers and genetic data are also included in the analysis, helping to improve the predictive power of the models. In their work, Yilmaz and Yağin [21] consider the problem of diagnosing heart diseases using machine learning methods. They note that cardiovascular disease remains a significant challenge to healthcare systems around the world. In this regard, they conducted a study in which they compared the performance of three different models: random forest (RF), logistic regression (LR), and support vector machine (SVM) for classifying coronary heart diseases. Bhatt *et al.* [22] discusses the importance of diagnosing and prognosticating cardiovascular diseases to ensure correct classification and appropriate treatment for patients. The authors propose the use of the k-mode clustering method with Huang's initial approximation to improve classification accuracy using machine learning. Using models such as RF, decision tree (DT), multilayer perceptron (MP), and extreme gradient boosting (XGBoost) on a real Kaggle dataset, we achieved high accuracy rates exceeding 87%, and the best performance was demonstrated by Multilayer perceptron with cross-validation, achieving an accuracy of 87.28%.

This study [23] focuses on the development of a wearable biomedical prototype for predicting the presence of cardiovascular diseases. The results of this study will be especially important in countries where the doctor-to-patient ratio is extremely low, as wearable technology can be used to monitor patient parameters anywhere, not limited to the hospital environment. The goal is to predict the possibility of cardiovascular disease using machine learning algorithms based on the analysis of electrocardiograms. The proposed prototype with the random forest algorithm showed a prediction efficiency of 88%, which confirms its potential as an aid for older people. Li *et al.* [24] proposes an efficient and accurate system for diagnosing heart diseases based on machine learning techniques. The work uses various classification algorithms such as support vector machine, logistic regression, artificial neural networks, nearest neighbors, naive Bayes classifier, and decision tree, as well as feature selection algorithms to improve the classification accuracy and reduce the execution time of the system. A new feature selection algorithm based on fast conditional mutual information (FCMIM) has been proposed and shown to be effective when combined with a support vector machine classifier to create a high-level intelligent system for heart disease identification. Experimental results confirm that the proposed diagnostic system achieves good accuracy compared with previously proposed methods and can be easily implemented in healthcare to identify heart diseases. Ansari *et al.* [25] addresses the growing problem of heart disease, which is becoming one of the deadliest diseases nowadays. Researchers describe the need to develop early diagnosis of heart disease with greater accuracy, given that about 12 million people die from it every year. The paper applies various machine learning algorithms and proposes a modified algorithm using logistic regression with principal component analysis to predict heart disease with greater accuracy based on various characteristic features such as age, blood pressure, chest pain, blood cholesterol, and others.

In summary, this review highlights the importance of integrating machine learning techniques into cardiology practice and represents a contribution to the development of personalized medicine capable of providing accurate and timely solutions for cardiac health management. In our scientific work, we differ from other studies in that we place special emphasis on analyzing the relationships between biochemical parameters and their impact on health. We consider a wide range of factors, including age, cholesterol, ferritin, aspartate aminotransferase (AST), homocysteine, and a target variable that reflects the presence or absence of cardiovascular disease risk. This multi-pronged approach allows deeper correlations between these indicators to be identified and explored, providing a unique opportunity to understand and predict potential risks. Our focus on exploring these connections in detail makes our work especially useful for shaping prevention and treatment strategies.

2. METHOD

This study examined two different approaches to building and training predictive models: one based on the XGBoost algorithm and the other using a convolutional neural network (CNN) on the TensorFlow platform. The main goal was to compare their performance on classification and regression problems using data imported from Excel files. The following parameters were set for the XGBoost model. $'reg: squared\ error'$ was chosen as the loss function to minimize the squared error in regression problems. The proportion of features randomly selected for each tree ($colsample\_bytree$) was set to 0.3, which helps reduce overfitting and improve the generalization ability of the model. The learning rate was set to 0.1, ensuring smooth and controlled learning. The depth of each tree (max_depth) was limited to five levels to control model complexity, and the number of trees ($n\_estimators$) was set to 100 to achieve sufficient depth of data analysis. During the training process, both training and validation data sets were used, which made it possible to track overfitting using the root mean square error (RMSE) metric as shown in Figure 1.

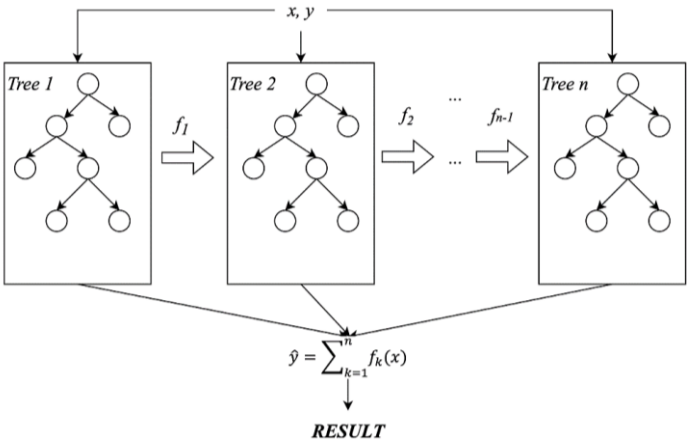


Figure 1. XGBoost model architecture

The parameters of the CNN model were chosen as follows. The architecture included convolutional layers, subsampling layers (MaxPooling), flatten layer (Flatten), and fully connected layers (Dense), which are the standard choice for processing images or structured data. ReLU was used as an activation function on the convolutional and fully connected layers to prevent shallow gradients and improve the learning process, and a sigmoid function was used on the output layer, suitable for binary classification problems. Adam was chosen as the optimizer due to its effectiveness in adjusting the learning rate based on the first and second moments of gradients. Ten epochs with validation data were used to train the model, which made it possible to track overtraining and evaluate the progress of the model as shown in Figure 2.

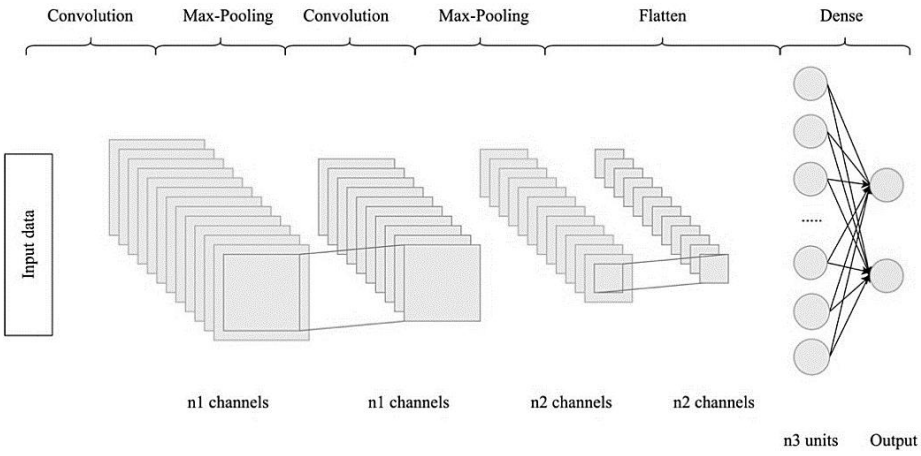


Figure 2. Vanilla CNN model architecture

The performance evaluation process for both models used MSE and R^2 metrics to analyze prediction accuracy and model fit to the data. In addition, the AIC and BIC metrics were used to compare the quality of models taking into account their complexity. The use of a variety of metrics allowed for an in-depth assessment of model performance, which made it possible to identify areas for further optimization and select parameters that are most suitable for solving research problems. This approach provided a more complete understanding of the effectiveness of the models and their applicability to specific problems, facilitating the development of methodology and improving research results.

3. RESULTS AND DISCUSSION

In this work, we used a dataset that contains 10,000 records, divided into six columns. The data represents various biomarkers and demographic information that can be used to analyze risk factors for heart disease. The first column indicates the patient's gender, with indicators 1 for men and 2 for women. The second column includes information about the patient's age in years. Cholesterol data presented in the third column indicates the level of this indicator in the blood, measured in mmol/l. The fourth column contains data on the level of ferritin in the blood, measured in ng/ml. This indicator serves as a marker of inflammation and is potentially associated with the risks of cardiovascular disease. The fifth column includes the blood level of aspartate aminotransferase (AST), measured in U/L, which is an enzyme involved in amino acid metabolism. The last column contains information about the level of homocysteine in the blood, measured in µmol/l. Homocysteine is an amino acid associated with the risk of cardiovascular disease. This dataset may be particularly useful for building and testing machine learning models that predict heart disease risk based on biomarkers and demographic characteristics. By analyzing these parameters, algorithms can be developed that can identify patients at increased risk at an early stage. This will allow us to focus our efforts on preventing diseases and improving the overall health of the population.

To establish predictive models of cardiovascular disease, it is critical to determine threshold values for key biomarkers. The following threshold levels were used in this work: ferritin above 100 ng/ml, homocysteine above 15 µmol/l, AST (aspartate aminotransferase) above 50 U/l, cholesterol above 8 mmol/l and glucose above 7 mmol/l. These thresholds were established based on literature and clinical guidelines that correlate with an increased risk of cardiovascular disease. Exceeding these thresholds allows patients to be classified according to their risk of developing cardiovascular disease using a binary system. “0” indicates normal (low risk) and “1” indicates risk (high risk). This approach not only simplifies the data analysis process but also increases its visibility and practical applicability in clinical settings. Data binarization allows the use of various statistical and machine methods to identify the most significant risk factors and forecasting models. This may include logistic regression, decision trees, and ensemble methods that provide not only classification accuracy but also the ability to interpret models in terms of the contribution of each biomarker to disease risk. Adjusting thresholds based on analysis of real patient data helps refine prediction models, making them more sensitive to early signs of pathology. This is especially important in personalized medicine, where individual patient characteristics may influence risks and require adjustments to standard thresholds. Thus, the application of established thresholds in machine learning models is a key step towards more accurate and personalized diagnosis of cardiovascular diseases.

Correlation analysis in Figure 3 is a fundamental statistical tool that allows one to examine the extent and nature of relationships between various variables. In the study of risk factors for cardiovascular diseases, special attention is paid to the analysis of connections between biochemical parameters and their impact on health status. This analysis examines age, cholesterol, ferritin, AST, homocysteine, and a target variable that reflects the presence or absence of cardiovascular disease risk.

The significant positive correlation between homocysteine and the target variable (0.48) indicates a possible direct effect of homocysteine levels on the risk of cardiovascular disease. High levels of this amino acid are associated with damage to the vascular wall and may contribute to the development of atherosclerosis. The positive correlation between ACT and the target variable (0.41) is also significant. AST is an enzyme found inside cells, but elevated levels in the blood can indicate cell death, indicating possible damage to the heart muscle or liver. Moderate correlations between ferritin and the target variable (0.15) and between cholesterol and the target variable (0.13) suggest that these biomarkers may be associated with the risk of developing cardiovascular diseases. However, high ferritin levels, although associated with risk, are not an independent predictor and likely reflect a general inflammatory state of the body. The cholesterol correlation highlights its impact on the cardiovascular system, but the relatively low coefficient indicates the importance of not only total cholesterol levels but also the ratio of its fractions.

The correlation between age and the target variable (0.11) was found to be low, which may be due to the variety of biological and social factors influencing the health of individuals. The low to moderate correlations between ferritin and AST (0.16) and between ferritin and cholesterol (0.20) demonstrate the complexity of the interactions in the biological pathways involved in the development of cardiovascular

disease. Correlation analysis allows us to identify potential biomarkers for further study and use in clinical practice. However, it is worth remembering that correlation does not mean causation. Additional research is needed to establish this, including prospective studies and clinical trials. Overall, this analysis contributes to a better understanding of disease mechanisms and can be used to develop personalized strategies for the prevention and treatment of cardiovascular diseases. Figure 4 shows the dynamics of changes in the RMSE of the XGBoost model during training by epoch for the training (solid line) and testing (dashed line) samples. RMSE is a key indicator that characterizes the average deviation of predicted values from actual values. The observed decrease in RMSE during training indicates an improvement in model quality.

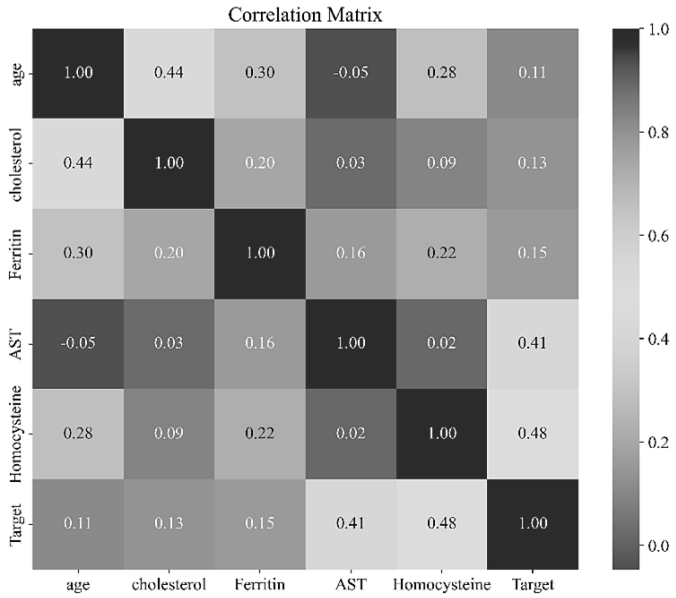


Figure 3. Correlation analysis of biochemical parameters

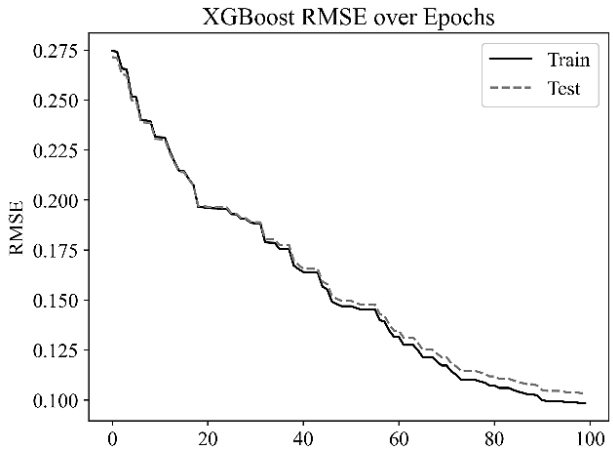


Figure 4. Dynamics of changes in the RMSE of the XGBoost model dynamics of changes in the RMSE of the XGBoost model

The learning curve plot shows a gradual and consistent decrease in RMSE on both the training and test sets, indicating good learning and generalization ability of the model. In the initial stages, the decrease in RMSE occurs quickly as the model actively adapts to the structure of the data. Subsequently, the curves gradually converge, indicating the stabilization of the learning process and the achievement of an optimal balance between bias and dispersion. Model performance metrics highlight its accuracy and predictive power. The mean square error (MSE) value is 0.0107, which indicates that the model is highly accurate as

MSE measures the mean square error of the predictions relative to the actual values. A low MSE value means that the deviations of the predictions from the actual data are minimal. The coefficient of determination (R^2) is 0.865, which indicates that the model effectively describes the dependence of the data and has high predictive power. The mean absolute error (MAE) is 0.0474, reflecting the average absolute value of the errors between the predicted and actual values. A low MAE value emphasizes that the model makes predictions with minimal deviations.

The Akaike information criterion (AIC) is 2.7625 and the Bayesian information criterion (BIC) is 36.347. These evaluation criteria take into account not only the quality of the fit but also the number of parameters in the model to prevent unnecessary complexity. Low AIC and BIC values indicate that the model achieves a good balance between accuracy and complexity. The training results of the XGBoost model indicate its high efficiency and ability to adequately predict, confirmed by a decrease in RMSE, high R^2 values, and low error rates. This makes XGBoost a powerful tool for solving regression problems, especially in the context of large and complex data. Such a model can be applied in various application areas where reliable forecasting of numerical variables is required, including finance, economics, and natural sciences. Figure 5 illustrates the dynamics of the loss (loss) of the convolutional neural network model (vanilla CNN) during training by epoch for the training (solid line) and validation (dashed line) samples. The training graph shows how the loss function decreases for the training set and stabilizes at a certain level for the validation set.

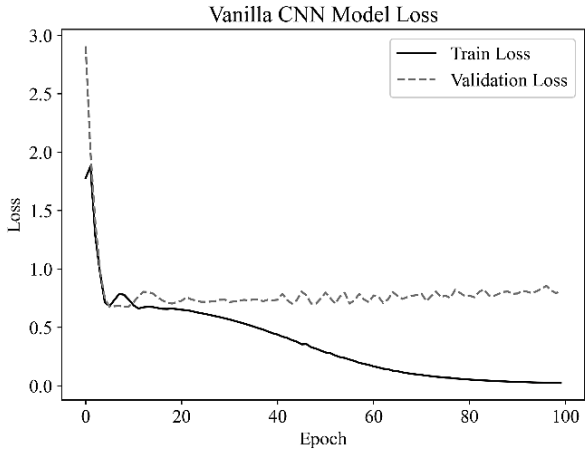


Figure 5. Dynamics of loss (loss) of a convolutional neural network model (vanilla CNN)

The learning curve graph starts with a high level of loss, which decreases quickly, especially in the initial stages of learning. This indicates that the model is effective at extracting features from the data and adapting to the task. However, after an initial period of decline, the validation loss curve exhibits volatility and does not decrease further, indicating the possible onset of overfitting. At the same time, the losses on the training set continue to decrease, while the losses on the validation set stabilize and even increase slightly. This highlights the performance gap between training and validation data. The mean square error (MSE) is 0.45, which is higher than expected for a well-performing regression model. This indicates significant deviations in forecasts from actual values, indicating problems with model accuracy. The coefficient of determination (R^2) is -0.875, which is a negative value. This is unusual because it indicates that the model performs worse than the baseline model which simply predicts the mean of the dependent variable. This result may be due to problems in the model architecture or data quality. The AIC is 1562.0 and BIC is 2342.65. Although these metrics are less common in evaluating machine learning models, they provide information about the relationship between the number of model parameters and its ability to adequately describe the data. High AIC and BIC values in this case confirm the presence of overfitting problems or suboptimal model specification. Training vanilla CNN showed that, despite the initial reduction in the loss function on the training set, the model faces difficulties in generalizing the results to new data. A negative R^2 value and a high level of loss on the validation set indicate problems in the architecture or model training process that require revision and optimization. It is recommended to conduct additional data analysis, revise the network architecture, increase the amount of training data, or use regularization methods to prevent overfitting. This approach will help improve the generalization ability of the model and its performance on new data.

When comparing the performance of the models, several key differences can be identified. The XGBoost model has a mean square error (MSE) of 0.0107, indicating a very small prediction error. The coefficient of determination (R^2) is 0.865, indicating that the model fit the data well, explaining 86.5% of the variance in the target variable. The AIC and BIC of this model are significantly lower compared to vanilla CNN, indicating a good trade-off between model complexity and goodness of fit. On the other hand, the vanilla CNN model performs significantly worse. The mean square error (MSE) is 0.45, which is much higher compared to XGBoost, indicating less accurate predictions. The coefficient of determination (R^2) is -0.875, reflecting an extremely poor result. This means that the model is not able to accurately predict and does not even reach the level of a basic mean prediction model. The AIC and BIC information criteria are also much higher, indicating overfitting or inadequate specification of the model. As a result, XGBoost demonstrates significantly better performance across all key metrics. Due to its high accuracy and generalization ability, this model outperforms vanilla CNN, which suffers from overfitting and an inability to adequately generalize data, which is reflected in high validation loss and negative R^2 value.

4. CONCLUSION

In conclusion, research on cardiovascular disease risk factor assessment using machine learning techniques is important in selecting the right algorithms and parameters to ensure accurate prediction. A comparison of the XGBoost and CNN models revealed clear advantages of XGBoost when working with tabular data such as biomarkers and demographic characteristics. This model has demonstrated high predictive accuracy due to its ability to efficiently model complex nonlinear relationships and its robustness to sparse data. On the other hand, the CNN model, despite its potential advantages in analyzing spatially correlated data, showed signs of overfitting and underperformance in biomarker classification tasks. This indicates the need for further optimization of the architecture and training algorithm, as well as the use of regularization methods and increasing the size of the training set.

The use of MSE metrics, coefficient of determination (R^2), AIC, and BIC provided in-depth analysis and comparison of models, revealing their strengths and weaknesses. The XGBoost approach has demonstrated robust prediction ability by effectively balancing bias and variance, while the CNN model requires adjustment and parameter optimization. These findings highlight the importance of proper choice of model architecture and training parameters in cardiovascular risk prediction tasks. In the future, research should continue in the direction of personalized medicine, optimizing models for identifying risk factors and developing strategies for early diagnosis and prevention of cardiovascular diseases.

REFERENCES

[1] A. C. O'Kelly *et al.*, "Pregnancy and reproductive risk factors for cardiovascular disease in women," *Circulation Research*, vol. 130, no. 4, pp. 652–672, Feb. 2022, doi: 10.1161/circresaha.121.319895.

[2] J. L. Silveira Rossi, S. M. Barbalho, R. de Araujo, M. D. Bechara, K. P. Sloan, and L. A. Sloan, "Metabolic syndrome and cardiovascular diseases: going beyond traditional risk factors," *Diabetes/Metabolism Research and Reviews*, vol. 38, no. 3, Oct. 2021, doi: 10.1002/dmrr.3502.

[3] J. Hu *et al.*, "Social isolation, social support, loneliness and cardiovascular disease risk factors: a cross-sectional study among older adults," *International Journal of Geriatric Psychiatry*, vol. 36, no. 11, pp. 1795–1809, Jul. 2021, doi: 10.1002/gps.5601.

[4] A. Shaito *et al.*, "Herbal medicine for cardiovascular diseases: efficacy, mechanisms, and safety," *Frontiers in Pharmacology*, vol. 11, Apr. 2020, doi: 10.3389/fphar.2020.00422.

[5] Y. Deng *et al.*, "Application of the nano-drug delivery system in treatment of cardiovascular diseases," *Frontiers in Bioengineering and Biotechnology*, vol. 7, Jan. 2020, doi: 10.3389/fbioe.2019.00489.

[6] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: an effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/access.2020.3010511.

[7] R. TR, U. K. Lilhore, P. M, S. Simaiya, A. Kaur, and M. Hamdi, "Predictive analysis of heart diseases with machine learning approaches," *Malaysian Journal of Computer Science*, pp. 132–148, Mar. 2022, doi: 10.22452/mjcs.sp2022no1.10.

[8] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques," *ICT Express*, vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.icte.2021.08.021.

[9] S. Subramani *et al.*, "Cardiovascular diseases prediction by machine learning incorporation with deep learning," *Frontiers in Medicine*, vol. 10, Apr. 2023, doi: 10.3389/fmed.2023.1150933.

[10] O. Taylan, A. Alkabaa, H. Alqabbaa, E. Pamukçu, and V. Leiva, "Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods," *Biology*, vol. 12, no. 1, Jan. 2023, doi: 10.3390/biology12010117.

[11] Y. Sethi *et al.*, "Precision medicine and the future of cardiovascular diseases: A clinically oriented comprehensive review," *Journal of Clinical Medicine*, vol. 12, no. 5, Feb. 2023, doi: 10.3390/jcm12051799.

[12] J. Tussupov *et al.*, "Analysis of formal concepts for verification of pests and diseases of crops using machine learning methods," *IEEE Access*, vol. 12, pp. 19902–19910, 2024, doi: 10.1109/access.2024.3361046.

[13] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2022.100130.

[14] B. Salman Shukur and M. Mohsin Mijwil, "Involving machine learning techniques in heart disease diagnosis: a performance analysis," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 2177–2185, Apr. 2023, doi:

10.11591/ijece.v13i2.pp2177-2185.
[15] S. Saeedbakhsh, M. Sattari, M. Mohammadi, J. Najafian, and F. Mohammadi, "Diagnosis of coronary artery disease based on machine learning algorithms support vector machine, artificial neural network, and random forest," *Advanced Biomedical Research*, vol. 12, no. 1, Feb. 2023, doi: 10.4103/abr.abr_383_21.
[16] E. Dritsas and M. Trigka, "Efficient data-driven machine learning models for cardiovascular diseases risk prediction," *Sensors*, vol. 23, no. 3, Jan. 2023, doi: 10.3390/s23031161.
[17] S. Asif, Y. Wenhui, Q. ul Ain, Y. Yueyang, and S. Jinhai, "Improving the accuracy of diagnosing and predicting coronary heart disease using ensemble method and feature selection techniques," *Cluster Computing*, vol. 27, no. 2, pp. 1927–1946, Jun. 2023, doi: 10.1007/s10586-023-04062-2.
[18] A. Naizagarayeva *et al.*, "Detection of heart pathology using deep learning methods," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6673–6680, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6673-6680.
[19] G. Abdikerimova *et al.*, "Detection of lung pathology using the fractal method," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6778–6786, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6778-6786.
[20] G. Abdikerimova *et al.*, "Detection of chest pathologies using autocorrelation functions," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4526–4534, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4526-4534.
[21] R. Yilmaz and F. H. Yağin, "Early detection of coronary heart disease based on machine learning methods," *Medical Records*, vol. 4, no. 1, pp. 1–6, Jan. 2022, doi: 10.37990/medr.1011924.
[22] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, Feb. 2023, doi: 10.3390/a16020088.
[23] S. V Jansi Rani, K. R. S. Chandran, A. Ranganathan, M. Chandrasekharan, B. Janani, and G. Deepsheka, "Smart wearable model for predicting heart disease using machine learning: Wearable to predict heart risk," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 9, pp. 4321–4332, Mar. 2022, doi: 10.1007/s12652-022-03823-y.
[24] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/access.2020.3001149.
[25] M. F. Ansari, B. Alankar, and H. Kaur, "A prediction of heart disease using machine learning algorithms," in *Image Processing and Capsule Networks*, Springer International Publishing, 2020, pp. 497–504.

BIOGRAPHIES OF AUTHORS

Natalya Maxutova lecturer at the Department of Informational systems of the Eurasian National University named after L.N. Gumilyov. In 2013, she received a bachelor's degree in information system and in 2015, she received a master's degree in computer science from the Eurasian National University named after L.N. Gumilyov. Since 2023, doctoral student of the specialty "8D06103"-Informational systems of the Eurasian National University named after L.N. Gumilyov. Her research interest's information and computer technologies and systems, artificial intelligent, machine learning. She can be contacted at email: natalya.maxutova@gmail.com.

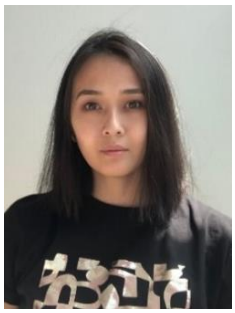
Jamalbek Tussupov graduated from Karaganda State University at the Faculty of Mathematics with a degree in mathematics (1974-1979). He continued his education in graduate school at KarSU and was sent to the Institute of Mathematics named after. S. L. Soboleva, Siberian Branch of the USSR Academy of Sciences, Novosibirsk (1983-1986). He defended his PhD thesis in mathematics in 1990 at Novosibirsk State University. He defended his doctoral dissertation in 2007 at Al-Farabi Kazakh National University on the topic "Problems of definability and algorithmic complexity of relations over algebraic structures." He completed a scientific internship at the University of Notre Damme in Indiana, USA, in 2011, being a scholarship recipient of the Bolashak Presidential Program for this period. In 2011, he was awarded the title of "Professor" by the Ministry of Education and Science of the Republic of Kazakhstan. He can be contacted at email: tussupov@mail.ru.

Kamilya Kedelbayeva In 2007 she graduated from the Kazakh National Medical University named after S.D. Asfendiyarov, specialty "general medicine", qualification "doctor". In 2014, she graduated residency, qualification "cardiologist, including children's". In 2022, she graduated from doctoral studies at the Kazakh National Medical University named after S.D. Asfendiyarov with a specialty 8D10103 - "Medicine". From 2014 to the present, he has been assistant, head of the educational part of the Department of Cardiology, KazNMU named after S.D. Asfendiyarov. She is the author of more than 40 works. Participant and winner in international, republican and city conferences. Based on the results of scientific work, he has 7 copyright certificates. Her research interests include at optimizing diagnostics, developing and evaluating cardiovascular disease prevention programs, as well as rehabilitation programs after myocardial infarction. Interest in developing data processing algorithms to improve the diagnosis, treatment and management of cardiovascular diseases. She can be contacted at email: kamilla-km@mail.ru.

Assemgul Tynykulova has teaching experience at leading universities in Kazakhstan. The main research interests of the master Tynykulova A.S. are optimization methods, decision-making, expert systems. Over the past 5 years, he has published 10 scientific articles, including 1 in a journal included in the Scopus database. Currently, Master of Science Tynykulova A.S. works as a senior lecturer at the Higher School of Information Technology and Engineering Astana International University, Astana. She can be contacted at email: asem_110981@mail.ru.

Zulfiya Balabayeva graduated in 2020 from NAO "Astana Medical University" with a bachelor's degree in healthcare, specialty "General medicine". From 2020 to 2022, she completed her training as a general practitioner at the same university. Currently, she is a resident in adult and pediatric cardiology at the Medical Center Hospital of The President's Affairs Administration of the Republic of Kazakhstan. She can be contacted at email: S.a.b-1994@mail.ru.

Zauresh Yersultanova in 1989 she graduated from the Kazakh State University named after S.M. Kirov, qualifications "mechanical", "applied mathematician". In 2001, she graduated from the graduate school at Al-Farabi Kazakh National University with a degree in 02.05.18 theory of mechanisms and machines. From 2001 to the present, she is a candidate of technical sciences (Doctor of Philosophy-Ph.D.) in the specialty 02.05.18 Theory of mechanisms and machines of the Kostanay Regional University named after Akhmet Baitursynuly. She is the author of more than 70 works. Her research interests include the development of information intelligent systems, mobile applications in the field of education, information and communication technologies. She can be contacted at email: ersul_67@mail.ru.



Zhainagul Khamitova lecturer at the Department of Informational systems of the Eurasian National University named after L.N. Gumilyov. In 2013, she received a bachelor's degree in information system and in 2015, she received a master's degree in computer science from the Eurasian National University named after L.N. Gumilyov. Her research interest's information and computer technologies and systems, AI, machine learning. She can be contacted at email: khamitova_zhzh@mail.ru.

Kamila Zhunussova in 2014, she graduated from the Queen Mary University of London with a degree in environmental science. In 2017, she received a master's degree in engineering management at Nazarbayev University. In 2020, finished courses at University of Educational Medicine (UOM) as a Health coach. In 2024 established National Association of Nutritionists and Health coaches and was chosen as a president. Her research interests include healthy food, nutrition, and supplements. She can be contacted at email: k.zhunussova1991@gmail.com.