Research and Analysis of Heart Disease Risk Prediction Model Based on Machine Learning

Jiali Sun[1*]

[1]The Sino-British College, University of Shanghai for Science and Technology, Shanghai, China

Abstract. This review looks at recent progress in using artificial intelligence and machine learning for heart disease prediction, focusing on three main model types: Gradient Boosted Trees, Transformer-Based Sequence Models and Artificial Neural Networks. The analysis shows that each method has its own advantages. Gradient Boosted Trees work well with structured data tables, achieving good accuracy numbers and giving clear feature importance information. Transformer models, which use attention mechanisms, perform exceptionally on large electronic health record data, capturing long-term patterns for multi-disease prediction. Artificial Neural Networks model complex relationships effectively, sometimes reaching near-perfect results on certain datasets. The review covers the datasets, data preparation steps, and evaluation metrics including accuracy, precision, recall, F1-score and AUC-ROC used in the studies. While the results seem promising, there remain significant challenges before clinical use. That is to say, people need better model understanding, reduced algorithmic bias to be fair across different groups, smooth integration into medical workflows, and strong data privacy protections. To put it simply, future success depends not just on accuracy but on developing practical, clear, fair, and secure AI decision tools to truly fight heart disease.

1 Introduction

Heart disease continues as the main cause of death globally, presenting big challenges for health systems everywhere. Early and correct prediction of heart problems is crucially important, that is to say, because it allows help to be given sooner, improves what happens for patients, and helps use resources in the best way possible. Traditional ways of diagnosis, while useful, can often take too long, be expensive, and depends on how doctors see things. Given this situation, the emergence of Artificial Intelligence (AI) and Machine Learning (ML) brings new approaches to change how heart care works through looking at lots of information to find patterns.

The use of ML in medicine involves using past patient information to find complex, hidden patterns that normal methods might miss. In recent years, there has been growth in computer models designed for health, for instance, ranging from those that explain how they work to deep systems that find patterns automatically. These models offer not just possibly better predictions but also the chance to make things automatic and help doctors decide better.

* Corresponding author's email: lucy_135321@outlook.com

Putting together things like cleaning up the data properly, exploring what the data shows, and using advanced modeling steps is very important for making tools that can go from the lab into real hospitals or clinics reliably.

Looking at what research has been done recently, there are quite a few ways to tackle this issue, some of them working pretty well. For example, in one particular study, they put together a whole setup that included examining the data first, then using older computer learning methods, and finally adding a neural network thing, that is to say a complicated learning system, all tested on a common online dataset [1]. This approach managed to get very high scores, almost perfect, pointing out that you need to be really careful with your method to avoid the computer learning just the specific examples and not being able to handle new stuff, which is a common problem. Then, another piece of research was mainly about how easy it is to understand and how well these tree-based boosting methods perform, especially a type called XGBoost [2]. They used it on another well-known medical dataset and found it did much better than other classifiers, like the simpler Logistic Regression or Random Forest, achieving a high level of correctness, showing it's really good for organized patient data tables. Going even deeper, a really important study introduced BEHRT, which is a special kind of model using patient history data [3]. Simply put, they trained it on a huge amount of information from over a million people tracked over time. This BEHRT model did significantly better than existing deep learning models at guessing lots of different future health problems someone might get, proving how powerful this sequence modeling and the attention mechanism are for seeing patterns that unfold over many years in a person's medical records.

2 Overview of Heart Disease Prediction Models

Heart disease prediction using machine learning models can be broadly grouped into three main types: Gradient Boosted Trees, Transformer-Based Sequence Modelling, and Artificial Neural Networks. Each type has its own way of working and applications in clinical or research settings.

2.1 Gradient Boosted Trees (XGBoost)

XGBoost implements a method that improves step by step, where decision trees are added one after another. Each new tree tries to fix the mistakes made by the previous ones put together. The goal here combines how wrong the predictions are with a part that discourages making the trees too complicated. Training happens by working out first and second changes in prediction errors, then fitting a tree to these changes. To put it simply, it’s a way to make the model better gradually. Hyperparameters get tuned usually with tools like Optuna, which is to say, settings that control the learning process need adjusting for best results.

Artificial Neural Networks work by mimicking how the human brain processes information through layers of connected nodes. They are good at finding hidden patterns in complex data like medical images or test results. That said, they often act like a "black box," meaning it’s hard to see exactly why they make certain predictions. Training involves adjusting connections between nodes to reduce errors, which can take significant time and resources. Despite this, they are widely used for their ability to handle varied types of health information.

In a heart disease study by researchers like Narayana et al., a process involving cleaning data, then using SMOTE for balancing, selecting features with ANOVA, and training XGBoost with Optuna tuning, and finally cross-validation, achieved 92% accuracy, which is better than other common classifiers [4]. Similarly, Meti and Lingraj used XGBoost on the Cleveland dataset after handling categorical variables and missing values [2]. XGBoost

performed best, getting 93% accuracy and showing better scores in F1, precision, recall, and AUC. Earlier work by Doki et al. on smaller groups showed XGBoost with grid search could get an AUC around 0.853 and accuracy near 86% [5].

XGBoost works really well with different kinds of data features, automatically understanding how features interact. It has built-in ways to stop overfitting, and building trees in parallel speeds things up. The model also gives good scores for how important each feature is. However, it sees each record separately, which means it isn't good for time-based information patterns. Finding the best settings can take a lot of computer time, and the final model can be a bit hard to explain.

2.2 Transformer-Based Sequence Modelling

Transformers use self-attention, that is to say, they figure out how relevant each piece of data is in a sequence, capturing long connections better than older methods like RNNs which had fading signal problems. To add order, they use positional codes. For patient records, systems like BEHRT adapt BERT for medical terms, putting together patient age, visit identifiers, and diagnostic codes. Li et al. showed a BEHRT-style model could predict over 1,000 diseases, including things like heart failure, with high AUROC scores [3].

Kim and colleagues demonstrated that a self-attention model trained beforehand on electronic health record information from multiple hospitals could effectively predict mortality risk for heart patients [6]. Mateescu et al. proposed an explainable transformer model using diffusion-based time pattern analysis to anticipate heart failure incidents, achieving strong predictive performance while providing attention visualisations ways to see which parts of the data were important [7].

The typical process involves, to put it simply, turning medical events over time into tokens, pre-training with some codes hidden, adjusting the model for a specific outcome, and checking it on external data. Self-attention is really important because it learns connections between different visits, different medications, and different diagnoses. Pre-training on large unlabeled datasets creates useful representations, meaning you don't need as much labeled data. Explainability methods like attention visualisation can show influential past events, which is helpful. Weak points include needing a lot of computer power, requiring big training groups to prevent overfitting, and the model being a bit like a black box, which can make doctors hesitant to trust it. Also, turning data into tokens might not work well for uncommon medical codes.

2.3 Artificial Neural Networks (ANNs)

A basic feedforward ANN for predicting heart disease usually has an input layer, one or more hidden layers using non-linear activation functions, and an output layer with a sigmoid neuron. Training involves reducing a binary loss function, often with techniques to prevent overlearning such as randomly dropping connections during training and stopping early.

Rashedi Gazari and others developed a three-layer ANN model using about 1,200 patient cases [1]. After handling missing data and adjusting variables, plus doing grid searches and cross-validation, the model showed high accuracy around 88% with an AUC score near 0.91. Another research project using the Kaggle dataset reported very strong performance with ANN. More recently, Hussainy and team combined ANN with explainable AI methods training it on three public data sources [8]. It achieved high accuracy across different datasets – 94% for Cleveland, 98% for Hungarian, and 87% for Switzerland – while also showing which features mattered most through visualizations.

ANNs are relatively straightforward to set up, flexible enough to handle complex relationships, and don't require too much computing power when making predictions. With

proper regularization techniques, they can work well even on smaller datasets. The weaknesses include that they don't naturally handle time-series information well, meaning manual work is needed for temporal features. They tend to overfit when there's too many parameters compared to data points, and understanding why they make decisions requires extra tools like SHAP or LIME, to put it simply.

3 Results

3.1 Datasets, Sources, and Preprocessing

3.1.1 Gradient Boosted Trees (XGBoost)

The XGBoost model underwent testing using the “UCI Heart Disease Dataset”, which holds records for 303 patients, each with 14 clinical information features. These features cover information like the patient’s age, whether male or female, the type of chest pain experienced, blood pressure at rest, cholesterol numbers, levels of fasting blood sugar, results from resting ECG tests, the highest achieved heart rate, the presence of exercise-induced angina, and ST depression measurements along with thalassemia information [2]. Missing information, particularly within the "thal" and "ca" features, was addressed by either filling in the missing pieces or removing those entries entirely. Different kinds of information, such as chest pain type (cp), thalassemia (thal), and ST slope were handled in a special way for processing. Other numerical features were adjusted properly through scaling techniques to help the model learn better to make the information more uniform and easier to analyze. Helpful ways to select features were used to pick the most important predictors; for instance, key ones identified included the type of chest pain, the maximum heart rate achieved, and ST depression measurements [2].

Later studies expanded this approach to other heart disease datasets. Narayana and others used a well-known heart study dataset, which included about 3,650 records and 16 different risk factors such as age, blood pressure measurements, cholesterol levels, body mass index, glucose readings, and whether people smoked. They cleaned missing data points, adjusted the scale of the data, and applied a balancing method before model training. A selection method reduced the number of features, and an optimized model achieved about 92% accuracy with good sensitivity and specificity. Another study by Dal and Sezgin worked with a different dataset containing approximately 70,000 records and 14 clinical and lifestyle features. After handling missing values and adjusting parameters like learning rates and depth settings, their model reached over 97% accuracy, around 96% sensitivity, and 95% precision—that is to say, performing better than other models.

3.1.2 Transformer-Based Sequence Modelling

Upadhyayula & Pothugunta carried out tests on FT-Transformer, SAINT, and TabTransformer using three public datasets: the Framingham Heart Study which has around 4,300 records and 15 variables, the UCI Cleveland dataset (303 records, 13 attributes), and a Kaggle cardiovascular dataset with 1,025 or 303 samples [9, 10]. They applied a standard preprocessing approach using the median to fill missing values, encoding categorical fields with one-hot method, and scaling based on training data statistics. For time-series information such as ECG data, they used sinusoidal positional embeddings. Across these datasets, transformer models consistently achieved accuracy between 97.8% and 99.7%, with precision and recall above 97% and F1-scores over 0.96.

In another study, Rahman et al. utilized a self-attention transformer on the Cleveland dataset, which has 303 patients and 76 recorded attributes like demographics, lab tests, and ECG results. They removed features having more than 20% missing values and normalized numeric variables using training-set statistics. The transformer was then fine-tuned with binary cross-entropy loss. On a held-out test set, it reached 98.4% accuracy, 98.1% precision, 98.6% recall, and an F1-score of 98.3%. These outcomes indicate that transformer-based models are highly accurate and interpretable for classifying heart conditions, to put it simply.

3.1.3 Artificial Neural Networks (ANNs)

Artificial neural networks were assessed using a Kaggle dataset with 1,025 training samples and 303 test samples, containing 13 predictive features and one binary target. The features included age, sex, chest pain type, resting blood pressure, cholesterol levels, and maximum heart rate among others [1]. Data preparation involved normalization using the training data's average and spread to avoid data leakage issues. Sheta & Modi utilized another public heart disease dataset from Kaggle, which also had about 14 features such as age, sex, whether they smoke, blood pressure medication usage, cholesterol measurements, and so on, with approximately 300 samples [11]. They processed the categorical variables using encoding methods, that is to say, one-hot encoding, filled in missing data points with average values from the columns, and scaled the data to fall between 0 and 1. Their neural network structure involved two hidden layers, specifically 128 neurons followed by dropout, then 64 neurons followed by dropout, and finally 32 neurons. Akter et al. worked with a medical survey dataset that was imbalanced, having 33 healthy cases and 41 disease cases [12]. They addressed the imbalance using synthetic data generation techniques, normalized all the features, and divided the data 70% for training and 30% for testing purposes. Their ANN model incorporated three dense layers with 64, 32, and 16 neurons respectively, using ReLU activation functions and a sigmoid output layer. All three investigations employed back-propagation learning with the Adam optimization method and binary cross-entropy loss function. The training-test splits were 80%/20% for models based on the UCI data, and 70%/30% for the survey-based study after applying balancing techniques.

3.2 Evaluation Metrics

The XGBoost model was evaluated using several different metrics, that is to say, for a complete assessment. Accuracy, which tells us about overall correctness, was 93%. Precision, showing how many positive identifications were correct, was 0.92, while recall, which is sensitivity, was 0.94. The F1-score, which balances precision and recall, was 0.93. The Area Under the Curve, an overall performance measure, was 0.96, meaning the model was good at telling things apart [2]. Experiments done on the Framingham data and the Kaggle datasets reported similarly strong results, with accuracy ranging from 92% to 97.6%, high sensitivity, and good precision, demonstrating that Gradient Boosted Trees work well across various heart disease datasets.

Transformer models were looked at with accuracy, sensitivity (which is recall), specificity, F1-score, and the overall performance measure. Sensitivity was particularly important because missing a positive case in medical diagnosis is very costly. Specificity measured the model's ability to correctly identify negative cases. The F1-score gave a balanced view of precision and recall rate, while the overall performance measure evaluated how well the model could tell things apart across different settings. These metrics ensured a thorough evaluation of the models' predictive performance.

The ANN model demonstrated extremely high accuracy, precision, recall, and F1-score on the test data [1]. A method for reducing data dimensions showed that cutting down from

13 features to 11 features kept most of the information, about 93.84%, though actually not reducing dimensions worked best for performance. Sheta & Modi found validation accuracy around 92%, with sensitivity near 94% and specificity approximately 90% on their balanced dataset [11]. Akter et al. achieved accuracy of about 0.92, precision of 0.926, sensitivity of 0.926, and specificity of 0.909 on uneven survey data after applying a technique to balance data [12]. The error rate went up from roughly 17.1% on smaller samples to nearly 44.6% on bigger samples, which was observed across all three experiments.

3.3 Experimental Results and Analysis

The strong points of XGBoost, Transformers, and ANNs were different. The XGBoost performed quite successfully with structured data tables i.e. due to its regularization and boosting. It was comparable among different datasets and, therefore, a good starting point model. Transformers had been really good at big and clean data, achieving a high level of accuracy and easier interpretation, but failed when the classes are disproportionate. ANNs achieved flawless precision in certain instances, demonstrating that they are capable of dealing with complicated designs, however, they are not simple to read and require higher processing speed, which is a disadvantage.

To conclude, the model selection does truly depend upon the dataset itself. There is one particular boosting algorithm that is an effective baseline to structured data. Sequential information designed models that is to say more appropriate to very big datasets that are clean. Artificial neural networks which are an example are more favourable with non-linear complex patterns though they require additional computer power. The results are useful in informing the choice of models to be used when predicting heart disease activities.

4 Challenges and Prospects

In the future, Artificial Intelligence (AI) one of the areas where the future use of heart disease prediction will hold considerable promise is through the use of AI. There are various challenges that should be addressed before the successful utilization of AI in predictive heart disease in a practical and widespread setting in healthcare.

A critical issue includes getting these powerful black-box models intelligible and amenable to the doctors. Although in most cases AI may be able to forecast heart disease with a high level of accuracy, the medical practitioners ought to be aware of the rationale behind the model arriving at a specific conclusion to make effective choices. In other words, a misguided prediction may lead to extremely dire implications on a patient. Thus, further work should be oriented towards the development and standardization of the way such decisions are made, such as visualization techniques of some models, which represent the main component of any clinical AI tool. The doctors can be able to visualize what influences most in the risk score that was generated by the model of a particular pattern in ECG readings or a trend in cholesterol levels etc.

The other major challenge is the viability of these models to all people. Numerous AI models are trained using information that derives specifically, and frequently, hospital populations in western nations. This increases the possibility that they may fail to deliver where patients who do not belong to their ethnic background, gender, or region and hence health inequalities may even escalate. Even the models in future must be created and properly tested on the data that is diverse and representative worldwide. Algorithms to identify and rectify any unfairness should be a routine aspect of algorithms, which should be utilized in a novel community context before any model is applied.

Moreover, to transition AI out of the research paper to a hospital ward it must integrate smoothly with the current clinical processes. The model must be very accurate but not so

slow to the point of not allowing the data that is not easily available by the data doctors and must not disrupt their routine. The third-generation AI tools must be built as a viable clinical decision support system. This implies that they not only need to process data in close real-time, but act on already existing information in electronic medical records and deliver simple and easily understandable insights, such as a basic risk score with justifications, right to the physician at the point of care.

Lastly, the aspect of data privacy and security is still at the frontline. The AI training necessitates huge amounts of personal patient data. The next stage of development should be focused on the privacy-saving methods, like federated learning, when the models may be trained in several hospitals without access to the original patient data itself. To establish trust among the masses and the medical professionals, there is also a need to have effective governance and well-defined rules regarding the process of development and validation of medical AI.

5 Conclusion

To summarize, the review shows that artificial intelligence has a lot of potential to alter how heart disease is predicted and detected at its initial stages. Through a multimodal examination of intricate data with a complex structure, AI models are able to recognize nuanced patterns that the conventional credit of things might have easily overlooked, and as such risk is assessed more correctly and promptly. The performance of Gradient Boosted Trees such as XGBoost is one of the most effective and dependable methods to handle the structured clinical data and it always demonstrates a high level of accuracy on various data sets. Transformer-based architectures are good processors of sequential or large data that provide strong predictive accuracy and better interpretability, given their attention execution. In the meantime, ANNs are very effective in higher order, non-linear correlations in the data and sometimes they can even be almost perfect when there is control.

According to the experimental results, such types of AI solutions have a capacity to perform better than conventional risk scores in key measurements of performance, such as accuracy, sensitivity, and predictive power in general. This better work indicates that the introduction of AI tools may indeed benefit doctors particularly in the regions that lack adequate specialists, in other words, simply put. Using AI systems to identify individuals at risk in advance and with greater accuracy can enable timely initiation of treatment, which could result in better patient outcomes and reduce the global hearing regarding heart disease.

Although, these research successes must be worked into the mainstream of clinics and that too in a large scale before several big problems can be sorted out. One prime concern is the "black-box" manner in which most complex models operate in other words, it is difficult to have a look inside them, which may lead to mistrust by the medical personnel. Generalizing clear features that render things very important so that the doctor can follow the rationale making each prediction by the AI. Moreover, the challenges of the fairness and good work in various groups have to be addressed immediately. Models that are trained on data of only some types of people may not perform well on all people, which may lead to the worsening of existing health gaps. Ensuring that AI tools are tested on actual patient data that conceptualizes all types of patients is one of the points that should be making sure healthcare is equitable.

The further development also requires the creation of AI systems that can be easily incorporated into doctors and nurses' daily practices and activities and fit into the procedures and tools that they already have in place. Lastly, the moral application of patient data should continue to be among the priorities. Privacy preserving techniques, as well as powerful regulation structures are essential in keeping the trust of the populace. Going forward, it would be more important to concentrate on constructing realistic models rather than to

construct viable, explainable, equitable and safe decision-support systems. With these challenges overcome, AI has the potential to turn out to be a revolutionary agent in cardiovascular care by helping increase their personalization of prevention and deliver more desirable outcomes in the field of heart health across the globe.

References

1. A. Rashedi Gazari, S. Rokhva, T. Khatibi, E. Akhondzade Noughabi, & B. Teimourpour. A combination of EDA, machine learning, and artificial neural networks for the accurate prediction of heart disease. Annals of Healthcare Systems Engineering, **2**(1), 38-46. (2025)
2. M. Meti, & R. Ling. Heart boost: Clinical data-driven heart disease prediction using XGBoost. International Research Journal on Advanced Engineering Hub (IRJAEH)*,* **3**(9), 3517–3525. (2025)
3. Y. Li, S. Rao, J. R. A. Solares, A. Hassaine, R. Ramakrishnan, D. Canoy, Y. Zhu, K. Rahimi, & G. Salimi-Khorshidi. BEHRT: Transformer for Electronic Health Records. *Scientific reports*, **10**(1), 7155. (2020) https://doi.org/10.1038/s41598-020-62922-y
4. V. M. L. Narayana, Y. S. Ram, D. L. Ganapathi et al. Prediction of heart disease using Xgboost algorithm. International Journal on Science and Technology (IJSAT), **16**(1). (2025)
5. S. Doki, S. Devella, G. Tallam Reddy et al. Heart disease prediction using XGBoost. In Proceedings of the Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) (pp. 1317 1320). (2022)
6. Y. Kim, H. Kang, H. Seo, et al. Development and transfer learning of self-attention model for major adverse cardiovascular events prediction across hospitals. Scientific Reports, **14**, 23443. (2024) https://doi.org/10.1038/s41598-024-74366-9
7. A. Mateescu, I. Hadarau, I. Anghel. A Laplace diffusion-based transformer model for heart rate forecasting within daily activity context. arXiv preprint, arXiv:2508.16655v1-3. (2025)
8. F. S. A. Hussainy, J. Jayapradha, M. Roslee. Improved interpretation model for heart disease diagnosis using artificial neural networks. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), **16**(2), 136-153. (2025) https://doi.org/10.58346/JOWUA.2025.I2.009
9. S. Dal, & N. Sezgin. Heart attack classification with a machine learning approach based on the random forest algorithm. Balkan Journal of Electrical & Computer Engineering, **13**(2). (2025) https://doi.org/10.17694/bajece.1691905
10. S. K. Upadhyayula, & R. Pothugunta. Benchmarking transformer-based and conventional machine learning models for cardiovascular disease prediction on datasets of varying scale and complexity. medRxiv. https://doi.org/10.1101/2025.08.03.25332878 (2025)
11. P. V. Sheta, & P. J. Modi. An efficient artificial neural network for coronary heart disease prediction. International Journal for Research in Applied Science and Engineering Technology, **9**(12), 1904–1910. (2021) https://doi.org/10.22214/ijraset.2021.39559
12. S. B. Akter, S. Akter, T. S. Pias, D. Eisenberg, & J. F. Fernandez. Identification of myocardial infarction (MI) probability from imbalanced medical survey data: An artificial neural network (ANN) with explainable AI (XAI) insights. medRxiv.

https://www.ijraset.com/research-paper/efficient-artificial-neural-network-for-coronary-heart-disease-prediction (2024)