Received: 8 July 2025 | Revised: 15 December 2025 | Accepted: 5 March 2026 | Published online: 28 April 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 5(2) 354–370
DOI: 10.47852/bonviewJCCE62026735

BON VIEW PUBLISHING

Enhancing Cardiovascular Disease Risk Prediction: A Comparative Analysis of Machine Learning Techniques

Esraa Eldesouky[1,2,*], Walaa H. Elashmawi[2,3], Ahmed S. Salama[4,5,6], Ahmed Ali[1,5] and Magi Mahfouz[7]

[1] *Department of Computer Science, Prince Sattam bin Abdulaziz University, Saudi Arabia*

[2] *Department of Computer Science, Suez Canal University, Egypt*

[3] *Faculty of Computer Science, Misr International University, Egypt*

[4] *Faculty of Computers and Information Technology, Innovation University, Egypt*

[5] *Higher Future Institute for Specialized Technological Studies, Egypt*

[6] *Department of Computer Engineering and Electronics, Cairo Higher Institute for Engineering, Computer Science and Management, Egypt*

[7] *School of Computing and Digital Technology, ESLSCA University, Egypt*

Abstract: Cardiovascular disease (CVD) remains a global health threat. Accurately assessing CVD risk is crucial for preventative measures and interventions. This paper explores advancements in CVD risk prediction by examining several approaches and trials designed with the aim of diagnosing myocardial infarction (MI), which is generally known as a heart attack. MI is a critical medical condition in which a blocked artery cuts off blood flow and oxygen to a specific part of the heart muscle. This starves the heart tissue, leading to permanent damage and ranking as a top cause of mortality globally. By analyzing vast amounts of patient data, machine learning (ML) algorithms can predict the probability of a heart attack, pinpointing high-risk individuals. This allows for preventative measures and early intervention. The study in this paper utilizes tabular data for risk factors and includes an examination of multiple ML models to improve diagnostics, particularly for high-risk MI. These ML models include random forest classifiers, decision tree classifiers, support vector machines, logistic regression, and gradient boosting (GB). The experimental results reveal that GB has achieved higher accuracy than other models. This provides insights into enhancing cardiovascular health monitoring and diagnosis in clinical settings.

Keywords: machine learning (ML), cardiovascular disease, myocardial infarction (MI), coronary heart disease (CHD), classification approaches

1. Introduction

Myocardial infarction (MI) is a critical life-threatening event that is marked by the insufficient blood supply to the heart, commonly referred to as a heart attack. This results in a lack of oxygen delivery to the heart, leading to the death of heart tissue. This dangerous illness is brought on by coronary artery blockage, which is usually caused by the rupture of an atherosclerotic plaque. MI is a medical emergency that needs to be treated right away to stop further cardiac damage and restore blood flow [1]. According to the definition provided by the World Health Organization (WHO), the presence of myocardial cell necrosis brought on by severe and prolonged ischemia is what characterizes MI. It is usually, but not always, an immediate indication of atherosclerosis-related coronary heart disease (CHD).

MI is caused by two main obstructing mechanisms: CHD, which is defined as a blood flow block caused by plaques in the coronary arteries, or other blocking processes, for example, spasm of plaque-free arteries. The latter is far less common. Every case of plaque development results from atherosclerosis. Underlying stable or unstable plaques may be related to CHD. The vascular wall's triggered inflammation at the plaque site is a characteristic of unstable plaques. The plagues could be eroding, cracking,

***Corresponding author:** Esraa Eldesouky, Department of Computer Science, Prince Sattam bin Abdulaziz University, Saudi Arabia and Department of Computer Science, Suez Canal University, Egypt. Email: em.eldesouky@psau.edu.sa

or even rupturing. At the location of an active plaque, platelets might build up, further preventing blood flow and resulting in unstable angina. Overt MI or acute coronary syndromes are typically the result of atherosclerotic plaque rupture. Atherosclerotic plaques typically expand in stages rather than slowly. The healing process adds a layer to the plaque after platelets build up on the surface; this layer may eventually become hardened, brittle, and lipid laden. When a minor coronary event develops clinically, it may escalate to potentially fatal conditions or even cause sudden death. Those who make it through the first incident are susceptible to MI attacks in the future. As previously mentioned, standardizing data on the distribution of MI in a community can yield valuable insights into the prevalence of coronary artery disease (CAD) within that population. If standardized information on incidents, recurring episodes, and sudden cardiac death can be gathered for MI, the entire burden can then be calculated [2]. The WHO had a significant role in the creation and acceptance of the recognized guidelines for the diagnosis of MI and CHD [2]. Risk factors are used by doctors all over the world to diagnose heart attacks. While risk factors by themselves cannot determine whether a patient has had a heart attack or not, they are a very good indicator that additional evaluation of the patient's condition is needed. Patient demographics, medical history, and risk factors such as genetics, diabetes, and hypertension are commonly included in these statistics. This data helps understand the circumstances of MI and evaluate the overall risk. Additionally, tabular datasets contain biomarkers frequently used to diagnose MI. In tabular datasets, excessive levels of these markers may indicate myocardial injury. Also, heart tissue loss is the direct result of MI, but several other issues can develop both during and after the event. The patient's prognosis and general cardiovascular health are greatly affected by these problems [1, 2].

A coronary artery that supplies the heart with blood and oxygen gets blocked, resulting in a heart attack. These arteries develop plaques due to the buildup of fatty cholesterol-containing deposits over time. A blood clot that forms because of a plaque rupture may obstruct an artery and result in a heart attack. The cardiac muscle tissue dies due to the reduced blood supply [1, 3]. The sophisticated relationship among acute risk circumstances, plaque vulnerability, and the advancement of atherosclerosis highlights the complexity of acute cardiovascular events, especially myocardial infarction (MI). According to the information supplied, atherosclerosis develops gradually over time under the influence of both acute and chronic risk factors, leading to the formation of fragile plaques prone to breaking or erosion. These weak plaques provide the foundation for intravascular thrombosis, which causes acute events like MI. In addition, a wide range of triggers can cause plaque instability and ensuing thrombotic events, including biomechanical pressures, mental stress, and drug usage. To fully understand the mechanisms underlying acute cardiovascular events, it is imperative to comprehend the dynamic processes determining plaque vulnerability and the significance of acute risk factors [4–6]. In developed and developing countries, one of the primary causes of death for women is coronary heart disease, or CHD. The incidence of CHD in women is significantly lower than in males before the age of fifty, but as women age, this tendency changes, and by the eighth decade of life, CHD rates in women begin to resemble those in men. The Framingham study identified women's CHD risk factors, but it only included white Americans of Caucasian descent, and it did not explain why women get MI later than men. The Framingham study's assessment of a restricted set of risk factors may be the cause of this disparity. However, it is generally accepted that women's delayed onset of MI may be caused by significant roles that could also be played by dietary changes, smoking behaviors, and the protective effects of female hormones [7]. Also in Reference [8], the study showed that MI in young adults (<45 years) is distinct from older patients, featuring different plaque characteristics, risk issues, and genomic variations. Heart attack symptoms usually vary greatly, from mild to severe, and other people may not even have any symptoms at all. Chest pain or discomfort, which might feel like pressure, tightness, squeezing, or aching, is a common symptom. Moreover, different body parts such as the shoulder, arm, upper abdomen, neck, back, jaw, or teeth may all experience radiating pain. Cold sweat, exhaustion, heartburn, dizziness, nausea, and shortness of breath are some other symptoms. Unusual symptoms in women could include sudden, severe pain in the back, arm, or neck. A first indication of a heart attack, which may occasionally occur, is the sudden cardiac arrest [9]. While numerous people exhibit symptoms or warning signs from days to weeks before enduring a heart attack, others do not. Angina, or continuous pain in the chest that persists even after rest, may indicate a higher risk of having a heart attack or might lead to an acute reduction in the blood flow of the heart. The variables that increase the likelihood of having a heart attack can be listed as follows:

1) Age: Males aged 45 years and older and females aged 55 and older are at higher risk compared to younger individuals.
2) Smoking: Smoking and experiencing secondhand nicotine increases the risk. Quitting smoking can reduce this risk.
3) High blood pressure: Over time, high blood pressure can damage the heart's arteries. Risk increases when combined with obesity, high cholesterol, or diabetes.
4) High triglycerides and low-density lipoprotein (LDL) cholesterol: Higher levels of triglycerides and LDL ("bad") cholesterol increase the risk, while higher levels of high-density lipoprotein (HDL) ("good") cholesterol may reduce it.
5) Obesity: Obesity raises the risk of several conditions including diabetes, high blood pressure, and high cholesterol.
6) Diabetes: Uncontrolled blood sugar increases the risk of heart attacks.
7) Metabolic syndrome: Conditions such as central obesity, high blood pressure, high cholesterol, and elevated blood sugar greatly increase the risk.
8) Family history: A family history of early heart attacks increases an individual's risk.
9) Physical inactivity: Regular exercise improves heart health, in contrast to an unhealthy lifestyle, which raises the risk.
10) Unhealthy diet: Diets high in processed foods, sweets, trans and saturated fats, and salt raise the risk.
11) Stress: Emotional stress, especially intense rage, may increase the risk of a heart attack.
12) Illegal drug use: Stimulants like cocaine and amphetamines can cause coronary artery spasms, leading to heart attacks.
13) Past preeclampsia: Pregnancy-related high blood pressure increases the probability of heart disease.
14) Immunity disorders: Disorders like lupus or rheumatoid arthritis can increase the risk of heart attacks.

Machine learning (ML) is an expansive era that includes many diverse tools and strategies used in several industries, including healthcare. Its usage in medicine has been shown to be extremely beneficial, particularly in resolving diagnostic problems in various medical fields, such as wearable sensors, cancer

diagnosis, and medical imaging. The ML assists in the prediction and classification of diseases and the evaluation of their developmental stages by analyzing important clinical indicators and extracting medical insights. This allows for more efficient patient care and assistance. Additionally, ML facilitates more effective healthcare tracking by analyzing data and sending out intelligent notifications as needed [10]. Patient diagnoses are recorded as medical records in specialty healthcare settings, where they serve as the basis for ML systems. The accurate encoding of patient data plays an effective role in the use of ML techniques. While encoding might appear simple, it is a crucial initial step in ML systems, allowing them to automatically analyze data and make connections with issues that have already been solved. As a result, the ML enables doctors to quickly and accurately diagnose new patients. Additionally, ML democratizes medical diagnosis by making its advantages available to both nonspecialists and students, improving access to high-quality healthcare services [10]. A wide range of medical and healthcare issues can be addressed with the use of ML. The next section will discuss some related works in which ML has been successfully implemented. These fields will include works that researched medical imaging analysis, others that implemented wearable sensor technologies, and others that offer perspectives based on the types of medical problems addressed and ML techniques used, which are derived from earlier research contributions. This thorough investigation is included to clarify the various and significant ways that ML advances healthcare solutions in several fields [10]. The purpose of this study is to give doctors a tool as a second opinion to help them detect MI by utilizing ML algorithms, and it is not intended for public usage without a professional's help as this tool cannot and will not reduce the doctor's role or replace him at all; it is just there to help professionals as ML algorithms when applied to tabular datasets have showed promise in improving the diagnostic accuracy of MI and predicting consequences accurately. The rest of this paper is as follows: Section 2 discusses some of the research and technologies used for heart attack diagnostics. Section 3 discusses the proposed model architecture for the classification of MI. Section 4 shows the results obtained from the model. Section 5 discusses the results of the model and the further work to be done.

2. Related Works

Considering the intense interest in artificial intelligence (AI), this paper will showcase and analyze several ML and AI applications in the healthcare industry. More specifically, it will highlight studies and publications that have used the benchmark dataset underlying our work. The quantum fusion model (QFM) is presented by Pallavi et al. [11] to increase the quality of bioactivity forecasting for treatments of heart disease. The encodings of molecular data into quantum states were completed through the use of quantum random forests on the ChEMBL database. Consequently, the logistic regression (LR) classifiers were trained on the encodings of this data. The QFM merges the quantum-inspired algorithms along with conventional ML, which achieved an accuracy of 92.7% in classifying bioactive compounds. It thus outperformed both the separate models and existing methods. The QFM demonstrated considerable precision, recall, and F1-score of 0.92, 0.93, and 0.92, respectively. Moreover, the receiver operating characteristic (ROC) area under the curve (AUC) was 96.1%, and the precision–recall AUC was 95.9%. This enhances the ability of the model to accurately recognize the sophisticated structures of suitable molecular groups. The study by Ahamad et al. [12] employs six ML algorithms on two datasets—UCI Kaggle Cleveland dataset and a comprehensive dataset spanning many locations—to handle the difficult challenge of early cardiac disease prediction. Notably, the gradient boosting classifier with tuned hyperparameters attained an accuracy of 99.03% for dataset-II, whereas a support vector machine (SVM) with tuned hyperparameters achieved 87.9% for dataset-I. Using grid search cross-validation to improve model performance in the training and testing stages is new. The best factors for predicting heart disease were found via experimentation. Comparisons with past research demonstrate how well this method performs, outperforming earlier findings in the prediction of heart disease. To successfully create sparse and intelligible scoring systems for multiclass classification problems, the research study by Grzeszczyk et al. [13] provides a novel technique called Multiclass Interpretable Scoring Systems (MISS), which offers a totally data-driven approach. While more conventional techniques, such as Supersparse Linear Integer Model (SLIM), were limited to binary classification, MISS expands to multiclass domains with ease, and the SoftMax function makes it simple to convert scores into class probabilities. The research describes optimization and dimensionality reduction strategies that preserve model optimality while improving training effectiveness. The suggested method has been thoroughly tested on a variety of datasets and has shown competitive classification performance metrics in addition to offering well-calibrated class probabilities. This establishes it as a useful tool for decision support in numerous fields where readability and usability are crucial.

Paikaray et al. [14] tackle the pressing need for more accurate early detection of heart failure, given the condition's complex clinical characteristics and the rising global incidence of heart failure. The development of more effective diagnostic technology is critical, as a correct diagnosis has the potential to save lives. The work employs the UCI ML heart illness dataset for analysis and focuses on building a fundamental ensemble learning model to improve the identification of cardiac diseases. The efficacy of the model in correctly diagnosing cardiac illness is highlighted in the research through a thorough examination and comparison of several classification criteria. The goal of the research is to encourage data scientists to investigate and use comparable strategies for improving healthcare diagnostics by demonstrating the efficacy of ML models in biomedical analysis. The medical community in Reference [15] has paid considerable attention in recent years to the use of AI and ML approaches in the prediction of cardiovascular illnesses. By comparing two popular classifiers, LR and naïve Bayes (NB), using two different datasets—the Cleveland and Hungarian datasets obtained from the UCI repository—this work contributes to the increasing literature of research in this area. The research assesses the predictive power of these classifiers using five performance metrics: accuracy, recall, precision, F1-score, and the AUC. Additionally, the incorporation of the Synthetic Minority Over-Sampling Technique (SMOTE) is investigated as a means of mitigating class imbalance and ultimately improving model accuracy. The data show encouraging trends, with accuracy ranging from 81% to 92% throughout the various classifiers and datasets. Furthermore, correlations with extant literature highlight the importance of our study in propelling the domain of ML approaches for heart disease prediction. Elshafee and Hassan [16] incorporated the ML models into the HEMA Reason Engine, a robotic assistant they made, which represents a noteworthy progression in healthcare technology. They used TensorFlow to construct five different ML models. These models—Models 2 through 6 in particular—are built as four-layered deep learning (DL) neural networks, where

each individual layer includes a Dense Layer made up of 128 neurons. Interestingly, there are just four neurons with sigmoid activation functions in the output layer. Such a design shows a targeted strategy adapted to determined healthcare responsibilities handled by HEMA. Furthermore, Model 1 clusters mixed data using the k-nearest neighbor (K-NN) algorithm to divide patients into two groups: normal or those who suffer from long-term illnesses. These models' exceptional precision and accuracy, as demonstrated by the assessment metrics on the Cleveland dataset, highlight their value for medical diagnostics. These models function well, with reported accuracy and precision of 92% and 94%, respectively, indicating their potential to transform illness diagnosis and care for marginalized groups. The focus on model building and performance draws attention to how cutting-edge ML approaches may significantly improve healthcare outcomes and accessibility. Xu et al. [17] proposed a computer-aided detection system for predicting the risk level of MI. They address the challenge of diagnosing MI in its early stages, which can be challenging due to subtle indications and varying risk factors. To tackle this issue, the authors propose a supervised classifier-based approach. Specifically, they develop an MI prediction system utilizing three different DL algorithms: SVM, Feed-Forward Neural Network (FFNN), and Cascade Correlation Neural Network. Additionally, they employ Genetic Optimized Neural Network and Particle Swarm Optimized Neural Network algorithms to optimize the efficiency of the prediction system. Performance is evaluated among various metrics, which likely include measures such as accuracy, sensitivity, specificity, and AUCROC. These metrics assess the prediction system's effectiveness and reliability in identifying individuals at risk of MI. Similarly, Bertsimas et al. [18] address the critical global health concern: cardiovascular diseases (CVDs) challenges by proposing a cloud-based heart disease prediction system leveraging ML techniques. This approach aims to develop a system that identifies individuals that highly can face heart disease before symptoms manifest, enabling timely interventions and improving patient outcomes. The study comprehensively evaluates various ML algorithms using the WEKA platform, a popular Java-based open-access data mining tool. The goal is to determine the most efficient algorithm for heart illness detection. Through rigorous analysis and comparison. To complement the predictive aspect of the system, the researchers developed a real-time patient monitoring system using Arduino. This system is proposed to continuously aggregate and transmit key physiological parameters such as body temperature, blood pressure, humidity, and heartbeat to a centralized server. By leveraging IoT technology, the system enables remote monitoring of patients' health status, facilitating timely interventions and proactive healthcare management. One notable feature of their system is its ability to trigger alerts when any monitored parameter exceeds predefined thresholds. This proactive notification mechanism ensures that healthcare providers are promptly alerted to deviations from normal physiological ranges, enabling them to intervene swiftly and provide timely medical assistance.

This notification functionality is facilitated through GSM technology, allowing prescribed doctors to receive instant alerts even when they are not actively monitoring the system. In essence, their work represents a significant step toward improving early detection and management of heart diseases through integrating advanced ML techniques and real-time patient monitoring technologies. By combining predictive analytics with continuous remote monitoring capabilities, the proposed system offers a comprehensive solution for enhancing cardiovascular healthcare delivery and reducing the burden of CVDs on individuals and healthcare systems. A new discriminant function model for assessing the probability of angiographic coronary disease is presented in Reference [19]. The probability derived via the novel Cleveland algorithm is compared to that acquired using the Bayesian CADENZA approach on the identical dataset. Both systems, CADENZA and the Cleveland algorithm, frequently overstate the probability of disease in the US and Hungarian centers. However, CADENZA tends to overestimate the change more than the Cleveland algorithm. On the other hand, CADENZA somewhat overestimates disease risk in the Swiss group, whereas the discriminant function underestimates it. In comparison to CADENZA, the new discriminant function performs somewhat better in the Hungarian group and comparably in the American and Swiss groups, based on the proportion of patients correctly diagnosed. According to the study, results show that discriminant function-derived coronary disease probabilities, like the newly developed Cleveland algorithm, are helpful in clinical practice, particularly for patients with chest pain syndromes and intermediate disease prevalence. Hossain et al. [20] developed a method that can serve as an auxiliary tool for predicting heart illness in the initial stages. In this study, various factors in the patient data are examined for accurate heart disease prediction. The usage of the correlation-based Feature Subset Selection algorithm, along with the Best First Search, determines the key characteristics for predicting heart diseases. The most important characteristics associated with heart diseases are found to be age, sex, smoking, overweight, nutrition, physical exercises, stress, type of chest pain, previous chest pain, diastolic blood pressure, diabetes, troponin, electrocardiogram (ECG), and the target. Different AI methods (LR, NB, K-NN, SVM, decision tree, random forest, and Multilayer Perceptron) are applied and compared for two kinds of datasets concerning heart diseases (all features included and selected features). The random forest with the use of selected features for heart disease provides the highest accuracy (90%) as compared with using all input features and other AI techniques. A system that integrates a fuzzification module, knowledge base, inference engine, and defuzzification module was proposed in Reference [21], whereas seven parameters such as type of chest pain, HbA1c (hemoglobin A1c), HDL, LDL, heart rate, age, and blood pressure are considered as inputs to the system. The knowledge database has been augmented with a number of IF-THEN rules aggregated through extensive consultation with medical experts in the field of heart illness diagnostics. Each input after fuzzifications fuzzifies utilizing adequate membership function, and the inference engine activates the right IF-THEN rule from the knowledge base and evaluates the output value with the help of a suitable defuzzification technique. Furthermore, the application of web technology makes it cost-effective and renders it suitable for the prognosis of heart disease, allowing the patient to take decisions based on the status of his heart. Moreover, it can help a medical practitioner in providing more accurate treatment for a patient with heart disease. The results have been evaluated with the help of the Mamdani inference method. The system has been tested with the Cleveland database and has been cross-checked with the infield database. The accuracy of the proposed method when compared with other existing expert systems works out to 98.08% and can take decisions accurately for diagnosing heart diseases. Lee [22] illustrated how ML has the potential to reveal hidden information from medical datasets to generate usable information regarding cardiovascular health and assist the clinical decision-making process. This research then provides a specialized ML framework regarding the prediction of the severity of heart disease from the UCI Heart Disease Data

from Kaggle, a multivariate dataset obtained from the Cleveland database. The dataset consisted of 14 predictor variables including clinical and demographic variables that were utilized to train and test many supervised learning algorithms. Of note were the use of LR, decision trees, random forests, and gradient boosting machines. The results demonstrate that the XGBoost model yielded the highest efficiency with an accuracy of 62.5%. Performance assessment depended on the metrics precision, recall, and F1-score, which were improved by systematic hyperparameter optimization. Advanced feature engineering and cross-validation techniques contributed to an improved predictive capability of the model. In the study by Izonin et al. [23], the binary classification is addressed using small and unbalanced medical dataset, in which the implementation of the Improved Probabilistic Neural Network (I_PNN) has been improved. This mainly relied on generating the results of the summation layer of the Probabilistic Neural Network (PNN) that retains the condition of ensuring the complete system of events as in the analog (i.e., formation of a set of probabilities of belonging to each class that in the sum equal 1). Unlike in the analogs, the addressed method considered the exaggerated representation of all classes in the utilized dataset. Thus, the possibility to effectively realize the decision of classification tasks was increased, especially in the unbalanced dataset. All the algorithmic realization I-PNN steps of the proposed method were described. The experimental results enhanced the accuracy of the I-PNN, up to 5% for the F1-score, compared to the literature PNN. These benefits give rise to the prerequisites for the practical application of I-PNN in the case of processing unbalanced datasets in various areas of medical diagnostics. Kolukisa and Bakir-Gungor [24] aimed to diagnose CADs with the help of computer-aided systems. These systems can analyze some physical and chemical parameters with reduced economic burden. This study evaluated seven feature selection (FS) methods, including one knowledge-based FS method, as well as various classification methods, and an exhaustive ensemble FS method and a probability ensemble FS method were proposed. The approach has been evaluated on three publicly available CAD datasets (the Z-Alizadeh Sani, Statlog, and Cleveland) using six different classification methods and four different complementary methods. The results have been comparatively analyzed as well as various combinations of classifier and FS methods. The multilayer perceptron classifier gave sufficient answers on three datasets. Performance analysis revealed that the proposed method resulted in accuracy of 91.78%, 85.55%, and 85.47% for the Z-Alizadeh Sani, Statlog, and Cleveland datasets.

This underscores the importance of FS and classifier choice in achieving accurate diagnoses for both CAD and MI, showcasing the potential of ML methods in improving healthcare outcomes. In Table 1, a summary of the prior literature is addressed.

3. Proposed Model

The proposed model for this classification problem consists of a preprocessing phase for the data and then training the data with several ML algorithms to yield results, as shown in Figure 1.

3.1. Dataset description

The Cleveland Heart Disease dataset is an extensively used benchmark dataset in the field of ML and healthcare research. It contains clinical and demographic information of patients collected from four different institutions. Compiled in 1988, this dataset serves as a valuable resource for studying cardiovascular health and developing predictive models for heart disease [19]. Number of Instances: The dataset consists of a total of 303 instances, each representing a patient. Number of Attributes: There are 13 attributes in the dataset, including demographic information, medical measurements, and diagnostic test results. Target Variable: The target variable in the dataset represents the level of heart illness and is categorized into five classes, ranging from 0 to 4. Each class corresponds to a different level of heart disease severity or the number of major vessels exhibiting narrowing. This study considers only the binary classification of heart disease (i.e., the presence or the absence of the disease) [19]. As aforementioned in Reference [18], the Cleveland dataset will be merged with another dataset that closely resembles it. This database consists of 270 instances and has the same features and target variables. The key attributes of the merged datasets are described in Table 2, including a mix of categorical and numerical variables. The Cleveland Heart Disease dataset has been extensively utilized in published research studies to develop predictive models, evaluate algorithms, and extract insights into cardiovascular health. Researchers use this dataset to explore the relationships between various patient attributes and the likelihood of heart disease, as well as to develop tools for early detection and risk stratification [19]. In conclusion, the Cleveland Heart Disease dataset provides a valuable resource for studying heart disease and developing predictive models to assist in clinical decision-making.

Table 1
Summary of prior research in the medical field

Refs.	Year	Description	Technique
[22]	2025	This paper builds a machine learning framework for predicting the severity of cardiac disorders based on a public domain multivariate clinical database	Supervised learning algorithms (logistic regression, decision trees, random forests, and gradient boosting machines (with XGBoost))
[13]	2024	Introduces MISS, a data-driven method for creating user-friendly scoring systems for multiclass classification tasks, showcasing competitive performance and well-calibrated class probabilities across diverse datasets	SLIM expanded to multiclass
[21]	2024	This research describes the design and implementation of a fuzzy logic-based expert system for the prediction and diagnosis of CVD	Fuzzy logic

(*Continued*)

Table 1
(*Continued*)

Refs.	Year	Description	Technique
[14]	2023	Presents a basic ensemble learning model designed to enhance the diagnosis of heart conditions	Extreme gradient boosting
[12]	2023	Explores the prediction of heart disease using six machine learning algorithms, achieving high accuracy rates through the optimization of hyperparameters and comparative analyses against existing methods	Ensemble learning
[15]	2023	Heart disease prediction using SMOTE and machine learning algorithms	Logistic regression and naive Bayes
[16]	2023	A robot healthcare system that includes heart disease prediction	Ensemble learning of deep neural networks
[24]	2023	Feature selection and engineering	Data analysis
[20]	2023	An assistant framework for the prediction of the early stages of heart disease is going to be developed by finding out which are the most significant attributes and applying various techniques related to artificial intelligence (AI)	Correlation-based Feature Subset Selection with Best First Search. AI classification techniques (logistic regression, naïve Bayes, K-NN, SVM, decision tree, random forest, and multilayer perceptron)
[23]	2022	This study focuses on the problem of binary classification in medical diagnosis, especially when performing it on short and unbalanced series (in which one class predominates over the other)	Improved Probabilistic Neural Network (I-PNN)
[18]	2018	Cloud-based heart disease prediction system leveraging machine learning techniques	SVM
[17]	2016	Computer-aided detection system for predicting the risk level of myocardial infarction	FFNN
[19]	1989	This paper proposes a new discriminant function model for angiographic coronary disease probability estimation	Bayesian algorithm

Figure 1
Diagram model

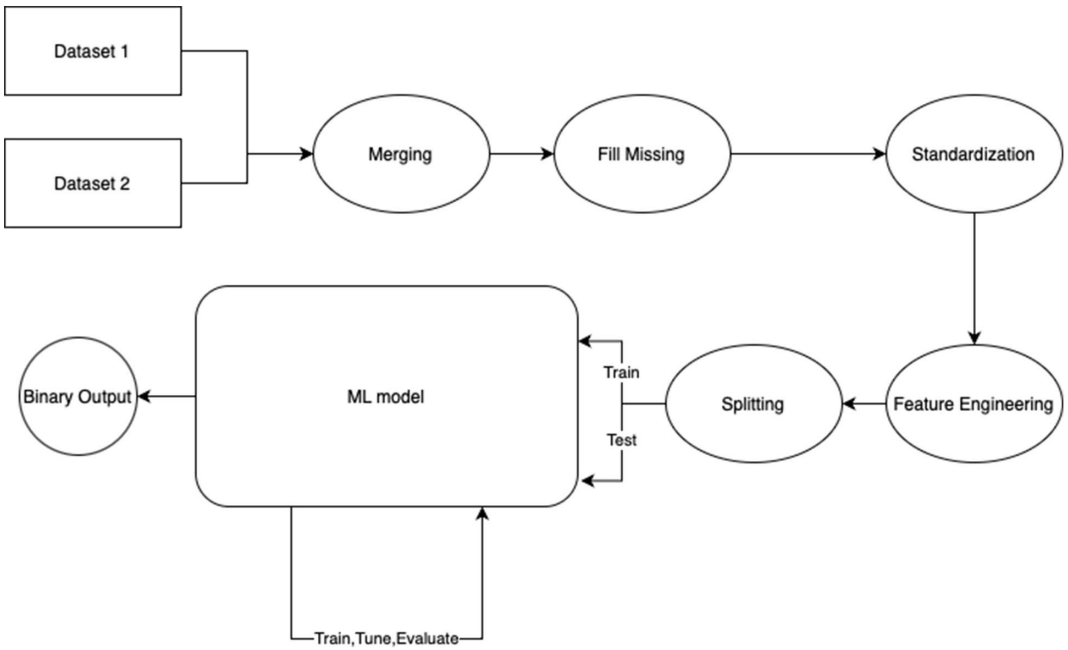


Table 2
Description of dataset attributes

Attribute	Description	Type
Age	The age in years	Continuous
Sex	The gender binarized (1 = male; 0 = female)	Categorical
Chest pain	The type of chest pain	Categorical
Resting blood pressure	The resting blood pressure in mm Hg	Continuous
Cholesterol	The serum cholesterol level in mg/dl	Continuous
Fasting blood sugar	Fasting blood sugar binarized by > 120 mg/dl (1 = true; 0 = false)	Categorical
Resting electrocardiographic results	Results of the resting ECG (e.g., normal, abnormal)	Categorical
Maximum heart rate achieved	The maximum heart rate achieved during exercise	Continuous
Slope	The slope of the peak exercise ST segment (values are 1.0, 2.0, and 3.0)	Categorical
ST depression	ST depression induced by exercise relative to rest	Continuous
Number of major vessels colored by fluoroscopy	The number of major vessels exhibiting significant narrowing (0–3)	Categorical
Thallium stress test result	Results of the thallium stress test (e.g., normal, reversible defect)	Categorical
Num	The target variable represents the level of heart illness, and its classes are Class 0: No heart disease. Classes 1–4: Increasing severity of heart disease	Categorical

3.2. Preprocessing

Data preprocessing is a very important step in data analysis, involving preparing and transforming raw data using different methods into suitable data for analysis and modeling. This preparatory phase typically includes cleaning, transforming, and organizing the data. By addressing inconsistencies, missing values, outliers, and other issues in the data, preprocessing is the starting step for accurate analysis and leads to better insights and decision-making. The following data preprocessing techniques are utilized in our model:

3.2.1. Data merging

In this study, two similar datasets are merged to increase the number of records and enhance the accuracy of the proposed model. In the literature, the average accuracy of baseline models on the Cleveland dataset was 82% [18, 25]. By performing data augmentation, the number of records is increased from 303 records to 573 records, as shown in Figures 2 and 3. The merged process commenced directly after verifying that both datasets share the same features, such as names and data types. Later, each dataset is loaded and checked against the presence of inconsistent values. These two datasets are concatenated row-wise using a manual merging procedure in Python, which results in the formation of a unified dataset containing all the records. The resulting dataset is shuffled to randomize the order of the records to prevent sequential bias. The combined dataset improves the model's robustness during both the training and testing phases.

3.2.2. Fill missing values

The presence of missing data poses a significant risk for the trained models, and the way in which this issue is addressed would be crucial to the performance of our models. Since there is not a lot of missing data, the missing values are substituted with the average of that column that the missing value belongs to [25].

3.2.3. Data shuffle

Data shuffling is a common practice in data processing and ML involving randomizing the order of data instances or samples within a dataset. The primary drive of data shuffling is to introduce randomness into the data, which helps prevent any inherent order or patterns from influencing the learning process during model training.

The main aim is to ensure the model is exposed to a diverse and objective representation of the dataset throughout the training phase. For this reason, the arrangement of the data is modified. This introduces unpredictability and lowers the possibility of overfitting and helps the model improve in generalization to new data. Due to the combination of two datasets, it is necessary to guarantee that there will be no bias when training and testing the groups [25, 26]. In Figures 4 and 5, a sample of the shuffled dataset and the standardized dataset is shown.

3.2.4. Standardization

Standardization is a fundamental data preprocessing technique aimed at ensuring consistency and comparability across different variables or features in a dataset. This process involves transforming the scale of each variable to have a mean of 0 and a standard deviation of 1, often referred to as z-score normalization [25]. The data is primarily standardized to efficiently eliminate the effects of scale differences between variables, which can disproportionately influence certain analyses, such as distance-based algorithms like KNNs or clustering algorithms like k-means. By standardizing the data, each variable is guaranteed to equally contribute to the analysis process, preventing variables with larger scales from dominating the results. Standardization also makes it easier to interpret the coefficients or weights of variables in certain statistical models, as they are now directly comparable in terms of their influence [27]. Mathematically, standardization can be represented as:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Where:

- Z is the standardized value,
- X is the original value,
- μ is the mean of the distribution, and
- σ is the standard deviation.

Figure 2
Before augmentation description

Statistic	age	sex	cp	trestbps	chol	fbs	restecg
count	303.0000	303.0000	303.0000	303.0000	303.0000	303.0000	303.0000
mean	54.36634	0.678899	3.158415	131.6898	246.6931	0.148515	0.528053
std	9.082101	0.467298	0.961465	17.59975	51.77692	0.356198	0.52586
min	29.0000	0.0000	1.0000	94.0000	126.0000	0.0000	0.0000
25%	48.0000	0.0000	3.0000	120.0000	211.0000	0.0000	0.0000
50%	55.0000	1.0000	3.0000	130.0000	240.0000	0.0000	1.0000
75%	61.0000	1.0000	4.0000	140.0000	275.0000	0.0000	1.0000
max	77.0000	1.0000	4.0000	200.0000	564.0000	1.0000	2.0000
Statistic	**thalach**	**exang**	**oldpeak**	**slope**	**ca**	**thal**	**num**
count	303.0000	303.0000	303.0000	303.0000	303.0000	303.0000	303.0000
mean	149.6073	0.326733	1.039604	1.39934	0.729373	2.313531	0.544554
std	22.875	0.469194	1.161075	0.616226	1.022606	0.612277	0.498835
min	71.0000	0.0000	0.0000	1.0000	0.0000	1.0000	0.0000
25%	133.5000	0.0000	0.0000	1.0000	0.0000	2.0000	0.0000
50%	153.0000	0.0000	0.8000	2.0000	0.0000	2.0000	0.0000
75%	166.0000	1.0000	1.6000	2.0000	1.0000	3.0000	1.0000
max	202.0000	1.0000	6.2000	3.0000	4.0000	3.0000	1.0000

Figure 3
After augmentation description

Statistic	age	sex	cp	trestbps	chol	fbs	restecg
count	573.0000	573.0000	573.0000	573.0000	573.0000	573.0000	573.0000
mean	54.54084	0.679398	3.157298	131.7517	246.7093	0.145286	0.524956
std	9.084734	0.467136	0.961173	17.51751	51.75277	0.352836	0.52476
min	29.00000	0.000000	1.000000	94.00000	126.0000	0.000000	0.000000
25%	29.00000	0.000000	3.000000	120.0000	211.0000	0.000000	0.000000
50%	29.00000	1.000000	3.000000	130.0000	240.0000	0.000000	1.000000
75%	29.00000	1.000000	4.000000	140.0000	275.0000	0.000000	1.000000
max	29.00000	1.000000	4.000000	200.0000	564.0000	1.000000	2.000000
Statistic	**thalach**	**exang**	**oldpeak**	**slope**	**ca**	**thal**	**num**
count	573.0000	573	573	573	573	573	573
mean	149.4947	0.324795	1.036608	1.396855	0.718559	2.305571	0.542087
std	22.85305	0.468472	1.158038	0.613878	1.019859	0.610278	0.498207
min	71.00000	0.000000	0.000000	1.000000	0.000000	1.000000	0.000000
25%	133.5000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	153.0000	0.000000	0.800000	2.000000	0.000000	2.000000	0.000000
75%	166.0000	1.000000	0.800000	2.000000	1.000000	3.000000	1.000000
max	202.0000	1.000000	0.800000	3.000000	4.000000	3.000000	1.000000

Standardization facilitates the comparability of variables' contributions in statistical models, as coefficients or weights become directly interpretable in terms of their relative influence. This normalization step enhances the robustness and interpretability of analytical outcomes [25, 27].

3.2.5. Feature engineering

Feature engineering commonly includes the development of new features or the modification of some features to improve the accuracy of ML models. Because features have a direct impact on how well ML models perform, feature engineering is crucial. By creating features that accurately represent significant correlations, patterns, and subtleties found in the data, the model's learning processes are facilitated and improve its prediction capabilities. Furthermore, well-designed features can boost overall model performance, alleviate problems like overfitting, and increase the interpretability of the model. To put it simply, feature engineering is an essential phase in the data science workflow that helps us create more reliable and accurate predictive models by maximizing the value of the data that is already available, so the engineered features are:

1) Age category
2) Total risk
3) Exercise angina
4) Cholesterol HDL ratio

Figure 4
Dataset after shuffling

age	sex	cp	trestbps	chol	fbs	restecg
-0.3	0.45	1.1	1.37	0.04	1.1	0.67
1.51	-1.79	1.1	-1.22	-1.54	-0.73	-1.57
-1.21	0.45	-0.73	-0.67	1.17	-0.73	-0.45
0.3	0.45	-0.73	0.07	0.47	-0.73	0.67
-0.3	0.45	-0.73	0.44	-0.14	1.1	0.67
thalach	**exang**	**oldpeak**	**slope**	**ca**	**thal**	**num**
0.93	1.79	1.35	1.1	-0.73	1.33	1.1
-1.39	-0.45	-0.08	-0.73	-0.73	-0.72	-0.73
0.46	-0.45	-1.31	-0.73	1.1	-0.72	-0.73
-0.7	-0.45	0.49	-0.73	1.1	0.82	1.1
0.7	-0.45	-0.46	1.1	-0.73	-0.72	-0.73

Figure 5
Data after standardization

age	sex	cp	trestbps	chol	fbs	restecg
71.0000	0.0000	4.0000	112.0000	149.0000	0.0000	0.0000
53.0000	1.0000	3.0000	130.0000	197.0000	1.0000	2.0000
44.0000	1.0000	3.0000	118.0000	242.0000	0.0000	1.0000
53.0000	1.0000	4.0000	140.0000	203.0000	1.0000	2.0000
59.0000	1.0000	3.0000	126.0000	218.0000	0.0000	2.0000
thalach	**exang**	**oldpeak**	**slope**	**ca**	**thal**	**num**
125.0000	0.0000	1.6000	2.0000	0.0000	3.0000	0.0000
152.0000	0.0000	1.2000	3.0000	0.0000	3.0000	0.0000
149.0000	0.0000	0.3000	2.0000	1.0000	3.0000	0.0000
155.0000	1.0000	3.1000	3.0000	0.0000	7.0000	1.0000
134.0000	0.0000	2.2000	2.0000	1.0000	6.0000	1.0000

In Table 3, these new features are introduced. According to Figure 6, a sample of the engineered features is presented.

3.2.6. Train–test splitting

Train–test splitting is the division of data into two groups, training and testing, and is employed to assess the performance of predictive models. By training the model on a certain set of data and evaluating its performance on a different group, its performance can be evaluated on data that has not been exposed to during training. This helps us understand if the model has learned meaningful patterns from the data or if it's simply memorizing the training examples (overfitting). A ratio of 80:20 is utilized to split the data due to the limited number of records available [25].

3.3. Modeling

The ML is selected over DL methodologies for their interpretability, efficiency with smaller datasets, and ability to mitigate overfitting. ML models, such as decision trees and LR, offer transparency in decision-making, making them suitable for this research, which prioritizes understanding model behavior.

The examined dataset size and computational resources favored ML algorithms, which demonstrate efficacy with modest data volumes and require fewer computational resources compared to DL models. ML techniques also provide effective means to combat overfitting, ensuring robust model performance in scenarios with small or noisy datasets.

Moreover, the structured nature of this data and existing domain expertise reduced the need for complex feature engineering, aligning well with ML algorithms' operational efficiency on structured data. Additionally, deployment considerations favored ML algorithms, as they offer faster inference times and lower resource requirements compared to DL models.

Overall, the decision made to utilize ML algorithms over DL methodologies was guided by considerations of interpretability, data efficiency, overfitting mitigation, and deployment constraints, enabling the development of robust and interpretable models tailored to the proposed research objectives. Three different and popular classification algorithms are used to obtain the best results [17, 25].

3.3.1. Logistic regression

A statistical technique that is employed in binary classification, with the aim of predicting the likelihood of an occurrence. Despite its name, LR predicts the likelihood of a binary result, making it a classification technique as opposed to a regression one. Fundamentally, the logistic function or sigmoid function is utilized to represent the connection between one or more independent variables (i.e., features) and a binary dependent variable

Table 3
Model evaluation metrics

Feature name	Description
Age category	Age can be an important factor in predicting heart disease risk. Ages are categorized into groups ("Young," "Middle-aged," "Senior"). Thus, the potential differences can be captured in heart disease prevalence among different age groups and visualized. Type of Feature Engineering: This is an example of binning, where a continuous numerical feature (age) is transformed into categorical groups
Total risk	Both resting blood pressure ("trestbps") and serum cholesterol ("chol") are risk factors for heart disease. Type of Feature Engineering: This is an example of feature combination, where existing features are combined or aggregated to create a new feature
Cholesterol HDL ratio	The ratio of total cholesterol to HDL cholesterol is considered a better predictor of heart disease risk than total cholesterol alone HDL cholesterol is often referred to as "good" cholesterol, and a lower ratio may indicate increased risk. Type of Feature Engineering: This is an example of feature transformation, where existing features are transformed using mathematical operations to create a new feature
Exercise-induced angina	Exercise-induced angina (chest pain or discomfort during physical exertion) is a symptom of coronary artery disease. By combining the presence of exercise-induced angina ("exang" == 1) with the maximum heart rate achieved during exercise ("thalach"), additional information may be captured about the severity of coronary artery disease. Type of Feature Engineering: This is an example of feature combination

Figure 6
Image of the new features

age_category	total_risk	exercise_angina	cholesterol_hdl_ratio
Senior	347	TRUE	1.4375
Middle-aged	474	TRUE	2.171779
Senior	401	FALSE	1.713287
Middle-aged	405	FALSE	2.968421
Senior	394	FALSE	1.739726
Middle-aged	316	FALSE	1.426573
Senior	261	FALSE	1.192
Senior	401	FALSE	2.728155
Middle-aged	317	TRUE	1.092593
Middle-aged	345	TRUE	1.558442

(target). Any real integer can be mapped to a probability value between 0 and 1 using it [25].

The anticipated probability of the positive class is obtained by the LR model by applying a linear combination of the feature values, weighted by coefficients, and passing this result via the logistic function. This can be stated mathematically as:

$$P(y = 1|x) = \frac{1}{1 + e - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)} \quad (2)$$

Where:

1) $P(y = 1 \mid x)$ represents the probability of the positive class given the features x.
2) $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ *are the coefficients (parameters) of the model.*
3) $x_1, x_2, \ldots, x_n$ are the feature values.

LR is appropriate for binary classification problems because the logistic function guarantees that the predicted probabilities are restricted between 0 and 1. The two groups are divided by the decision boundary, which is normally fixed at $P(y = 1 \mid x) = 0.5$.

3.3.2. Decision tree

A flexible and easy-to-understand ML approach that is frequently implemented for both regression and classification applications. Fundamentally, a decision tree divides the feature space into regions in a recursive manner using a sequence of binary judgments, or "splits," which are dependent on the feature values. This procedure creates a structure like a tree, with each leaf node representing a class label or regression value and each inside node representing a choice based on a particular attribute.

In order to maximize the homogeneity (or purity) of the resultant subsets, the algorithm at the root node of a decision tree chooses the characteristic that best divides the dataset into discrete groups. Usually, this splitting criterion is based on metrics that measure the level of disorder within each group, like information gain [25].

The process of fine-tuning the feature space partitioning continues as the tree expands, with succeeding nodes being chosen and divided recursively using the same principles until certain halting conditions are satisfied. Achieving a minimal number of samples per leaf node, a maximum tree depth, or the absence of further purity improvement with subsequent splits are a few examples of these requirements.

Once the tree is done, classification choices are performed by following the route determined by the input data as you navigate the tree from the root node to a leaf node. The anticipated class label for the input instance is then set to the class label linked to the leaf node that this traversal reached [25].

3.3.3. Random forest classifier

An ensemble method that uses decision trees as a base during training. Random forest builds several decision trees as opposed to depending just on one, then a random set of the original dataset is used to train each tree in the forest, and a random subset of characteristics is considered for splitting at each node. The robustness of the model, along with its generalization, is made stronger by the unpredictability that is brought by this way, thus lowering the possibility of overfitting. By averaging the predictions of each individual tree during prediction, the algorithm determines its

output; for classification tasks, voting is used, and for regression tasks, it is done by averaging [25].

3.3.4. Gradient boosting

Another effective ensemble ML approach that is frequently used for classification applications is the gradient boosting classifier. This is a member of the boosting algorithm family, which establishes powerful prediction models by repeatedly combining weak learners. Gradient boosting, which Friedman first presented in 2001, has grown to be one of the most widely used and successful algorithms for handling both structured and unstructured data.

The gradient boosting classifier's basic principle is to train a sequence of weak learners—usually decision trees—one after the other until a strong prediction model is produced. Gradient boosting optimizes the model by lowering the loss function values using gradient descent, in contrast to classic boosting techniques that concentrate on minimizing errors directly. The approach adapts a new weak learner to the remaining errors of the preceding learners at each iteration, placing more emphasis on the instances that were misclassified or poorly predicted. By iteratively improving upon the shortcomings of the previous models, gradient boosting gradually converges to a strong ensemble model with superior predictive performance.

One of the key advantages of the gradient boosting classifier is its flexibility and robustness to overfitting. By using shallow decision trees as weak learners and implementing techniques like regularization, gradient boosting effectively mitigates overfitting and generalizes well to unseen data.

Additionally, gradient boosting is highly interpretable, allowing users to understand the importance of features in the prediction process.

Mathematically, gradient boosting can be described by the objective function, gradient calculation, pseudo-residual calculation, fitting the weak learner, and updating the model equations. At each iteration, the algorithm minimizes the loss function by reducing the residuals and improving the predictions, resulting in a final ensemble model optimized for accurate classification.

Many classification challenges have been effectively tackled by the gradient boosting classifier such as customer churn prediction, fraud detection, recommendation systems, and natural language processing. Its ability to handle heterogeneous data types, handle missing values, and capture complex relationships between features makes it a versatile and indispensable tool in the ML practitioner's toolkit.

3.3.5. Support vector machines

SVMs are robust supervised ML algorithms that are applied to regression and classification problems. Introduced by Vapnik and Cortes in the 1990s, SVMs have since gained widespread popularity because they are effective with big data and their ability to find complex decision boundaries.

In simpler terms, an SVM aims to find the best straight line (or plane in higher dimensions) to separate data points into two categories. This line is not randomly chosen; the SVM specifically picks the line that creates the largest gap between itself and the closest data points from each category. These closest data points are known to be the support vectors because they essentially "support" the margin (the gap) between the line and the rest of the data.

SVMs' capacity to handle nonlinear separable data by utilizing core functions is one of their main advantages.

The initial feature space is mapped onto a higher-dimensional space. SVMs are known for their robustness to overfitting, especially in high-dimensional spaces, and their ability to generalize well to unseen data [25]. SVMs are used in image classification, text categorization, bioinformatics, finance, and many other areas [25].

4. Experimental Results

In this section, the results of the proposed model are discussed and analyzed. The criteria used to evaluate the model are as follows:

Table 4 uses the metrics shown and used in References [18, 28]. Having established the criteria for evaluating the models, the training results are shown in Table 5, while Table 6 displays the testing outcomes.

Based on the results listed in Tables 5 and 6, the performance of the proposed model is demonstrated via the confusion matrices, as in Figure 7, which explains the performance of the gradient boosting model. Training the decision tree and an extreme gradient boosting resulted in a 100% training accuracy, precision, recall, and even F1-score, which is a great indication that decision trees and XGB are overfitting to the training data, especially since their testing numbers are much lower than

Table 4
Model evaluation metrics

Criteria	Description
Accuracy	The degree to which a measurement aligns with the real or accepted value Accuracy = (True Positives + True Negative)/Total Prediction (3) where Total Prediction equals the total number of predictions (positive and negative)
Precision	The proportion of predicted positive cases that were correctly identified by the model Precision = True Positives/(True Positives + False Positives) (4)
Recall	The proportion of actual positive cases that were correctly identified by the mode Recall = True Positives/(True Positives + False Negatives) (5)
F1-score	The harmonic mean of precision and recall, providing a balanced measure of the model's performance. F1-Score = 2 ·Precision · Recall/(Precision + Recall) (6)
Support	The number of real instances of the class in the given dataset. It shows how many instances of each class there are with that class label in the true target variable
Macro Average	The performance statistic (such as recall, accuracy, or F1-score) averaged for each class separately, regardless of the classes' imbalance in the dataset

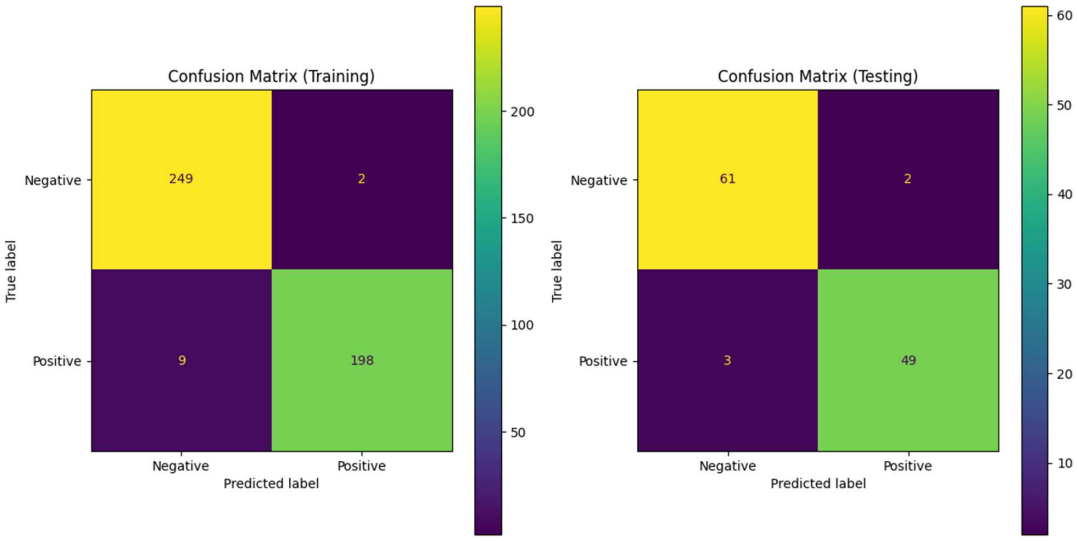
Table 5
Training results table

Model	Training accuracy	Precision	Recall	F1-score
Random forest	0.95	0.96	0.95	0.95
Decision tree	1.00	1.00	1.00	1.00
SVM	0.85	0.85	0.85	0.85
Gradient boosting	0.98	0.98	0.97	0.98
Logistic regression	0.85	0.86	0.83	0.88
XGB	1.00	1.00	1.00	1.00

Table 6
Testing results table

Model	Training accuracy	Precision	Recall	F1-score
Random forest	0.91	0.92	0.91	0.91
Decision tree	0.94	0.94	0.94	0.94
SVM	0.88	0.88	0.87	0.88
Gradient boosting	0.96	0.96	0.96	0.96
Logistic regression	0.88	0.88	0.88	0.88
XGB	0.93	0.93	0.93	0.93

Figure 7
Gradient boosting confusion matrices



their training numbers, as shown in their confusion matrices and Figures 8 and 9. In addition, the random forest confusion matrix classifies the true negatives and positives (as shown in Figure 10). The LR performed poorly compared to other algorithms. This is evident from the baseline training accuracy that was depicted in Figures 11 and 12, which shows the confusion matrices of the LR and SVM models. Additionally, the ROC curves in Figure 13 support this finding. Random forest classifiers and gradient boosting algorithms yield the best results as they do not overfit while maintaining a great testing and training accuracy as well as a better F1-score.

Figures 7, 8, 9, 10, 11, and 12 are all the confusion matrices of all the models. Furthermore, Figure 13 contains all the ROC curves of all the models.

As it is displayed, the gradient boosting confusion matrix in Figure 7 correctly classifies the true negatives by a value of 249 in the training and 61 in the testing matrices. Also, it correctly classifies 198 and 49 true positives in the training and testing matrices, respectively; the models have 2 false positives in the training and testing phases while having 9 and 3 false negatives.

As for the decision tree confusion matrices in Figure 8, it is noticeable that signs of overfitting are present as the confusion matrices are perfect with no false negatives or false positives in the training phase while having 3 false negatives and 4 false positives in the testing matrix.

For extreme gradient boosting confusion matrix in Figure 9, it is depicted that signs of overfitting for having less than 0.5% of the training examples being false negatives and false positives—2 false negatives to be exact—while having 4 false negatives and 4 false positives in the testing confusion matrix compared to 251 true negatives and 205 true positives in the training examples while having 59 and 48 true negatives and true positives, respectively.

Figure 8
Decision tree confusion matrices

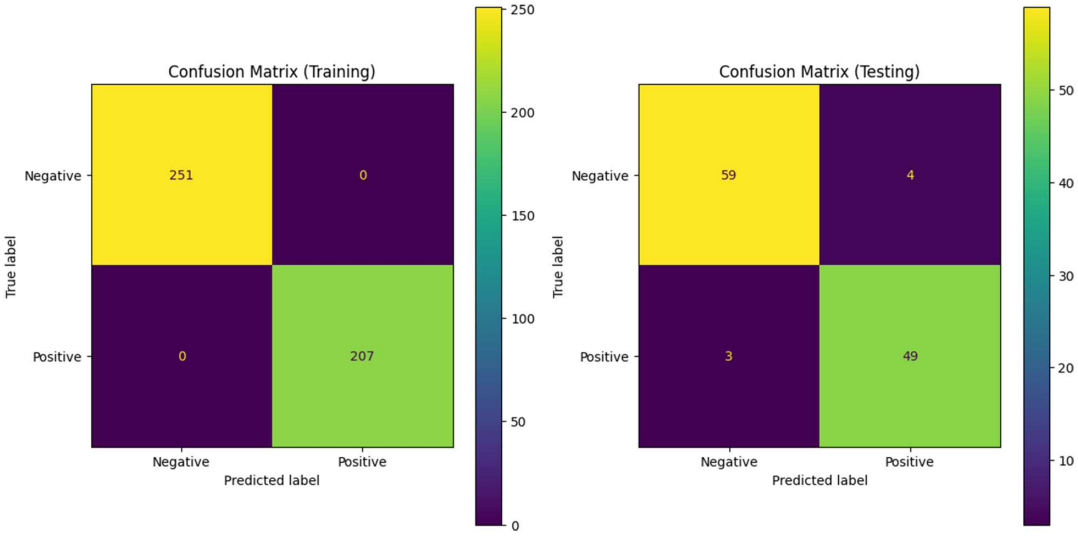


Figure 9
Extreme gradient boosting confusion matrices

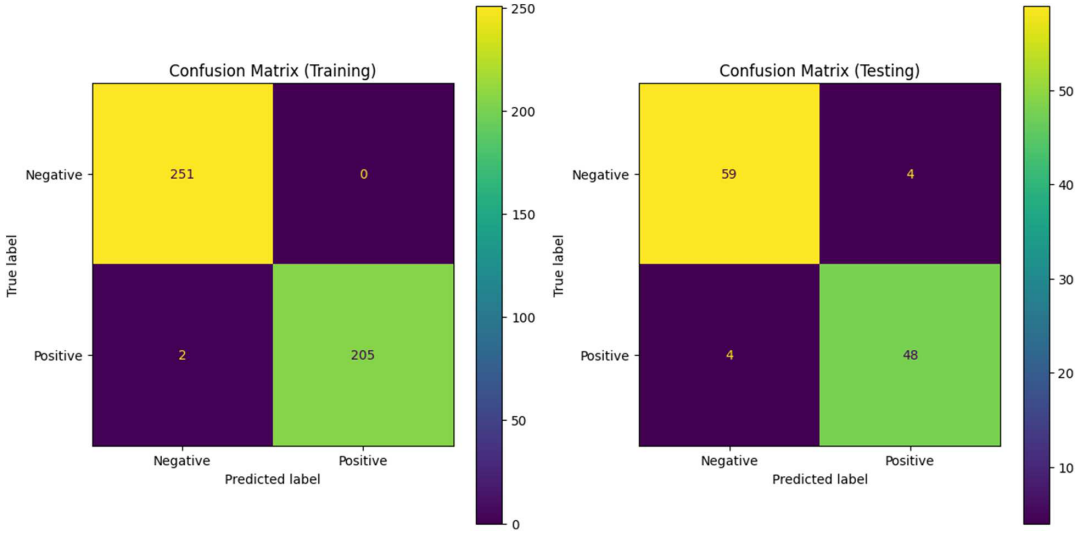


Figure 10
Random forest confusion matrices

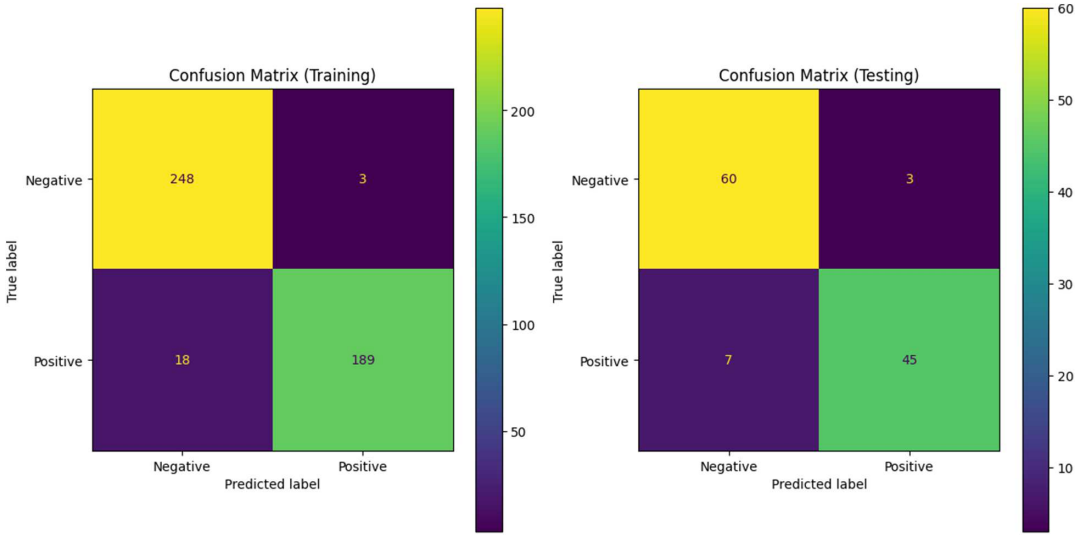


Figure 11
Logistic regression confusion matrices

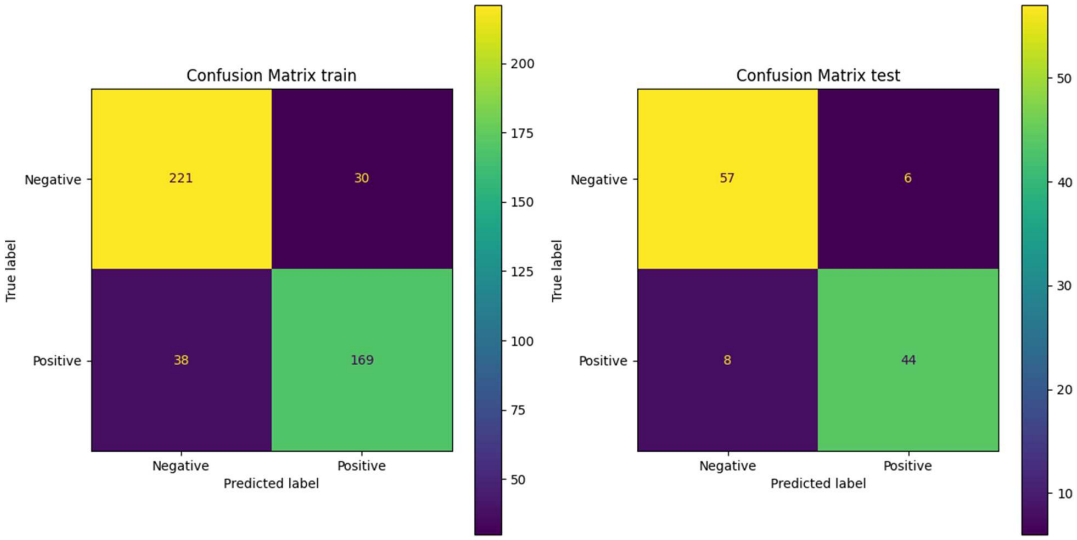
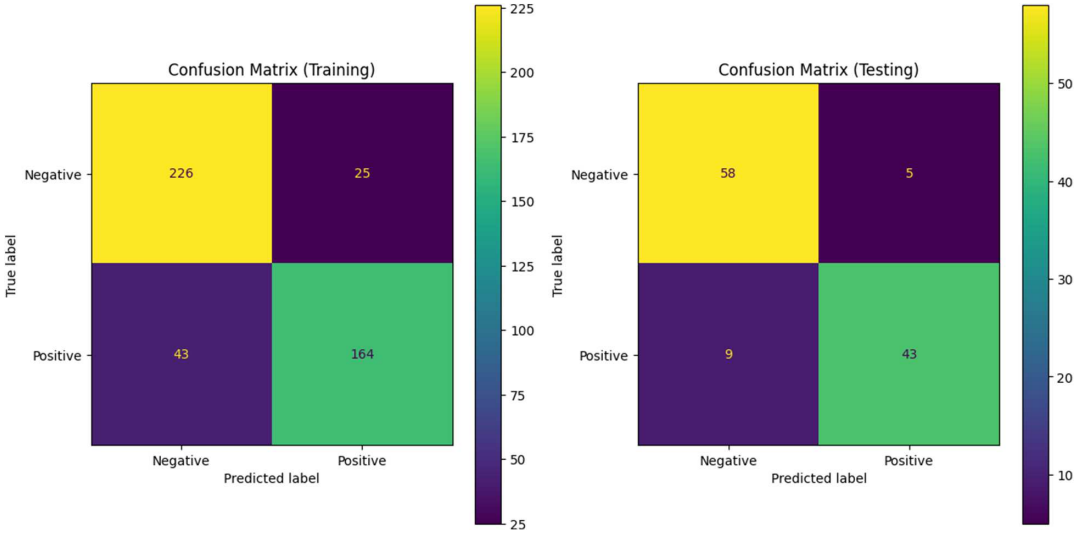


Figure 12
SVM confusion



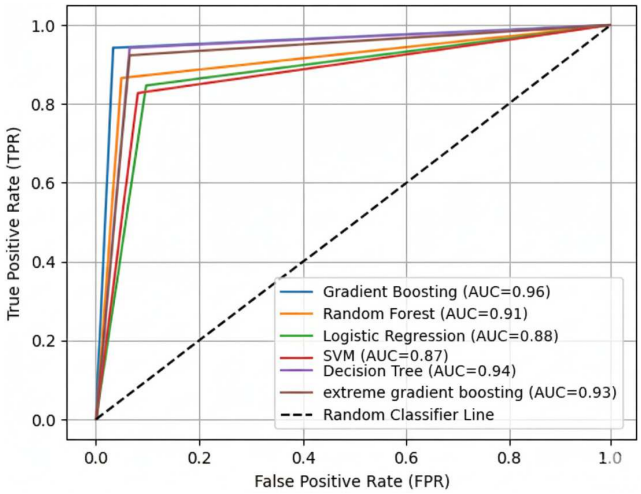
The random forest confusion matrix in Figure 10 correctly classifies the true negatives by a value of 248 in training and 60 in the testing matrices, which are close numbers to the gradient boosting. In addition, it correctly classifies 189 and 45 true positives in the training and testing matrices, respectively. The gradient boosting confusion matrices have better numbers on the testing dataset. The models have 3 false positives in the training and testing phases while having 18 and 7 false negatives.

In Figure 11, the LR shows the following numbers in the training datasets: 221 true negatives and 38 false negatives; the model has 169 true positives and 30 false positives, and its testing confusion matrix is better than its training one, having 57 true negatives, 8 false negatives, 44 true positives, and 6 false positives. The SVM model, in Figure 12, predicted 226 true negatives, 43 false negatives, 25 false positives, and 164 true positives on the training set; for the testing set, the model predicted 58 true negatives, 9 false negatives, 5 false positives, and 43 true positives, which are not as bad as LR model but still has much room for improvement compared to random forest and gradient boosting algorithms in Figure 13, which contains the ROC curves for all the mentioned algorithms, showing that the gradient boosting algorithm has the best AUC by 0.96.

The gradient boosting exhibits the best ROC curves among all algorithms, as shown in the ROC curves figure (Figure 13). For that reason, gradient boosting is utilized as the primary classifier for having better performance in general, as addressed in Tables 7 and 8.

The proposed ML model results are encouraging, as presented in Tables 7 and 8, to indicate a well-performing model; the training accuracy of 0.98 suggests the model effectively learned the patterns within the training data unlike the LR model. This is a positive initial step, but the true test of a model lies in its ability to handle unseen data. The testing accuracy of 0.96 indicates good generalizability, meaning the model can perform well on new data it wasn't explicitly trained on. A slight decrease in accuracy from training to testing is typical, and a difference of

Figure 13
ROC curves



2% is generally acceptable, so there is no overfitting and can adapt to unseen examples unlike the other models such as the decision tree and SVM model.

An improved representation of each class's performance is shown in Table 7 for the training set and Table 8 for the testing set. For both classes, the accuracy, recall, and F1-score values over 0.94 show a well-balanced performance. The training and testing accuracies obtained are 0.98 and 0.96, respectively. The precision of a model indicates its capacity to distinguish between true positives and false positives. Recall measures how well the model prevents false negatives and catches all real positive situations. A harmonic mean average of both measures is provided by the F1-score; also, these metrics are reflected in the confusion matrices displayed in Figure 7.

There is room for improvement in the importance of optimizing for specific classes depending on the problem. If, for instance, misclassifying class 1 has more severe consequences, prioritize improving the model's recall or F1-score for that class. This could involve techniques like cost-sensitive learning.

Table 7
Training classification report of gradient boosting model

	Precision	Recall	F1-score	Support
0	0.97	0.99	0.98	251
1	0.99	0.96	0.97	207
Accuracy			0.98	458
Macro Avg	0.98	0.97	0.98	458
Weighted Avg	0.98	0.98	0.98	458

Table 8
Testing classification report for the gradient boosting model

	Precision	Recall	F1-score	Support
0	0.95	0.97	0.96	63
1	0.96	0.94	0.95	52
Accuracy			0.96	115
Macro Avg	0.96	0.96	0.96	115
Weighted Avg	0.96	0.96	0.96	115

Gradient boosting is often considered better than random forest in combating overfitting due to several reasons. First, gradient boosting optimizes the ensemble model by sequentially fitting weak learners to the residual errors of the previous models. This iterative process focuses on reducing the errors made by the ensemble, leading to a more robust and generalized model. In contrast, random forest builds multiple decision trees independently and combines their predictions through averaging or voting, which may result in overfitting, especially when the individual trees are deep and highly correlated.

Additionally, gradient boosting utilizes a technique called shrinkage or learning rate, which controls the contribution of each weak learner to the ensemble. By scaling down the contribution of each tree, gradient boosting reduces the risk of overfitting and allows for a more gradual learning process. On the other hand, random forest does not incorporate such a mechanism to control the complexity of the ensemble, potentially leading to overfitting, especially in high-dimensional or noisy datasets.

Also, gradient boosting typically uses shallow decision trees as weak learners, which are less prone to overfitting compared to deep trees used in random forest. Lastly, gradient boosting offers regularization techniques such as tree pruning, which further help in controlling model complexity and mitigating overfitting so that the ensemble focuses on important features and avoids fitting noise in the data.

5. Conclusions and Further Works

In conclusion, this study has explored the use of ML algorithms, with a particular focus on gradient boosting, for diagnosing cardiovascular conditions, specifically heart attacks, and using tabular risk factor data. Through thorough investigation and experimentation, the effectiveness of gradient boosting is demonstrated in accurately identifying cardiovascular phenomena indicative of heart attack risk.

The results of these experiments have shown that gradient boosting, among other ML algorithms evaluated, exhibits promising performance in diagnosing heart attack risk based on tabular risk factor data. By iteratively optimizing weak learners and controlling model complexity, gradient boosting achieves robust predictive capabilities while mitigating the risk of overfitting.

Furthermore, the proposed study underscores the importance of leveraging tabular risk factor data in cardiovascular health monitoring and diagnosis. By systematically analyzing risk factors and employing advanced ML techniques, the diagnostic capabilities are enhanced, potentially leading to early detection and intervention for individuals at high risk of cardiovascular events.

However, acknowledging the limitations of the proposed study is crucial, including the need for further research to optimize gradient boosting models for low-power environments and explore additional medical surveillance applications. Additionally, ongoing efforts in data collection and feature engineering may further improve the performance of ML models in cardiovascular risk prediction, while trying DL using neural networks may yield better results with reduced susceptibility to overfitting.

In summary, this study contributes to advancing the understanding and diagnosis of heart attack risk using tabular risk factor data and the gradient boosting algorithm. The findings offer valuable insights into improving cardiovascular health monitoring and diagnosis in clinical settings, ultimately benefiting

patients and healthcare professionals by enabling more effective risk assessment and personalized interventions.

Funding Support

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2026/R/1447).

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/ritwikb3/heart-disease-cleveland.

Author Contribution Statement

Esraa Eldesouky: Conceptualization, Project administration, Funding acquisition. **Walaa H. Elashmawi:** Conceptualization, Software, Formal analysis, Resources, Supervision. **Ahmed S. Salama:** Data curation, Writing – review & editing. **Ahmed Ali:** Methodology, Investigation, Writing – original draft. **Magi Mahfouz:** Validation, Visualization.

References

[1] Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., . . . , & Liu, L. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. *The Lancet*, *364*(9438), 937–952. https://doi.org/10.1016/S0140-6736(04)17018-9

[2] Mendis, S., Thygesen, K., Kuulasmaa, K., Giampaoli, S., Mähönen, M., Ngu Blackett, K., & Liu, L. (2011). World Health Organization definition of myocardial infarction: 2008–09 revision. *International Journal of Epidemiology*, *40*(1), 139–146. https://doi.org/10.1093/ije/dyq165

[3] Nandal, N., Goel, L., & Tanwar, R. (2022). Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis. *F1000Research*, *11*, 1126. https://doi.org/10.12688/f1000research.123776.1

[4] Bertelli, E. V. M., Ribeiro, G. J. S., de Barros, F. R. B., de Oliveira, J. V., Marinho, M. L., & de Araújo Pinto, A. (2025). Myocardial infarction and traditional cardiovascular risk factors in older patients in primary care. *Revista Brasileira de Enfermagem*, *78*(5), e20240535. https://doi.org/10.1590/0034-7167-2024-0535

[5] Moysidis, D. V., Papazoglou, A. S., Anastasiou, V., Daios, S., Karagiannidis, E., Fyntanidou, B., . . . , & Vassilikos, V. (2025). Acute myocardial infarction in patients without standard modifiable risk factors–A state-of-the-art review. *Trends in Cardiovascular Medicine*, *35*(8), 506–516. https://doi.org/10.1016/j.tcm.2025.06.005

[6] Wereski, R., Kimenai, D. M., Bularga, A., Taggart, C., Lowe, D. J., Mills, N. L., & Chapman, A. R. (2022). Risk factors for type 1 and type 2 myocardial infarction. *European Heart Journal*, *43*(2), 127–135. https://doi.org/10.1093/eurheartj/ehab581

[7] van Oortmerssen, J. A. E., Ntlapo, N., Tilly, M. J., Bramer, W. M., den Ruijter, H. M., Boersma, E., . . . , & Roeters van Lennep, J. E. (2024). Burden of risk factors in women and men with unrecognized myocardial infarction: A systematic review and meta-analysis. *Cardiovascular Research*, *120*(14), 1683–1692. https://doi.org/10.1093/cvr/cvae188

[8] Sagris, M., Antonopoulos, A. S., Theofilis, P., Oikonomou, E., Siasos, G., Tsalamandris, S., . . . , & Tousoulis, D. (2022). Risk factors profile of young and older patients with myocardial infarction. *Cardiovascular Research*, *118*(10), 2281–2292. https://doi.org/10.1093/cvr/cvab264

[9] Vasan, R. S., Pan, S., Larson, M. G., Mitchell, G. F., & Xanthakis, V. (2021). Arteriosclerosis, atherosclerosis, and cardiovascular health: Joint relations to the incidence of cardiovascular disease. *Hypertension*, *78*(5), 1232–1240. https://doi.org/10.1161/HYPERTENSIONAHA.121.18075

[10] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, *145*, 105458. https://doi.org/10.1016/j.compbiomed.2022.105458

[11] Pallavi, G., Altalbe, A., & Rangarajan, P. K. (2025). QFM-BioPred: Quantum fusion model for bioactivity prediction in cardiovascular disease drug discovery. *Journal of Computational and Cognitive Engineering*. Advance online publication. https://doi.org/10.47852/bonviewJCCE52025138

[12] Ahamad, G. N., Shafiullah, Fatima, H., Imdadullah, Zakariya, S. M., Abbas, M., . . . , & Usman, M. (2023). Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease. *Processes*, *11*(3), 734. https://doi.org/10.3390/pr11030734

[13] Grzeszczyk, M. K., Trzciński, T., & Sitek, A. (2024). MISS: Multiclass Interpretable Scoring Systems. In *Proceedings of the 2024 SIAM International Conference on Data Mining*, 55–63. https://doi.org/10.1137/1.9781611978032.7

[14] Paikaray, B. K., Nayak, S. K., Pramanik, J., Thapa, R., Swain, D., & Samal, A. K. (2023). Application of ensemble learning in predicting heart failure. *AIP Conference Proceedings*, *2981*(1), 020017. https://doi.org/10.1063/5.0182674

[15] Bouqentar, M. A., Terrada, O., Lamrani, D., Ouhmida, A., Cherradi, B., & Raihani, A. (2023). Primary prediction of heart disease using machine learning algorithms and SMOTE. In *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology*, 1–7. https://doi.org/10.1109/IRASET57153.2023.10153064

[16] Elshafee, A. K., & Hassan, M. K. (2023). HEMA: A proposed robot for improving healthcare access in underserved communities. *Journal of Engineering Research*, *7*(5), 208–219.

[17] Xu, W., Wang, L., Wang, B., & Cheng, W. (2022). Intelligent recognition algorithm of multiple myocardial infarction based on morphological feature extraction. *Processes*, *10*(11), 2348. https://doi.org/10.3390/pr10112348

[18] Bertsimas, D., Mingardi, L., & Stellato, B. (2021). Machine learning for real-time heart disease prediction. *IEEE Journal of Biomedical and Health Informatics*, *25*(9), 3627–3637. https://doi.org/10.1109/JBHI.2021.3066347

[19] Gulati, S., Guleria, K., & Goyal, N. (2022). Classification and detection of coronary heart disease using machine learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering*, 1728–1732. https://doi.org/10.1109/ICACITE53722.2022.9823547

[20] Hossain, M. I., Maruf, M. H., Khan, M. A. R., Prity, F. S., Fatema, S., Ejaz, M. S., & Khan, M. A. S. (2023). Heart disease prediction using distinct artificial intelligence techniques: Performance analysis and comparison. *Iran Journal of Computer Science*, *6*(4), 397–417. https://doi.org/10.1007/s42044-023-00148-7

[21] Ali, M. L., Sadi, M. S., & Goni, M. O. (2024). Diagnosis of heart diseases: A fuzzy-logic-based approach. *PLOS One*, *19*(2), e0293112. https://doi.org/10.1371/journal.pone.0293112

[22] Lee, J. (2025). HeartSense: Leveraging machine learning to predict cardiovascular risk. *Journal of Advanced Artificial Intelligence, Engineering and Technology*, *1*(4), 44.

[23] Izonin, I., Tkachenko, R., & Greguš, M. (2022). I-PNN: An improved probabilistic neural network for binary classification of imbalanced medical data. In *Database and Expert Systems Applications: 33rd International Conference*, 147–157. https://doi.org/10.1007/978-3-031-12426-6_12

[24] Kolukisa, B., & Bakir-Gungor, B. (2023). Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis. *Computer Standards & Interfaces*, *84*, 103706. https://doi.org/10.1016/j.csi.2022.103706

[25] Bharadwaj, Prakash, K. B., & Kanagachidambaresan, G. R. (2021). Pattern recognition and machine learning. In K. B. Prakash & G. R. Kanagachidambaresan (Eds.), *Programming with TensorFlow: Solution for edge computing applications* (pp. 105–144). Springer International Publishing. https://doi.org/10.1007/978-3-030-57077-4_11

[26] Kieu, T., Yang, B., Guo, C., Jensen, C. S., Zhao, Y., Huang, F., & Zheng, K. (2022). Robust and explainable autoencoders for unsupervised time series outlier detection. In *2022 IEEE 38th International Conference on Data Engineering*, 3038–3050. https://doi.org/10.1109/ICDE53745.2022.00273

[27] Golub, G. H., & van Loan, C. F. (1996). *Matrix computations* (3rd ed.). USA: Johns Hopkins University Press.

[28] Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., . . . , & Holzinger, A. (2023). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 329–341. https://doi.org/10.1109/TPAMI.2022.3145392

How to Cite: Eldesouky, E., Elashmawi, W. H., Salama, A. S., Ali, A., & Mahfouz, M. (2026). Enhancing Cardiovascular Disease Risk Prediction: A Comparative Analysis of Machine Learning Techniques. *Journal of Computational and Cognitive Engineering*, *5*(2), 354–370. https://doi.org/10.47852/bonviewJCCE62026735