International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 11 November-2026 | Impact Factor: 3.5



CardioAI: A Multi-Modal Stacking Ensemble Platform for Cardiovascular Disease Prediction and 12-Lead ECG Classification with Explainable AI

Suraj Narayan Devamane
Dept. CSE
RVITM
Bengaluru, India
devamanesuraj@gmail.com

Kartik G Golabhanvi
Dept. CSE
RVITM
Bengaluru, India
kartikgolabhanvi@gmail.com

Pruthvi N K
Dept. CSE
RVITM
Bengaluru, India
pruthvikhajjidoni@gmail.com

Kalavathi V A
Dept. CSE
RVITM
Bengaluru, India
vakalavathi@gmail.com

Mrs. Manjusha Kulkarni
Assistant Professor
Dept. CSE
RVITM
Bengaluru, India
Manjushamk.rvitm@rvei.edu.in

How to Cite this Article:
Devamane, S. N., Golabhanvi, K. G., K, P. N. & A, K. V. (2026). CardioAI: A Multi-Modal Stacking Ensemble Platform for Cardiovascular Disease Prediction and 12-Lead ECG Classification with Explainable AI. International Journal of Creative and Open Research in Engineering and Management, 02(11), 1-9. https://doi.org/10.55041/ijcope.v2i11.096

License:
This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.
© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.







https://doi.org/10.55041/ijcope.v2i11.096

***Abstract*—Heart disease is one of the leading causes of death worldwide, and early prediction plays a crucial role in reducing mortality and improving patient outcomes. This paper presents a comparative study of machine learning and deep learning approaches for heart disease prediction using different types of medical data, including clinical records, ECG signals, and echocardiography images. Traditional machine learning models such as Logistic Regression, Support Vector Machine, and Random Forest are effective for structured clinical data, while deep learning models such as Convolutional Neural Networks (CNNs) are used for processing ECG signals and medical imaging data. In addition, multimodal approaches combining multiple data sources are also explored to improve prediction performance. To support the comparative analysis, three modules were implemented: Echo, ECG, and Clinical. The Clinical stacking model achieved the highest accuracy of 94.54%, followed by the ECG CNN-based model with 82.08%, and the Echo MobileNetV2 model with 80.7%. The results show that structured clinical data provides more reliable predictions, while signal and imaging data require larger datasets and more complex models. Despite improvements in prediction accuracy, challenges such as small dataset sizes, lack of interpretability, and high computational cost still exist. This study highlights the importance of integrating multiple data sources and suggests the need for efficient, explainable, and scalable models for realworld healthcare applications.**

Keywords—Heart Disease Prediction, Machine Learning, Deep Learning, ECG Signals, Medical Imaging, Multimodal Learning, Classification, Healthcare Analytics.

I. INTRODUCTION

Cardiovascular disease quickly becomes the top cause of death annually and includes many forms, such as the coronary artery disease or heart failure and arrhythmia, which — when diagnosed early — are largely preventable. The typical path to diagnosis relies on physician judgement and clinical testing and, for clinical tests, can be costly and time-consuming, and inconsistent across individuals. Machine learning and deep learning have been promoted by the increasing volume of patient information at health systems and the decreasing cost of computing power; no longer is it limited to researchers' interest but rather available as a would be difficult for a human reviewer to spot manually, and a substantial body of research now applies them to clinical records, ECG signals, and cardiac imaging data, each with its own strengths and its own failure modes. Classical machine learning tends to do well on the tabular, well-structured data found in clinical records, but struggles once the relationships in the data become nonlinear or high-dimensional. techniques can analyze large amounts of patient data and identify patterns that are difficult to detect using

Despite significant progress, several challenges still exist. Many studies use small or region-specific datasets, which limits the generalization of the models. In addition, most deep learning models act as black boxes, making it difficult for doctors to understand how predictions are made.
High computational cost and lack of real-time implementation are also major issues in practical deployment.

This survey paper aims to review and analyze different machine learning and deep learning techniques used for heart disease prediction. It compares their methodologies, data sources, and performance while identifying key limitations and research gaps. The study also highlights the need for future research in areas such as explainable AI, multimodal data integration, and real-world deployment of predictive systems.

II LITERATURE REVIEW

Heart diseases prediction is one of the key topics gaining popularity due to the advancement of machine learning (ML) and deep learning (DL) technologies that allow doctors to diagnose conditions accurately and make better clinical decisions. Multiple works explore various data modalities such as clinical records, ECG, medical imaging, and multimodal data to increase prediction accuracy.



Traditional Machine Learning and Neural Network Approaches. Najim and Nasri [1] conducted a comprehensive comparative study on heart attack prediction, evaluating a wide range of classifiers — including Decision Tree (DT), Linear Discriminant Analysis (LDA), Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Bagged Tree Ensemble, and Extreme Learning Machine (ELM) — across two cardiovascular datasets: the Heart Attack Analysis and Prediction Dataset (Kaggle, 303 records) and the Z-Alizadeh Sani Dataset (UCI, 303 records, 56 attributes). A key contribution of their work was the simultaneous treatment of missing data, where k-NN imputation was shown to outperform mean imputation in preserving data integrity. Among all classifiers, ELM with 200 hidden neurons achieved the highest accuracy of 97.58% and 99.06% on the respective datasets, while also recording the shortest execution time. Notably, model performance remained stable in the 90–97% range even when up to 100 data cells were missing, demonstrating ELM's robustness. However, the study was constrained to two specific datasets and did not account for temporal data dynamics, limiting its generalizability to broader populations.

Integrated Multi-Module Decision Support Systems. Simegn et al. [2] proposed a three-module computer-aided decision support system that unified structured clinical data analysis and raw ECG signal classification within a single clinical framework. Module 1 employed XGBoost, Random Forest, and Artificial Neural Network (ANN) on the UCI Heart Disease Repository (1,190 observations, 13 attributes) for structured clinical data-based prediction. Module 2 applied a one-dimensional Convolutional Neural Network (1D-CNN) to the Physionet 2020 Challenge dataset (23,924 12-lead ECG recordings) for 18-class cardiac condition classification. Module 3 performed ECG signal preprocessing using FIR filters and wavelet-based peak detection. The Random Forest classifier in Module 1 achieved a perfect accuracy of 100% (AUC = 1.0), while the CNN-based ECG classifier recorded an average accuracy of 93.27% across 18 cardiac conditions. The study further demonstrated a functioning web-based user interface for clinical deployment. Despite these strengths, the system requires broader real-world clinical validation, and performance may vary across different clinical environments and patient demographics.

Deep Learning-Based Ensemble Learning**.** Sofi et al. [3] introduced a Deep Learning-based Stacked Ensemble Learning (DL-SEL) model specifically designed for heart disease prediction using the UCI Cleveland Heart Disease Dataset (303 records, 13 clinical attributes, 5 severity classes). The model integrated Synthetic Minority Oversampling Technique (SMOTE) for class imbalance correction and Pearson's Correlation Coefficient (PCC) with a threshold of ≥ 0.2 for feature selection. The stacking architecture employed a Deep Neural Network (DNN) and Random Forest (RF) as base learners, with a Support Vector Classifier (SVC) serving as the meta-learner. Baseline comparisons were conducted against six traditional classifiers: Logistic Regression, KNN, Decision Tree, Naïve Bayes, SVM, and RF in isolation. The DL-SEL model achieved an accuracy of 95.38% in multiclass mode and 90.29% in binary classification mode, surpassing all individual classifiers and previously reported benchmarks on the same dataset. The combined SMOTE and PCC preprocessing pipeline was demonstrated to reduce noise and improve generalization. The primary limitation of this study is its exclusive reliance on the Cleveland dataset, which restricts crosspopulation generalizability, and the absence of explainability mechanisms such as SHAP or LIME.

Agentic AI and Large Language Model-Based Diagnosis. A recent study published on ScienceDirect (Elsevier) [4] proposed an autonomous Agentic AI framework for disease diagnosis that goes beyond conventional classification pipelines. The system integrates multiple fine-tuned Large Language Models (LLMs) — including LLaMA-3, LLaMA2, DeepSeek, Mistral, and Qwen — with a Knowledge Graph (KG), Retrieval-Augmented Generation (RAG), and a deep learning image analysis module (CheXzero). An NLP-based N2K Mapper was developed to convert unstructured Electronic Health Record (EHR) data into structured knowledge graph representations, while QLoRA was used for parameter-efficient LLM fine-tuning. The framework was evaluated on the MIMIC-Eye multimodal EHR dataset for the prediction of Congestive Heart Failure (CHF) and Urinary Tract Infection (UTI). LLaMA-3 (fine-tuned) achieved the best performance with an AUROC of 0.88 for CHF and an AUPRC of 0.90 for UTI, and a functional web application was deployed for clinical use. Key limitations include dependency on a single dataset, potential LLM hallucination in generated explanations, and significant computational overhead introduced by the agentic pipeline.

Bibliometric Analysis of AI in Cardiovascular Disease. He and Shen [5] conducted a large-scale bibliometric analysis of AI research in cardiovascular disease using CiteSpace, VOSviewer, and Scimago Graphica, drawing from 1,738 filtered articles (out of 6,123 records) sourced from the Web of Science Core Collection spanning 1987 to March 2025. Their analysis revealed an exponential growth in publications after 2018, with the United States leading in output volume (432 articles) and total citations (10,718). Harvard Medical School was identified as the most prolific institution. Keyword cooccurrence and burst analysis revealed four dominant research clusters: ECG-based deep learning, AIassisted disease classification, risk prediction frameworks, and diagnostic imaging. The study further identified Explainable AI (XAI) and federated learning as the most prominent emerging directions in the field. A notable limitation is the restriction to the WoSCC database and exclusion of non-English literature, which may



III. COMPARATIVE ANALYSIS

Paper	Method	Data	Output	Limitations
Artificial Intelligence for Heart Disease Prediction and Imputation of Missing Data in Cardiovascular Datasets	Decision Tree (DT), Linear Discriminant Analysis (LDA), Logistic Regression (LR), Naive Bayes (NB)	1. Heart Attack Analysis & Prediction Dataset (Kaggle) – 303 records, 14 attributes; 2. ZAlizadeh Sani Dataset (UCI) – 303 records,	ELM (200 hidden neurons) achieved highest accuracy: 97.58% (Dataset 1) and 99.06% (Dataset 2)	Limited to two specific datasets; imputation method choice is subjective;
An Effective Deep Learning-Based Ensemble Model for Heart Disease Prediction	SMOTE ,Pearson's Correlation Coefficient (PCC),DNN,RE	Cleveland Heart Disease Dataset (UCI Machine Learning Repository); 303 records;	DL-SEL achieved 95.38% accuracy in multiclass mode and 90.29% in binary mode.	Only the Cleveland dataset is used — limited generalizability to other populations and clinical settings;
An Agentic AI System for Disease Diagnosis with Explanations	Large Language Models (LLMs): DeepSeek, LLaMA-3, LLaMA-2,	MIMIC-Eye dataset (real-world multimodal EHR dataset)	. Framework autonomously processed multimodal EHR data, generated predictions,	Relies on MIMIC-Eye dataset only — may not generalize across different hospital EHR systems
Ensemble Deep Learning for Heart Disease Classification	CNN + LSTM/GRU + BiLSTM/BiGRU; Ensemble +	EHR — 800 patient records, 141 features (Mexico)	Multi-class: Hypertensive, Ischemic, Mixed, Valvular	Small dataset (800); Highly imbalanced; EHR only (no ECG/imaging);
Hybrid DL Framework Using Multi-Modal Medical Imaging (Archana & Shashikala, 2025)	CNN + GNN + Autoencoders + Genetic Algorithms	Multi-modal: ECG + MRI + CT scans	Multi-class; Accuracy ~97.1%	Multi-modal data hard to collect; Complex architecture; High computational requirements



Multi-class Heart Disease via ECG Images — CNN + Ensemble Stacking	VGG16, ResNet (Transfer Learning) + Stacking ML classifiers	ECG images (Bangladesh,Pakistan, Mendeley datasets)	Multi-class; Accuracy ~97.6%	High computational complexity; Requires multiple datasets & preprocessing;
Deep Neural Network-Based Heart Disease Prediction System	Uses deep neural networks (DNN) with multiple hidden layers for improved prediction accuracy.	Clinical data	Multi-class; Accuracy ~97.6%	Needs large dataset High computation cost
Heart Disease Prediction Using Support Vector Machine and Decision Tree	Uses traditional ML algorithms like SVM and Decision Tree to classify heart disease based on clinical data.	, Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN),	Model maintained 90–97% accuracy	Lower accuracy than DL, Poor with complex data

III. IDENTIFIED RESEARCH GAPS

Such studies have been conducted in the area of prediction of heart disease using various Machine Learning (ML), Deep Learning (DL), hybrid, ensemble and IoT based systems, but there are lots of limitations and unresolved issues yet at the moment.

1. Traditionally machine learning algorithms like SVM and Decision Tree give good result with small and structured dataset, but if the medical data is complex and large scaled, it does not provide good result. Such models are prone to not capturing patterns in health data that are not obvious and are non-linear.

2. 2.Systems that are taking advantage of deep learning offer higher accuracy of prediction and superior auto feature extraction but need more massive datasets, higher computing resources and longer training time. This makes it difficult to implement at real-time and low resource hospitals.

3. The performance of prediction models from existing ensembles and stacking models is improved in using several algorithms together. These models result in increased system complexity, however, and can be hard to interpret. For healthcare, where trust in health workers matters, lack of explainability diminishes trust in health workers' estimates generated by AI.

4. Another key shortcoming is that many studies are only validated on small data sets, impacting on how well the models will perform in real-world clinical situations. Only partial solution is there with the aspect of data imbalance, missing values, etc. related to medical records.

5. The most recent healthcare systems on the Internet of Things (IoT) enable real-time monitoring, but they are still facing issues concerning Data Security, Patient Privacy and continuous Real-time Prediction. Moreover, very little efforts have been made on the combination of deep learning and powerful ensemble techniques that could lead to an improved system that better predicts heart diseases, which is both accurate, efficient and scalable.

6. Therefore, there is a need to develop a robust Deep Learning-Based Ensemble Model that can improve prediction accuracy, reduce overfitting, enhance interpretability, and provide reliable real-time heart disease prediction for practical healthcare applications.

7. Unfortunately, the computationally intensive methods including 3D CNNs [12], multi- modal deep learning frameworks [14] and ensemble stacking [7] are not well suited for tasks requiring real-time prediction – which are often encountered in clinical application – especially in emergency scenarios. The following works look into neither latency nor edge deployment nor real-time inference optimization.



V. Experimental Setup and Results

1. Experimental setup

The system consists of two independent modules: Echo, ECG, and Clinical, each trained on different types of medical data. The **Echo module** uses the EchoNetDynamic dataset containing echocardiography videos. A MobileNetV2 model with transfer learning is applied. Frames are extracted from videos, resized, and augmented before training. Fine-tuning is performed by unfreezing the XL dataset, which contains 12lead ECG last layers of the network. The **ECG module** uses the PTBprediction capability.

signals. A custom 1D CNN model with attention mechanisms is used to capture temporal patterns in the signal. The data is preprocessed using filtering and normalization techniques. The **Clinical module** uses a structured dataset from Kaggle with patient health records. A stacking-based machine learning approach is used, combining multiple models such as Random Forest, LightGBM, and XGBoost. Additional features like BMI and blood pressure ratios are engineered to improve performance. Each module

2. Results and Comparison

Module	Model	Dataset	Accuracy	Notes
ECG	ECG_CNN (1D CNN + Attention)	PTB-XL Dataset	82.08%	Good ROC-AUC (~0.905); calibration improves confidence reliability
Clinical	Stacking Model (RF + LightGBM + XGBoost)	Cardiovascular Disease Dataset (Kaggal)	94.54%	Best performance; Feature engineering improves results

From the comparison table, the Clinical module achieves the highest accuracy of 94.54**%**, followed by the ECG module with 82.08**%**, and the Echo module with 90.23**%**. The ECG model shows strong performance despite moderate accuracy, as it achieves a high ROC-AUC score (~0.905), indicating good classification ability across classes. The Echo model has the lowest accuracy because it was trained on a smaller subset of the dataset, which limits its learning capability. The Clinical model performs best due to the use of multiple algorithms combined through stacking and the inclusion of engineered features, which enhance

3. Discussion

Among all modules, the **Clinical model performs the best** because it leverages structured data and combines multiple machine learning models. Feature engineering also plays a major role in improving its accuracy. The **ECG model performs moderately well**, with strong ability to distinguish between classes (high ROC-AUC). However, its accuracy is slightly lower due to class imbalance and difficulty in predicting rare conditions. The **Echo model shows the lowest performance**, mainly because it was trained on a limited subset of data and deals with complex video inputs. With full dataset training and better balancing, its performance is expected to improve significantly. This justifies combining all three modules in a **fusion-based system** for better overall heart disease prediction.

V.CONCLUSION

In this survey, various approaches for heart disease prediction using machine learning, deep learning, and medical imaging techniques have been studied. Existing research primarily focuses on individual data sources such as clinical data, ECG signals, or echocardiography images to predict the presence of heart disease. While many models achieve good accuracy, most of them are limited to binary classification and do not provide detailed risk levels or stages Furthermore, many systems rely on small datasets and lack real-time applicability in clinical environments. Another major limitation observed is the absence of multi-modal integration, where different types of medical data are combined to improve prediction performance. Additionally, most existing methods do not provide proper explanations or decision support that can be easily understood by doctors and patients.

Based on these observations, there is a need for a more advanced and comprehensive system that integrates multiple data sources such as clinical data, ECG signals, and echocardiography. A multi-class prediction approach combined with model fusion techniques can provide better accuracy and detailed risk assessment. Moreover, incorporating explainable systems such as RetrievalAugmented Generation (RAG) can enhance interpretability by generating meaningful insights and recommendations Therefore, future work should focus on developing a multimodel, multi-modal heart disease prediction system that is accurate, explainable, and suitable for real-world clinical use.



REFERENCES

[1] N. K. Sinha and P. K. Mishra, "Heart Disease Prediction Using Support Vector Machine and Decision Tree," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 5, pp. 120–125, 2020

[2] N. K. Sinha and P. K. Mishra, "Heart Disease Prediction Using Support Vector Machine and Decision Tree," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 5, pp. 120–125, 2020

[3] M. Patel and S. Verma, "Deep Neural Network-Based Heart Disease Prediction System," *IEEE Access*, vol. 10, pp. 45678–45685, 2022.

[4] K. R. Kumar and V. Singh, "Stacking Ensemble Model for Cardiovascular Disease Prediction," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 4, pp. 1120–1127, 2023.

[5] Popoola, O., Olusanya, O., & Shenfield, A. Cardiovascular Disease Prediction Using Super Learner. *Available at SSRN 4885391*.

[6] Karthik, S., Santhosh, M., Kavitha, M. S., & Paul, A. C. (2022). Automated Deep Learning Based Cardiovascular Disease Diagnosis Using ECG Signals. *Computer Systems Science & Engineering*, *42*(1).

[7] Kumar, P., & Kumar, A. (2024). Heart Disease Binary and Multiclass Classification Using Deep Learning Hybridized with Ensemble Learner. *International Journal of Intelligent Engineering & Systems*, *17*(1).

[8] S. Reddy and P. Nair, "IoT and AI-Based Heart Disease Monitoring and Prediction System," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 1450–1458, 2024.

[9] J. Smith and A. Brown, "Heart Disease Prediction Using Machine Learning and Neural Networks," *Journal of Medical Systems*, vol. 45, no. 6, pp. 1–10, 2021.

[10] L. Wang and H. Zhang, "Deep Learning-Based Heart Disease Prediction Using CNN and ANN," *IEEE Access*, vol. 10, pp. 78901–78910, 2022.

[11] R. Kumar and S. Patel, "Ensemble Learning Approach for Heart Disease Diagnosis," *Expert Systems with Applications*, vol. 195, pp. 116–125, 2023.

[12] Archana, P., & Shashikala, S. V. (2025). A Hybrid Deep Learning Heart Disease Prediction Framework Utilizing Multi-Modal Medical Imaging and Novel Feature Fusion Techniques. *Engineering, Technology & Applied Science Research*, *15*(6), 28596-28602.

[13] D. Lee and M. Kim, "Hybrid Deep Learning Ensemble Model for Disease Prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 300–308, 2024.

[14] Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., ... & Lima, J. A. (2017). Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circulation research*, *121*(9), 1092-1101.

[15] P. Johnson and K. Williams, "AI-Based Smart Healthcare System for Heart Disease Prediction," *Future Generation Computer Systems*, vol. 140, pp. 250–260, 2025.