Machine Learning in Heart Disease Prediction: A Comprehensive Investigation and Future Prospects

Ce Zhang

College of mathematics and systems science (big data college of alibaba cloud), Shandong University of Science and Technology, Shandong, China

Abstract. This paper provides a systematic review of machine learning (ML) and deep learning (DL) applications in heart disease prediction, highlighting advancements, challenges, and future prospects. Traditional ML models, including Random Forests, Support Vector Machines, and ensemble methods, have demonstrated strong performance in feature extraction and risk classification. Deep learning approaches, such as Graph Neural Networks (GNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures, further enhance predictive accuracy by capturing complex relationships in multi-dimensional medical data. Despite their success, critical limitations persist, including the "black-box" nature of neural networks, which hampers clinical interpretability; data heterogeneity across regions and institutions, limiting model generalizability; and privacy risks associated with centralized medical data training. To address these challenges, emerging solutions like interpretable rule-based systems, domain adaptation frameworks, and federated learning with differential privacy are proposed. The review underscores the need for interdisciplinary collaboration to integrate clinical expertise with advanced AI techniques, ensuring robust, transparent, and ethically compliant tools for early heart disease diagnosis. Future research should prioritize model interpretability, cross-institutional adaptability, and secure data-sharing mechanisms to bridge the gap between theoretical innovation and clinical implementation.

1 Introduction

Heart disease refers to abnormal heart structure or function, often caused by arteriosclerosis, hypertension or congenital defects. Symptoms include chest pain, dyspnea, palpitations, etc., which may lead to serious consequences such as myocardial ischemia and heart failure. According to the Global Health Estimates 2019 Report released by the World Health Organization (WHO), about 17.9 million people worldwide died from cardiovascular disease in 2019, with ischemic heart disease (coronary heart disease) being the deadliest type of cardiovascular disease. This accounts for approximately 47% (approximately 8.4

202211080232@sdust.edu.cn

million) of total cardiovascular disease deaths [1]. Common types of heart disease include coronary artery disease, heart failure, arrhythmia, cardiomyopathy, etc.

At present, the traditional diagnosis method for heart disease in the medical field has limitations. Generally, it is manually diagnosed by doctors. In this way, it has certain challenges, such as slow efficiency, a certain misdiagnosis rate, and high labor costs. In this case, it is necessary to explore or explore better diagnostic methods. Artificial intelligence, which has strong feature extraction and prediction capabilities, can be considered to be combined with traditional heart disease diagnosis methods to innovate more powerful and efficient methods to help diagnose heart disease.

Artificial intelligence quickly began to attract people's attention after the intelligent Go program AlphaGo defeated the world-renowned player Lee Sedol in 2016. Since then, artificial intelligence has ushered in explosive development. Machine learning is the core field of artificial intelligence. It refers to the automatic learning of rules from data and building models, so that the system can complete tasks such as prediction, classification or decision-making without explicit programming. Against the background of the rapid development of artificial intelligence, many models have emerged in machine learning, from simple linear regression, decision trees, random forests to today's neural networks.

In the past few years, these models have been widely used in various fields, such as economics and chemistry [2, 3]. Especially in the field of medicine, many people have carried out related works. For example, Sanju et al. proposed a hybrid feature selection and deep learning framework combining random forests, Principal Component Analysis (PCA) and L1 regularization for predicting thyroid diseases [4]; Islam et al. used extreme tree classifiers combined with tree structured Parzen estimators to optimize hyperparameters and predict liver diseases [5].

At the same time, there is another very important direction in the field of medicine, which is the prediction of heart disease that this article mainly studies. In the past, many scientists and scholars have devoted themselves to research in this direction. For example, Francisco Mesquita and Goncalo Marques used Random Forest (RF), CatBoost and SHapley Additive Explanation (SHAP) interpretability methods to predict Coronary Heart Disease (CHD) based on multi-center combined data sets [6]; Yue Huang et al. combined Coronary Artery Calcium Score (CACS) and clinical factors to predict coronary heart disease (CHD) through machine learning models such as RF and Support Vector Machine (SVM) [7]. Due to the widespread application and rapid development of machine learning in the medical field, many new algorithms have been proposed in recent years, so it is necessary to make a comprehensive review of this direction.

The rest of this paper contains three parts. The second part mainly describes the method, focusing on how other scientists and scholars use machine learning models to predict heart disease. The third part will discuss some shortcomings of the current model, some limitations of the field, and future development prospects. The fourth part is the conclusion and summary of the whole paper.

2 Method

2.1 Traditional machine learning models

2.1.1 Random Forest (RF)

Rana et al. used the random forest algorithm to pre-process, train, and classify a dataset of heart disease provided by Kaggle, and showed that the method achieved the highest accuracy of 81% in a comparative study [8]. Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training. Each tree is trained on random subsets of data (bootstrap sampling) and features. Final predictions are determined by majority voting (classification) or averaging (regression), enhancing accuracy and reducing overfitting. Its robustness in handling noisy/high-dimensional data makes it ideal for tasks like disease prediction.

2.1.2 Ensemble machine learning

Patra et al. proposed a two-step hybrid ensemble learning approach. First, Feature Weighted Meta-Models were constructed using a forward feature selection algorithm to screen key features. Then, the best performing meta-model was combined into a Hybrid Voting Model through soft and hard voting strategies, effectively combining the advantages of multiple classifiers, significantly reducing the misdiagnosis rate of ten-year risk prediction of CHD and increasing the accuracy rate to 95.87% [9]. Integrated machine learning improves the accuracy, stability and generalization ability of the overall model by combining the prediction results of multiple base models (such as decision trees and SVM), using voting, weighted average or stacking strategies, and effectively reducing overfitting and classification errors. Typical methods include random forests, AdaBoost, and XGBoost.

2.1.3 Support Vector Machine (SVM)

The paper develops a decision support system based on Least Squares Support Vector Machine (LS-SVM), using Radial Basis Function (RBF) kernel method for normal/abnormal classification of cardiac Doppler signals: After the signal features were extracted by wavelet transform and short-time Fourier transform, LS-SVM was used to replace the ANN classifier in the previous study, and the classification sensitivity and specificity were evaluated by combining the ROC curve. Meanwhile, the training efficiency and classification performance of LS-SVM and Backpropagation Neural Network (BP-ANN) were compared [10]. SVM is a supervised learning model, which constructs the optimal hyperplane by maximizing the classification interval and uses kernel function to process nonlinear data. It is suitable for high-dimensional small sample classification.

2.2 Deep learning models

2.2.1 ANN

The study presented by Reddy et al. used an artificial neural network (ANN) approach to predict heart disease risk due to anxiety disorders. Although the survey showed that healthcare professionals were more likely to use random forests (RF, 29%) and support vector machines (SVM, 28%) for anxiety classification, traditional ANN was less accurate (7%). This study creatively improves ANN feature extraction capabilities by integrating Stacked Autoencoders (SAEs), and constructs a hybrid SAE-ANN model to achieve 90% accuracy in heart disease prediction (traditional ANN is only 70-80%). The model integrates physiological indicators (BMI, lipoprotein) and psychological data (anxiety characteristics), combined with the investigation and clinical data analysis of 75 machine learning experts, breaking through the single dimension limitation of traditional diagnostic methods, and providing a new AI-driven solution for the joint prediction of mental health and cardiovascular diseases [11].

2.2.2 GNN

Rangisetti et al. proposed a heart disease prediction model based on Graph Neural Networks (GNN), comparing four architectures: Graph Attention Network (GAT), Graph Convolutional Network (GCN), GraphSAGE and Graph Isorphic Network (GIN). By integrating the heterogeneous medical data such as patient demographics and clinical characteristics, the "patient relationship map" is constructed, and the graph structure is innovatively used to capture the complex correlation among medical parameters. GIN model has the best performance, with a test accuracy of 94.15%, significantly surpassing traditional machine learning methods. Research innovations include: GNN was introduced into the field of cardiovascular disease prediction for the first time, combining graph structure with multi-dimensional medical data, breaking through the limitations of traditional table data modeling, improving the model's ability to analyze medical relationship networks, and providing an interpretable graph learning framework for personalized diagnosis and treatment [12].

2.2.3 LSTM

Djerioui et al. proposed an LSTM-based heart disease prediction model using the UCI heart disease dataset and comparing it with the multi-layer perceptron (MLP). This study innovatively applies the LSTM network traditionally used for time series analysis (which deals with long-term dependence and gradient problems through a gating mechanism) to structured clinical data (13 static medical features) to construct an intelligent diagnostic system. The main contributions include: to verify the effectiveness of LSTM in tabular data classification based on the characteristics of non-temporal medical data, the test accuracy rate is 96.5%, significantly better than MLP (89.18%), and the training time is reduced to 3 seconds (183 seconds for MLP). This method is a breakthrough combination of deep learning temporal model and static feature engineering, providing new ideas for medical binary classification problems, demonstrating the potential advantages of LSTM in non-temporal medical data analysis, and providing a more efficient AI solution for the early warning of heart disease [13].

2.2.4 Transformer

Fan et al. proposed a Transformer model with embedded learning filters for heart murmur detection. The model combines CNN, frequency domain adaptive learnable filter layer and Transformer encoder. After embedding layer, a learnable filter layer is added. Through FFT/IFFT transformation and parameter optimization, frequency domain adaptive noise reduction and periodic feature extraction are realized. The Transformer encoder captures long-term dependencies on heart sound signals. The innovation lies in combining learnable filters with Transformer to enhance feature expression through frequency domain modulation. The model used a majority voting strategy to aggregate fragment-level prediction results. On the implicit test set of the 2022 PhysioNet Challenge, the noise detection task ranked 39th (weighted accuracy 0.402) and the clinical result recognition task ranked 35th (cost score 15,982), confirming the advantages of its anti-noise and timing modeling [14].

2.2.5 Autoencoder

Iftee et al. propose an unsupervised heart disease classification model that combines autoencoders with AdaBoost. In this model, a custom autoencoder with SIGMOIDactivated neurons is used. The last layer of the encoder generates one-dimensional potential features, realizes unsupervised binary classification (presence or absence of heart disease) by threshold division, and input the classification results into the AdaBoost algorithm to iteratively adjust the weight of misclassified samples to improve accuracy. The innovations include: L1 regularized autoencoder is used as AdaBoost base classifier for the first time to realize the fusion of labelless learning and complex feature extraction; By combining multi-source data set (CHLS) training and testing on unknown data set (Statlog), the strong generalization ability of the model is verified, and the accuracy rate is up to 97.23%, showing the superiority of unsupervised representation learning combined with integrated enhancement technology [15].

3 Discussion

3.1 Limitations and challenges

3.1.1 Interpretability

As stated in the paper, in medical scenarios, the performance of AI machine learning models is quite excellent, but poor interpretability is one of its main challenges. Due to the high complexity of models (such as deep neural networks), the decision-making process is often regarded as a "black box", which makes it difficult for doctors to clearly explain the basis of the prediction in the face of patient questions [16]. For example, when the model diagnoses a disease or recommends treatment, the patient cannot intuitively understand its reasoning logic, which can lead to a loss of trust in the patient's doctor. In addition, medical decision-making involves ethical and legal responsibilities, and if the model is misdiagnosed due to unexplainability, it is challenging to trace the root of the error, which may lead to doctor-patient disputes. The lack of interpretability also hinders clinical validation and model optimization, limiting the practical application of AI in precision medicine.

3.1.2 Applicability

The poor applicability of machine learning models in medical scenarios is another challenge and often caused by differences between countries and hospitals. Different countries have different disease profiles, demographics and medical standards: cancer screening models developed for populations in Europe and the US, for example, can be invalidated by genetic or morbidity differences in Asian populations. The data collection methods (such as the model of image equipment and the standard of medical record) in different hospitals are not uniform, which leads to the poor generalization of the model across institutions. The reliability of the model is further weakened by low data quality and insufficient labeling in developing countries. In addition, differences in local medical policies and cultural preferences (such as treatment path selection) also affect the landing of AI models.

3.1.3 Privacy

The privacy risks of AI machine learning models in medical scenarios cannot be ignored, especially when the model is trained, it is easy to cause data leakage [17]. Medical data contains sensitive information (such as medical history and genetic data), which needs to be shared across institutions during the centralized testing and training of the model. If encryption or anonymization is insufficient, the data may be maliciously restored, resulting in the disclosure of patients' personal privacy information. Meanwhile, data protection regulations vary from country to country: strict regulations in Europe and the United States (such as GDPR), while weak regulations in developing countries also exacerbate privacy risks.

3.2 Future prospects

3.2.1 Interpretable machine learning

To solve the problem of black box decision-making of medical AI model, interpretable methods based on tropical geometry and fuzzy reasoning system can be considered. The latest research successfully visualized the rules (e.g., the clinically explainable path of "low ejection score and low motor capacity to heart failure") by dynamically constructing the synergistic mechanism of "low/medium/high" semantic rules with the attention matrix (A) and the connection matrix (M), and established the mapping relationship between the rules and classification results through the inference matrix (W). However, such research still has problems such as the high complexity of dynamic rule generation and dependence on expert experience for parameter initialization. Lightweight rule pruning algorithms should be further developed to explore automatic rule discovery mechanism based on reinforcement learning, especially in the cross-dimensional rule fusion of multi-modal medical data [18].

3.2.2 Domain Adaptation

In order to solve the distribution migration problem of medical image cross-domain migration, Adrenal Domain adaptation Network (ADAN) has tried to achieve interdomain alignment through dual feature extractor architecture and mixed loss function and achieved 95.59% cross-domain recognition accuracy in the gastric cancer early screening task. However, most of the existing studies are limited to the static migration scenario from single source domain to single target domain, and lack effective solutions to the complex situations such as multi-source domain and dynamic incremental domain, which are common in clinical practice. Therefore, the adaptive domain weight allocation mechanism based on meta-learning should be further studied, and the feature space elastic mapping technology for continuous domain drift should be developed. In particular, breakthroughs are needed in the field of real time domain adaptation of endoscopic images [19].

3.2.3 Federated Learning

In order to deal with the contradiction between privacy protection and fragmented utilization of medical data, the differential privacy federated learning framework has achieved privacy-performance balance (10% accuracy improvement) in stress detection tasks through Laplacian noise injection and transfer learning pre-training strategies. However, the existing methods have a trade-off dilemma between noise sensitivity and model accuracy, and the model performance deteriorates exponentially when the privacy budget is less than 2%, and the problem of data heterogeneity among medical devices is not fully considered. It is urgent to develop a hybrid differential privacy mechanism based on adaptive noise threshold and explore a federated knowledge distillation scheme for non-independent and equally distributed medical data, especially in the multi-center joint modeling scenario across institutions, a more robust parameter aggregation paradigm needs to be built [20].

4 Conclusion

In this paper, the application of machine learning and deep learning in the prediction of heart disease is comprehensively reviewed, the latest technological progress is reviewed and the future research direction is prospected. This review systematically analyzes traditional machine learning models (such as random forests, support vector machines, ensemble learning) and deep learning models (such as artificial neural networks, graph neural networks, long short-term memory networks, Transformer, and autoencoders), focusing on the advantages of these models in feature extraction and complex medical data modeling. Innovative approaches include hybrid frameworks that incorporate domain knowledge, graph-based modeling of patient relationship networks, and timing models optimized for static clinical data.

Through discussion, this paper reveals the core challenges in the current field: 1) The lack of interpretability of "black box" models such as neural networks, which affects clinical trust and responsibility traceability; 2) Limited applicability due to cross-regional and cross-institutional heterogeneity of data; 3) Medical privacy risks caused by centralized training. To address these issues, future research should focus on the following directions: the development of interpretability techniques (such as SHAP, rule-based expert systems), the use of domain adaptive methods to improve cross-institutional generalization capabilities, and the combination of differential privacy federated learning frameworks to achieve secure distributed training.

References

1. WHO. ‘Global health estimates: Leading causes of death’, https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death, accessed 8 April 2025
2. Jia, J., Zhang, X.: 'A human-scale investigation into economic benefits of urban green and blue infrastructure based on big data and machine learning: A case study of Wuhan', J. Cleaner Prod., 2021, 316, 128321
3. Choi, J., Yun, J.-I.: 'Optimization of water chemistry to mitigate corrosion products in nuclear power plants using big data and multiple linear regression in machine learning', Prog. Nucl. Energy, 2025, 183, 105685
4. Sanju, P., Ahmed, N.S.S., Ramachandran, P., Sajid, P.M., Jayanthi, R.: 'Enhancing thyroid disease prediction and comorbidity management through advanced machine learning frameworks', Clin. eHealth, 2025, 8, pp. 7–16
5. Islam, R., Sultana, A., Tuhin, M.N.: 'A comparative analysis of machine learning algorithms with tree-structured parzen estimator for liver disease prediction', Healthc. Anal., 2024, 6, 100358
6. Mesquita, F., Marques, G.: 'An explainable machine learning approach for automated medical decision support of heart disease', Data Knowl. Eng., 2024, 153, 102339
7. Huang, Y., Ren, Y., Yang, H., Ding, Y., Liu, Y., Yang, Y., Mao, A., Yang, T., Wang, Y., Xiao, F., He, Q., Zhang, Y.: 'Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment', Comput. Biol. Med., 2022, 151 (Part B), 106297

8. Rana, M., Ur Rehman, M.Z., Jain, S.: 'Comparative Study of Supervised Machine Learning Methods for Prediction of Heart Disease', Proc. 2022 IEEE VLSI DCS, Kolkata, India, 2022, pp. 295–299
9. Patra, S.C., Maheswari, B.U., Pati, P.B.: 'Forecasting Coronary Heart Disease Risk With a 2-Step Hybrid Ensemble Learning Method and Forward Feature Selection Algorithm', IEEE Access, 2023, 11, pp. 136758–136769
10. Çomak, E., Arslan, A., Türkoğlu, İ.: 'A decision support system based on support vector machines for diagnosis of the heart valve diseases', Comput. Biol. Med., 2007, 37 (1), pp. 21–27
11. Reddy, A.R.P., Jweeg, M.J., Das, S., Kumar, M.S., Pant, K.: 'The ANN Method for Better Living's Method of using Artificial Neural Networks to Predict Heart Attacks Caused by Anxiety Disorders', Proc. 2023 ICACITE, Greater Noida, India, 2023, pp. 457–460
12. Rangisetti, L.S., Shaik, F., Bhagavatula, A., Kommareddy, L.S.: 'Heart Disease Detection Using Graph Neural Networks (GNNs)', Proc. 2024 ICESC, Coimbatore, India, 2024, pp. 1436–1441
13. Djerioui, M., Brik, Y., Ladjal, M., Attallah, B.: 'Heart Disease prediction using MLP and LSTM models', Proc. 2020 ICEE, Istanbul, Turkey, 2020, pp. 1–5
14. Fan, P., Shu, Y., Han, Y.: 'Transformer Embedded with Learnable Filters for Heart Murmur Detection', Proc. 2022 Computing in Cardiology (CinC), Tampere, Finland, 2022, pp. 1–4
15. Iftee, M.A.R., Das, S.: 'Unsupervised Binary Classification of Heart Diseases Using an Autoencoder Model with Boosting Algorithm', Proc. 2023 ICCCNT, Delhi, India, 2023, pp. 1–7
16. Huang, J., Cai, Y., Wu, X., Huang, X., Liu, J., Hu, D.: 'Prediction of mortality events of patients with acute heart failure in intensive care unit based on deep neural network', Comput. Methods Programs Biomed., 2024, 256, 108403
17. Petersen, E., et al.: 'Responsible and Regulatory Conform Machine Learning for Medicine: A Survey of Challenges and Solutions', IEEE Access, 2022, 10, pp. 58375–58418
18. Yao, H., et al.: 'A Novel Tropical Geometry-Based Interpretable Machine Learning Method: Pilot Application to Delivery of Advanced Heart Failure Therapies', IEEE J. Biomed. Health Inform., 2023, 27 (1), pp. 239–250
19. Jie, L., Liang, P., Zhao, Z., Chen, J., Chang, Q., Zeng, Z.: 'ADAN: An Adversarial Domain Adaptation Neural Network for Early Gastric Cancer Prediction', Proc. 2022 IEEE EMBC, Glasgow, UK, 2022, pp. 2169–2172
20. Wang, Z., Yang, Z., Azimi, I., Rahmani, A.M.: 'Differential Private Federated Transfer Learning for Mental Health Monitoring in Everyday Settings: A Case Study on Stress Detection', Proc. 2024 IEEE EMBC, Orlando, FL, USA, 2024, pp. 1–5