Predicting heart disease using machine learning: A review

Heshan Sheng

The Chinese University of Hong Kong, Greenburg Building, Sha Tin, Hong Kong, China

1155157130@link.cuhk.edu.hk

Abstract. Heart disease remains the leading cause of global mortality, necessitating precise and nuanced detection strategies. Recent advancements in machine learning and deep learning have significantly enhanced predictive capabilities for heart disease, utilizing diverse patient data such as electrocardiograms, historical diagnoses, and genomic information. This review examines various computational models—including decision trees, support vector machines, and neural networks—assessing their applications, strengths, and limitations in heart disease prediction. Despite these advancements, challenges persist, including algorithmic safety, ethical considerations, and the need for prospective clinical validation. Future research should focus on integrating multimodal data, enhancing model interpretability, and addressing privacy concerns to further improve predictive accuracy and clinical utility.

Keywords: Machine Learning, Heart Disease, Healthcare AI, Ethical AI Use, Data Privacy.

1. Introduction

According to the World Health Organization (WHO), heart disease remains the leading cause of death globally, accounting for over 17.9 million deaths annually, or approximately 32% of global mortality. The diversity in heart disease types necessitates specific and nuanced detection and prevention strategies to effectively manage and mitigate risk across different patient populations. Recent advancements in machine learning, particularly deep learning, have significantly impacted the medical field by providing innovative solutions for predicting heart disease with unprecedented accuracy. These advancements leverage vast amounts of patient data—including electrocardiograms, historical diagnoses, lab results, and demographic information—to develop predictive models that can assist clinicians in identifying at-risk individuals and tailoring personalized treatment plans.

Predicting heart disease through machine learning involves the use of a range of sophisticated algorithms designed to analyze and interpret complex medical datasets [1-4]. Logistic regression, for instance, calculates the probability of heart disease based on various input features, allowing for straightforward interpretation of risk factors. Decision trees and random forests enhance prediction accuracy by creating decision rules through recursive data splitting, which helps in handling a wide range of data characteristics and interactions. Support vector machines (SVMs) are employed to find the optimal hyperplane that separates different disease categories, thus improving classification precision. Neural networks, with their multiple layers of interconnected nodes, are capable of learning complex, non-linear relationships within the data, making them particularly useful for identifying subtle patterns that may be indicative of heart disease.

This review explores the application of these computational technologies in predicting heart disease, summarizing current advancements, challenges, and future directions. It examines key methodologies, datasets, and real-world case studies to underscore the transformative potential of machine learning and deep learning in improving heart disease prediction and management.

2. Background on Heart Disease

Heart disease encompasses a range of conditions that affect the heart's structure and function, each presenting distinct challenges for diagnosis, management, and treatment:

Coronary Artery Disease (CAD): This is the most common form of heart disease, arising from the buildup of plaque in the coronary arteries which supply blood to the heart muscle itself. Plaque buildup causes the arteries to narrow and harden, a condition known as atherosclerosis. This reduces blood flow to the heart muscle, leading to symptoms such as chest pain (angina), shortness of breath, or, in severe cases, a heart attack. Risk factors include high cholesterol, hypertension, smoking, diabetes, sedentary lifestyle, and family history.

Congestive Heart Failure (CHF): CHF occurs when the heart's pumping power is weaker than normal, causing blood to move through the heart and body at a slower rate, and pressure in the heart increases. As a result, the heart cannot pump enough oxygen and nutrients to meet the body's needs. CHF can involve the left side (affecting the left ventricle which pumps oxygen-rich blood to the body), the right side (affecting the right ventricle which pumps blood to the lungs), or both sides of the heart.

3. Common Models and Their Applications

3.1. Decision Trees

The decision tree method is highly effective for classification, prediction, interpretation, and data manipulation, with several applications in medical research [5]. Classification and Regression Trees (CART) [6] are a subset of decision trees commonly used in heart disease prediction.

Melillo et al. [7] developed an automatic classifier based on the CART algorithm to assess the risk of congestive heart failure through long-term heart rate variability. Patients were categorized into high-risk (NYHA III and IV) and low-risk (NYHA I and II) groups.

Amiri et al. [8] proposed an automatic method for segmenting heart sounds using CART on a dataset of 116 heart sound recordings. This method aimed to reduce unnecessary echocardiograms and prevent the release of newborns with undetected heart disease, achieving a classification accuracy of 99.14% and 100% sensitivity.

Ozcan et al. [9] applied the CART algorithm to a dataset containing patient health information, including age, sex, and chest pain type. They extracted rules from the decision tree, which corresponded to the decision tree diagram, achieving an accuracy of 87%.

However, the decision tree method has limitations in complexity and accuracy. Bharti et al. [10] compared various machine learning methods on a public health dataset, finding that the highest accuracy for the decision tree method was 82.3%, while deep learning methods reached 94.2%. Thus, decision trees may not always be the optimal choice for heart disease prediction.

3.2. Support Vector Machines (SVM)

Support Vector Machines (SVM) frequently appear alongside other machine learning algorithms in recent research [11-13], either for comparative studies on heart disease prediction accuracy or as part of hybrid models.

Anggoro et al. [11] compared the performance of K-Nearest Neighbors (KNN) and SVM in predicting heart disease using a dataset with 304 samples and 14 attributes. Normalization improved SVM's accuracy to 90.10%, significantly higher than KNN's 81.31%. Without normalization, SVM and KNN achieved accuracies of 84.61% and 64.83%, respectively.

Ahmed et al. [12] introduced a hybrid KNN-SVM model that uses KNN for probability prediction, which is then used as input for SVM. This approach addresses the high computational load of SVM and

the sensitivity of KNN to noisy data. The hybrid model achieved an accuracy of 81%, outperforming KNN (75%) and SVM alone (76%).

Although SVM remains a powerful tool for heart disease prediction, much of the research employing SVM as a classification model was conducted in the 2010s [14-16]. Gokulnath et al. [14] evaluated binary particle swarm optimization (BPSO) and genetic algorithms (GA) for feature selection in coronary heart disease detection using SVM. They found that BPSO outperformed GA in identifying the presence of coronary heart disease when used with SVM.

Parthiban and Srivatsa [13] tested SVM against the Naive Bayes method using a dataset of 500 diabetic patients. The Naive Bayes method achieved a classification accuracy of 74% with a recall rate of 74%. In contrast, SVM achieved a classification accuracy of 94.6%.

Gokulnath and Shantharajah [14] integrated SVM with GA and compared it to other algorithms like Relief, CFS, and filtered subsets, demonstrating that the SVM-GA approach performed well.

3.3. Neural Networks

Recent research has explored various neural network architectures for predicting heart disease, each demonstrating distinct strengths.

Multi-layer Perceptron (MLP) models, as demonstrated by Kompella and Boddu [15], achieve high accuracy (98.58%) and specificity in heart disease prediction, outperforming traditional methods like decision trees and SVM.

Back-Propagation Neural Networks (BPNN), utilized by Olaniyi et al. [16], offer a robust prediction system with an accuracy of 85%, performing better than naive Bayes and decision trees.

Cascade Forward Neural Networks, compared by Awan et al. [17], excel with an accuracy of 97.7% after applying Principal Component Analysis (PCA), surpassing other neural network models in performance metrics.

Deep Neural Networks (DNNs) have shown significant improvements with optimized configurations. Darmawahyuni et al. [18] achieved 96% accuracy with their deep network, while Ali et al. (2019) [19] and Ramprakash et al. [20] enhanced performance to 93.99% and 93.33%, respectively, by combining feature selection methods with exhaustive hyperparameter tuning.

Convolutional Neural Networks (CNNs), as introduced by Mehmood et al. [21], reached a high accuracy of 97% using temporal data modeling to predict cardiovascular disease early, demonstrating superior performance compared to existing methods.

Overall, these neural network models highlight the evolving sophistication and effectiveness in heart disease prediction, with deep and convolutional networks leading the advancements.

4. Challenges

Despite the significant potential and value of AI in healthcare, its implementation presents several challenges and concerns:

Algorithmic Safety and Multidisciplinary Collaboration: Keskinbora et al. [22] emphasize the need for robust algorithmic safety measures in AI systems. They advocate for the involvement of experts from diverse fields such as biomedicine, psychology, ethics, economics, law, and policy in AI development to address these concerns effectively.

Ethical, Legal, and Social Implications (ELSI): Cartolovni et al. [23] identify key ELSI issues in AI medical decision support tools, including patient safety, algorithm transparency, inadequate supervision, responsibility attribution, impacts on the doctor-patient relationship, and governance of AI-enabled medicine.

Ethical Issues and Legal Responsibilities: Naik et al. [24] highlight critical ethical issues such as informed consent, data security and transparency, algorithm fairness, bias mitigation, and data privacy. They also point out challenges related to legal responsibilities and obligations when dealing with opaque AI "black box" systems in clinical decision-making.

5. Future Directions

Despite significant progress in machine learning and deep learning for predicting heart disease, further advancements are needed to fully exploit their potential. Key improvements include integrating multimodal data sources—such as medical images, genomic information, and wearable device data—to build more comprehensive models; developing interpretable AI technologies to enhance model transparency and clinical decision-making; and exploring joint learning methods to expand training datasets while safeguarding privacy. Additionally, rigorous prospective clinical validation is essential for assessing the model's performance and safety in practical medical contexts.

6. Conclusion

Machine learning and deep learning technologies have revolutionized heart disease prediction, offering promising improvements over traditional methods. Decision trees, support vector machines, and neural networks each contribute uniquely to this field, with neural networks showing particularly high accuracy and effectiveness. However, to fully realize the potential of these technologies, ongoing research must address key challenges such as algorithmic safety, ethical implications, and the need for comprehensive clinical validation. Future directions should prioritize the integration of multimodal data, the development of interpretable AI systems, and the implementation of privacy-preserving techniques to enhance predictive capabilities and ensure clinical applicability.

References

[1] Ibomoiye, D. M., Sun, Y., & Wang, Z. (2019). Prediction performance of improved decision tree-based algorithms: A review. *Procedia Manufacturing*, 35, 698-703.

[2] Pisner, D. A., & Schnyer, D. M. (2020). Machine learning. In *Machine Learning* (pp. 101-121).

[3] Walczak, S. (2018). Advanced methodologies and technologies in artificial intelligence, computer simulation, and human-computer interaction. *Advanced Methodologies and Technologies in Artificial Intelligence*, 14.

[4] Kelleher, J. D. (2019). *Deep learning*.

[5] Song, Y.-Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 85-93. PMC4466856.

[6] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1986). *Classification and Regression Trees*. New York: Chapman and Hall.

[7] Melillo, P., De Luca, N., Bracale, M., & Pecchia, L. (2013). Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. *IEEE Journal of Biomedical and Health Informatics*, 17(3), 638-645.

[8] Amiri, A. M., & Armano, G. (2013). Early diagnosis of heart disease using classification and regression trees. In *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN)*, 1-8.

[9] Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3, 100130.

[10] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience*, 2021, Article ID 8387680.

[11] Anggoro, D. A., & Kurnia, N. D. (2020). Comparison of accuracy level of support vector machine (SVM) and k-nearest neighbors (KNN) algorithms in predicting heart disease. *International Journal of Emerging Trends in Engineering Research*, 8(5), 1692-1700.

[12] Ahmed, R., Bibi, M., & Syed, S. (2023). Improving heart disease prediction accuracy using a hybrid machine learning approach: A comparative study of SVM and KNN algorithms. *International Journal of Computations, Information and Manufacturing (IJCIM)*, 3(1), 18-25.

[13] Parthiban, G., & Srivatsa, S. K. (2012). Applying machine learning methods in diagnosing heart disease for diabetic patients. *International Journal of Applied Information Systems (IJAIS)*, 3(7), 11-15.

[14] Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, 22, S14777-S14787.

[15] Kompella, S., & Boddu, V. (2019). Neural network based intelligent system for predicting heart disease. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(5), 251-256.

[16] Olaniyi, E. O., Oyedotun, O. K., & Helwan, A. (2015). Neural network diagnosis of heart disease. In *Proceedings of the 2015 International Conference on Advances in Biomedical Engineering (ICABME)*.

[17] Awan, S. M., Riaz, M. U., & Khan, A. G. (2018). Prediction of heart disease using artificial neural network. VFAST Transactions on Software Engineering, 6(1), 51-61.

[18] Darmawahyuni, A., Nurmaini, S., & Firdaus. (2019). Coronary heart disease interpretation based on deep neural network. *Computer Engineering and Applications*, 8(1), 23-30.

[19] Ali, L., Rahman, A., Khan, A., Zhou, M., Javeed, A., & Khan, J. A. (2019). An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network. *IEEE Access*, 7, 122834-122844.

[20] Ramprakash, P., Sarumathi, R., Mowriya, R., & Nithyavishnupriya, S. (2020). Heart disease prediction using deep neural network. *IEEE Access*, 8, 120021-120032.

[21] Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., Masood, M. (2021). Prediction of heart disease using deep convolutional neural networks. *Arabian Journal for Science and Engineering*, 46, 8375-8384.

[22] Keskinbora, K. H. (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, 64, 277-282.

[23] Cartolovni, A., Tomicic, A., & Lazic Mosler, E. (2022). Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. *International Journal of Medical Informatics*, 161, 104738.

[24] Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? Frontiers, 2022.