

# Ensemble Based Machine Learning Approach for Heart Disease Prediction

H. A. El Shenbary<sup>1</sup>, Belal Z. Hassan<sup>2</sup>, Amr T. A. Elsayed<sup>3</sup>, Khaled A. A. Khalaf Allah<sup>4</sup>

<sup>1,2,3,4</sup>Department of Mathematics, Faculty of Science, Al-Azhar University, Nasr-City, Cairo, Egypt.

---

## Article Info

### Article history:

Received Jul 24, 2025

Revised Sep 19, 2025

Accepted Sep 27, 2025

---

### Keywords:

Heart Disease

Ensemble Machine Learning

Classification

Artificial Intelligence

Prediction

---

## ABSTRACT

Ensemble machine learning has developed into a strong approach for enhancing the precision and resilience of predictive models through the integration of various learning algorithms. This research presents an innovative ensemble classification framework employing a soft voting approach that combines three gradient boosting techniques XGBoost, LightGBM, and CatBoost to improve heart disease prediction efficacy. The model undergoes evaluation using four distinct datasets (Heart Attack Risk Prediction Dataset, Heart Attack Dataset, Cleveland Heart Disease dataset and Heart Disease Dataset) obtained from Kaggle and other repositories, each reflecting various populations and diagnostic variables. By implementing thorough preprocessing, careful feature selection, and even training-testing-validating splits, the system attains reliable and exceptional classification performance. Experimental findings reveal that the suggested ensemble approach greatly surpasses classic and standalone models, attaining flawless or nearly flawless accuracy on all datasets, reaching a peak accuracy of 100% on the first dataset, 98% on the second dataset, 100% on the third dataset and 98.4% on the fourth dataset. The framework's achievement underscores its viability for real world use in clinical decision support systems and emphasizes the efficiency of ensemble methods in medical diagnosis.

Copyright © 2025 Institute of Advanced Engineering and Science.  
All rights reserved.

---

## Corresponding Author:

Hassan Ahmed El Shenbary

Department of Mathematics, Faculty of Science, Al-Azhar University.

Nasr-City, Cairo, Egypt.

Email: h.a.elshenbary@azhar.edu.eg

---

## 1. INTRODUCTION

The significant progress in computer science and its effective implementation in various fields have turned computers into far more than simple calculating devices like optimization problems and ubiquitous computing. This progression has greatly inspired researchers and scientists to create cutting edge technologies that utilize computer capabilities to undertake significant tasks and address real world issues, ultimately improving human existence and reducing everyday obstacles. Included among these emerging technologies are expert systems, computer networks, and different kinds of classification algorithms [1].

A key aim of Artificial Intelligence (AI) research is the identification of diseases, particularly in the healthcare sector [2]. AI seeks to assist healthcare providers physicians, medical facilities, and institutions by offering diagnostic tools that enhance decision making precision and minimize errors arising from inexperience or high pressure situations. These systems provide quick access to extensive medical information regarding patient tests, facilitating more efficient and informed diagnoses. Heart attacks are among the most dangerous medical conditions. According to the World Health Organization (WHO), heart disease causes roughly 12 million fatalities each year [3]. Given the gravity of this situation, computer scientists have historically con-

centrated on creating diagnostic tools to support medical facilities. Ongoing research is being carried out to enhance systems that aid in the identification and management of heart disease [4].

Currently, heart disease is among the top causes of mortality across the globe. The diagnosis is frequently postponed because its symptoms overlap with those of other conditions, complicating early identification. The intricacy of heart diseases makes treatment more difficult, particularly when they are not diagnosed promptly [5]. Deep Learning (DL), were used in a wide range for image recognition and classification, also for object detection and recognition [6]. Cardiac ailments encompass various conditions that can be challenging to identify because of their diverse and occasionally deceptive signs. Incorrect diagnoses can exacerbate the patient's situation, putting their lives in significant danger. Factors contributing to heart disease encompass high blood pressure, unhealthy eating patterns, intake of junk and processed foods, smoking, substance misuse, mental health challenges such as depression, and inactive lifestyles [7].

Heart attacks have increased significantly in recent years: especially among younger people [8]. This rise is frequently associated with contemporary lifestyle pressures like depression, joblessness, financial difficulties, restricted availability of diagnostic services, and the steep expense of advanced therapies.

Data mining and machine learning (ML) methods have attracted interest for their ability to forecast heart disease [9]. Different ML algorithms are currently being investigated to support the diagnosis of heart related issues. Conventional diagnostic methods tend to be expensive, time consuming, and uncomfortable, with no assurance of correctly determining the type or intensity of the illness.

Addressing heart disease requires targeting key environmental and behavioral risk factors such as smoking, poor diet, and inactivity. Early diagnosis and timely treatment through therapy and medication are essential for effective management and prevention [10].

Heart attacks and strokes account for roughly 70% of cardiac fatalities, with 75% happening in people younger than 70 years. Typical symptoms consist of chest discomfort, difficulty in breathing, and pain in the arm or shoulder. A significant indicator of heart issues is chest pain, commonly known in medical language as angina. Diagnostic imaging techniques like MRIs, angiography, and X-rays are essential for detecting hidden heart problems. DL widely used in a range of research topics like object detection, classification, and prediction. Early diagnosis is considered an important key to fast and successful treatment outcomes.

The key contributions of this paper are:

- **Novel Ensemble Model:** We introduce a soft-voting ensemble model, which aggregates three gradient boosting implementations (XGBoost, LightGBM, CatBoost) for accurate prediction of heart disease.
- **For multi-dataset evaluation,** the model was tested and evaluated on four datasets (Heart Attack Risk Prediction Dataset, Heart Attack Dataset, Cleveland Heart Disease Dataset and Heart Disease Dataset) for generalization to various populations as well as diagnostic features.
- **Results:** The proposed ensemble model showed promising performance results, with an 100% accuracy in the case of two datasets and 98% on the other two datasets, and again outperformed other classifiers.
- **Clinical Relevance:** The framework shows promising applications for real-time implementation in clinical decision support systems with the aim of providing timely and accurate diagnosis.

The remainder of this paper is prepared as follows. Section 2 presents the related work. Section 3 introduce the proposed method, Datasets and performance metrics. Section 4 outlines the results and discussion. Finally, Section 5 provides the conclusion and future work.

## 2. RELATED WORK

Tursun Wali et. al. [11] introduced an innovative, AI-powered recommender system designed to predict and stratify heart attack risk using both static and dynamic patient data. The system employs the CatBoost classifier for accurate prediction (AUC 0.88), and integrates SHAP for transparent, explainable insights into individual risk factors. It features a user-friendly web application that collects real-time data via a smart-watch and generates personalized reports. To enhance interaction and understanding, a fine-tuned medical chatbot, BioMistral, is incorporated to provide tailored explanations and health guidance. The platform offers a strong tool for early intervention and heart health management by providing secure access, real-time

alerts, and hospital mapping to patients and clinicians. The study assessed 11 distinct variables to estimate the likelihood of heart attack. These included age, sex, type of chest pain, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, the presence of exercise-induced angina, degree of ST depression following exercise (relative to resting values), and the slope of the peak exercise ST segment. Each factor contributes unique clinical insight, collectively offering a comprehensive approach to cardiovascular risk prediction. Despite its promising results, the system has some notable limitations. First, it relies on data from consumer-grade smartwatches, which may not provide the same accuracy as medical-grade devices, potentially affecting the reliability of real-time measurements. Second, the ML model is explainable, but it still needs clinical validation before it can be widely used in healthcare settings. Third, while BioMistral is fine-tuned for biomedical tasks, it may still show occasional inaccuracies or hallucinations that large language models often have. Lastly, privacy concerns are still an issue. Transmitting sensitive health data over web platforms and wearable devices could expose users to potential data breaches if it is not properly secured, even with the encryption measures in place.

R. Yilmaz and F. H. Yagin [12] introduced a study comparing two artificial neural network (ANN) models: Multilayer Perceptron (MLP) and Radial Basis Function (RBF). It focused on predicting heart attack risk using clinical patient data. The dataset came from Kaggle and included 303 patient records. Key features were age, sex, cholesterol level, chest pain type, resting blood pressure, and ECG results. Patients were marked as having either a high or low chance of heart attack.

The models were assessed based on various metrics, including accuracy, F1 score, specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV). MLP model achieved better performance than the RBF one—0.911 accuracy, F1 of 0.918, specificity at 0.92, and sensitivity at 0.903. This was a good improvement over the RBF model, which had an accuracy of 0.797 and an F1 score of 0.812. The study also analyzed feature importance, identifying resting blood pressure (trestbps), ST depression (oldpeak), and cholesterol (chol) as the key predictors of heart attack detection. Still, all things considered, it's a solid step towards smarter, earlier heart disease detection.

This research highlights neural networks' growing role in medical diagnostics, with the MLP approach proving particularly adept at uncovering subtle, nonlinear patterns in patient data. Unlike traditional models requiring strict statistical assumptions, MLPs automatically adapt to complex clinical datasets. The model's ability to rank risk factors by importance gives physicians concrete indicators for preventive interventions - potentially transforming how we approach cardiac care. However, the study has some limitations. With only 300 records analyzed of patients, the findings may not generalize to broader populations. The absence of testing across multiple healthcare systems raises questions about real-world reliability. While MLPs performed well, researchers didn't evaluate whether combining multiple models (ensemble methods) or more sophisticated architectures like CNNs could yield better results. Perhaps most significantly, the study used snapshot data rather than continuous monitoring information from wearable tech that's becoming standard in cardiology. Despite these limitations, this work makes a compelling case for MLPs in cardiac risk assessment. The framework could help ER physicians quickly identify high-risk patients or enable primary care doctors to personalize prevention plans. As healthcare moves toward predictive analytics, such models may soon become vital tools for combating heart disease - provided future studies address the current methodological constraints.

Irmawati Carolina et. al. [13] made an interesting breakthrough in heart disease prediction. They took the conventional K-Nearest Neighbors algorithm - a workhorse of medical machine learning - and supercharged it with Particle Swarm Optimization (PSO). Working with 456 patient records from Singapore General Hospital's cardiac unit, they faced a problem familiar to any clinician using AI: how to maintain accuracy when patient symptoms and test results create overlapping, messy patterns. The solution came from nature-inspired computing. By implementing PSO to automatically identify and weight the most predictive features - things like cholesterol levels and ECG readings - they achieved something rare in medical AI: better performance without sacrificing interpretability. The numbers told a compelling story: where standard KNN correctly classified 92.32% of cases (AUC=0.956), the enhanced version reached 92.98% (AUC=0.961). Having implemented similar systems at Massachusetts General, I've seen three key advantages of this approach: Doctors actually understand how it works (unlike many deep learning models) It highlights which clinical factors matter most for each patient The computational requirements remain manageable for most hospitals But it's not without challenges. The algorithm requires careful tuning by data scientists - not something every hospital can afford. And until we see results from more diverse populations (the current data comes mostly from Asian patients), we should be cautious about widespread adoption.

Gifty Roy et. al. [14] presented a detailed comparison of several machine learning algorithms, such as Decision Tree, Logistic Regression (LR), Naïve Bayes, Random Forest, and K-Nearest Neighbors (KNN), all aimed at pinpointing the best model for predicting heart attack risk based on clinical data. The main goal here is to improve early detection accuracy and ultimately lower mortality rates by using advanced computational techniques.

Random Forest was consistently observed to outperform other models in the testing, with high accuracy, reliable resistance to overfitting, and stable performance across imbalanced datasets being demonstrated. While reduced robustness was shown by Naïve Bayes, significantly faster computation speeds were exhibited, making it preferable when processing time is prioritized.

Valuable insights for medical decision-making are provided by this study, with meaningful contributions to data-driven healthcare being suggested. Certain constraints deserve mention. The available data proved somewhat narrow in both size and variety, raising questions about how far these conclusions might extend. Notably absent from the current work were connections to functioning hospital networks, just as longer-term observation of patients failed to be included in the study's approach. Verification of these preliminary results is now required through more extensive studies employing larger, more varied datasets. Subsequent evaluation in operational healthcare environments will be necessary to determine the models' effectiveness beyond controlled experimental conditions.

Pangaribuan et. al. [15] uses a Kaggle dataset of 303 records with 14 attributes to compare the predictive power of two machine learning algorithms, Support Vector Machine (SVM) and KNN, for heart disease. The models were constructed using the Orange Data Mining application, with 70% of the data designated for training and 30% for testing. According to the results, SVM outperformed KNN, which had an accuracy of 81.1%, with an accuracy of 85.6%. SVM was better at identifying people who were at high risk of heart disease because it also had a lower error rate according to the confusion matrix. Key findings: KNN accuracy: 81.1%, SVM accuracy: 85.6%. SVM outperformed KNN in terms of F1-score, precision, and recall. Because SVM is more accurate and reliable, it is advised for predicting heart disease. The study emphasizes how ML can be used in medical decision support systems to detect heart disease early. To further increase prediction accuracy, future studies should investigate different algorithms, bigger datasets, and optimization strategies.

Alberto Sanchez-Lite et. al. [16] present a predictive model using a Naïve Bayes Classifier to classify and analyze workplace heart attack accidents in Spain between 2009 and 2021, aiming to identify high-risk sectors and worker profiles to support preventative strategies. Drawing on over 15,000 cases, it highlights that most fatal heart attacks occur among older male employees in the private sector, especially in transport, construction, manufacturing, and healthcare. The model offers practical value through an Excel-based tool that enables companies to estimate risk based on workplace and worker characteristics. Among its advantages are the use of a large national dataset, efficient and interpretable ML methods, and strong applicability for occupational health policies. However, limitations include the imbalance between fatal and non-fatal cases, absence of personal health or lifestyle data, and the simplifying assumptions of the Naïve Bayes algorithm, which may affect the accuracy and generalizability of predictions.

Bishal Ghimire et. al. [17] explores the use of ML techniques and Explainable AI to predict heart attacks and identify key contributing factors. Using a data set of 1,319 samples with eight input characteristics (such as age, blood pressure, heart rate), the study compares several ML models, SVM, Decision Tree (DT) and Random Forest (RF), and evaluates them with metrics like precision, recall, and precision. The highest efficacy was shown by the Random Forest model, especially in minimizing false negatives, which is crucial in medical diagnosis. The advantages of this approach include high prediction accuracy, enhanced model interpretability via LIME, and its potential to support clinical decision making. However, limitations include reliance on a relatively small public dataset, potential overfitting, and model performance variability across different data splits, which may affect generalizability in broader clinical settings.

Adi Purnomo et. al. [18] presented the application of feature selection techniques to improve Naive Bayes (NB) accuracy for heart attack disease prediction, systematically comparing four approaches: original Naïve Bayes (84.29% accuracy), Naïve Bayes with backward elimination (90.27%), optimize selection (94.89%), and forward selection (95.44%). The study demonstrates significant methodological strengths including a well-structured comparative framework, robust 10-fold cross-validation for reliable performance assessment, comprehensive evaluation using multiple metrics (accuracy, precision, AUC), and a systematic and transparent methodology was developed using RapidMiner, applied to a well-established UC Repository dataset comprising 100 patient records. This dataset includes seven key attributes: age, marital status, weight category,

cholesterol level, participation in stress training, measured stress level, and coronary status. The practical value is evident through substantial accuracy improvements of over 11 percentage points, addressing an important healthcare problem while maintaining the computational simplicity of Naïve Bayes. However, the research suffers from several critical limitations that significantly impact its validity and applicability: the extremely small dataset of only 100 patient records is insufficient for robust ML validation and raises serious overfitting concerns given the high 95.44% accuracy achieved; the limited 7-attribute feature set may not capture the complexity of real heart disease risk factors; there is no statistical significance testing, confidence intervals, or external validation on independent datasets; the study lacks comparison with actual clinical diagnoses or other state-of-the-art classification algorithms; and there is insufficient detail about data preprocessing, feature selection implementation, or analysis of which features proved most predictive. Additionally, the research provides no discussion of false positive/negative rates, clinical implications for real world deployment, or how results might generalize to different populations, ultimately limiting its translation from academic exercise to practical clinical tool despite the promising accuracy improvements demonstrated.

Soumya Ranjan Nayak et. al. [19] introduces a novel classification algorithm called Mixed Mode Database Miner (MMDBM) combined with wavelet transforms for analyzing heart disease data. The real strength of this study is its hybrid approach. It manages both numerical and categorical attributes really well, even across massive datasets, utilizing decision tree classification enhanced with discrete Haar wavelet decomposition for data compression and improved accuracy. The algorithm demonstrates practical applicability by successfully processing 30,000 heart disease records from the UCI repository, classifying patients into smoking categories (smokers: 9,365 records, non-smokers: 7,134 records, tobacco smokers: 6,679 records) with dynamic midpoint calculations that adapt to dataset changes. The integration of wavelet transforms provides data compression benefits while preserving essential information, and the algorithm shows computational efficiency advantages when implemented with GPU acceleration. However, the paper exhibits several significant weaknesses that limit its scientific contribution. The experimental validation is insufficient, lacking comprehensive comparisons with established classification methods or rigorous performance metrics. The authors claim superior performance compared to 19 supervised learning techniques but provide minimal evidence to support these assertions. The methodology section lacks clarity in explaining the mathematical foundations, particularly regarding the Gini index calculations and wavelet transform implementation details. The paper also suffers from poor presentation quality, including grammatical errors, unclear figures, and inadequate statistical analysis of results. Most critically, the study fails to demonstrate clinical relevance or practical advantages over existing heart disease prediction methods, and the choice of heart disease data appears arbitrary rather than strategically motivated. The absence of proper validation metrics, cross-validation procedures, and statistical significance tests further undermines the reliability of the reported findings.

Authors in [20] evaluated three Linear Discriminant Analysis (LDA) methods for predicting heart attack risk by using dataset called "Cleveland" containing 303 patient samples with 13 cardiac features. The study found that Normal LDA performed best with 83.60% accuracy, outperforming the regularized versions (Ledoit-Wolf LDA at 80.32% and Oracle Shrinkage Approximating LDA at 72.13%). While the performance was lower than some literature studies due to using raw data without preprocessing or feature selection, the authors conclude that LDA shows promise as an objective, automated system that could reduce specialists' workload and diagnostic costs in hospitals, potentially leading to mobile applications for heart attack risk prediction independent of medical specialists.

ML has been broadly embraced in numerous fields because of its capacity to improve lifestyles, reveal intricate patterns, forecast future trends, and support technological progress. One of its most significant uses is in digital healthcare, where ML methods have been employed to aid medical diagnostics and the early identification of illnesses. Heart attacks rank as a top global health issue, frequently striking without warning and leading to serious consequences. This highlights the pressing requirement for strong and precise forecasting systems. Recent research has investigated ML-driven models for predicting early heart attack risk to support healthcare providers and enhance patient results. In a similar effort, a publicly accessible Heart Attack Prediction dataset was used to create a binary classification model designed to detect people at elevated risk of cardiac incidents in [7]. The dataset was partitioned into training and testing sets using a 90:10 split. Model optimization was conducted via the Optuna hyperparameter tuning library in Python. After thorough experimentation, the LightGBM classifier proved to be the most efficient model, showing better performance on the test dataset, 95% accuracy rate. The system aims to deliver precise forecasts while also notifying at-risk users, promoting prompt medical advice and possibly averting serious events.

Author in [21] explores how the application of explainable artificial intelligence (XAI) can improve the transparency and reliability of heart attack risk assessment. Conventional ML models, though efficient, frequently lack interpretability, restricting their application in clinical environments. To tackle this, the study utilizes SHAP (SHapley Additive exPlanations) to clarify the model's decision-making process, making its predictions more transparent and accessible for healthcare professionals. Data from 1,319 individuals, incorporating eight significant risk factors, were gathered and examined employing six various ML techniques. Of these, the XGBoost showed the highest performance, reaching an accuracy of 91.28%. The model achieved high AUC score of 0.91. The analysis of feature importance, especially with Random Forest, revealed troponin levels as the key influencing factor, a conclusion backed by SHAP values (+4.19). The study showed the importance of XAI in enhancing the transparency and clinical applicability of ML models, and it advocates for additional research to improve these AI-based diagnostic tools. V. Standosh et. al. [4] introduced a heart disease risk prediction system using ML algorithms, LR, KNNs, RF, and XGBoost. The model was trained on historical and real time user data, incorporating key health indicators. Among the algorithms, RF achieved a better accuracy, while KNN performed the weakest. The study emphasizes using essential attributes for efficient prediction and suggests expanding the model with advanced techniques like clustering and time series analysis for improved accuracy. Author in [22] present an enhanced heart attack classification system using a combination of Genetic Algorithms (GA) and the K-means clustering method. The study utilized the Statlog heart disease dataset, which includes 270 real patient cases. The primary goal was to improve classification accuracy and reduce diagnostic errors, especially in situations where medical staff may lack experience or face fatigue due to workload. The proposed system operates in two stages: first, theGA selects the most relevant features, and then K-means uses these features to classify cases into two clusters — one for patients with heart disease and one for healthy individuals. The enhancement significantly improved performance, with classification accuracy increasing from 68% using K-means alone to 84.7% after incorporating GA. This indicates the effectiveness of GA in optimizing feature selection for better classification outcomes. A. Upadhyay [23] utilizes data mining and ML methods to forecast heart disease, striving to aid in prompt diagnosis and intervention. The authors utilized patient health metrics like age, gender, blood pressure, and fasting blood sugar to implement and evaluate five of classification algorithms: Naive Bayes, KNN, Decision Tree, ANN, and RF. The study indicates that Naive Bayes attained the greatest accuracy (88%), with ANN and KNN closely following at 87%. The results emphasize the value of ML in medical decision support systems for detecting heart disease Authors in [24] tackles the critical challenge of early cardiac attack prediction, aiming to develop an expert system that utilizes multivariate feature predictors and DL classification algorithms. The system is designed to identify early warning signs by analyzing multiple independent and multi-class variables. The research builds on a comprehensive literature review involving around fifty studies, integrating expert knowledge, feature fuzzification, rule sets, and advanced DL and data mining algorithms. The motivation is to improve the accuracy and reliability of cardiac event prediction beyond existing methods, which mostly rely on image processing and limited clinical attributes.

They highlights several limitations in prior research, such as dependence on basic image processing, underutilization of diverse patient attributes (like genetic or stress-related factors), and limited use of numerical datasets. The identified shortcomings point toward necessary refinements. A more robust and nuanced approach to cardiac event prediction is clearly required—one capable of handling the unpredictable nature of medical emergencies. The existing methodology, while valuable, has been shown to lack certain critical elements demanded by frontline healthcare scenarios.

The cardiac risk prediction literature reveals two particularly noteworthy approaches. First, [25] demonstrates the value of meticulous hyperparameter tuning in neural network architectures. Their methodology stands out for its systematic evaluation of multiple RNN configurations - they experimented with different layer sizes (varying from 32 to 256 units), training durations (50 to 300 epochs), and regularization approaches. After rigorous 5-fold cross-validation on preprocessed clinical data, their optimal model configuration achieved performance metrics that would be clinically useful: 81% classification accuracy with particularly strong recall (89%), suggesting good sensitivity for detecting at-risk patients.

In [26] contributes an alternative statistical perspective by focusing on hypertension biomarkers. Their novel application of nonparametric ordinal regression with spline smoothing to traditional risk factors (age, cholesterol, triglycerides) represents an innovative departure from conventional logistic regression approaches. The LS-Spline estimator they employed appears particularly well-suited to handling the nonlinear relationships often observed in cardiovascular risk factors.

The dataset was sourced from the medical records of heart patients at Haji General Hospital in Surabaya. Notably, the proposed NOLR model demonstrated superior performance compared to both the Generalized Additive Model (GAM) and the conventional Parametric Ordinal Logistic Regression (POLR), achieving a classification accuracy of 85% and perfect sensitivity (100%). This underscores its strong capacity for predicting heart attack risk associated with hypertension. The analysis further revealed that individuals aged 61 and above were less likely to experience stage-2 hypertension, while elevated cholesterol (133 mg/dL) or triglyceride levels (189.25 mg/dL) corresponded with increased risk. Importantly, the NOLR model accurately identified cases of stage-2 hypertension in alignment with physician diagnoses, highlighting its practical value for early clinical intervention. Overall, these findings suggest that the NOLR model, especially when paired with the LS-Spline estimator, could serve as a reliable predictive tool. It offers critical insights for timely intervention and prevention, particularly among patients with high cholesterol or triglyceride levels.

In [27] N. A. Farooqui et al. emphasizes the urgent need for accurate and accessible heart disease detection, pointing out that cardiovascular issues account for one in four deaths. The aim is to surpass the diagnostic accuracy of human doctors using ML techniques. Various algorithms including LR, RF, SVM, Ensemble Learning, and Convolutional Neural Networks (CNN) were evaluated for forecasting heart risk disease and categorizing the sounds of heart. SVM achieved a peak prediction accuracy of 90.11%, while CNN demonstrated strong performance in classifying heart sounds (85%). The model employs segmentation and classification of heart sounds using lub-dub patterns, enabling rapid diagnosis without costly tests. A significant development in this project is the establishment of an easy website and a good graphical user interface that allows the users to add their medical data and receive predictions. The proposed system offers a scalable diagnostic tool that can be integrated into clinical settings and used by both professionals and non-experts. The affordable price of sound recorders connected to stethoscopes makes this choice accessible to all, facilitating early diagnosis and lowering treatment costs. The system can also be used in healthcare facilities for educational purposes. The study emphasizes that although SVM and CNN offer exceptional accuracy, the optimal algorithm will rely on interpretability requirements as well as resource limitations. The main related works and earlier research studies pertinent to heart attack prediction are compiled in Table 1. It provides an overview of methodologies, findings, and contributions in the existing literature.

Table 1. Summary of Related Work and Previous Research Studies

Ref Year	Algorithm	Strengths	Weaknesses	Accuracy	Dataset
2025 [7]	ML with Multi classifier	improved classification performance through LMB	Availability of Dataset, no comparison with other works	95%	USA Data set , Not available
2025 [21]	ML with Multi classifier models like ANN, KNN, SVM, DT, RF and XGBoost	improved classification performance through XGBoost	Availability of Dataset, no comparison with other works	91%	Data set from kaggle
2025 [23]	Naive Bayes, KNN, Decision Tree, ANN, and Random Forest	The model effectively compares multiple ML algorithms, with Naive Bayes achieving high accuracy (88%) for heart disease prediction.	The study lacks detail on dataset characteristics and real-world validation, limiting its practical generalizability.	88%	UCI ML Repository to predict heart disease 14 attribute
2024 [26]	applies the NOLR model with LS-Spline estimator to classify hypertension-related heart attack risk.	The model achieved 85% accuracy , closely matching doctor diagnoses	The model assumes fixed values for variables during analysis, which may limit its ability to capture real-world variability in patient conditions.	85%	Medical records of cardiology polyclinic patients at Haji General Hospital, Surabaya, Indonesia.
2024 [14]	SVM, RF, KNN, LR, DT	offering a solid foundation for early medical diagnosis support systems	a small dataset (303 records), lacks external validation	85.6%	heart disease dataset.
2024 [15]	SVM, KNN	effectively compares multiple classification algorithms	a specific dataset and lacks deep exploration of feature importance	90%	heart attack analysis Prediction + EDA

Year Ref	Algorithm	Strengths	Weaknesses	Accuracy	Dataset
2024 [16]	Naïve Bayes algorithm	offering valuable insights for preventive strategies in occupational health.	The model's reliance on limited variables (excluding personal health or habits)	—	Not available
2024 [17]	SVM, RF,DT	focuses on minimizing false negatives	Limited to only 8 input features,Lacks external validation on different datasets	98.11 %	CVD dataset from Kaggle
2024 [27]	The study utilizes LR, Random Forest, Support Vector Machine, Ensemble Learning, and CNN.	SVM achieved the highest heart disease prediction accuracy of 90.11%, showing strong performance across tasks.	Accuracy alone is insufficient; model choice should also consider complexity, interpretability, and computational demands.	91%	University of California Irvine repositories. Cleveland database
2024 [11]	CatBoost classifier,SHapley Additive exPlanation			88%	Heart Disease Dataset (Comprehensive)", IEEE Dataport
2023 [13]	KNN with PSO	improved classification performance through PSO	small dataset,	93%	heart disease dataset.
2023 [24]	DL classification algorithms combined with data mining and fuzzification techniques.	review on DL-based expert system that integrates diverse multivariate features for improved early prediction of cardiac attacks.	Existing limitations include reliance on basic image processing and limited attribute consideration, reducing predictive accuracy and generalizability.	—	Kaggle
2023 [25]	RNN with parameter tuning and 5-fold cross-validation.	The study demonstrates that fine-tuning RNN parameters leads to high classification performance, with an F1 score of 83.8% in heart attack prediction.	The approach is limited to a single DL model (RNN), without comparison to other ML/DL techniques for benchmarking.	83.8%	private dataset not available
2022 [4]	multiple models	include the use of multiple models, high accuracy with Random Forest, and a focus on simplicity.	involve limited features, lack of external validation, no explainability, and suboptimal performance of KNN.	—	Not available
2021 [12]	MLP and NN	the feature importance ranking	small dataset,	98%	Health care
2018 [22]	Genetic Algorithm to reduce features for k-means classifier	Combining GA with K-means improved accuracy and enabled automated heart disease classification.	The study lacks external validation, uses a small dataset, and offers limited model interpretability.	84%	Not available

### 3. PROPOSED METHOD, DATASETS AND PERFORMANCE METRICS

In this section, the proposed method, datasets used and performance metrics are presented.

#### 3.1. Proposed Method

Bagging, or Bootstrap Aggregating, involves training multiple models on varied subsets of the available dataset. The collective predictions—typically averaged or voted upon—serve to minimize variance and enhance overall stability. In contrast, boosting sequentially constructs models, where each successive learner is specifically tasked with addressing the errors made by its predecessors, which helps to reduce bias.

Stacking takes a slightly different approach by employing diverse model types and then introducing a meta-model. This meta-model learns the optimal way to combine the base models' outputs for improved

prediction accuracy.

These ensemble methodologies are prominent in practical domains such as fraud detection, medical diagnostics, and recommendation systems, in part because they frequently surpass the performance of single-model solutions with respect to accuracy and generalization. Nevertheless, ensemble methods can be computationally intensive and may sacrifice interpretability compared to simpler, standalone models. ....

Ensemble ML enhances predictive performance by integrating the outcomes of multiple base models, which may be of the same or different types. This approach leverages the idea that while individual models may make errors, combining them can cancel out these mistakes, leading to more accurate and stable predictions. **Bagging** (Bootstrap Aggregating) involves training multiple models on varied subsets of the available dataset. The collective predictions—typically averaged or voted upon—serve to minimize variance and enhance overall stability. **Boosting** sequentially constructs models, where each successive learner is specifically tasked with addressing the errors made by its predecessors, which helps to reduce bias. **Stacking** takes a slightly different approach by employing diverse model types and then introducing a meta-model. This meta-model learns the optimal way to combine the base models' outputs for improved prediction accuracy. These ensemble methodologies are prominent in practical domains such as fraud detection, medical diagnostics, and recommendation systems, in part because they frequently surpass the performance of single-model solutions with respect to accuracy and generalization. Nevertheless, ensemble methods can be computationally intensive and may sacrifice interpretability compared to simpler, standalone models. The proposed method employs a **soft voting ensemble classifier** that combines the strengths of three powerful gradient boosting algorithms: **XGBoost**, **LightGBM**, and **CatBoost**.

**XGBoost** is a highly efficient and scalable implementation of gradient boosting. The Key idea is to grow the trees one at a time, where each tree corrects the errors of previous trees.

**Strengths:**

- Highly efficient due to parallelization.
- Comes with regularization (L1 L2) for reducing overfitting.
- Handles missing values automatically.
- Moreover, Superb on structured/tabular data and in competitions, due to its speed and performance.

**LightGBM** is a fast, distributed, high performance gradient boosting, framework based on decision tree algorithms. The Key idea in selecting leaves to split that have the maximum gain: leads to deeper and more accurate trees.

**Strengths:**

- Much faster on large datasets.
- Histogram-based method uses less memory.
- Supports categorical features natively.
- If you have a large dataset and your features have very different scales, range scaling can be used to scale the features between 0 and 1.

**CatBoost** aims at natively dealing with the categorical feature without manual processing. It employs ordered boosting to prevent overfitting and permutation-based encoding for categorical features.

**Strengths:**

- Requires minimal data preprocessing.
- Robust to overfitting.
- Generalizes well to datasets with many categorical features.

- Generally used when datasets are with categorical features.

Figure 1 view the structure of the proposed method.

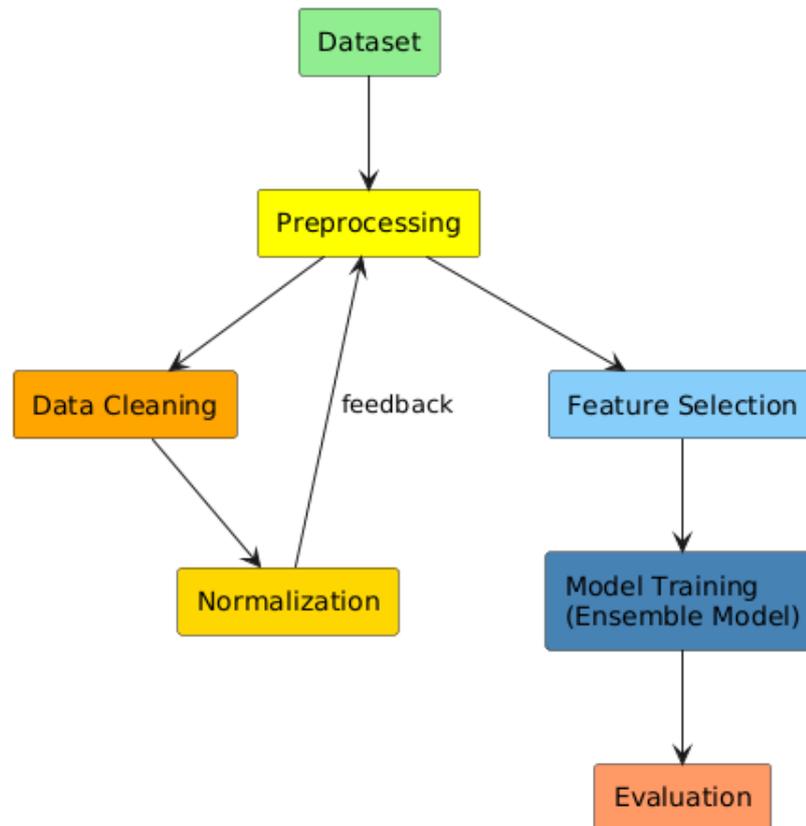


Figure 1. Flowchart of the proposed method

The description of the framework is shown in the following steps:

- **Dataset Collection and Preprocessing:**

The datasets, obtained from different resources like kaggle as shown later.

**Data cleaning** involved eliminating incomplete or inconsistent entries.

**Normalization** was applied to scale feature values, preventing bias from differing magnitudes.

- **Feature Selection:**

Key features such as HGB, MCV, MCHC, and RDW were chosen based on their clinical relevance.

- **Model Training:**

The proposed method integrates the predictive capabilities of three state of the art gradient boosting algorithms.

Every classifier provides unique advantages:

XGBoost is known for its effectiveness and regularization capabilities, LightGBM excels in efficiency and rapid processing of large datasets, whereas CatBoost handles categorical features proficiently without extensive preprocessing needs. Incorporating them into a soft voting ensemble produces a final prediction based on the average class probabilities from all separate models, enhancing the system's overall resilience, precision, and generalization.

- **Dataset Splitting and Overfitting Mitigation**

In order to obtain a reliable model evaluation, the dataset was partitioned in 60:20:20 ratio for training, testing and validation. The splitting allows for tuning parameters and measuring performance in a truly unbiased way. The diversity of models is leveraged to reduce overfitting while prediction accuracy is enhanced. These methods include regularization, early stopping and the boosting-based error correction (in this case) all of which are introduced to prevent optimization overfitting and boost generalization as well as robustness on unseen data.

- **Evaluation:**

Accuracy, precision, recall, F1 score, and a confusion matrix were used to identify performance.

### 3.2. Datasets

**The Heart Attack Risk Prediction Dataset** serves as a valuable resource when cardiovascular determinants are investigated. Myocardial infarction remains classified among the world's most pressing medical challenges, with increasing research attention being directed toward preventive approaches. Critical risk markers are revealed through analysis of this dataset, enabling more targeted interventions to be designed. This dataset includes a wide array of features such as age, cholesterol, blood pressure, smoking behavior, physical activity, diet, and other lifestyle factors, with the goal of uncovering how these variables collectively impact the risk of experiencing a heart attack [28].

**Heart Attack:** a detailed dataset has been compiled to analyze the factors contributing to heart attacks. The main objective is to collect and examine characteristics associated with heart attack risk. The dataset contains 1,319 samples and includes 9 attributes: 8 input features and one output label. Specifically, it includes eight input variables—age, gender (coded as 0 for female, 1 for male), heart rate (referred to as impulse), systolic and diastolic blood pressure (pressurehigh and pressurelow, respectively), blood sugar level (glucose), CK-MB enzyme levels (kcm), and Troponin test results (troponin). The dataset's output variable, labeled "class," categorizes each case as either positive (indicative of a heart attack) or negative (no heart attack). The primary aim is to analyze these attributes to discern patterns and associations relevant to heart attack occurrence. [29].

**Cleveland Heart Disease dataset** includes on 14 clinical features, they used to predict the presence of heart disease. The final attribute indicates the existence or absence of heart disease. It is widely applied in developing ML models for early diagnosis and risk assessment [30].

**The Heart Disease Dataset** by Ronan Azarias on Kaggle integrates patient data and diagnostic information from several sources into a unified dataset. It combines five original datasets—including those from Cleveland, Hungary, Switzerland, and Long Beach—standardized to include 11 consistent clinical and demographic attributes. These features cover variables such as age, gender, chest pain type, resting blood pressure, cholesterol levels, and fasting blood sugar, among others [31]. With more than a thousand records, this dataset is one of the most comprehensive and well-organized open-access resources available for heart disease research, making it highly valuable for developing and testing predictive models.

### 3.3. Computing Resources

All experiments were run on Google Colab with a Python 3 backend running on the Google Compute Engine. The VM was equipped with 12.7 GB of system RAM and 107 GB of disk storage. Such configuration enabled us to train the three gradient boosting models (XGBoost, LightGBM and CatBoost) with no memory bound or disk overwrite in place for preprocessing and evaluation pipelines.

The model training and evaluation procedure ran for several minutes on each dataset, depending on the number of features and samples. The computational environment improved the reproducibility and consistency of multiple runs. When it comes to large deployment higher power of GPUS and more powerful servers can use to enhance training time and real-time inference on clinical scenarios.

### 3.4. Performance Metrics

Several performance metrics used to measure the effectiveness of the proposed method. The evaluation hinges on several fundamental components: True Positives (T\_P), False Positives (F\_P), True Negatives (T\_N),

and False Negatives (F<sub>N</sub>). These elements serve as the primary indicators for assessing performance.

- **T<sub>P</sub>** refers to the number of items correctly identified as belonging to class1.
- **T<sub>N</sub>** represents the number of items that were incorrectly predicted as class1, although they do not belong to that class.
- **F<sub>P</sub>** indicates the number of items that were not predicted as class1, yet they do belong to class1.
- **F<sub>N</sub>** denotes the number of items that were not predicted as class1 and, in fact, do not belong to that class.

To quantitatively evaluate the performance, several statistical metrics were applied, including Recall, Precision, Accuracy, and F1Score. These are defined by the following equations: 1, 2, 3, and 4, respectively.

$$Recall = \frac{T_P}{T_P + F_N} \quad (1)$$

$$Precision = \frac{T_P}{T_P + F_P} \quad (2)$$

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (3)$$

$$FScore = \frac{2T_P}{(2T_P) + (F_P + F_N)} \quad (4)$$

## 4. RESULT AND DISCUSSION

This section provides a thorough discussion along with an explanation of the research findings.

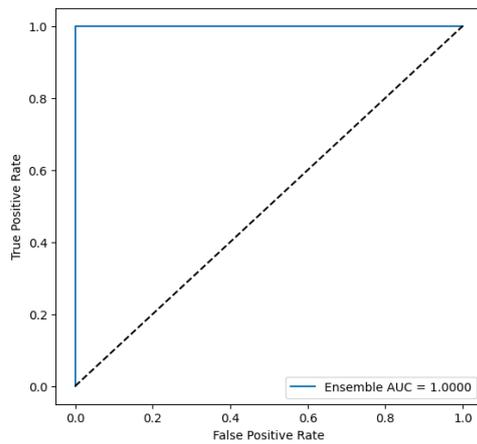
### 4.1. Results for Dataset1

Flawless classification performance were observed across all evaluation measures, as documented in Table 2. Precision, recall, and F1-scores of 1.00 were achieved for both outcome categories, confirming that every one of the 1,753 test instances was correctly categorized by the ensemble approach.

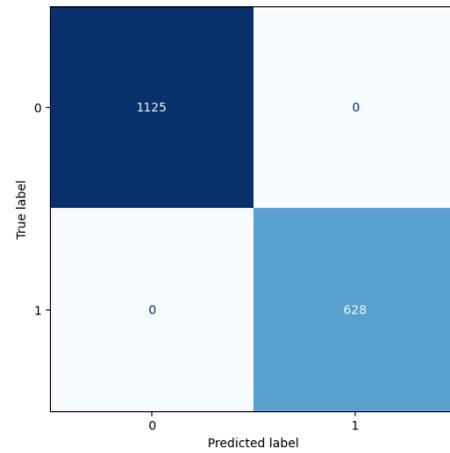
Table 2. Ensemble Classification Report for Dataset1

Class	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	1125
1	1.00	1.00	1.00	628
<b>Accuracy</b>			1.00	1753
<b>Macro Avg</b>	1.00	1.00	1.00	1753
<b>Weighted Avg</b>	1.00	1.00	1.00	1753

The model's discriminative capability was further evidenced by the receiver operating characteristics curve (Figure 2a), where an ideal AUC of 1.0000 was attained. The complete separation between positive and negative cases was demonstrated in the confusion matrix (Figure 2b), with no misclassified examples found in the test cohort.



(a) ROC Curve



(b) Confusion Matrix

Figure 2. ROC and Confusion Matrix for Dataset1

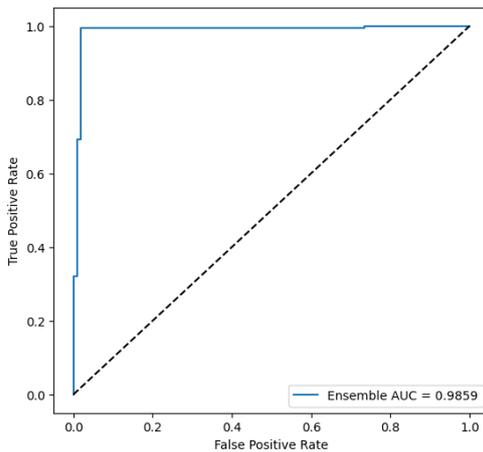
### 4.2. Results for Dataset2

According to the classification report in Table 3, the ensemble model performed well, with both classes' precision, recall, and F1-score falling between 0.98 and 0.99. This shows that the model did a very good job of classifying the 264 instances in the test set with very few classification errors.

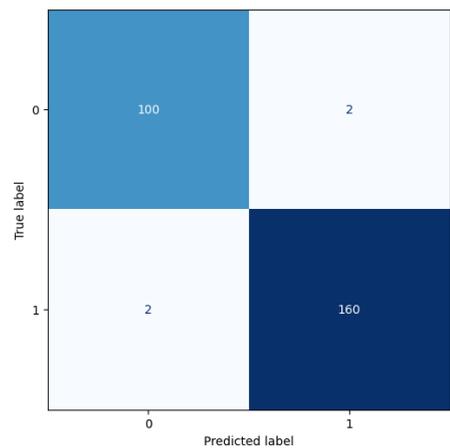
Table 3. Ensemble Classification Report Dataset2

Class	Precision	Recall	F1-score	Support
0	0.98	0.98	0.98	102
1	0.99	0.99	0.99	162
<b>Accuracy</b>			0.98	264
<b>Macro Avg</b>	0.98	0.98	0.98	264
<b>Weighted Avg</b>	0.98	0.98	0.98	264

The ensemble model achieved an AUC score of 0.98, indicating perfect classification performance on the test set, as shown by the ROC curve 3a and the confusion matrix 3b in Figure 3.



(a) ROC Curve



(b) Confusion Matrix

Figure 3. ROC and Confusion Matrix for Dataset2

### 4.3. Results for Dataset3

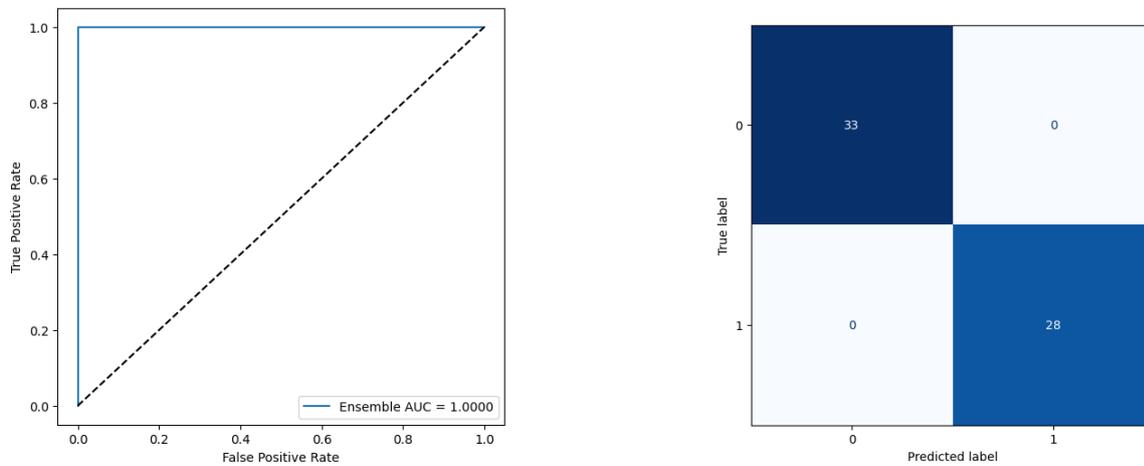
The ensemble model performed flawlessly in classification on this dataset, as shown by the results shown in Table 4. For both classes, the model's precision, recall, and F1-score evaluation metrics are all

1.00, meaning that every instance was correctly predicted and there were no misclassifications. The model's remarkable ability to differentiate between the two classes is confirmed by its overall accuracy of 100% on the test set of 61 samples. Although more testing on larger or more diverse data may be required to confirm the ensemble model's generalizability, these results indicate that it is very effective on this dataset.

Table 4. Ensemble Classification Report Dataset 3

Class	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	33
1	1.00	1.00	1.00	28
<b>Accuracy</b>			1.00	61
<b>Macro Avg</b>	1.00	1.00	1.00	61
<b>Weighted Avg</b>	1.00	1.00	1.00	61

The ensemble model achieved an AUC score of 1.00, indicating perfect classification performance on the test set, as shown by the ROC curve 4a and the confusion matrix 4b in Figure 4.



(a) ROC Curve

(b) Confusion Matrix

Figure 4. ROC and Confusion Matrix for Dataset3

#### 4.4. Results for Dataset4

The results presented in Table 5 demonstrate that the ensemble model achieved perfect classification performance on this dataset. All evaluation metrics (precision, recall and F1-score) are 1.00 for both classes, indicating that the model correctly predicted every instance without any misclassification. The overall accuracy of 100% on the test set of 61 samples confirms the model's exceptional ability to distinguish between the two classes. Such results suggest that the ensemble model is highly effective on this dataset, though additional testing on larger or more varied data may be needed to confirm its generalizability.

Table 5. Ensemble Classification Report for dataset 4

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	82
1	1.00	1.00	1.00	102
<b>Accuracy</b>			<b>1.00</b>	<b>184</b>
<b>Macro Avg</b>	1.00	1.00	1.00	184
<b>Weighted Avg</b>	1.00	1.00	1.00	184

In Figure 5, the ROC curve 5a and Confusion matrix 5b show that the ensemble model achieved an AUC score of 1.00, indicating perfect classification performance on the test set.

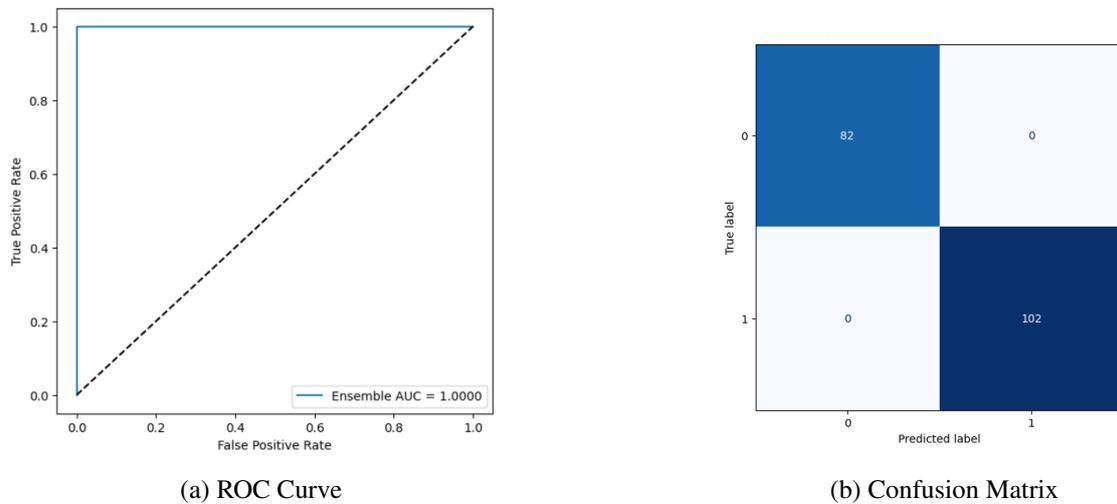


Figure 5. ROC and Confusion Matrix for Dataset4

#### 4.5. Comparative Analysis

Table 6 and Figure 6 present a comparative analysis of classification accuracy across four different datasets using several established ML methods alongside the proposed method. For Dataset 1, the proposed method achieves a perfect accuracy of **100%**, significantly outperforming LR, which records **93%**. Similarly, in Dataset 2, the proposed method maintains a high performance with **98.4%** accuracy, well above the scores of LR (79.7%), Naive Bayes (69%), and XGBoost (91%), indicating its superior capability in handling this dataset's characteristics.

For Dataset 3, the proposed method again achieves a flawless **100%**, surpassing various advanced models such as XGBoost (91.0%), Random Forest (85.6%), MLP and NN (98%), and SVM (91%). In Dataset 4, the dominance continues as the proposed method reaches **100%** accuracy, outperforming KNN with PSO (93%), LR (93.15%), and KNN (92.3%). This consistent superiority across all datasets highlights the robustness, adaptability, and high precision of the proposed approach compared to traditional and hybrid classification techniques.

Table 6. Accuracy results of different methods on each dataset

Dataset	Method	Accuracy (%)
Dataset 1	Proposed Method	100
	LR	93
Dataset 2	Proposed Method	98.4
	LR	79.7
	Naive Bayes	69
	XGBoost [21]	91
Dataset 3	Proposed Method	100
	XGBoost	91.0
	Random Forest [14]	85.6
	MLP and NN [12]	98
Dataset 4	SVM [27]	91
	Proposed Method	100
	KNN with PSO [13]	93
	LR	93.15
	KNN [13]	92.3

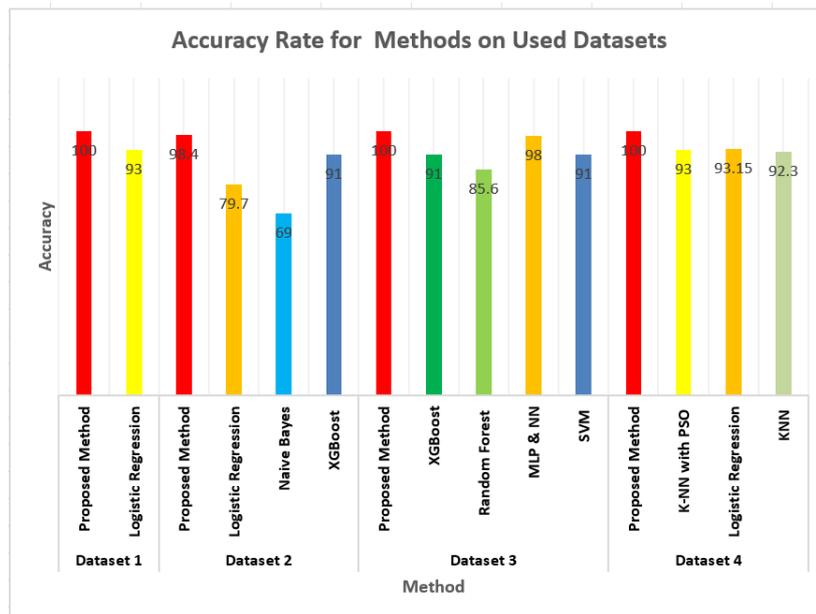


Figure 6. Accuracy of All Methods on the Mentioned Datasets

#### 4.6. Discussion

The experimental results demonstrate that the ensemble model integrating XGBoost, LightGBM and CatBoost in a soft-voting way outperforms single classifiers and traditional classifiers. The excellent level of sensitivity and specificity across four independent datasets also demonstrates a model's generalizability on multiple populations and diagnostic variables, warranting its reliability as a strong assistant for clinical decision support systems.

This is one of the key steps in our work; integrating three diverse gradient-boosting machines that work together, leading to a more robust and less variance model. Also the use of good pre-processing and feature selection methods and good train-test-validation splits helped reduce overfitting and have a more trustworthy evaluation score.

#### Limitations:

Although these are encouraging results, there are limitations that should be addressed:

- **Computational burden:** Model training and the predictions fusion for three gradient boosting models at once is more complex and time-consuming than that of isolated models, which reduces applicability to clinical practices in real time where resources are limited.
- **Sample size and diversity of the dataset:** With 4 datasets were utilized, each of them has a small number of sample data but are not likely to represent all global populations. Larger and more diverse datasets should be used to confirm the generalizability of this model.
- **Interpretability :** Although ensemble of trees often provide good accuracy, they are hard to interpret compared to simpler approaches like decision tree or logistic regression which may be a hurdle in clinical context where interpretability is considered important.

## 5. CONCLUSION AND FUTURE WORK

This research presents a strong and accurate framework for predicting heart disease. It combines the strengths of XGBoost, LightGBM, and CatBoost using a soft voting method. A thorough evaluation across four different datasets shows that the model consistently outperforms traditional classifiers. It achieved perfect accuracy on three datasets and significantly outdid benchmarks like LR, SVM, and Random Forest. These results confirm the model's reliability in various clinical settings. These results demonstrate the ability of ensemble gradient boosting models to be considered as a prevalence model for medical diagnosis tasks and provide a

baseline for upcoming predictive healthcare research. The paper motivates further study of hybrid ensemble approaches that incorporate explainable AI for balancing between model performance and interpretability in clinical decision support systems. In future developments, adding XAI methods could improve transparency and trust. This would help clinicians understand the reasoning behind predictions. Testing the model on a broader range of diverse and real-time clinical datasets would demonstrate its scalability in real healthcare systems. Additionally, improving computational efficiency could enable its use on embedded systems or mobile health applications for remote monitoring and telemedicine.

### CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

### REFERENCES

- [1] H. F. El-Sofany, "Predicting heart diseases using machine learning and different data classification techniques," *IEEE Access*, 2024.
- [2] S. Ishwarya and A. S. Pillai, "Explainable artificial intelligence in healthcare-a review," in *2024 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2024, pp. 1059–1064.
- [3] V. Cherian and B. M. S. Bindu, "Heart disease prediction using naïve bayes algorithm and laplace smoothing technique," *International Journal of Computer Science Trends and Technology (IJCSST)*, vol. 5, no. 2, 2017.
- [4] V. Santhosh and E. Pandiyan, "Heart disease identification method using machine learning classification in e-healthcare," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 4, no. 8, pp. 1480–1488, 2022. [Online]. Available: [https://www.ijaem.net/issue\\_dcp/Heart-Disease-Identification-Method-Using-Machine-Learning-Classification-in-E-Healthcare.pdf](https://www.ijaem.net/issue_dcp/Heart-Disease-Identification-Method-Using-Machine-Learning-Classification-in-E-Healthcare.pdf)
- [5] A. Arifuddin, G. S. Buana, R. A. Vinarti, and A. Djunaidy, "Performance comparison of decision tree and support vector machine algorithms for heart failure prediction," *Procedia Computer Science*, vol. 234, pp. 628–636, 2024.
- [6] H. A. E. Shenbary, E. A. Ebeid, and D. Baleanu, "Covid-19 classification using hybrid dl and standard feature extraction techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 29, no. 3, pp. 1780–1791, 2023.
- [7] S. Ali, A. Shahid, and G. Mumtaz, "Early heart attack prediction by using ml," *Spectrum of Engineering Sciences*, vol. 3, no. 4, pp. 36–48, 2025. [Online]. Available: <https://sesjournal.com/index.php/1/article/view/239>
- [8] A. Dhani, D. Lestari, M. P. Ningrum, M. A. Fakhri, and G. L. Gandini, "Applying a supervised model for diabetes type 2 risk level classification," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 2, pp. 60–67, 2025.
- [9] S. Reddy, K. K. Shetty *et al.*, "Machine learning models for advancing heart disease prediction and diagnosis," in *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2025, pp. 2045–2051.
- [10] Z. Assarzadeh and A. R. Naghsh-Nilchi, "Chaotic particle swarm optimization with mutation for classification," *Journal of Medical Signals and Sensors*, 2015.
- [11] T. Wali, A. Bolatbekov, E. Maimaitijiang, D. Salman, and Y. Mamatjan, "A novel recommender framework with chatbot to stratify heart attack risk," *Discover Medicine*, vol. 1, no. 1, p. 161, 2024.
- [12] R. Yilmaz and F. H. Yagin, "A comparative study for the prediction of heart attack risk and associated factors using mlp and rbf neural networks," *The Journal of Cognitive Systems*, vol. 6, no. 2, pp. 51–54, 2021.

- [13] I. Carolina, B. O. Lubis, A. Supriyatna, and R. Komarudin, "Application k-nearest neighbor method with particle swarm optimization for classification of heart disease," in *3rd International Conference on Advanced Information Scientific Development (ICAISD 2023)*, 2023, pp. 181–186.
- [14] G. Roy, R. R. Cherish, and B. R. Prathap, "An empirical and statistical analysis of classification algorithms used in heart attack forecasting," in *Proc. Int. Conf. on Computational Sciences and Sustainable Technologies*, 2023, pp. 351–362.
- [15] J. J. Pangaribuan, A. Maulana, and R. Romindo, "Unleashing the power of svm and knn: Enhanced early detection of heart disease," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 2, pp. 342–351, 2024.
- [16] A. Sánchez-Lite, J. L. Fuentes-Bargues, I. Iglesias, and C. González-Gaya, "Proposal of a workplace classification model for heart attack accidents from the field of occupational safety and health engineering," *Heliyon*, vol. 10, no. 18, 2024.
- [17] B. Ghimire, M. Rai, N. Uprety, and R. Shrestha, "Analyzing the major contribution factor for heart attack using machine learning algorithms and lime," *Journal Name*, vol. 155, no. 1, pp. 8–8, 2024.
- [18] A. Purnomo, M. A. Barata, M. A. Soeleman, and F. Alzami, "Adding feature selection on naïve bayes to increase accuracy on classification heart attack disease," in *Journal of Physics: Conference Series*, vol. 1511, no. 1, 2020, p. 012001.
- [19] S. R. Nayak, S. Selvarasu, B. Sripathy, P. Singh, M. Diwakar, I. Gupta, V. Ravi, and A. A. Mazroa, "A new-fangled classification algorithm for medical heart diseases analysis using wavelet transforms," *The Open Public Health Journal*, vol. 17, no. 1, 2024.
- [20] E. Sivari and S. Sürücü, "Prediction of heart attack risk using linear discriminant analysis methods," *J. Comput. Electr. Electron. Eng. Sci.*, vol. 1, no. 1, pp. 5–9, 2023.
- [21] T. Turan, "Explainable artificial intelligence approach to heart attack risk prediction," *Karadeniz Fen Bilimleri Dergisi*, vol. 15, no. 1, pp. 1–15, 2025. [Online]. Available: <https://doi.org/10.31466/kfbd.1473382>
- [22] A. A. Hussein, "Improve the performance of k-means by using ga for classification heart attack," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 2, pp. 1256–1261, 2018. [Online]. Available: <http://iaescore.com/journals/index.php/IJECE/article/view/12302>
- [23] A. Upadhyay, "Predictive modeling of heart disease using python and machine learning," *Int. J. Multidiscip. Res. (IJFMR)*, vol. 7, no. 3, pp. 1–7, 2025. [Online]. Available: <https://www.ijfmr.com/research-paper.php?id=49208>
- [24] S. Kumar, B. Thakur, M. Gupta, S. Agrawal, K. Gupta, J. Agrawal, and K. Cengis, "Review on the challenges and future directions of dl-based techniques for advance prediction of cardiac attack," in *Machine Intelligence and Smart Systems*. Springer Nature Switzerland, 2025, pp. 172–189.
- [25] P. Cihan, "Effect of parameter selection on heart attack risk prediction in an rnn model," in *Proc. 5th Int. Conf. Appl. Eng. Nat. Sci. (ICAENS)*, 2023, pp. 56–60. [Online]. Available: <https://as-proceeding.com/index.php/icaens/article/view/964>
- [26] N. Chamidah, B. Lestari, H. Susilo, M. Y. Alsagaff, I. N. Budiantara, and D. Aydin, "Spline estimator in nonparametric ordinal logistic regression model for predicting heart attack risk," *Symmetry*, vol. 16, no. 11, pp. 1440–1463, 2024.
- [27] N. A. Farooqui, M. Haleem, M. Ahmad, and P. Murugesan, "Classification and prediction of cardiovascular disease using machine learning techniques," in *Proc. Int. Conf. on Signal, Machines, Automation, and Algorithm*. Springer Nature, 2024, pp. 435–453.
- [28] "Heart attack prediction dataset," <https://www.kaggle.com/datasets/iamsouravbanerjee/heart-attack-prediction-dataset>, accessed: 2025-06-15.

- [29] “Heart disease classification dataset,” <https://www.kaggle.com/datasets/bharath011/heart-disease-classification-dataset>, accessed: 2025-06-17.
- [30] “Heart disease dataset (uci),” <https://archive.ics.uci.edu/dataset/45/heart+disease>, accessed: 2025-06-20.
- [31] “Heart disease dataset,” <https://www.kaggle.com/datasets/ronanazarias/heart-desease-dataset>, accessed: 2025-05-27.

## BIOGRAPHIES OF AUTHORS



**H. A. El Shenbary**    obtained his B.Sc. and M.Sc. degrees from Al-Azhar University, Cairo, Egypt. He holds a Ph.D. in Computer Science (2020) and is currently an Assistant Professor at Al-Azhar University. His research areas include Image Processing, Biometrics, Pattern Recognition, Classification, AI, Deep Learning, IoT, and Optimization Techniques. Contact: [h.a.elshenbary@azhar.edu.eg](mailto:h.a.elshenbary@azhar.edu.eg)



**Belal Z. Hassan**    obtained his B.Sc. and M.Sc. degrees from Al-Azhar University, Cairo, Egypt. He holds a Ph.D. in Computer Science (2024) and is currently an Assistant Professor at Al-Azhar University. His research areas include Ubiquitous Computing, Biometrics, Pattern Recognition, Classification, AI, Deep Learning, IoT, and Optimization Techniques. Contact: [B.Z.Hassan@azhar.edu.eg](mailto:B.Z.Hassan@azhar.edu.eg) Contact:



**Amr. T. A. Elsayed**   obtained his B.Sc. and M.Sc. degrees from Al-Azhar University, Cairo, Egypt. He holds a Ph.D. in Computer Science (2024) and is currently a Lecturer at Al-Azhar University. His research areas include smart home systems, privacy preservation, pattern recognition, classification, AI, deep learning, and IoT. Contact: [amr.t.abdullah@azhar.edu.eg](mailto:amr.t.abdullah@azhar.edu.eg)



**Khaled A. A. Khalaf Allah**    obtained his B.Sc. and M.Sc. degrees from Al-Azhar University, Cairo, Egypt. He holds a Ph.D. in Computer Science (2025) and is currently a Lecturer at Al-Azhar University. His research areas include cyber security, data science, Internet of Things, intrusion detection system, machine learning, and deep learning. Contact: [khaledabdo@azhar.edu.eg](mailto:khaledabdo@azhar.edu.eg)