Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.56, p.240-247. November/December, 2025.
DOI: https://doi.org/10.5935/jetia. v11i56.2734

RESEARCH ARTICLE OPEN ACCESS

MACHINE LEARNING MODELS ANALYSIS USING MULTI-FACTORS FOR HEART ATTACK RISK PREDICTION

Rahul Nigam*[1], Hare Ram Jha[2] and Ravi Ranjan[3]

[1, 2] EC Department, Medicaps University, Indore, Madhya-Pradesh, India.
[3] CS Department, Muzaffarpur Institute of Technology, Muzaffarpur, Bihar, India.

[1]http://orcid.org/0000-0001-8555-5731, [2]http://orcid.org/0000-0002-5547-0523, [3]http://orcid.org/0009-0000-5631-6555

Email: *transferrn@gmail.com, hrjha.ece@gmail.com, ravi.bce.bit@gmail.com

ARTICLE INFO

Article History
Received: September 24, 2025
Revised: November 20, 2025
Accepted: December 1, 2025
Published: December 31, 2025

Keywords:
Machine Learning,
Heart Attack,
Supervised ML,
Risk Prediction,
Nural Network.

ABSTRACT

In human population heart attacks is a leading cause of dying in world, so the need of accurate and timely risk assessment strategies is critically required. This comparative study presents an analysis of heart attack risk prediction by applying different machine learning methods such as Logistic Regression, Decision Tree, Random Forest etc. These methods were trained the model by using dataset of clinical features and then tested. The prediction of heart attack show based on various parameters such as accuracy, precision, recall, F1 score, training time, ROC-AUC, and interpretability. The prediction results indicates that XGBoost and Random Forest with selected top features provide higher predictive accuracy score of 98.49% and 99.01%. Decision trees provide faster processing and neural networks model give complex relationships. These models comparatively less stable and require greater resources. This study shows that the application of different machine learning techniques with appropriate feature selection as an effective solution for prediction of heart attack risk.

I. INTRODUCTION

The human heart is a very important organ of body and it is responsible for flowing blood through the systemic and pulmonary circulations, supplying tissues with oxygen and essential nutrients. Cardiovascular diseases (CVDs) represent a special type of pathologies affecting the heart and blood flow system. It also includes coronary artery problem, cerebrovascular accidents, Surrounding artery problem, rheumatic heart problems. When flow of blood block due to blood clot, resulting in block oxygen to cardiac tissues it is an emergency condition and timely treatment is very critical. The arterial blockage occurs because of accumulation semisolid mass and blood thicken. Heart diseases, also known as CVDs are the main cause of death around the world. In 2019, nearly 18 million people died from these diseases, which is about one out of every three deaths globally. Most of these deaths around 85% were caused by heart attacks and strokes. Unfortunately, more than three-quarters of these deaths happen in countries with lower or middle incomes, where healthcare resources may be limited. Among people who die prematurely (before age 70) from noncommunicable diseases, about 38% die because of heart-related problems.

In India, the risk of having a heart attack is much higher compared to many other parts of the world. In India a large portion of the deaths caused by heart attack. High blood pressure is one of the biggest risk factors for heart attack, especially in India. The number of people dying from heart diseases in India has been rising rapidly from about 2.26 million deaths in 1990 to nearly 4.8 million deaths in 2020 [1]. Studies show that coronary heart disease affects between 1.6% and 7.4% of people living in rural areas, while in cities, it affects between 1% and 13.2% of the population [2]. A heart attack is a very critical medical emergency where timely treatment is very essential. Even a small delay in treatment can cause to an irrecoverable damage to the heart or result in destructive outcomes. The ability to know the risk of cardiovascular disease is very important for these reasons. To do this it is require to know about the accurate individual risk profiles, so more effective diagnostic, and treatment possible that cure patient with reduce healthcare costs. Furthermore, the early identification of heart conditions can decrease the risk of myocardial infarction.

II. RISK FACTORS

Most heart diseases can be cure by early detection or prediction by using risk factors, so proper management can be possible. These factors include use of tobacco, unhealthy diet routine, obesity, physical inactivity hours, excessive alcohol consumption and sometimes air pollution.

II.1 AGE AND GENDER

The risk of heart attack increases with increase in age. With men this risk becomes more after the age of 45 while in women it generally rises after the age of 50.

II.2TOBACCO CONSUMPTION

Smoking and consumption of tobacco products are major contributors to cardiovascular disease. Discontinuing use of tobacco has reduces the risk of heart attack.

II.3 HIGH BLOOD PRESSURE

Hypertension can damage coronary arteries, result in interrupt blood flow to the heart. This risk is further increase with other conditions like obesity, raised cholesterol levels or diabetes.

II.4 CHOLESTEROL AND TRIGLYCERIDES

Raise level of LDL (bad cholesterol) can blocked blood vessels. Similarly, high triglyceride levels increase risk. Oppositely, normal levels of HDL (good cholesterol) can lower the chance of heart attack.

II.5 OBESITY

Too much body weight with high blood pressure, diabetes, high triglycerides amount, LDL and low value of HDL increase the heart attack probability.

II.6 DIABETES

In this condition either human organs does not produce enough insulin or may be not utilise it effectively. Both situations increasing heart attack risk.

II.7 METABOLIC SYNDROME

A situation of increase in abdominal fat, high BP, low HDL, raise triglycerides level and diabetes create metabolic syndrome, which doubles the risk of heart disease.

II.8 FAMILY HISTORY

If a close relative suffered from heart attack (early age) may increase heart attack risk.

II.9 PHYSICAL INACTIVITY

Having a sedentary lifestyle is tend towards a higher chance of heart attacks. Regular physical activity helps to improve heart health condition.

II.10 UNHEALTHY DIET

Diet with high sugars, saturated fats, fast foods, low fibre foods and lack nutrition foods are strongly leads to increase risk of heart attack. Oppositely, a balanced diet with fruits, vegetables, whole grains, and high fibre foods leads to support heart health.

II.11 STRESS

Stress either in body or mental, particularly involving intense anger or anxiety, can increase the risk of high BP and high heart rate, ultimately leads towards a heart attack condition.

II.12 DRUG USE

The use of certain substances such as alcohol, cocaine, and other stuff, can affect coronary artery or block the flow of blood into and out from the heart and set off a condition of heart attack.

II.13 AUTOIMMUNE DISEASES

Conditions like rheumatoid arthritis or lupus are increasing risk of heart attacks.

III. REVIEW OF MACHINE LEARNING APPLICATIONS IN HEART DISEASE DETECTION AND PREDICTION

Recent developments shows that an early identification and management of heart disease plays an important role. According to [3] introduced an application that combines algorithmic decision-making with real-time sensor data that allow physicians to monitor patients continuously. The system also has live video streaming facility and sends alerts to doctors via GSM technology, when a monitored parameter exceeds from reference limit. In another work, developed a diagnostic system using heart disease datasets to classify individuals and convey the classification process [4]. Similarly, explored heart disease attributes using several supervised classification methods. Their dataset consisted of 303 records and 76 features, from these 14 important attributes selected to improve accuracy of model [5]. For previous authors, conducted a detailed survey of existing work in heart disease prediction. They give review in three broad areas: traditional classification and data mining methods, contemporary statistical learning models and emerging deep learning techniques [6]. Employed models such as logistic regression and K-Nearest Neighbors to improve accuracy for heart attacks prediction. They show these models can also support decision making in process [7].

Performed a comparative study using the UCI Heart Disease dataset, this study includes 14 critical features. They used deep learning model to demonstrate predictive capability with 94.2% accuracy and suggesting practical applications in GSM based health monitoring systems [8]. According to [9], examined multiple heart disease risk factors and compared the predictive performance of various statistical models [9]. Introduced a clustering-based approach using k-modes with Huang initialization to improve classification accuracy. They used multiple models and the result shows that MLP model achieved the best performance with an accuracy of 87.28% [10]. According to [11], proposed a healthcare monitoring that integrates sensor-based physiological data and clinical records stored in the cloud. The system employed a Bi-Directional Long Short-Term Memory model and achieved an impressive 98.86% accuracy, outperforming other heart disease prediction methods. In a related study, [12] designed a diagnostic model for Coronary-Artery-Disease using Recurrent Convolutional Neural Networks (RCNN), aimed at supporting clinicians in making informed treatment decisions. According to [13], evaluated six commonly used machine learning algorithms using two datasets—Cleveland and IEEE Dataport.

Their findings showed that LR delivered the 90% accuracy and AdaBoost provide 90% accuracy with respective datasets. In [14], reviewed trends in heart disease diagnostic research conducted between 2014 and 2022 and give suggestion for future researches. According to [15] utilized deep learning to predict four major cardiac conditions using public ECG image datasets. Their CNN model gives 98.23% accuracy. When it is combined with Naive Bayes for feature extraction give 99.79% accuracy. Tested several machine learning models on heart disease datasets and shows that highest accuracy of model is 96% [16]. According to [17] conducted a study of several machine learning methods to forecast heart attack risk. Their comparative results showed Random Forest give best accuracy of 88.52%. For [18], analyzed data from 9,499 patients using tree-based models like GLMM trees and GMERF. The result shows that GMERF achieved the best predictive performance.In [19] used SMOTE to handle data imbalance, finding XGBoost achieved best performing accuracy of 97.57% with good sensitivity. In [20] reviewed ML methods to assist doctors in early heart disease prediction and treatment planning.

According to [21] used several machine learning models alongside LSTM neural networks, model and shows that Gradient Boosting and LSTM achieved the results in terms of accuracy and ROC-AUC scores. For [22] proposed a model that combine ML methods and achieved the best performance with multiple datasets for CVD prediction. In [23] suggested a hybrid architecture combines different ML algorithms and achieved an accuracy of 99.7% in CVD classification using ECG data. According to [24] highlight the benefits of applying machine learning in CVD treatment, so timely interventions and ongoing monitoring of cognitive function is possible. According to [25] performed experiments with and without geographical features on CVD datasets. They find XGBoost achieved the highest accuracy of 95.24% when geographical data included. In [26], introduced an integrated machine learning model combining Particle Swarm Optimization and Neural Networks for improved CVD prediction. In [27] developed a system utilizing 13 risk parameters from the Cleveland dataset and tuned machine learning models processed for different stroke stages classification. In summary, machine learning and deep learning technologies become essential for early diagnosis and prediction of heart related diseases.

Various studies have made different tools and models for real-time patient monitoring data, health records and datasets to accurately and timely prediction the risk of CVDs. Most of these methods include supervised algorithms such as logistic regression, decision trees, random forests, Naive Bayes and KNN with more complex deep learning models like CNN, LSTMs, and hybrid models. Some of these models demonstrated high percentage of accuracy up to 95% to predict risk of heart disease and lead towards heart attacks. The Internet of Things (IoT) devices with these models created cloud-based healthcare systems that further enhanced by patient monitoring and providing quick medical services. Additionally, methods like feature selection and data balancing particularly SMOTE have played an important role in improving machine learning model stability and reducing error rate. In India where CVDs problem becomes a major public health issue, these techniques create a valuable support for early detection and risk prediction. Overall advanced ML algorithms lead to strong potential in decision-making, reduce death rate and support patient treatment. The hybrid machine learning methods on diverse dataset provide further improvement in prediction accuracy and patient health.

IV. METHODOLOGY

Machine learning gained popularity in almost every field and specially in healthcare because of its ability to analyse massive volume of complex dataset, so early diagnosis of patient is possible. These methods used here enable to identify relationships between various risk factors and heart diseases. Machine learning methods are broadly classified into three categories: supervised learning, unsupervised learning and reinforcement learning.

IV.1 SUPERVISED ML

Supervised machine learning methods are the most widely used in heart related diseases research. In these methods models are trained on datasets where inputs are labelled with features and with a known outcome. In this method the algorithm learns the relation between inputs and outputs and after learning phase over the model can predict outcomes for unseen data. Most popular techniques used in supervised learning are probabilistic models like Naive Bayes, Logistic Regression, kernel-based classifiers Support Vector Machines and methods like Random Forest and Gradient Boosting and Neural networks. Among these methods hybrid models and SVMs have shown strong performance in recent cardiovascular diseases due to their capability of handle non-linear relationships datasets.

IV.2 UNSUPERVISED ML

Unsupervised learning methods based on finding hidden formation in unlabelled data. These methods include clustering algorithms and dimensionality reduction techniques. These methods are especially used in finding subgroups within patient groups, detecting statics or inspect unknown patterns in datasets. In CVDs diseases unsupervised machine learning models have been applied to classify patients based on risk factors, thereby supporting more focused interventions.

V. MACHINE LEARNING MODELS USED FOR HEART ATTACK RISK PREDICTION

Following ML models used in this research for Prediction of risk of heart attack: -

V.1 LOGISTIC REGRESSION

Logistic regression is a basic model primarily used for binary classification works and commonly used here to predict the presence or absence of heart disease. It calculates the probability of an event based on a linear combination of input variables using the logistic function. This method is valued for its intelligibility, clear perception about each factor contributes to the overall risk. It performs best when relationships among features and outcomes are approximately linear.

V.2 SUPPORT VECTOR MACHINES (SVM)

SVM is a powerful classification technique based on supervised learning method that creates a marginal boundary to separate data points belongs to different classes. It is especially effective when datasets with many dimensional features and non-linear partitions. In this type of cases kernel functions used to transform the input range. SVMs are best known for its robustness and this property successfully applied to various heart disease datasets.

V.3 DECISION TREES

Decision Trees operate on the principle of splitting data repeatedly into subgroups based on feature values, resulting in an illustratable tree structure that handles decision-making processes. By applying this method, it is easy to understand and visualize data with patients features. However sometimes they are subjected to overfitting, that can limit generalizability of model on complex datasets.

V.4 RANDOM FORESTS

Random Forest is an improved version of decision trees method. RF method combines multiple trees into a group to increase predictive accuracy and reduce the overfitting. It is used for both categorical and numerical data efficiently. It also provides importance statics of all features, so in this way it helps to know about the feature that influence heart attack risk most.

V.5 NAIVE BAYES

Naive Bayes is also a type of probability-based classifier that mainly uses Bayes' theorem for prediction. It assumes that features are independent, which simplifies calculations and makes it highly efficient. Due to this independence characteristic, it often performs extremely well specially on clinical datasets of patient. It is especially useful if computational resources are limited.

V.6 GRADIENT BOOSTING MODELS (XGBOOST)

Gradient Boosting models such as XGBoost have become more popular in classification tasks due to high performance. These models combine weak models (e.g. decision trees) in a consecutive way and optimize for unsolved errors. XGBoost adjust features to optimize specific problem enhance speed and efficiently handle categorical data. It well-suited for complex datasets in heart disease prediction tasks and often appear as best performs among different machine learning models.

V.7 NEURAL NETWORKS

Neural Networks made of interconnected layers of nodes called neurons that mimic the complex structure of human brain neurons. They can express complex, non-linear relationships among clinical features such as blood pressure, cholesterol levels and heart rate. They require more computational resources and larger datasets compared to traditional models. This model predictive capabilities make it valuable for heart diseases risk modelling.

V.8 K-NEAREST NEIGHBORS (KNN)

KNN is a simple sample-based classifier that assigns labels based on the nearest examples in the training data. It is well on smaller and well-balanced datasets but may not get good yield with high-dimensional or noisy data. Its performance is also sensitive with choice of distance metric and the number of neighbors (K).

V.9 DEEP NEURAL NETWORKS (DNN):

DNN machine learning method is an improvement of traditional neural networks include multiple layers network. These layers allow model to be suitable for complex feature hierarchies. In field of heart attack prediction DNNs can learn sophisticated relations across all features of big and high-dimensional datasets. These models have strong predictive power. These models require careful tuning, so that large volumes of data not be go in overfitting zone. They are also less illustratable than simpler models, although recent methods have improved their importance in medical applications.

Table 1: Comparisons of ML Algorithms.

Algorithm	Type	Strengths	Limitations	Interpretability	Suitability for Clinical Use
Logistic Regression	Linear, Supervised	Simple, fast, interpretable; works well with linearly separable data [28]	Poor at capturing non-linear patterns	High	Excellent (widely used)
Decision Tree	Non-linear, Supervised	Easy to interpret; handles both categorical & numerical data [29]	Prone to overfitting	High	Good
Random Forest	Ensemble, Supervised	Robust, reduces overfitting; handles missing data well [30]	Less interpretable than single trees	Moderate	Very Good
Gradient Boosting (XGBoost)	Ensemble, Supervised	High accuracy; handles complex patterns and large datasets [31]	Requires careful tuning; slower to train	Moderate	Very Good
Support Vector Machine (SVM)	Supervised	Effective in high-dimensional spaces; good with small-to-medium datasets [32]	Poor interpretability; not ideal for large datasets	Low	Moderate
K-Nearest Neighbors (KNN)	Instance-based, Supervised	Simple; no training phase [33]	Sensitive to noisy data and scale; computationally expensive at inference	Low	Limited
Naive Bayes	Probabilistc Supervised	Fast, efficient with high-dimensional data [34]	Assumes feature independence; not ideal for correlated medical data	Moderate	Fair
Neural Network (NN)	Deep Learning	Captures complex relationships; adaptable to many data types [35]	Requires large data; risk of overfitting; harder to interpret	Low	Moderate
Deep Neural Network (DNN)	Deep Learning	Handles highly complex and high-dimensional data [36]	Requires large datasets and computational power; less interpretable	Low	Good (with explainability tools)

Source: Authors, (2025).

VI. RESULT AND DISCUSSIONS

This Prediction of heart attack used different machine learning algorithm on labelled dataset obtain from Mendeley Contain 1319 dataset with top 8 features. The evaluation of various machine learning models indicates that Random Forest, XGBoost, and Gradient Boosting are the top performers in terms of prediction accuracy and reliability. Among them, the XGBoost model using selected key features stands out, offering a high accuracy of 98.48%, balanced precision and recall, and a very fast training time of just 0.09 seconds. The Random Forest model with top features follows closely behind, delivering strong results with slightly more training time. Gradient Boosting also maintains high performance, although it takes comparatively longer to train. Table 2 shows the comparisons of various machine learning algorithms models we simulated using all features available in dataset.

Table 2: Comparisons of ML Algorithms (All Features included).

Algorithm	Accuracy	Precision	Recall	F1-Score	Training time	ROC-AUC	Interpretability (qualitative)
Logistic Regression	0.7727	0.8643	0.7469	0.8013	0.0109	0.8701	High - Easily explainable
Decision Tree	0.9811	0.9816	0.9877	0.9846	0.0070	0.9791	High - Easily explainable
Random Forest	0.9811	0.9876	0.9815	0.9845	0.3600	0.9953	Moderate - Feature importance available
Gradient Boosting	0.9848	0.9877	0.9877	0.9877	0.1097	0.9846	Moderate - Tree ensemble, partially explainable
XGBoost	0.9848	0.9877	0.9877	0.9877	0.1097	0.9846	Moderate - Can extract feature importance
Support Vector Machine (SVM)	0.8030	0.9167	0.7469	0.8231	0.2383	0.8945	Low - Hard to interpret kernel decisions
K-Nearest Neighbors (KNN)	0.6402	0.7680	0.5926	0.6690	0.0070	0.7142	Moderate - Based on nearby instances
Naive Bayes	0.6705	0.9870	0.4691	0.6360	0.0040	0.8456	Moderate - Probabilistic interpretation
Neural Network (NN) MLP Classifier	0.8939	0.9295	0.8951	0.9119	6.1477	0.9468	Low - Complex and black -box
Deep Neural Network (DNN)	0.8447	0.8854	0.8580	0.8715	12.8110	0.9152	Low - Black-box neural model

Source: Authors, (2025).

Decision Tree classifiers, despite their simplicity, yield robust results with an accuracy of 98.1% and an exceptionally fast training speed, making them a practical choice for situations where interpretability and speed are crucial. Neural networks, including deep learning models, demonstrate good predictive strength, but their longer training times and slightly lower performance place them behind ensemble methods for this task. Table 3 shows the comparisons of various machine learning algorithms models we simulated using selected top 5 features get by using ML algorithms available in dataset.

Table 3: Comparisons of ML Algorithms (Top 5 Features included).

Algorithm	Accuracy	Precision	Recall	F1-Score	Training time	ROC-AUC	Interpretability (qualitative)
Logistic Regression	0.7879	0.8897	0.7469	0.8121	0.0069	0.8748	High-Easily explainable
Decision Tree	0.9811	0.9816	0.9877	0.9846	0.0070	0.9791	High-Easily explainable
Random Forest	0.9848	0.9877	0.9877	0.9877	0.4428	0.9910	Moderate-Feature importance available
Gradient Boosting	0.9848	0.9877	0.9877	0.9877	0.3833	0.9815	Moderate-Tree ensemble, partially explainable
XGBoost	0.9848	0.9877	0.9877	0.9877	0.0937	0.9849	Moderate-Can extract feature importance
Support Vector Machine (SVM)	0.7955	0.9286	0.7222	0.8125	0.1805	0.8968	Low-Hard to interpret kernel decisions
K-Nearest Neighbors (KNN)	0.7462	0.8800	0.6790	0.7666	0.0040	0.8463	Moderate - Based on nearby instances
Naive Bayes	0.6818	0.9875	0.4877	0.6529	0.0030	0.8485	Moderate-Probabilistic interpretation
Neural Network (NN) MLP Classifier	0.9167	0.9605	0.9012	0.9299	5.7250	0.9650	Low-Complex and black -box
Deep Neural Network (DNN)	0.8712	0.9211	0.8642	0.8917	11.9552	0.9401	Low-Black-box neural model

Source: Authors, (2025).

Support Vector Machines and Logistic Regression offer moderate performance, but they fall short when compared to ensemble-based models. Naive Bayes and K-Nearest Neighbors perform the weakest, particularly in recall and F1 score, which makes them less suitable for sensitive medical predictions such as heart attack risk.

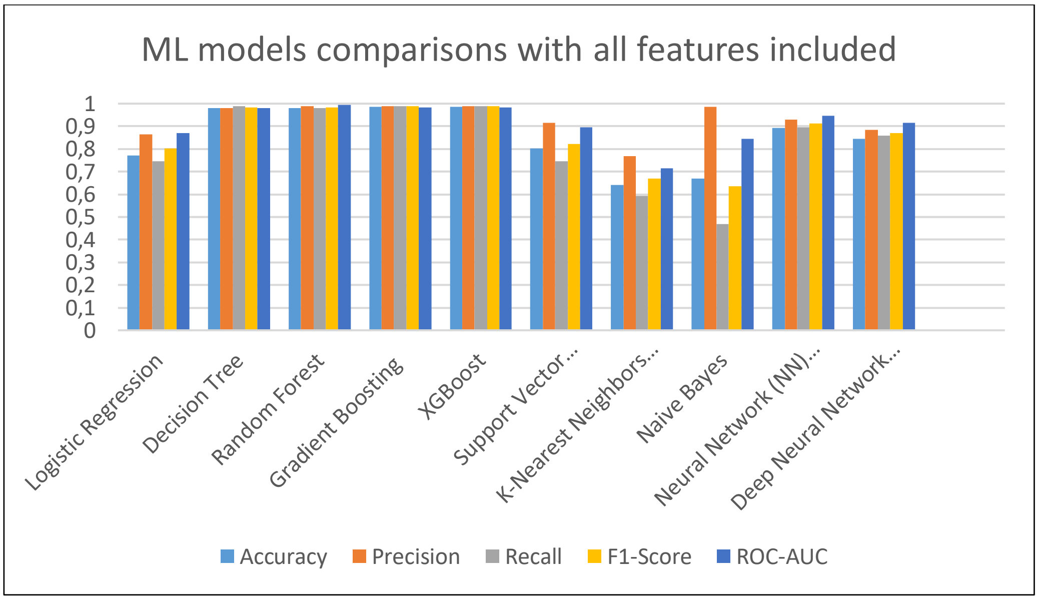

Figure 1: ML models comparisons based on all features selected.
Source: Authors, (2025).

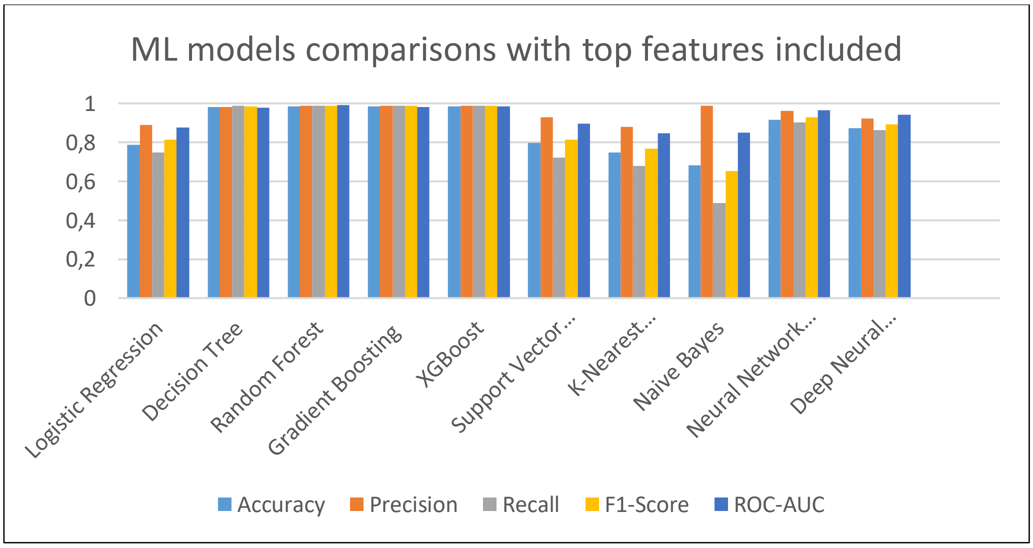

Figure 2: ML models comparisons based on only top features included.
Source: Authors, (2025).

Figure 1 and Figure 2 showing risk parameter for each model under both feature selection strategies. In summary, tree-based ensemble models, especially XGBoost and Random Forest with selected features are the most effective options for this application. While neural networks can be useful in certain contexts their computational cost and marginally lower effectiveness should be carefully considered. Applying feature selection enhanced the effectiveness of most models, highlighting the advantage of focusing on the most relevant attributes in the dataset.

VII. CONCLUSIONS

After evaluating different machine learning models for predicting heart attack risk, tree-based ensemble methods especially XGBoost and Random Forest using selected top features emerge as the most dependable and efficient. These models consistently deliver strong predictive performance while maintaining short training times, making them well-suited for both large-scale deployment and time-sensitive scenarios. While Decision Trees provide quick and interpretable results and neural networks capture complex data relationships, they tend to either lag slightly in accuracy or require more computational effort. Conventional models such as Support Vector Machines, Logistic Regression, Naive Bayes, and K-Nearest Neighbors generally perform less effectively and are therefore less suitable for critical healthcare applications. Overall, ensemble learning with carefully chosen features proves to be the most effective approach for heart attack prediction. It offers a reliable solution that balances accuracy with practical implementation, aligning well with the demands of real-world medical environments.

VIII. AUTHOR'S CONTRIBUTION

Conceptualization: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.
Methodology: Rahul Nigam.
Investigation: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.
Discussion of results: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.
Writing – Original Draft: Rahul Nigam.
Writing – Review and Editing: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.
Resources: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.
Supervision: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.
Approval of the final text: Rahul Nigam, Hare Ram Jha and Ravi Ranjan.

IX. REFERENCES

[1] C. J. Murray and A. D. Lopez, "Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study," Lancet, vol. 349, pp. 1498–1504, 1997, doi: 10.1016/S0140-6736(96)07492-2.

[2] R. Gupta, P. Joshi, V. Mohan, K. S. Reddy, and S. Yusuf, "Epidemiology and causation of coronary heart disease and stroke in India," Heart, vol. 94, pp. 16–26, 2008, doi: 10.1136/hrt.2007.132951.

[3] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," World Journal of Engineering and Technology, vol. 6, pp. 854–873, 2018, doi: 10.4236/wjet.2018.64057.

[4] A. Ul Haq, J. P. Li, M. Hammad Memon, S. Nazir, and R. Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," Mobile Information Systems, vol. 2018, Article ID 3860146, 21 pages, 2018, doi: 10.1155/2018/3860146.

[5] D. Shah, S. Patel, and S. Bharti, "Heart Disease Prediction using Machine Learning Techniques," SN Computer Science, vol. 1, 2020, doi: 10.1007/s42979-020-00365-y.

[6] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-vascular diseases (CVD) using Machine Learning and Deep Learning techniques," ICT Express, vol. 8, pp. 109–116, 2022, doi: 10.1016/j.icte.2021.08.021.

[7] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, p. 012072, 2021, doi: 10.1088/1757-899X/1022/1/012072.

[8] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," Computational Intelligence and Neuroscience, vol. 2021, Article ID 8387680, 2021, doi: 10.1155/2021/8387680.

[9] R. Katarya and S. K. Meena, "Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis," Health Technol., vol. 11, pp. 87–97, 2021, doi: 10.1007/s12553-020-00505-7.

[10] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, p. 88, 2023, doi: 10.3390/a16020088.

[11] A. Nancy, D. Ravindran, P. Vincent, K. Srinivasan, and D. Gutiérrez, "IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning," Electronics, vol. 11, no. 15, p. 2292, 2022, doi: 10.3390/electronics11152292.

[12] K. Saikumar and V. Rajesh, "A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset," Int. J. Syst. Assur. Eng. Manag., vol. 15, no. 1, pp. 135–151, Jan. 2024.

[13] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," Processes, vol. 11, no. 4, p. 1210, 2023, doi: 10.3390/pr11041210.

[14] P. Rani, A. Jain, R. Lamba, R. Sachdeva, K. Kumar, and M. Kumar, "An Extensive Review of Machine Learning and Deep Learning Techniques on Heart Disease Classification and Prediction," Archives of Computational Methods in Engineering, vol. 31, 2024, doi: 10.1007/s11831-024-10075-w.

[15] M. B. Abubaker and B. Babayiğit, “Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods,” IEEE Trans. Artif. Intell., vol. 4, no. 2, pp. 373–382, Apr. 2023, doi: 10.1109/TAI.2022.3159505.

[16] S. Sivakannan et al., “cardiovascular diseases prediction by machine learning incorporation with deep learning,” Frontiers in Medicine, vol. 10, 2023, doi: 10.3389/fmed.2023.1150933.

[17] A. A. Stonier, R. K. Gorantla, and K. Manoj, “Cardiac disease risk prediction using machine learning algorithms,” Healthc. Technol. Lett., vol. 11, no. 4, pp. 213–217, Nov. 2023, doi: 10.1049/htl2.12053.

[18] F. Asadi, R. Homayounfar, Y. Mehrali et al., “Detection of cardiovascular disease cases using advanced tree-based machine learning algorithms,” Sci. Rep., vol. 14, p. 22230, 2024, doi: 10.1038/s41598-024-72819-9.

[19] H. El-Sofany, B. Bouallegue, and Y. M. A. El-Latif, “A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method,” Sci. Rep., vol. 14, p. 23277, 2024, doi: 10.1038/s41598-024-74656-2.

[20] G. S. Bhavekar, A. Das Goswami, C. P. Vasantrao et al., “heart disease prediction using machine learning, deep Learning and optimization techniques—A semantic review,” Multimed. Tools Appl., vol. 83, pp. 86895–86922, 2024, doi: 10.1007/s11042-024-19680-0.

[21] S. Pathan and S. Imran, “Integrated Machine Learning and Deep Learning Models for Cardiovascular Disease Risk Prediction: A Comprehensive Comparative Study,” J. Intell. Learn. Syst. Appl., vol. 16, pp. 12–22, 2024, doi: 10.4236/jilsa.2024.161002.

[22] H. Sadr, A. Salari, M. T. Ashoobi et al., “cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models,” Eur. J. Med. Res., vol. 29, p. 455, 2024, doi: 10.1186/s40001-024-02044-7.

[23] K. K. Kumar, G. S. Suneetha, K. Reddy, P. Rao, S. K. Vududala, and A. Gupta, “Electro Cardio Gram Using Different Machine Learning Techniques for Early Heart Attack Prediction,” J. Neonatal Surg., vol. 14, no. 19S, Apr. 2025. [Online]. Available: https://www.jneonatalsurg.com/index.php/jns/article/view/3101

[24] S. Jain, H. Jha, K. Vishwakarma, and R. Nigam, "Predicting Cognitive Decline in Heart Failure Patients Using ML Based Multi Parameter Risk Scoring," Journal of Neonatal Surgery, vol. 14, pp. 1741–1748, 2025, doi: 10.63682/jns.v14i32S.7653.

[25] B. E. Sianga, M. C. Mbago, and A. S. Msengwa, “Predicting the prevalence of cardiovascular diseases using machine learning algorithms,” Intelligence-Based Medicine, vol. 11, 2025, Art. no. 100199, doi: 10.1016/j.ibmed.2025.100199.

[26] S. R. Reddy and G. V. Murthy, “Cardiovascular Disease Prediction Using Particle Swarm Optimization and Neural Network Based an Integrated Framework,” SN Comput. Sci., vol. 6, no. 2, Jan. 2025, doi: 10.1007/s42979-025-03723-w.

[27] N. Joshi and T. Dave, “Improved Accuracy for Heart Disease Diagnosis Using Machine Learning Techniques,” JIWE, vol. 4, no. 1, pp. 42–52, Feb. 2025.

[28] B. X. Liew, F. M. Kovacs, D. Rügamer, and A. Royuela, “Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain,” Eur. Spine J., vol. 31, no. 8, pp. 2082–2091, 2022.

[29] H. A. Abdulqader and A. M. Abdulazeez, “A review on decision tree algorithm in healthcare applications,” Indones. J. Comput. Sci., vol. 13, no. 3, 2024.

[30] M. Fascia, “Machine learning applications in medical prognostics: a comprehensive review,” arXiv preprint arXiv:2408.02344, 2024.

[31] N. H. Alhumaidi, D. Dermawan, H. F. Kamaruzaman, and N. Alotaiq, “The use of machine learning for analyzing real-world data in disease prediction and management: systematic review,” JMIR Med. Inform., vol. 13, no. 1, p. e68898, 2025.

[32] S. Karamizadeh, S. M. Abdullah, M. Halimi, J. Shayan, and M. J. Rajabi, “Advantage and drawback of support vector machine functionality,” in Proc. 2014 Int. Conf. Comput., Commun., Control Technol. (I4CT), Sep. 2014, pp. 63–65.

[33] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, “Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications,” J. Big Data, vol. 11, no. 1, p. 113, 2024.

[34] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, “Variable selection for Naïve Bayes classification,” Comput. Oper. Res., vol. 135, p. 105456, 2021.

[35] R. Saleem, B. Yuan, F. Kurugollu, A. Anjum, and L. Liu, “Explaining deep neural networks: A survey on the global interpretation methods,” Neurocomputing, vol. 513, pp. 165–180, 2022.

[36] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. I. Lee, “Explainable AI for trees: From local explanations to global understanding,” arXiv preprint arXiv:1905.04610, 2019.