Traditional and Novel Predictive Models of the Heart Disease

Jiayi Liang[a]
Faculty of Art and Science, University of Toronto, Toronto, Ontario, Canada

Keywords: Heart Disease Prediction, Decision Trees, Logistic Regression, Support Vector Machines, Weighted Associative Rule Mining.

Abstract: Considering that the leading cause of mortality worldwide is still cardiovascular disease, prompt risk assessment might facilitate proactive treatment strategies that enhance patient outcomes and lessen the financial burden on healthcare systems. Predicting heart disease is therefore critical to reducing mortality, reducing complications, and improving patient outcomes through early intervention. Predictive algorithms can identify high-risk patients before symptoms appear, but traditional diagnostic techniques frequently miss early-stage disease. Machine learning techniques like K-nearest Neighbors (KNN) and Decision Trees (DT) have become more and more popular for the prediction of cardiac illness. The interpretability, accuracy, and computing efficiency of these models varies. Prediction accuracy has been further enhanced by recent developments like Weighted Associative Rule Mining (WARM) and ensemble learning approaches (XBoost, Adaboost, and random subspace classifiers). These approaches still have issues with generalization, overfitting, interpretability of the model, and computational complexity. In order to produce more precise, individualized, and interpretable forecasts, future advancements in cardiac disease prediction are probably going to concentrate on hybrid models, explainable artificial intelligence (XAI), and multimodal data integration. With the goal of improving heart disease risk assessment with AI-driven healthcare solutions, this paper examines both conventional and innovative predictive models, their limitations, and potential future paths.

1 INTRODUCTION

Cardiovascular disease (CVD) claims one life every 33 seconds, making it the world's biggest cause of mortality, according to the WHO (Centers for Disease Control and Prevention, 2024). Heart disease, which encompasses illnesses like coronary artery disease, heart failure, and arrhythmias, is thought to be the cause of 17.9 million deaths globally, with low and middle-income countries accounting for more than three-quarters of cardiovascular disease deaths (World Health Organization, 2021). About one-fifth of deaths in the US are caused by heart disease; in 2022, 702,880 deaths were reported. The most prevalent kind of heart disease among them is coronary heart disease. In addition to mortality, heart disease also brings a huge economic burden, with losses of approximately $252.2 billion from heart disease between 2019 and 2020, including medical services, medicines, and productivity losses due to death (Centers for Disease Control and Prevention, 2024). Reducing the morbidity and mortality linked to heart disease requires early detection and prevention. Accurate risk assessment and early identification of heart disease continue to be significant obstacles despite significant advancements in cardiovascular medicine. Timely lifestyle changes, medication management, and focused therapy can be made possible by identifying high-risk populations before severe symptoms appear. Nevertheless, conventional diagnostic techniques depend on clinical history, medical competence, and common biomarkers like electrocardiograms, stress tests, and clinical risk assessments. These indicators are frequently arbitrary, inconsistent, or have little predictive ability to identify diseases in their early stages. To solve these constraints, advanced predictive models are needed for better risk stratification and tailored medication, hence predictive models based on machine learning and artificial intelligence (AI) have become useful tools

[a] https://orcid.org/0009-0004-6893-9142

Liang, J.
Traditional and Novel Predictive Models of the Heart Disease.
DOI: 10.5220/0013690400004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 355-359
ISBN: 978-989-758-765-8

in healthcare. However, constructing reliable, interpretable, and efficient prediction models entails overcoming hurdles such as data quality, model generalizability, computing limits, and clinical uptake. The upcoming sections will discuss four prevalent pathological conditions of heart disease along with the key risk factors influencing its development. The focus of this article is on heart disease prediction. It will discuss the four most frequently used prediction techniques in conventional models, as well as new models that combine two conventional models with other advanced technologies. In addition, this article will also talk about the shortcomings and limitations of current technologies, as well as prospects for the development of future prediction models.

2 PATHOGENESIS OF HEART DISEASE

2.1 Types of Heart Disease

Coronary artery disease (CAD), often known as coronary heart disease, is one of the four most common types of heart disease. It reduces the flow of oxygen-rich blood to the heart muscle by affecting the coronary arteries, which are the primary blood vessels that supply the heart with blood. A complete blockage of blood flow can lead to a heart attack, while atherosclerosis occurs due to the buildup of fat, cholesterol, and other substances inside and along the artery walls. Breathlessness and chest pain are among the symptoms of this kind of illness, which typically lasts for years (Mayo Foundation for Medical Education and Research, 2024). The second type is called the Heart valve disease. It is a disorder where one or more heart valves are not functioning correctly. In order to guarantee that blood flows through the heart in the proper direction, the heart has four valves. Occasionally, though, the valves fail to fully open or close, which can alter how blood moves from the heart to the body. The type and severity of heart valve disease, as well as the afflicted heart valve, determine the course of treatment. Surgery may be required to replace or repair the damaged heart valve (Mayo Foundation for Medical Education and Research, 2023). The third type of heart failure, also known as congestive heart failure, is characterized by poor heart muscle pumping, frequent blood clots, and fluid accumulation in the lungs, which results in dyspnea. High blood pressure and cardiac artery constriction are two heart disorders that may eventually make the heart too weak or inflexible to pump and fill blood adequately. Patients with severe heart failure symptoms may require a heart transplant or a device to help the heart pump blood (Mayo Foundation for Medical Education and Research, 2025). The final one, arrhythmia, is an abnormality in the heartbeat's rhythm or timing. An arrhythmia is characterized by an irregular heartbeat or a heartbeat that is too rapid or too sluggish. While some arrhythmias are not life-threatening, others can cause cardiac failure, fainting, or even unexpected death.

2.2 Risk Factors for Heart Disease

There are numerous risk factors for CAD, some of which are under your control and some of which are not. Modifiable risk factors include high blood pressure, high blood cholesterol, diabetes, smoking, being overweight or obese, not exercising, eating badly, and stress. The following are uncontrollable: race, gender, age, and family history. Specifically, men are typically more susceptible to CAD, and the risk rises with age (Hajar, 2017).

3 PREDICTIVE MODELS FOR HEART DISEASE

Predicting heart disease requires the analysis of large amounts of patient data to assess the likelihood of cardiovascular disease under different conditions and influencing factors in order to improve the accuracy and validity of predictions. Various types of predictive models and methods have been used for this purpose in the literature, each with different characteristics, and the vast majority of these models are machine learning algorithms. Machine learning techniques belong to a branch of artificial intelligence that has been widely used in many scientific fields, but their application in the medical literature has been limited, partly due to technical difficulties. Therefore, most of the machine learning models utilized in medical research focus on several techniques, of which the Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), and K-nearest Neighbour (KNN) are the most commonly used. These models can be categorized into traditional and novel models.

3.1 Traditional Models

Traditional heart disease prediction methods rely on well-established machine learning algorithms that analyze clinical and diagnostic data to assess a patient's risk. These models include Decision Trees, Support Vector Machines, and Logistic Regression, among others. They are widely used due to their interpretability and effectiveness in handling structured medical data.

A Decision Tree is a flowchart-like structure commonly used for classification and regression tasks and belongs to a nonparametric supervised learning algorithm. The root node, branches, internal nodes, and leaf nodes make up its hierarchical tree structure. To generate predictions, they recursively divide the dataset according to feature values. The decision tree will start from the root node without any branches at the root node. Whereas, the branches from the root node flow into the internal nodes which are also known as decision nodes. Both node types are assessed to create a homogeneous subset based on the features that are accessible, and are represented as either terminal or leaf nodes. All potential outcomes in the data set are thus represented by leaf nodes (IBM, 2025). Decision trees are also frequently utilized for the prediction of heart disease because of their interpretability. However, since decision trees have a tendency to overfit, it is necessary to carefully tune them. Furthermore, by forecasting the likelihood of results, occurrences, or observations, logistic regression provides a straightforward and efficient statistical technique for binary classification tasks. Using a logistic function that limits the output to values between 0 and 1, the model simulates the likelihood that a given input falls into a particular category. Data is categorized into distinct groups using logistic regression, which examines the relationship between one or more independent variables (Singh & Kumar, 2020). Due to its importance in predictive modeling, which calculates the statistical likelihood that an occurrence falls into a particular category, this prediction technique is frequently used in heart disease research. However, when handling nonlinear relationships in clinical data, its performance might be constrained.

A supervised machine learning technique commonly used for classification and regression tasks is the Support Vector Machine (SVM). It classifies data points by determining the optimal hyperplane in an N-dimensional space and maximizing the separation between the closest points of different classes. By defining the maximum margin, these closest points, also referred to as support vectors, improve classification accuracy and the model's potential to generalize to new data (IBM, 2024). Due to its robustness in high-dimensional spaces and effectiveness in handling both linear and nonlinear classification tasks, SVM has been extensively applied in cardiology prediction, making it particularly suitable for medical datasets with numerous features. Comparable to this, the K-Nearest Neighbor (KNN) algorithm is a simple, instance-based, supervised, and nonparametric machine learning method that classifies or predicts outcomes based on the proximity of data points. It is commonly used in classification and regression assignments due to its ease of implementation and efficacy. By calculating the distance between the input data point and other points, the method finds the K nearest neighbors. The average or weighted average of these neighbors' goal values is used by regression to predict the value, whereas classification places the input data point in the most prevalent category among its K nearest neighbors (Srivastava, 2025). However, the distance measure and K selection affect KNN performance, therefore parameter adjustment is necessary for best outcomes.

A dataset from the UCI database was used for training and testing in a study on machine learning algorithms for heart disease prediction. The accuracy of diagnosing cardiac disease was assessed and predicted using a variety of computational machine learning models. These algorithms included the K-nearest neighbor algorithm, the decision tree algorithm, the linear regression algorithm, and the support vector machine algorithm (Singh & Kumar, 2020). Using a variety of machine learning algorithms, such as logistic regression and KNN, the authors of another article developed a system for predicting and classifying patients with heart disease. The system also uses a patient's medical history to assess the likelihood that the patient will be diagnosed with heart disease (Das & Biswas, 2018). Mythili and her team introduce a rule-based model that evaluates the effectiveness of applying rules to the individual predictions generated by logistic regression, decision trees, and support vector machines. By incorporating these machine learning approaches into a more accurate predictive model, their method seeks to improve the accuracy of heart disease prediction using the Cleveland Heart Disease Database (Mythili et al., 2013). From these studies, it is clear that these four models are used very frequently in heart disease prediction research.

3.2 Novel Models

In addition to the traditional models mentioned above, which have been adopted by many studies, there are several studies that have been drilling into new types of models. Most of these new forecasting methods are based on traditional models, and on the basis of these valid models that have been verified countless times, other new methods are introduced and their feasibility and accuracy are verified. They do this in order to explore better forecasting methods and to further enhance the predictive accuracy of the models.

Arunachalam and Rekha tried novel methods in their research. The baseline classifier in this study is k-Nearest Neighbor; the heart disease features are predicted using a set of X-boost, Adaboost, and stochastic subspace classifier models; and the cardiovascular disease features are predicted using linear support vector feature measures. To improve classification, the model takes into account different feature combinations. The clinical decision support system demonstrates the model's exceptional accuracy and performance. The MATLAB 2020b simulation environment was used to run the simulation, and the outcomes were compared to other approaches that have already been used. The result suggested that this proposed model outperforms current methods with a prediction accuracy of 96% (Arunachalam & Rekha, 2022). Yazdani and colleagues propose a method to assess the significance of key features contributing to heart disease prediction. Their study focuses on forecasting heart disease using Weighted Associative Rule Mining, leveraging the scores of essential variables. By analyzing the widely used UCI open dataset for heart disease research, they aim to identify a set of critical feature scores and diagnostic rules for improved prediction accuracy. They also conferred with cardiologists to verify the validity of these guidelines. Weighted Associative Rule Mining was utilized to derive strength scores for important predictors, leading to the development of significant rules for heart disease prediction with a maximum confidence score of 98%. In order to calculate strength scores for important factors in the prognosis of cardiac disease, this study is essential (Yazdani et al., 2021). The above mentioned studies are good evidence that these new hybrid models have improved the accuracy of prediction in comparison to the previous conventional models and hence can provide better help to researchers in the field of heart disease prediction.

4 CURRENT LIMITATIONS AND FUTURE PERSPECTIVES

4.1 Challenges in Current Models

Traditional models offer several benefits, including the fact that logistic regression is straightforward to use and relatively simple, decision trees have good interpretability, SVM performs well in difficult high-dimensional data, and KNN is also efficient. However, these models do have certain limitations such as the inability to apply logistic regression for nonlinear relationships, the necessity to prune decision trees to avoid overfitting, the computational cost of SVMs, and the poor performance of KNN on large datasets and the need for careful feature selection. The same holds true for newer models. The first model which used a set of X-boost, Adaboost, and stochastic subspace classifier models, is highly accurate and robust to overfitting but is computationally expensive, necessitates extensive hyperparameter tuning, and is not interpretable. The second model that applies the Weighted Associative Rule Mining finds important predictors and offers interpretable rule-based insights, but struggles with generalization, is computationally intensive, and is not flexible enough to adjust to new healthcare data. Further research is needed to attenuate or eliminate these limitations, which may be achievable at some point in the future through advancing technological techniques and integration between different models.

4.2 Future Directions

Through a number of significant advancements, future breakthroughs in heart disease prediction models should concentrate on enhancing generalization, efficiency, accuracy, and interpretability. To improve expected performance while maintaining interpretability, hybrid models should incorporate the best features of both machine learning and deep learning. Furthermore, enhancing model transparency would require explainable artificial intelligence (XAI), particularly in clinical applications where doctors must be able to understand the judgments made. Improved interpretability techniques will help make AI-generated insights more actionable for decision trees and rule-based models like weighted association rule mining. In order to deliver more precise forecasts and treatment suggestions, personalized AI models will make use of lifestyle factors, genetic data, and real-time patient monitoring. Finally, the use of multimodal data will considerably enhance the future of heart disease prediction by enabling comprehensive, patient-specific risk assessments. Therefore, future developments should concentrate on enhancing model generalization, interpretability, and efficiency through autonomous hyperparameter tuning, hybrid techniques, and deep learning integration. Real-world clinical applications will be improved by technologies like explainable artificial intelligence and personalized AI models, which will make forecasts more visible, scalable, and patient-specific.

5 CONCLUSIONS

Heart disease is the leading cause of death globally, placing a significant financial and medical burden on economies and healthcare systems. Early diagnosis and precise risk assessment are crucial to minimizing mortality and improving patient outcomes. However, conventional diagnostic techniques like electrocardiograms (ECGs) and clinical risk assessments frequently have poor predictive accuracy and are unable to identify cardiac disease early on. Predictive models based on machine learning are being investigated more and more as a way to improve early detection and customize healthcare in order to overcome these constraints. K-nearest Neighbors, Logistic Regression, Decision Trees, and Support Vector Machines are some of the most popular models, and each has advantages and disadvantages of its own. LR is straightforward and easy to understand, but it has trouble understanding nonlinear relationships. SVM works well on high-dimensional data but uses a lot of processing power, DT offers transparency but is prone to overfitting, and KNN is ineffective with big datasets despite its versatility. Novel hybrid models, like X-Boost, Adaboost, random subspace classifiers, and Weighted Associative Rule Mining, have been created to enhance these conventional techniques. Although these models have demonstrated increased accuracy in predicting heart disease, issues with generalizability, interpretability, and processing needs remain.Future advancements in heart disease prediction will focus on deep learning and hybrid machine learning models, enhancing transparency through explainable artificial intelligence, and integrating multimodal patient data, including behavioral, genetic, and real-time monitoring inputs. These enhancements will preserve clinical applicability while raising prediction accuracy. Even though there are still issues with model validation and implementation, Predictive techniques powered by AI could enhance cardiovascular healthcare through promoting improved patient outcomes, tailored treatment regimens, and early diagnosis, ultimately lowering the worldwide burden of heart disease.

REFERENCES

Arunachalam, S. K., & Rekha, R. 2022. A novel approach for cardiovascular disease prediction using machine learning algorithms. Concurrency and Computation: Practice and Experience, 34(19), e7027.

Centers for Disease Control and Prevention. 2024, October 24. Heart disease facts. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/heart-disease/data-research/facts-stats/index.html

Das, G., & Biswas, S. 2018. IOP conference series: materials science and engineering. IOP Publishing, 338(1), 012056.

Hajar, R. 2017. Risk factors for coronary artery disease: historical perspectives. Heart views, 18(3), 109-114.

IBM. 2024, December 19. What is support vector machine?. IBM. Retrieved from https://www.ibm.com/think/topics/support-vector-machine

IBM. 2025, January 22. What is a decision tree?. IBM. Retrieved from https://www.ibm.com/think/topics/decision-trees

Mayo Foundation for Medical Education and Research. 2023, November 22. Heart valve disease. Mayo Clinic. Retrieved from https://www.mayoclinic.org/diseases-conditions/heart-valve-disease/symptoms-causes/syc-20353727

Mayo Foundation for Medical Education and Research. 2024, June 14. Coronary artery disease. Mayo Clinic. Retrieved from https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613

Mayo Foundation for Medical Education and Research. 2025, January 21. Heart failure. Mayo Clinic. Retrieved from https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142

Mythili, T., Mukherji, D., Padalia, N., & Naidu, A. 2013. A heart disease prediction model using SVM-decision trees-logistic regression (SDL). International Journal of Computer Applications, 68(16).

Singh, A., & Kumar, R. 2020, February. Heart disease prediction using machine learning algorithms. In 2020 International Conference on Electrical and Electronics Engineering (ICE3) (pp. 452-457). IEEE.

Srivastava, T. 2025, February 25. Guide to K-nearest neighbors algorithm in machine learning. Analytics Vidhya. Retrieved from https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

World Health Organization. 2021, June 11. Cardiovascular diseases (CVDs). World Health Organization. Retrieved from https://www.who.int/ news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Yazdani, A., Varathan, K. D., Chiam, Y. K., Malik, A. W., & Wan Ahmad, W. A. 2021. A novel approach for heart disease prediction using strength scores with significant predictors. BMC Medical Informatics and Decision Making, 21(1), 194.