A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR HEART DISEASE PREDICTION

RAVINDRA CHAUHAN
R.D Engineering College Ghaziabad
ravindra.cs@rdec.in

NITISH VISHWAKARMA
R.D Engineering College Ghaziabad
nitish022ds0040@rdec.in

SHIVAM YADAV
R.D Engineering College Ghaziabad
shivam022ds0051@rdec.in

DIVYANSH SHARMA
R.D Engineering College Ghaziabad
divyansh022ds0017@rdec.in

PALAK PANWAR
R.D Engineering College Ghaziabad
palak022ds0041@rdec.in

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain a leading contributor to global mortality, with the World Health Organization estimating nearly 17.9 million fatalities annually attributed to these conditions. CVDs encompass a broad spectrum of disorders affecting the heart and circulatory system, including conditions such as coronary artery disease, cardiac failure, and irregular heart rhythms. Timely and precise identification of these conditions is vital for enhancing treatment outcomes and lowering death rates. Conventional diagnostic approaches, encompassing clinical assessments, electrocardiographic analysis, and medical imaging, while widely utilized, encounter significant drawbacks such as interpreter subjectivity, inconsistency in results, and poor scalability in settings with limited healthcare resources.

Over the past decade, machine learning (ML) has emerged as a powerful tool within the medical domain, capable of processing large volumes of heterogeneous data to identify hidden patterns beyond human perceptual capacity. ML-based systems can handle a wide variety of input data formats, spanning electronic health records, radiological images, and wearable device outputs, to construct prediction frameworks that support early disease detection, risk classification, and individually tailored therapeutic strategies for cardiac conditions.

This paper undertakes a critical examination of four recently published studies that employ distinct ML methodologies for cardiac disease prediction. These works collectively represent a broad methodological spectrum, encompassing both traditional statistical learning models and modern deep learning and ensemble-based strategies. By investigating their data sources, preparation pipelines, model selection rationale, and performance assessment criteria, this study aims to deliver a thorough synthesis of where ML currently stands in the arena of cardiovascular diagnostics.

Additionally, this paper addresses prominent obstacles encountered in this field, including data reliability issues, skewed class distributions, limited model transparency, and the imperative for validation in clinical environments. Promising research avenues are also explored, including the fusion of multiple data modalities, the use of explainable AI frameworks, and privacy-preserving distributed learning approaches. Through this analytical synthesis, the study intends to equip researchers, healthcare providers, and decision-makers with a nuanced understanding of both the capabilities and constraints of ML-powered cardiac prediction systems, ultimately supporting better-informed clinical decisions and improved patient

management.

II. LITERATURE SURVEY

The studies examined in this review adopt varied methodological frameworks, spanning conventional approaches such as logistic regression and k-nearest neighbor classification to sophisticated deep learning models including LSTM networks and ensemble stacking architectures. A structured comparison of their key parameters is presented in the tables that follow.

This section conducts a systematic analysis of four noteworthy investigations that have explored the use of ML in predicting heart disease. Each study exhibits distinct choices in terms of datasets, data preparation strategies, model architectures, and assessment criteria, illustrating the multifaceted nature of research in this domain.

Regarding data sources, the reviewed studies draw upon datasets of considerably different scales and origins, ranging from locally collected patient records to large publicly accessible repositories including the UCI Machine Learning Repository and Kaggle. These differences in dataset origin and volume underscore the critical role of data representativeness and quantity in building dependable predictive frameworks.

Data preparation constitutes a pivotal phase in any ML pipeline. The selected studies deploy a diverse set of preprocessing strategies: substituting missing values using imputation techniques, applying feature scaling via normalization and standardization, and mitigating imbalanced class distributions through oversampling methods such as SMOTE. Techniques for reducing feature dimensionality, including Principal Component Analysis (PCA) and mode-based clustering, are also employed to streamline the input space and boost model efficiency.

The algorithmic approaches vary considerably, from foundational models such as Logistic Regression and KNN to more sophisticated ensemble architectures and recurrent neural network structures like Long Short-Term Memory (LSTM). This spectrum of methods illustrates how the field is evolving to accommodate increasingly complex data structures while striving to balance model performance with resource requirements.

Each study reports model effectiveness through a set of quantitative metrics including accuracy, precision, recall, F1-score, specificity, and the area under the ROC curve (AUC). Together, these measures offer a holistic perspective on model behavior, particularly in balancing the trade-offs between correctly identifying disease cases and avoiding incorrect classifications.

In summary, this literature survey captures both the strides made and the persistent challenges in applying ML to cardiac disease forecasting. It provides the contextual basis for the detailed tabular comparisons presented in the following subsections.

TABLE 1: DATASET CHARACTERISTICS

A notable degree of heterogeneity exists across the datasets used in the reviewed studies, both in terms of record counts and attribute types. Singh et al. (2024) rely on a compact locally assembled dataset of roughly 300 patient entries, incorporating basic demographic and laboratory test variables. In contrast, Bhatt et al. (2023) work with a substantially larger Kaggle-sourced corpus of 70,000 entries, enabling greater demographic coverage and model generalization potential. Ezat et al. (2025) and Louridi et al. (2021) both draw from established benchmark datasets available through the UCI Machine Learning Repository and Kaggle platforms, with record counts ranging from a few hundred to several thousand. Attribute counts across studies span 12 to 15 variables, incorporating clinical measurements, patient demographics, and lifestyle-related indicators tied to cardiovascular risk.

Paper	Dataset	Size	Features	Source
Singh et al. (2024)	HeartDisease.csv	~300 records	Demographics, clinical tests	Local dataset
Ezat et al. (2025)	Hungarian dataset	490 records	14 features	UCI Repository
Louridi et al. (2021)	UCI + Framingham	303 + 4240 records	14 + 15 features	UCI & Kaggle
Bhatt et al. (2023)	Kaggle CVD dataset	70,000 records	12 features	Kaggle

TABLE 2: PREPROCESSING TECHNIQUES

The data preprocessing workflows adopted in the four reviewed studies are fundamental to guaranteeing data integrity and maximizing model accuracy. Managing incomplete records presents a recurring obstacle. The solutions employed range from straightforward value substitution and proximity-based averaging to more refined procedures such as K-Nearest Neighbor imputation, Random Forest-guided estimation, and Multiple Imputation by Chained Equations (MICE), each designed to mitigate noise and compensate for missing data that

routinely appears in clinical environments. To ensure that attributes are evaluated on a uniform basis, the studies apply a variety of scaling procedures, including z-score standardization, min-max normalization, and feature binning. Dimensionality management is addressed through techniques such as PCA and K-modes clustering. To counteract uneven class distributions the SMOTE algorithm is applied to artificially generate minority class samples.

Paper	Missing Data Handling	Normalization	Feature Selection
Singh et al. (2024)	Imputation, outlier	Standardization	Correlation analysis
Ezat et al. (2025)	Neighbor averaging	Normalization	PCA
Louridi et al. (2021)	Mean, KNN, RF, MICE	Min-Max scaling	SMOTE for imbalance
Bhatt et al. (2023)	Outlier removal	Binning, categorical conversion	K-modes clustering

TABLE 3: ALGORITHMS AND MODELS

The selection of prediction algorithms across these studies reflects the rapidly evolving ML landscape within healthcare informatics. Each study takes a distinct modeling approach, calibrated to its specific dataset and research priorities. Singh et al. (2024) opt for well-established models — K-Nearest Neighbors (KNN) and Logistic Regression — appreciated for their transparency and low computational overhead. Such approaches are well-suited to smaller datasets and provide interpretable baselines for evaluating more sophisticated alternatives. Ezat et al. (2025) leverage Long Short-Term Memory (LSTM) networks augmented with Particle Swarm Optimization (PSO) for hyperparameter tuning. Louridi et al. (2021) investigate ensemble-based approaches, incorporating Support Vector Machines (SVM), XGBoost, AdaBoost, and stacking classifiers. Bhatt et al. (2023) apply a broad suite of algorithms, including Random Forest (RF), Decision Trees (DT), Multilayer Perceptrons (MLP), and XGBoost.

Paper	Algorithms Used	Best Performing Models
Singh et al. (2024)	KNN, Logistic Regression	Logistic Regression (83% accuracy)
Ezat et al. (2025)	LSTM + PSO Optimization	LSTM (99.3% accuracy)
Louridi et al. (2021)	SVM, XGBoost, AdaBoost, Stacking	Stacking (95.8% accuracy)
Bhatt et al. (2023)	RF, DT, MLP, XGBoost	MLP (87.3% accuracy)

TABLE 4: EVALUATION METRICS

Rigorous performance evaluation is essential for determining the clinical suitability of ML models in cardiac disease detection. The reviewed studies employ a combination of metrics to assess model behavior from multiple angles. Accuracy reflects the proportion of total correct predictions. Precision measures how reliably a model's positive predictions correspond to actual disease cases. Recall quantifies the model's sensitivity in identifying genuine disease instances. The F1-score consolidates precision and recall into a single harmonic mean. Specificity captures a model's ability to correctly classify healthy individuals. The AUC offers a threshold-independent summary of discriminatory power.

Paper	Metrics Reported
Singh et al. (2024)	Accuracy, Precision, Recall, F1-Score
Ezat et al. (2025)	Accuracy, Precision, Recall, F-Score
Louridi et al. (2021)	Accuracy, Specificity, Precision, Recall, F-Measure
Bhatt et al. (2023)	Accuracy, AUC

III. DISCUSSION AND RESULT

This section presents a structured analysis of each reviewed study's relative strengths and limitations, contextualizing their contributions to the broader field of ML-based cardiac disease prediction.

Singh et al. (2024) benefit from the use of interpretable, computationally lightweight classical models. Careful attention to data preparation, including missing value handling and anomaly elimination, bolsters prediction reliability within the constraints of a limited local dataset. However, the narrow dataset scope limits the extent to which findings can be extrapolated to broader patient populations. The comparatively

modest accuracy levels suggest that simple models may be insufficient for fully capturing the intricate patterns embedded in cardiac data.

Ezat et al. (2025) demonstrate that the integration of LSTM networks with PSO-driven optimization yields near-flawless classification results. The incorporation of PCA for dimensionality compression and feature normalization further enhances training efficiency and overall model effectiveness. Despite these impressive outcomes, the intensive computational demands of deep learning architectures may hinder real-world deployment. Additionally, the reduced transparency of these models compared to classical alternatives poses challenges for clinical interpretability and regulatory acceptance.

Louridi et al. (2021) show that the stacking classifier approach effectively combines multiple learning algorithms to achieve strong predictive performance. The application of SMOTE to address imbalanced training data further strengthens the model's ability to handle skewed class distributions. On the other hand, ensemble architectures entail higher computational overhead and reduced model transparency. Managing and maintaining multiple interdependent model components can complicate deployment workflows in practical clinical environments.

Bhatt et al. (2023) demonstrate that utilizing an extensive dataset substantially improves the model's capacity to generalize across heterogeneous patient populations. The MLP architecture demonstrates notable effectiveness in modeling complex, non-linear data relationships. Nevertheless, preprocessing steps involving binning and cluster-based transformation risk compressing or losing nuanced information present in continuous variables. Performance outcomes remain below those achieved by deep learning frameworks, indicating that further refinement may be warranted.

Collectively, the four studies illustrate the inherent tension between predictive accuracy, model transparency, dataset scope, and computational feasibility. While deep learning frameworks such as those applied by Ezat et al. (2025) deliver the highest classification accuracy, classical and ensemble-based alternatives present notable advantages in explainability and resource efficiency. Singh et al. (2024) establish the merit of traditional algorithms under constrained data conditions. Louridi et al. (2021) demonstrate that stacking classifiers surpass individual models in performance, while Bhatt et al. (2023) validate the effectiveness of MLP across large and heterogeneous datasets.

A recurring insight across studies is the central role of dataset volume and quality in shaping model efficacy. Limited datasets constrain generalizability, whereas large-scale repositories bolster model robustness. Preprocessing interventions such as SMOTE-based oversampling and PCA-driven compression are shown to be pivotal in managing class imbalance and high-dimensional feature spaces, respectively.

IV. CONCLUSION AND FUTURE SCOPE

Future Research Should Focus on:

- **INTEGRATION OF MULTIMODAL DATA:** Combining clinical, imaging, and wearable sensor data.
- **EXPLAINABLE AI (XAI):** Enhancing interpretability of complex models for clinical adoption.
- **FEDERATED LEARNING:** Addressing privacy concerns by enabling decentralized model training.
- **REAL-TIME PREDICTION SYSTEMS:** Deploying ML models in hospital settings for continuous monitoring.

V. REFERENCES

[1] H. S. Ezat, A. A. Rahim, Z. A. Kakarash, G. A. Omar, A. Al-Slemani, "A Machine Learning Approach for Accurate Diagnosis of Heart Disease: A Comparative Study," Malatya Turgut Ozal Univ. J. Eng. Nat. Sci., vol. 6, no. 1, pp. 13-22, 2025.

[2] S. K. Singh, P. Jain, G. Patil, V. M. K. Pandey, "Heart Disease Diagnosis Prediction Using Machine Learning and Data Mining Techniques," IRE Journals, vol. 7, no. 8, pp. 108-112, 2024.

[3] C. M. Bhatt, P. Patel, T. Ghetia, P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 88, pp. 1-14, 2023.

[4] N. Louridi, S. Douzi, B. El Ouahidi, "Machine learning-basd identification of patients with a cardiovascular defect," Journal of Big Data, vol. 8, no. 133, pp. 1-15, 2021.

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.