Metaheuristic Optimization Review

♦ ISSN: 2692-4048 Vol: 02, No. 01, 2024 ♦ PP. 17–20
Article DOI: https://doi.org/10.54216/MOR.020203

REVIEW ARTICLE

A Review of Machine Learning in Predicting Heart Disease Risk Based on Medical Data

Safa S. Abdul-Jabbar[1,*]

[1] Computer Science Department, College of Science for Women, University of Baghdad, Baghdad, Iraq

Emails: Safa.s@csw.uobaghdad.edu.iq

Received: June 05, 2024 Revised: September 19, 2024 Accepted: December 10, 2024 ⋆ Corresponding author

ABSTRACT

Heart diseases go on to be the primary cause of mortality all over the world and hence call for accurate and efficient diagnostic tools. Traditional diagnostics are not scalable and precise in analyzing large and complex datasets generated in healthcare. Machine learning has come as a revolutionary solution in the form of advanced prediction models in the diagnosis and risk assessment of heart diseases. This review presents machine-learning techniques such as Random Forest, Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, and hybrid models containing deep learning versions such as CNN and LSTM. These techniques use multi-source data from Cleveland, Statlog, and UCI repositories and combine feature selection methods with data preprocessing techniques to achieve improved accuracy, reliability, and scalability. Ensemble methods such as majority voting and boosting improve model robustness, while SMOTE addresses imbalanced data scenarios. Despite these developments, specific challenges remain, especially model interpretability, data diversity, and clinical integration. This review discusses progress, challenges, and future avenues in using machine learning for heart disease prediction, emphasizing explainable AI models, diverse datasets, and real-world validation for clinical applications and improved global healthcare outcomes.

Keywords: Machine learning ▪ Heart disease prediction ▪ Random Forest ▪ Feature selection ▪ Hybrid models ▪ Explainable AI

1. INTRODUCTION

Heart disease ranks among the prime causes of death worldwide, while factors such as high blood pressure, diabetes, high cholesterol, limited exercise, and lifestyle change increase cardiovascular risk. Machine learning has emerged as an important tool for healthcare because it supports efficient predictive diagnosis through the analysis of large-scale and complex medical datasets. This review discusses progress, applications, and challenges in using machine learning for heart disease prediction, with attention to algorithms, datasets, and future directions for improving cardiovascular healthcare.

1.1 Machine Learning for Heart Disease Prediction—A Comprehensive Overview

Cardiovascular diseases remain leading causes of morbidity and mortality globally, claiming millions of lives every year and placing substantial burdens on healthcare systems. Risk factors such as hypertension, diabetes, high cholesterol, and physical inactivity require close attention because these conditions are progressive and can lead to severe complications if not detected and treated early. Even with tremendous medical discovery, conventional diagnostic techniques cannot always meet the demand for accuracy and speed in identifying asymptomatic patients or assessing patient-specific risk. This situation has encouraged the incorporation of ma-

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

chine learning in cardiovascular diagnosis because it expands opportunities for early identification, correct treatment, and favorable prognosis [1].

1.2 Machine Learning Applications in Healthcare

Machine learning (ML), an application of artificial intelligence, has evolved as a critical tool in healthcare data analysis. Predictive algorithms for heart disease use heterogeneous datasets that include demographic factors, medical history, diagnostic tests, and lifestyle information. These algorithms identify relationships in data that may be difficult for clinicians to discern alone. For example, ML systems can process age, blood pressure, cholesterol, glucose levels, and other risk factors simultaneously and provide likely results in real time. Compared with legacy diagnostic measures, ML systems are more scalable and can handle large healthcare datasets. This scalability is crucial as the amount of medical data produced worldwide continues to increase [2].

1.3 Key Machine Learning Techniques in Heart Disease Prediction

Contemporary research on heart disease prediction uses supervised algorithms such as Support Vector Machines (SVM), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Decision Trees (DT). These models perform well for classification problems and can separate patients into risk categories. Random Forest often shows strong performance because it combines multiple decision trees and reduces overfitting. SVM is useful for high-dimensional feature spaces, while Logistic Regression provides interpretable baseline predictions. KNN and Decision Trees remain common because they are simple and easy to implement.

Recent studies also combine classical ML with deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Such hybrid systems are useful when medical imaging, time-series measurements, or multimodal health records are included. Ensemble learning, boosting, and majority voting are frequently used to enhance robustness, while feature selection and preprocessing help reduce noise, address irrelevant variables, and improve predictive reliability [3, 4, 5].

Despite these advances, challenges remain. Many high-performing models, especially deep neural networks, behave like black boxes and are difficult for clinicians to interpret. Dataset imbalance, limited diversity, and inconsistent data quality can also produce biased or unreliable predictions. These limitations highlight the need for explainable AI, diverse clinical datasets, and validation in real healthcare environments.

2. LITERATURE REVIEW

Cardiac complications continue to be among the key causes of morbidity and mortality throughout the world. Recently, research and healthcare industries have incorporated machine learning in disease diagnosis and prognosis, providing more accurate and less time-consuming approaches than traditional models. This section reviews studies that use machine-learning approaches to forecast and categorize heart disease based on different data types and models, focusing on model performance, accuracy, and the importance of early detection for patient recovery.

Medical diligence generates substantial volumes of data, with cardiovascular diseases emerging as a leading cause of mortality. Early prediction is critical, and heart disease datasets stored in large databases provide valuable resources for model development [6]. Comparative studies of machine learning algorithms show that models such as Naïve Bayes, Decision Trees, KNN, SVM, Random Forest, and Logistic Regression can predict heart disease risk using clinical attributes. Some works emphasize feature selection techniques and data balancing methods such as SMOTE to improve results when class distributions are unequal [2, 4].

Other studies evaluate advanced models and hybrid approaches. Deep learning models, including CNN and LSTM, have been integrated with KNN and XGBoost, while ensemble majority voting has been used to determine final output classes [7]. Machine learning has also been used to predict mortality in elderly patients with heart failure and hypertension; Random Forest achieved strong results and helped identify influential clinical factors [8]. In primary care, Random Forest models for acute myocardial infarction (AMI) and ischemic heart disease (IHD) outperformed standard linear prediction tools across several metrics [9].

The reviewed literature reveals strong possibilities for machine learning in heart disease prediction and diagnostic reliability. Random Forests, Support Vector Machines, and ensemble methods appear as reliable performers across datasets. However, data imbalance, feature selection, and interpretability remain important limitations. Future research should overcome these challenges and implement predictive models in clinical practice.

3. DISCUSSION

Machine learning has changed heart disease prediction by introducing sophisticated tools for accuracy, efficiency, and early diagnosis. Research using different datasets and algorithms such as Random Forest, SVM, and hybrid models indicates substantial potential for cardiovascular healthcare. The discussion below presents key contributions, dataset usage, scalability, clinical integration challenges, and future perspectives.

3.1 Machine Learning Key Contributions to Heart Disease Prediction

Machine learning models such as Random Forest, SVM, Logistic Regression, and ensemble models have improved accuracy and reliability across many studies. Feature selection methods, including chi-square and ANOVA, and preprocessing approaches such as Min–Max normalization improve model performance by emphasizing relevant attributes and standardizing data formats. Hybrid approaches that combine ML with CNN and LSTM models further improve prediction accuracy, especially for complex datasets with unstructured inputs. These models support early detection and assist clinicians in preparing personalized treatment plans [5, 3].

3.2 Dataset Usage and Algorithm Performance

The reviewed studies use several dataset sources, including Cleveland, Statlog, UCI, private, and local datasets, to diver-

Table 1. Summary of Literature Review

Ref.	Focus of Study	Key Algorithms	Key Outcomes
[6]	Heart disease prediction using machine-learning techniques.	Naïve Bayes, Decision Tree, KNN, Random Forest	Demonstrated high prediction accuracy for Random Forest in detecting heart disease.
[2]	Feature selection and classifiers for heart disease prediction with data imbalance handled by SMOTE.	Naïve Bayes, SVM, XGBoost, Random Forest	XGBoost achieved 97.57% accuracy, and SMOTE addressed imbalance effectively.
[4]	Prediction of heart disease using machine learning models.	Random Forest and other ML classifiers	Random Forest achieved strong predictive performance and reliability.
[8]	Prediction of in-hospital mortality in elderly patients with heart failure and hypertension.	Random Forest	Random Forest achieved an AUC of 0.850 and high accuracy in predicting mortality.
[7]	Integration of deep learning and machine learning for cardiovascular disease diagnosis.	CNN, LSTM, KNN, XGBoost	Ensemble methods demonstrated superior classification performance across datasets.
[9]	Prediction of AMI and IHD in primary care cardiovascular patients.	Random Forest, SMART algorithm	Random Forest models outperformed traditional SMART algorithms in predictive metrics.
[10]	Heart disease prediction and patient risk assessment.	Logistic Regression, Decision Tree	Improved risk assessment using logistic regression and decision trees.
[11]	Survey of machine learning algorithms for heart disease prediction.	Random Forest, SVM, Naïve Bayes	Random Forest consistently outperformed other models in prediction accuracy.
[12]	Comparative analysis of supervised classification algorithms using UCI datasets.	SVM, KNN, Logistic Regression, Random Forest	SVM demonstrated the highest performance under the reported conditions.
[13]	AI-based system using machine-learning algorithms for heart disease prediction.	Decision Tree, Logistic Regression, Naïve Bayes	Highlighted limitations of clinical diagnostic methods and improved detection accuracy.
[14]	Evaluation of Cleveland and Statlog datasets for heart disease prediction.	Random Forest, KNN, Logistic Regression	Random Forest outperformed other algorithms on Cleveland and Statlog datasets.
[5]	Prediction using different data classification techniques.	Random Forest, Gradient Boosting	Gradient Boosting and Random Forest showed high reliability in predictions.

sify model training and validation. Random Forest repeatedly performs well across studies, and ensemble techniques improve model predictions by reducing bias and strengthening classification. Data balancing techniques such as the Synthetic Minority Oversampling Technique (SMOTE) address imbalanced datasets, reducing prediction errors in minority classes [7, 8].

3.3 Complexity and Scalability of Data Resolution

Scalability is central as healthcare data becomes larger and more complex. Cross-validation techniques, including 5-fold and 10-fold validation, help evaluate generalization to future datasets. Feature selection and ensemble approaches reduce noise and focus on significant variables. By incorporating structured clinical records and unstructured data such as medical imaging, hybrid models increase the applicability of ML in real-world healthcare scenarios [12].

3.4 Implementation and Clinical Integration Challenges

Although progress is clear, several issues must be improved before machine learning becomes common in clinical practice. Complex deep learning networks are often described as black boxes, which reduces trust among clinicians and patients. Skewed outputs may result from imbalanced datasets, variable data quality, or limited availability of comprehensive and diverse datasets. Clinical integration also requires extensive evaluation, validation, and user-friendly interfaces that fit healthcare workflow.

3.5 Future Directions and Recommendations

Future research should focus on explainable artificial intelligence models with transparent decision-making processes so clinicians can understand and trust predictions. Model generalization can be enhanced by improving the diversity of training datasets and including underrepresented populations. Further diagnostic accuracy may also come from combining machine learning with domain knowledge. Translating such models into clinical workflows requires collaboration among researchers, technologists, and healthcare providers. Real-world trials are required to validate machine-learning models regarding performance and impact under actual healthcare conditions [6].

4. CONCLUSION

In recent years, machine learning has become a crucial technique for addressing heart ailments beyond traditional diagnosis. Ensemble methods, Logistic Regression, Random Forest, and Support Vector Machines have shown better precision and satisfactory scalability. These methods process large amounts of complex data and discover important patterns, thereby delivering practical insights for early detection and personalized treatment planning. Combining machine learning with deep learning methods such as CNN and LSTM can further improve predictive accuracy.

The incorporation of machine learning into cardiovascular healthcare has affected early diagnosis and patient risk assessment. Models trained on datasets such as Cleveland and UCI show flexibility across different populations and can support resource allocation and timely intervention for high-risk patients. Real-time patient data analysis positions machine learning as a foundation of predictive healthcare management, especially for chronic conditions such as heart disease.

Progress notwithstanding, several impediments continue to hamper effective clinical incorporation. Persistent problems include data imbalance, limited interpretability of complex models, and the need for diverse high-quality datasets. Clinicians may be reluctant to rely on black-box models during critical decision stages. Integration also requires extensive validation, user-friendly interfaces, and collaboration between technologists and healthcare providers.

Future research should establish explainable AI models that build trust among medical professionals. Broader training datasets comprising underrepresented populations would improve generalizability and fairness. Real-world validation through clinical trials is essential for reliability and efficacy. Better hybrid models and smooth integration into clinical workflows could reshape heart disease management and reduce the global burden of cardiovascular diseases.

REFERENCES

[1] V. V. Ramalingam, A. Dandapath, and M. Karthik Raja, "Heart disease prediction using machine learning techniques: A survey," *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 2.8 Special Issue 8, pp. 684–687, 2018.

[2] H. Arghandabi, "A comparative study of machine learning algorithms for the prediction of heart disease," *International Journal of Research in Applied Science and Engineering Technology*, vol. 8, no. 12, pp. 677–683, 2020.

[3] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Computer Science*, vol. 1, no. 6, 2020.

[4] B. Saju, V. Asha, A. Prasad, P. Harish Kumar, V. Rakesh, and A. P. Nirmala, "Heart disease prediction model using machine learning," in *International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 2022, pp. 723–729.

[5] H. F. El-Sofany, "Predicting heart diseases using machine learning and different data classification techniques," *IEEE Access*, vol. 12, pp. 106 146–106 160, 2024.

[6] M. M. A. Mary, "Heart disease prediction using machine learning techniques: A survey," *International Journal of Research in Applied Science and Engineering Technology*, vol. 8, no. 10, pp. 441–447, 2020.

[7] H. Sadr, A. Salari, M. T. Ashoobi, and M. Nazari, "Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models," *European Journal of Medical Research*, vol. 29, no. 1, p. 455, 2024.

[8] X. Liu *et al.*, "Machine learning for predicting in-hospital mortality in elderly patients with heart failure combined with hypertension: a multicenter retrospective study," *Cardiovascular Diabetology*, vol. 23, no. 1, pp. 1–13, 2024.

[9] N. Salet, A. Gökdemir, J. Preijde, C. H. van Heck, and F. Eijkenaar, "Using machine learning to predict acute myocardial infarction and ischemic heart disease in primary care cardiovascular patients," *PLoS One*, vol. 19, no. 7, p. e0307099, 2024.

[10] K. S. Jagadeesh and R. R, "Heart disease prediction using machine learning," *International Journal of Research in Applied Science and Engineering Technology*, vol. 10, no. 3, pp. 183–184, 2022.

[11] D. S. Bhadane, P. Bedadewar, S. Nalawade, S. Daphal, and S. Gaikwad, "Use of machine learning in heart disease prediction: A survey," *IJARCCE*, vol. 12, no. 4, 2023.

[12] I. Ufumaka, "Machine learning algorithms for heart disease prediction: A comparative analysis," in *New Visions in Science and Technology*, 2021, vol. 7, pp. 157–167.

[13] V. K. Sahu, T. Sahu, and R. G. Brajesh, "Heart disease prediction; a machine learning approach," *CSVTU International Journal of Biotechnology Bioinformatics and Biomedical*, vol. 7, no. 3, 2022.

[14] P. Agrawal, K. Soni, and M. K. Gupta, "A relative research of machine learning build heart disease forecast model," *SKIT Research Journal*, vol. 10, no. 2, pp. 1–7, 2020.