Advancements in Machine Learning for the Detection of Human Heart Diseases: A Comprehensive Review

M Vindhya, Krithick Narayan D, Dr. UMADEVI RAMAMOORTHY

Student, MCA, CMR University, Karnataka, India
Student, MCA, CMR University, Karnataka, India
Professor, MCA, CMR University, Karnataka, India

ABSTRACT

This review paper delves into the application of machine learning techniques in the early detection and diagnosis of heart disease, an area that continues to challenge healthcare professionals worldwide. With cardiovascular diseases leading as a major cause of mortality, timely and accurate detection can significantly impact patient outcomes. This paper provides an accessible overview of current ML methods used to analyse and interpret complex cardiovascular data. We explore various algorithms, including decision trees, support vector machines, and neural networks, which are instrumental in recognizing patterns and predicting heart disease risks with increasing precision. Key studies and breakthroughs in the field are summarized to present a clear picture of where we stand and what the future may hold for ML in cardiology. Through this paper, readers will gain insights into the state-of-the-art ML tools transforming heart disease detection and the ongoing quest for models that are as robust and interpretable as they are effective.

Keyword - *Machine Learning, Heart Disease Detection, Cardiology Informatics, Predictive Analytics, Health Data Analysis, Algorithmic Diagnosis, Cardiovascular Risk Prediction, Artificial Intelligence in Healthcare, Medical Decision Support Systems, Pattern Recognition in Cardiology*.

1. Introduction

Heart disease remains a leading cause of death globally, posing significant challenges to public health systems. Traditional diagnostic methods, though effective, have limitations in terms of speed, accuracy, and early detection capabilities. With the advent of machine learning, there has been a paradigm shift in how medical data is analysed and interpreted. ML offers the potential to revolutionize heart disease detection by providing tools that can identify subtle patterns and make predictions based on vast and complex datasets that are beyond human analytical capacity.

This review paper examines the current landscape of machine learning techniques in the detection and analysis of heart disease. We begin by outlining the epidemiological impact of heart disease and the necessity for improved diagnostic strategies. The paper then delves into various machine learning models that have been employed in this domain, such as supervised learning algorithms like random forests and support vector machines, and unsupervised approaches like clustering and neural networks, including deep learning. We discuss the integration of ML into clinical workflows, highlighting how these models are trained, validated, and tested using cardiovascular data.

Furthermore, we address the challenges involved in implementing ML solutions, including data quality and availability, algorithmic transparency, and interpretability, as well as ethical considerations. The paper concludes with a look at the future directions for ML in cardiology, emphasizing the importance of interdisciplinary collaboration to enhance the development and application of these technologies in clinical settings, ultimately aiming to improve patient outcomes and reduce the burden of heart disease.

2. Literature Review

Heart disease is a leading cause of death worldwide, and early detection is crucial. This study introduces a new machine learning method to improve heart disease prediction, addressing issues of imbalanced data and high dimensionality. It uses supervised infinite feature selection (Inf-FS) for selecting key features and an Improved Weighted Random Forest (IWRF) model, optimized with Bayesian techniques. Tested on two public datasets, the model showed superior accuracy and performance compared to other methods, significantly enhancing heart disease detection and patient survival prediction [1].

Heart failure is a major cause of death globally, and diagnosing it is particularly difficult in under-developed and developing countries due to a lack of experts and equipment. Many researchers have created intelligent systems for automated heart failure detection, but they often suffer from overfitting, performing well on test data but poorly on training data. This study presents a new diagnostic system that uses a random search algorithm (RSA) for feature selection and a random forest model for prediction, optimized with a grid search algorithm. Tested on the Cleveland dataset, this method achieved 93.33% accuracy and outperformed other models, using only 7 features and being more efficient than conventional random forest models [2].

Heart disease is a significant global health concern, emphasizing the importance of early detection. This study introduces PaRSEL, a novel stacking model combining four classifiers at the base layer and LogitBoost at the meta layer for heart disease prediction. Addressing imbalanced and irrelevant features, PaRSEL integrates dimensionality reduction techniques like Recursive Feature Elimination and data balancing methods including SMOTE and ADASYN. Comparative analysis reveals PaRSEL's superior performance with 97% accuracy, 80% F1-score, precision exceeding 90%, 67% recall, and 98% AUC-ROC score, outperforming standalone classifiers. SHAP analysis further elucidates model insights, aiding in understanding classifier influences on prediction outcomes [3].

Heart failure affects millions globally, necessitating efficient early detection methods. This study introduces a machine learning-based approach to enhance early heart failure detection using patient health data. While medications are common, exercise is also vital in managing heart failure. We compared nine machine learning algorithms and developed a novel Principal Component Heart Failure (PCHF) technique to select key features. By optimizing this method, we created a new dataset with the eight best features. Our proposed decision tree model achieved a perfect accuracy score of 100%, surpassing other models and state-of-the-art studies. This research significantly contributes to improving heart failure diagnosis in the medical field [4].

Diabetes affects multiple organs and requires timely detection and management to prevent complications, such as heart disease. This study introduces an Optimal Scrutiny Boosted Graph Convolutional LSTM (O-SBGC-LSTM), enhanced by the Eurygaster Optimization Algorithm (EOA), to tune hyperparameters for early diabetes detection. This method captures features in spatial and temporal configurations, exploring their co-occurrence relationships, and uses a hierarchical architecture to improve learning efficiency and reduce computational costs. The O-SBGC-LSTM achieved over 98% accuracy, outperforming classic machine learning methods. Additionally, fuzzy-based inference techniques were used to enhance preventive measures through a suggestion table [5].

6. Heart disease detection is critical in managing patient health, particularly on the Internet of Medical Things (IoMT) platform. This study presents the Recursion Enhanced Random Forest with an Improved Linear Model (RERF-ILM) for heart disease detection. RERF-ILM integrates advanced recursive techniques within a random forest algorithm and enhances it with a linear model to improve prediction accuracy. By leveraging the IoMT platform, this method ensures efficient, real-time monitoring and analysis of patient data. The proposed RERF-ILM system demonstrated superior performance in detecting heart disease, offering a robust and reliable tool for early diagnosis and intervention [6].

7. Coronary arteriography (CAG) is a precise but invasive technique for diagnosing coronary heart disease (CHD), making it unsuitable for annual physical exams. To address this, we developed a machine learning-based model that integrates multiple algorithms and validates feature selection using common clinical data from annual check-ups. Our two-level stacking model uses base-level classifiers to inform a meta-level classifier. By calculating Pearson correlation coefficients and maximum information coefficients, we identify classifiers with low correlation, then use an enumeration algorithm to combine the best-performing classifiers. Using the Z-Alizadeh Sani CHD dataset of 303 cases verified by CAG, our model achieved an accuracy of 95.43%, sensitivity of 95.84%, and specificity of 94.44%. This method effectively assists clinicians in distinguishing between normal coronary arteries and CHD [7].

8. Various automated decision support systems using artificial neural networks (ANN) have been proposed for heart disease detection, but they often focus solely on feature preprocessing. This paper addresses both feature refinement and predictive model issues, specifically overfitting and underfitting. Overfitting occurs due to inappropriate network configuration and irrelevant features. To tackle this, we use the χ^2 statistical model to eliminate irrelevant features and an exhaustive search strategy to optimize the deep neural network (DNN) configuration. The proposed hybrid model, χ^2-DNN, shows improved performance compared to conventional ANN and DNN models and other state-of-the-art methods, achieving a prediction accuracy of 93.33%. This suggests that our system can assist physicians in accurately predicting heart disease [8].

9. Heart disease is a major global health concern, with millions of deaths annually. A robust prediction system using Hybrid Deep Neural Networks (HDNNs) combines ANN, CNN, and LSTM architectures to achieve 98.86% accuracy on public datasets. This approach surpasses conventional methods, offering potential for improved medical diagnosis and patient care [9].

10. Early detection of cardiovascular diseases (CVD) like atrial fibrillation (AFib) and heart failure (HF) is vital to prevent complications. Utilizing mechanocardiography (MCG) through smartphone sensors, this study aimed to detect concurrent AFib and acute decompensated HF (ADHF). Employing supervised machine learning, logistic regression, random forest, and extreme gradient boosting achieved high accuracy, with AFib detection at 98% and ADHF at 85%. This approach shows promise for accurate CVD detection using readily available smartphone technology [10].

11. This study introduces the Heart Wave dataset, comprising recordings of nine distinct classes of common heart sounds from various cardiovascular diseases. With 1353 well-labelled recordings, including rare and difficult cases, it offers a reliable reference standard established by experienced cardiologists. The dataset's substantial length and high signal-to-noise ratio facilitate accurate heart sound analysis, making it a valuable resource for training and evaluating machine learning models for automated heart sound classification and diagnosis [11].

12. This study introduces the heart Wave dataset, featuring recordings of nine distinct classes of heart sounds from various cardiovascular diseases. With 1353 well-labelled recordings, including rare cases, it provides a reliable reference standard established by experienced cardiologists. The dataset's substantial length and high signal-to-noise ratio ensure accurate analysis, making it valuable for training and evaluating machine learning models for automated heart sound classification and diagnosis [12].

13. This study introduces LU-Net, a deep encoder-decoder architecture for denoising heart sound signals captured by digital stethoscopes. Trained on a large dataset mixed with physiological noise, LU-Net effectively suppresses ambient and lung sounds, improving signal-to-noise ratio (SNR) by 5.575 dB on average with just 1.32 M parameters. Outperforming current state-of-the-art models, LU-Net enhances classification accuracy by 38.93% on the PASCAL heart sound dataset, demonstrating its potential to enhance computer-aided auscultation systems in noisy, low-resource hospital settings and underserved communities [13].

14. Recent research has highlighted the potential of Data-Efficient Image Transformer (DeiT) models for tasks like speech and image recognition. However, their use in Valvular Heart Disease (VHD) detection remains unexplored. This paper introduces a transfer learning method employing DeiT models pre-trained on image datasets for VHD classification. Additionally, a hybrid Convolution-DeiT (Conv-DeiT) architecture is proposed, combining convolutional blocks with Squeeze-and-Excitation (SE) attention mechanisms to enhance input features before DeiT processing. Evaluation on the Heart Sound Murmur (HSM) database yielded impressive results, with the Conv-DeiT

model achieving an overall accuracy of 99.44%, showcasing the efficacy of DeiT-based transfer learning in heart sound classification, particularly when augmented with convolutional and attention mechanisms [14].

15. Congenital heart defects (CHDs) are a leading cause of infant mortality, with prenatal intervention offering potential risk reduction. However, diagnosis variability exists due to qualitative criteria. This study aims to detect morphological and temporal changes in fetal ultrasound videos of hypoplastic left heart syndrome (HLHS) using deep learning models. A small cohort was enrolled, and videos were processed and segmented into cardiac cycle videos. Five CNN-LSTM models were trained, with the top performers combined into a stacking CNN-LSTM model. This model achieved high accuracy (90.5%), precision (92.5%), sensitivity (92.5%), F1 score (92.5%), and specificity (85%) for both video-wise and subject-wise classification. The study showcases the potential of deep learning in objective CHD assessment during prenatal care [15].

3. Proposed Method

The methodology for this review paper is built upon a systematic survey of the literature concerning the use of machine learning techniques in the detection of heart disease. We adhere to the following steps to ensure a comprehensive and unbiased review:

Identification of Sources: We identify relevant databases and electronic sources, including PubMed, IEEE Xplore, ScienceDirect, and Google Scholar, to search for peer-reviewed articles, conference papers, and important white papers in the field. Selection Criteria: Articles are selected based on predefined inclusion criteria: relevance to ML techniques in heart disease detection, publication within the last ten years, and citation frequency to ensure we include foundational and cutting-edge studies in our review. Search Strategy**:** We use a combination of keywords derived from the heart disease domain and machine learning field to create a comprehensive search string. The keywords include "machine learning," "heart disease detection," "predictive modelling," "cardiovascular risk assessment," and related terms. Data Extraction: From the chosen studies, we extract data regarding ML algorithms used, types of heart diseases explored, datasets and features involved, evaluation metrics, and results obtained.
Quality Assessment: The selected studies undergo a quality assessment to evaluate the robustness of their methodology and the reliability of their findings. Synthesis of Information: We analyse and synthesize the information from the selected studies, comparing and contrasting different ML approaches, discussing their effectiveness, identifying trends, and highlighting significant achievements as well as existing gaps in the research.
Reporting: The findings are systematically presented to facilitate an understanding of the current state of ML in heart disease detection and provide insights into future research directions.

4. CONCLUSIONS

The review of machine learning techniques for the detection of human heart disease has highlighted significant advancements in the field, underscoring the potential of these technologies to augment clinical diagnostics and improve patient prognosis. We have examined a spectrum of ML algorithms, demonstrating their applicability in various scenarios, from electrocardiogram analysis to echocardiography and beyond. These approaches have shown promise in enhancing the accuracy, efficiency, and predictive power of heart disease detection, often outperforming traditional statistical methods.

Despite the promise shown by ML techniques, challenges remain, including the need for large, annotated datasets, the interpretability of complex models, and the integration of ML tools into clinical practice. Addressing these challenges requires a concerted effort that spans beyond technical development to include regulatory considerations, ethical implications, and cross-disciplinary collaboration.

Moving forward, research should focus on the refinement of machine learning algorithms to improve their explainability and reliability, the development of standardized protocols for dataset collection and algorithm validation, and the exploration of ML's role in personalized medicine for heart disease treatment. It is also essential to pursue longitudinal studies to evaluate the long-term efficacy and impact of ML applications in cardiology.

In conclusion, while the journey from data to decision support tools in cardiology is complex, the reviewed work underscores the transformative potential of machine learning in the early detection and ongoing management of heart disease. The continued evolution of these technologies promises to contribute significantly to advancing cardiovascular healthcare and patient outcomes.

5. REFERENCES

[1]. Abdallah Abdellatif, Hamdan Abdellatef, Jeevan Kanesan; Chee-Onn Chow, “Improving the Heart Disease Detection and Patients’ Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest”, Published in: IEEE Access (Volume: 10), Page(s): 67363 - 67372, Date of Publication: 21 June 2022.

[2]. Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor, “An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection”, Published in: IEEE Access (Volume: 7), Page(s): 180235 - 180243, Date of Publication: 07 November 2019.

[3]. Ayesha Noor, Nadeem Javaid, Nabil Alrajeh, Babar Mansoor, Ali Khaqan, “Heart Disease Prediction Using Stacking Model with Balancing Techniques and Dimensionality Reduction”, Published in: IEEE Access (Volume: 11), Page(s): 116026 - 116045, Date of Publication: 18 October 2023.

[4]. Azam Mehmood Qadri, Ali Raza, Kashif Munir, Mubarak S. Almutairi, “Effective Feature Engineering Technique for Heart Disease Prediction with Machine Learning”, Published in: IEEE Access (Volume:11), Page(s): 56214 - 56224, Date of Publication: 30 May 2023.

[5]. B. Ramesh, Kuruva Lakshmanna, “A Novel Early Detection and Prevention of Coronary Heart Disease Framework Using Hybrid Deep Learning Model and Neural Fuzzy Inference System”, Published in: IEEE Access (Volume: 12), Page(s): 26683 - 26695, Date of Publication: February 2024.

[6]. Chunyan Guo, Jiabing Zhang, Yang Liu, Yaying Xie, Zhiqiang Han, Jianshe Yu, “Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform”, Published in: IEEE Access (Volume: 8), Page(s): 59247 - 59256, Date of Publication: 16 March 2020.

[7]. Jikuo Wang, Changchun Liu, Liping Li, Wang Li, Lianke Yao, Han Li, Huan Zhang, “A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease”, Published in: IEEE Access (Volume: 8), Page(s): 37124 - 37133, Date of Publication: 20 February 2020.

[8]. Liaqat Ali, Atiqur Rahman, Aurangzeb Khan, Mingyi Zhou, Ashir Javeed, Javed Ali Khan, “An Automated Diagnostic System for Heart Disease Prediction Based on (x)2 Statistical Model and Optimally Configured Deep Neural Network”, Published in: IEEE Access (Volume: 7),Page(s): 34938 - 34945, Date of Publication: 13 March 2019.

[9]. Mana Saleh Al Reshan; Sam,na Amin, Muhammad Ali Zeb, Adel Sulaiman, “A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks”, Published in: IEEE Access (Volume: 11), Page(s): 121574 - 121591, Date of Publication: 31 October 2023.

[10]. Saeed Mehrang; Olli Lahdenoja; Matti Kaisti; Mojtaba Jafari Tadi; Tero Hurnanen; “Classification of Atrial Fibrillation and Acute Decompensated Heart Failure Using Smartphone Mechanocardiography: A Multilabel Learning Approach”, Published in: IEEE Sensors Journal (Volume: 20, Issue: 14, 15 July 2020), Page(s): 7957 - 7968, Date of Publication: 17 March 2020.

[11]. Sami Alrabie, Ahmed Barnawi, “HeartWave: A Multiclass Dataset of Heart Sounds for Cardiovascular Diseases Detection”, Published in: IEEE Access (Volume: 11), Page(s): 118722 - 118736, Date of Publication: October 2023.

[12]. Samiul Based Shuvo, Shams Nafisa Ali, Soham Irtiza Swapnil, Mabrook S. Al-Rakhami, "CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings", Published in: IEEE Access (Volume: 9), Page(s): 36955 - 36967, Date of Publication: 02 March 2021.

[13]. Shams Nafisa Ali, Samiul Based Shuvo, Muhammad Ishtiaque Sayeed Al-Manzo, "An End-to-End Deep Learning Framework for Real-Time Denoising of Heart Sounds for Cardiac Disease Detection in Unseen Noise",Published in: IEEE Access (Volume: 11), Page(s): 87887 - 87901, Date of Publication: 05 July 2023.

[14]. Talit Jumphoo, Khomdet Phapatanaburi, Wongsathon Pathonsuwan, Patikorn Anchuen,"Exploiting Data-Efficient Image Transformer-Based Transfer Learning for Valvular Heart Diseases Detection", Published in: IEEE Access (Volume: 12), Page(s): 15845 - 15855, Date of Publication: 24 January 2024.

[15]. Tawsifur Rahman, Mahmoud Khatib A. A. Al-Ruweidi, Md. Shaheenur Islam Sumon, "Deep Learning Technique for Congenital Heart Disease Detection Using Stacking-Based CNN-LSTM Models from Fetal Echocardiogram: A Pilot Study", Published in: IEEE Access (Volume: 11), Page(s): 110375 - 110390, Date of Publication: September 2023.