IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review On Machine Learning And Deep Learning Approaches For Cardiovascular Disease Prediction: Risk Factors, Diagnostic Approaches, Models, And Future Challenges

Bommaiah Boya[1], Dr.P.Devaraju[2]

1. Research Scholar in Computer Science & Technology, Sri Krishnadevaraya University, Ananthapuramu, Andhra Pradesh-515003, India.
2. Assistant Professor in Computer Science & Technology, Sri Krishnadevaraya University, Ananthapuramu, Andhra Pradesh-515003, India.

Correspondence: Bommaiah Boya

ABSTRACT

Cardiovascular diseases (CVDs) are the foremost cause of death worldwide, covering conditions such as coronary artery disease, arrhythmias, heart failure, congenital defects, and stroke. Their development is influenced by risk factors including hypertension, diabetes, obesity, smoking, high cholesterol, sedentary lifestyle, and genetics. Traditional diagnostic tools like ECG, echocardiography, angiography, and stress testing remain common but are often costly, labor-intensive, and may not detect disease at early stages. Advances in artificial intelligence have introduced machine learning and deep learning methods as efficient alternatives for disease prediction. These models are capable of handling diverse datasets, ranging from clinical records to medical images, and can identify patterns that aid in accurate and timely diagnosis. This paper reviews CVD types, major risk factors, diagnostic methods, and previously applied ML and DL approaches, highlighting their strengths, limitations, and challenges such as data imbalance, lack of interpretability, and limited generalizability. Future research directions emphasize multi-modal data fusion, explainable AI frameworks, and improved neural network designs to build robust and clinically reliable diagnostic tools for cardiovascular diseases.

Keywords: Cardiovascular diseases (CVDs), Electrocardiograms (ECG), Echocardiography, angiography, Machine Learning, Deep Learning.

1. INTRODUCTION

Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels that have become the leading cause of death globally, accounting for nearly one-third of all deaths. These diseases include coronary artery disease, stroke, rheumatic heart disease, and congenital heart defects. According to World Health Organization (WHO) report, in 2022 year 19.8 million people died from CVDs, accounting for approximately 32% of all global deaths. Of these, 85% were due to heart attacks and strokes. Over 75% of CVD deaths take place in low- and middle-income countries where raised blood pressure happens to be amongst the most important risk factors for CVDs [1][2]. In 2016, India reported 63% of total deaths due to NCDs (non-communicable diseases), of which 27% were attributed to CVDs. CVDs also account for 45% of deaths in the 40-69 year age group. Individuals at risk of CVD may demonstrate raised blood pressure, glucose, and lipids as well as overweight and obesity. The burden of CVDs is increasing due to urbanization, lifestyle changes, dietary patterns, and aging populations [3][4]. Despite significant advancements in treatment and prevention strategies, the early and accurate diagnosis of CVDs remains a challenge in clinical practice. Traditional diagnostic methods, such as blood tests, imaging techniques, and ECG monitoring, although effective, often require specialized equipment and skilled professionals, making them less accessible in low-resource settings. This gap has led to a growing interest in applying machine learning (ML) [3][16] and deep learning (DL) [6][11] approaches to complement conventional diagnostic methods. Risk factors for CVDs can be categorized as modifiable and non-modifiable. Modifiable factors include smoking, hypertension, diabetes, obesity, sedentary lifestyle, and unhealthy diet, while non-modifiable factors include age, gender, and genetic predisposition [20][21]. Machine learning algorithms can process and analyze clinical data containing these risk factors to provide predictive insights. By doing so, they help clinicians identify high-risk individuals at an earlier stage, potentially reducing the incidence and severity of cardiovascular events. Unlike traditional statistical approaches, ML models can learn non-linear relationships in data, which enhances predictive performance.

Over the past decade, artificial intelligence (AI) has significantly transformed the healthcare domain. In cardiology, ML and DL methods have been extensively applied to diverse tasks such as heart disease prediction, arrhythmia detection from ECG signals, automated analysis of echocardiograms, and risk stratification for stroke. For instance, convolutional neural networks (CNNs) [4][6] [11] have shown remarkable ability in processing ECG and echocardiogram images to detect structural abnormalities and arrhythmias, while recurrent neural networks (RNNs) have been employed to analyze time-series data from continuous monitoring devices. The use of ML and DL in CVD prediction has demonstrated impressive accuracy compared to traditional models. Studies have reported prediction accuracies ranging from 80% to over 98% using various neural architectures. These models not only predict the presence or absence of disease but also assist in identifying contributing risk factors, offering an interpretable framework for clinicians.

2. LITERATURE REVIEW

We studied various research articles on cardiovascular diseases (CVDs) prediction using artificial intelligence (AI) models as machine learning and deep learning. Here we provide some information the research what they provide in their articles:

Ahsan et al. [2] conducted a study to identify recent trends in heart disease diagnosis using machine learning with imbalanced data. They analyzed 49 papers and observed that deep learning dominates this field, while SMOTE remains a widely used method for handling class imbalance. GAN-based models were also noted for synthetic data generation, despite their high computational cost. The authors suggested that future research should expand experiments with real-time patient data and employ interpretable models to ensure reliable performance in clinical applications.

Azmi et al. [3] compared seven machine learning algorithms—KNN, Logistic Regression, ANN, Naïve Bayes, Decision Trees, SVM, and Random Forest—for cardiovascular disease prediction. Random Forest consistently achieved the highest accuracy. The study highlighted the potential of machine learning to enhance healthcare prediction and treatment processes but emphasized the need for larger datasets and more advanced models for better reliability and scalability.

Swathy et al. [4] presented a survey consolidating studies on cardiovascular disease prediction using machine learning and deep learning. They highlighted that lifestyle factors such as smoking, high blood pressure, high cholesterol, diabetes, poor diet, and stress significantly contribute to cardiovascular risk.

The survey also emphasized that current prediction models require larger datasets and long-term follow-up, suggesting future work should focus on personalized risk assessments using combined machine learning and deep learning approaches.

Bharti et al. [5] proposed comparative methods to evaluate machine learning algorithms for heart disease prediction using 13 clinical features. KNN performed best after preprocessing and normalization, while outlier detection via Isolation Forest further improved performance. The study noted the limitation of a small dataset, suggesting that larger datasets and optimization techniques could enhance predictive accuracy.

Abubaker et al. [6] developed a lightweight CNN model to classify four major cardiac conditions using an ECG image dataset. The model achieved high accuracy and could serve as a feature extractor for traditional machine learning classifiers. The study highlighted its potential to reduce manual errors and save time in clinical practice and suggested optimization of hyperparameters and application to other classification tasks.

Subramani et al. [7] introduced a stacking fusion model for cardiovascular disease prediction. Combining multiple classifiers improved prediction performance and clinical utility. The study emphasized early intervention and preventive healthcare but noted the need for larger, multi-institutional datasets to enhance performance further.

Khan et al. [8] evaluated five machine learning algorithms—Decision Tree, Random Forest, Logistic Regression, Naïve Bayes, and SVM—using patient data from two hospitals in Pakistan. Random Forest achieved the highest accuracy of 85.01% with strong sensitivity and ROC score. Limitations included small and non-diverse datasets, and the authors recommended larger, high-dimensional datasets for improved prediction.

Pasha et al. [9] performed a comparative study of multiple machine learning algorithms for heart attack prediction using a Kaggle dataset. Algorithms included SVM, KNN, Decision Trees, ANN, and TensorFlow-based models. Performance varied depending on preprocessing and model choice. The authors highlighted the need for larger, more diverse datasets to improve real-world applicability and generalization.

Sumwiza et al. [10] developed a classification model for cardiovascular disease using 13 key features selected through correlation and feature importance analysis. Random Forest achieved the highest accuracy of 96% before feature selection and 99% after. The study emphasized validating models on heterogeneous datasets to ensure generalizability.

Bhatt et al. [12] applied the k-modes clustering algorithm to classify heart disease in a dataset of 70,000 patients. Age and blood pressure were binned, and data was split by gender. MLP achieved the highest accuracy of 87.23%. Limitations included reliance on a single dataset and limited clinical variables, with future work suggested to compare with other clustering algorithms and improve generalizability.

García-Ordas et al. [13] proposed a framework combining Sparse Autoencoder and Convolutional Classifier for heart problem prediction using 918 patient records. Feature augmentation via SAE improved performance, and CNN outperformed MLP by 0.6%. The model achieved 90.088% accuracy, surpassing classical classifiers and other ensemble methods, while remaining computationally efficient. The study emphasized clinical importance and recommended further validation.

Hager Ahmed et al. [14] compared distributed machine learning algorithms for stroke prediction using the Healthcare Stroke Dataset. By applying hyperparameter tuning and cross-validation, they enhanced performance, where the Random Forest classifier achieved an accuracy of 90%.

Harsh Agrawal et al. [15] proposed an ensemble approach combining ten classification algorithms for heart failure prediction. Their ensemble model achieved a test accuracy of 85.2% and a recall of 87.5%, showing better performance than individual classifiers.

T. Subba Reddy et al. [16] presented a survey highlighting the real-world applications and principles of machine learning algorithms across multiple domains. Their work provided insight into practical applications but did not include experimental validations, instead offering theoretical directions.

Harshita Puri et al. [17] developed a heart attack prediction model using Support Vector Machine with linear, quadratic, and cubic kernels. Among these, linear and quadratic SVMs demonstrated superior performance, offering higher accuracy in predicting heart attack risk.

Table 1 presents various ML and DL models used predict cardiovascular diseases, used datasets, limitations identified and they given accuracy performances and also figure 1 show the ML and DL model and their accuracy graphically.

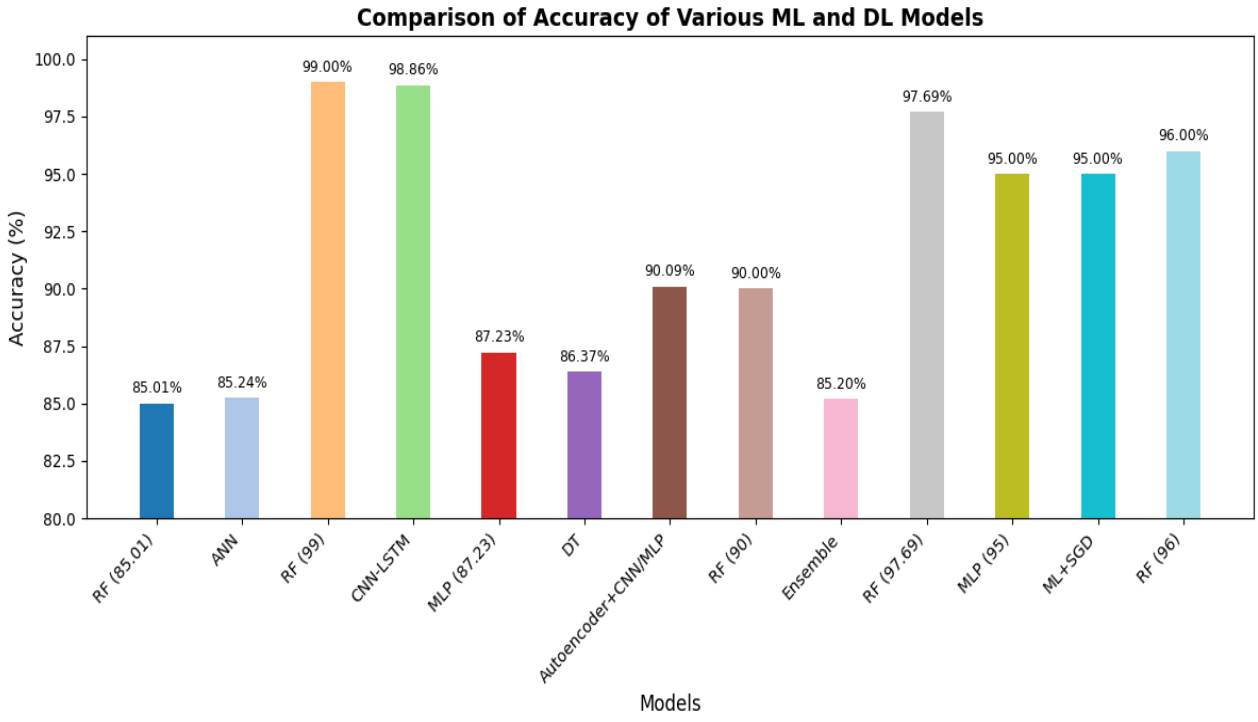

Fig 1: Presents comparison various ML and DL models studied and they given accuracy

Table 1: Presents various existing ML and DL models, used datasets, limitations and their accuracy for predicting CVDs

Author(s)	**Dataset Used**	**Models Used**	**Limitations**	**Accuracy (%)**
Khan et al. [8]	Patient data from 2 hospitals, Pakistan	DT, RF, LR, NB, SVM	Small and non-diverse datasets; limited generalizability	RF: 85.01
Pasha et al. [9]	Kaggle heart attack dataset	SVM, KNN, DT, ANN, TensorFlow	Only one dataset; may not reflect real-world data	ANN: 85.24
Sumwiza et al. [10]	Dataset with 13 selected features	RF, KNN, LR, SVM	External validation needed	RF: 96% before FS, 99% after
Al Reshan et al. [11]	Cleveland dataset + combined HD dataset	Deep ANN, LSTM, CNN, CNN-LSTM	High computational cost; complex architecture	CNN-LSTM: 97.75% (Cleveland), 98.86% (combined)
Bhatt et al. [12]	Kaggle cardiovascular dataset, 70,000 patients	k-modes clustering + MLP	Single dataset; limited clinical variables; no unseen test evaluation	MLP: 87.23%, DT: 86.37%
García-Ordas et al. [13]	918 patient records with 11 features	Sparse Autoencoder + CNN/MLP	Needs further validation; performance sensitive to SAE size	90.088%
Hager Ahmed et al. [14]	Healthcare Stroke Dataset	Random Forest	Focused only on one dataset; lacked external validation.	RF :90%

Harsh Agrawal et al. [15]	Clinical heart failure dataset	Ensemble model (10 classifiers combined)	Computational complexity; limited dataset size.	85.2%
Tarun Madduri et al. [18]	Heart stroke dataset	NB, DT, RF, KNN	Focused on small datasets;	97.69% (RF)
Aditi Galvane et al. [19]	Application-based (demographics + symptoms)	MLP, compared with others	Limited to symptom-based features; no imaging data included.	95% (MLP, SGD)
Govindarajan et al. [20]	Stroke disease dataset	Multiple ML models + SGD	Evaluation restricted to single dataset;	95%
Tahia Tazin et al. [21]	Physiological parameters dataset	RF, others	Limited feature diversity; tested on restricted dataset only	RF : 96%

3. CONCLUSION

Cardiovascular diseases remain the world's leading cause of death, largely influenced by factors such as hypertension, diabetes, obesity, smoking, and sedentary habits. Conventional diagnostic techniques like ECG, echocardiography, and angiography are effective but costly and often miss early detection. Recent progress in machine learning and deep learning has demonstrated remarkable improvements in prediction accuracy. Models including Random Forest, Support Vector Machines, XGBoost, Artificial Neural Networks, and CNN-LSTM hybrids have consistently achieved results between 85% and 99%. This study presents a detailed review literature and also cardiovascular diseases and their diagnosis approaches. Despite these advances, limitations exist. Many previous studies used small or homogeneous datasets like UCI Cleveland and MIT-BIH, reducing generalizability. Class imbalance and high computational requirements also pose challenges for real-world adoption. Future work should focus on building larger, diverse, multi-institutional datasets and integrating multimodal data such as clinical records, ECG signals, echocardiography, and genomic information. Optimization-based feature selection methods like PSO, GWO, and WOA can further enhance accuracy. Incorporating explainable AI will improve trust among clinicians, while lightweight models can extend applications to mobile and low-resource healthcare settings. Real-world validation through clinical trials will be essential to ensure reliability, scalability, and clinical acceptance of these predictive systems.

REFERENCES

1. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. Md Manjurul Ahsan , Zahed Siddique ,"Machine learning-based heart disease diagnosis: A systematic literature review", Artificial Intelligence In Medicine 128 (2022) 102289, March 2022, https://doi.org/10.1016/j.artmed.2022.102289.
3. Javed Azmi , Muhammad Arif , Md Tabrez Nafis , M. Afshar Alam , Safdar Tanweer , Guojun Wang, " A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data", Medical Engineering and Physics 105 (2022) 103825, May 2022, https://doi.org/10.1016/j.medengphy.2022.103825.
4. M. Swathy, K. Saruladha, "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques", ScienceDirect, ICT Express 8 (2022) 109–116, September2021,https://doi.org/10.1016/j.icte.2021.08.021.
5. Rohit Bharti, Aditya Khamparia, Mohammad Shabaz , Gaurav Dhiman, Sagar Pande, and Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Hindawi, Computational Intelligence and Neuroscience, Volume 2021, Article ID 8387680, https://doi.org/10.1155/2021/8387680.
6. Mohammed B. Abubaker and Bilal Babayigit, "Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods", IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, VOL. 4, NO. 2, APRIL 2023, DOI: 10.1109/TAI.2022.3159505.
7. Sivakannan Subramani, Neeraj Varshney, M. Vijay Anand, Manzoore Elahi M. Soudagar, Lamya Ahmed Al-keridis, Tarun Kumar Upadhyay, Nawaf Alshammari, Mohd Saeed, Kumaran Subramanian, Krishnan Anbarasu and Karunakaran Rohini, "Cardiovascular diseases prediction by machine learning incorporation with deep learning", Frontiers in Medicine, April 2017, doi: 10.3389/fmed.2023.1150933.
8. Arsalan Khan, Moiz Qureshi , Muhammad Daniyal, and Kassim Tawiah, "A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction", Hindawi, Health & Social Care in the Community, February 2023, Volume 2023, Article ID 1406060, https://doi.org/10.1155/2023/1406060.
9. Syed Nawaz Pasha, Dadi Ramesh, Sallauddin Mohmmad ,A. Harshavardhan and Shabana, "Cardiovascular disease prediction using deep learning Techniques", IOP Conf. Series: Materials Science and Engineering 981 (2020) 022006, doi:10.1088/1757-899X/981/2/022006.
10. Kellen Sumwiza, Celestin Twizere, Gerard Rushingabigwi, Pierre Bakunzibake, Peace Bamurigire, "Enhanced cardiovascular disease prediction model using random forest algorithm", Informatics in Medicine Unlocked 41 (2023) 101316, August 2023, https://doi.org/10.1016/j.imu.2023.101316.
11. Mana Saleh Al Reshan, Samina Amin, Muhammad Ali Zeb, Adel Sulaiman, Hani Alshahrani, and Asadullah Shaikh, , "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks", IEEE Access, November2023,DOI:10.1109/ACCESS.2023.3328909.

12. Chintan M. Bhatt, Parth Patel, Tarang Ghetia and Pier Luigi Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques”,MDPI, Algorithms 2023, 16, 88, February 2023,https://doi.org/10.3390/a16020088.
13. Marıa Teresa Garcıa-Ordas, Martın Bayon-Gutierrez, Carmen Benavides, Jose Aveleira-Mata, Jose Alberto Benıtez-Andrades, “Heart disease risk prediction using deep learning techniques with feature augmentation”, Multimedia Tools and Applications (2023) 82:31759–31773 , March 2023 , https://doi.org/10.1007/s11042-023-14817-z.
14. Hager Ahmed, Sara F Abd-elGhany, Eman Younis, Nahla Omran “Stroke Prediction using Distributed Machine Learning Based on Apache Spark” International Journal of Advanced Science and Technology Vol. 28, No. 15, (2019), pp. 89-97
15. Agrawal H, Chandiwala J, Agrawal S,Goyal Y (2021) “Heart failure prediction using machine learning with exploratory data analysis”. In 2021 International conference on intelligent technologies (CONIT), IEEE, pp 1-6.
16. T.Subba Reddy, N.Geethanjali, “Real-world Research Applications and Directions of Machine Learning Algorithms”, Design Engineering, 2021, ISSN: 0011-9342, Issue: 9, pp 14256-14275
17. H.Puri, J.Chaudhary, K.R. Raghavendra, R. Mantri and K. Bingi, “Prediction of Heart Stroke Using Support Vector Machine Algorithm,” 2021 8th International Conference on Smart Computing and Communications(ICSCC) ,Kochi,India,2021,pp:21-26
18. Tarun Madduri, Vimala Kumari Jonnalagadda, Jaswitha Sai Ayinapuru, Nivas Kodali & Vamsi Mohan Prattipati, “Heart Stroke Prediction Using Different Machine Learning Algorithms”, Proceeding of World Conference on Artificial Intelligence: Advances and Applications, Algorithms for Intelligent Systems.
19. A. Galvane, G. Kokkula, I. Pandya and K.Devadkar, “ Prediction of Heart Disease Using Machine Learning”, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp:1275-1278.
20. Govindarajan P, Soundarapandian RK, Gandomi AH, Patan R, Jayaraman P, Manikandan R (2020), “Classification of stroke disease using machine learning algorithms”, Neural Comput Appl 32:817-828.
21. Tahia Tazin, Md Nur Alam, Nahian Nakiba Dola, Mohammad Sajibul Bari, Sami Bourouis and Mohammad Monirujjaman Khan,” Stroke Disease Detection and Prediction Using Robust Learning Approaches”, Journal of Healthcare Engineering, Volume 2021, Article ID:7633381.