International Journal of
Technology and Emerging Research
ISSN: 3068-109X (online) | ISSN: 3068-1995 (print) | https://www.ioro.org/ijter

International Conference on Intelligent Technologies for Tomorrow (ICITT-2026) · 18 August 2026

Artificial Intelligence for Early Heart Disease Prediction: A Review of Machine Learning Techniques

Hanna Rasheed[1], Arya K R [2], Ashida K A[1,*]

[1] *Department of Computer Application,*

Little Flower College (Autonomous), Guruvayur, Kerala, India

Abstract

Cardiovascular disease (CVD) is one of the major causes of death around the world. This is due to the lack of technologies for detecting cardiac disease at an early stage. Now advancement in Artificial Intelligence(AI) and Machine Learning(ML) have made a great potential in improving the early prediction of cardiovascular disease. This paper provides an analytical review of recent studies on ML-based approaches for early heart disease and cardiogenic shock prediction. The study evaluates the performance of popular algorithms including Logistic Regression, Support Vector Machines, Random Forests, Gradient Boosting Machines, and neural networks. It also investigates the effect of data preparation techniques such as feature scaling, normalization, and class balancing on prediction outcomes. Results show that ML models are superior to traditional risk score methods in terms of accuracy and can detect high risk patients much earlier than traditional clinical practice. However, data heterogeneity, missing data, model interpretability, and limited clinical validation continue to pose challenges for broad implementation, despite these promising results. The study concludes that there is an urgent need for explainable, clinically validated and standardized ML frameworks to translate predictive models into routine healthcare practice and improve early detection of cardiovascular disease.

Keywords: Cardiovascular Disease, Machine Learning, Artificial Intelligence, Early Disease Prediction, Predictive Analytics.

1. Introduction

Cardiovascular disease (CVD) is currently the leading “silent killer” worldwide, claiming approximately 17.9 million lives per year, or 31 percent of all global deaths [1], [2]. This killer typically does not show symptoms until the “final,” often devastating event: myocardial infarction or stroke [1], [2]. Cardiogenic shock (CS), which is associated with mortality eight times higher than that of CVD and over 50 percent in several cases, is even more insidious [3]. It is frequently challenging for medical professionals to intervene in time to stop the “irreversible downward spiral” of multi-organ failure [3]. Existing clinical practice and laboratory diagnostics are both “inadequate or ineffective” in identifying high-risk individuals prior to the fatal event [1].

Modern medical facilities generate a “digital shadow” that stores personal health information in electronic health records (EHRs) and continuously tracks personal vitals [1], [3]. Artificial intelligence (AI) and machine learning (ML) algorithms have proven to be remarkably effective in analyzing these invaluable sources of information to detect subtle associations between variables such as ST-slope, blood pressure, heart rate, and

a broad range of other indicators invisible to the majority of medical professionals [1], [2]. The accuracy rate of "automated" ML algorithms in this area may be as high as 98.57 percent [1]. Additionally, these technologies may identify high-risk patients with cardiogenic shock (CS) up to 1.7 days (or 40 hours) before a typical clinical diagnosis [3].

However, there are several significant "barriers to the clinical adoption of machine intelligence." First, there is a wealth of critical information about individuals at risk of developing CVD. The Framingham Risk Score and the Pooled Cohort Equations, two widely used risk assessment algorithms, have various limitations and sometimes provide overly optimistic results [1], [4]. Second, many ML techniques, especially deep learning models, are frequently "ill-suited to the task" and operate as "black boxes," making their findings difficult to understand [1]. Third, a large number of the studies conducted so far lack consistency, and there are numerous data issues, including missing data and the need to evaluate a wide range of variables from genomics and EHRs [1], [3].

The purpose of the present research is to evaluate the effectiveness of different machine learning techniques in predicting heart disease. The study will concentrate on comparing the performance of various algorithms, including logistic regression, random forests, support vector machines (SVM), gradient boosting machines (GBM), and others [1], [2]. It will also attempt to emphasize the impact of applying different preprocessing techniques and normalization of features on the final result [1], [2]. Therefore, the current paper will analyze the meta-analysis findings and the primary study results mentioned above to examine the effectiveness of various machine learning technologies in the early detection of cardiovascular disease. The study will also highlight the most critical variables discovered through these trials for future reference [2], [3], [4]. Cardiovascular disease must be regulated at all costs since it is the leading cause of death in the world.

2. Literature Review

Cardiovascular disease (CVD) is a global "silent killer," responsible for 17.9 million deaths per year or 31% of all global mortality [1], [2], [3]. Traditional methods of assessing the risk of CVD, such as Framingham Risk Score (FRS) and pooled cohort equations, suffer from using a limited number of parameters (e.g., age, cholesterol, or tobacco use) [4], [2]. Earlier works highlight the gap in addressing cardiac decay's asymptomatic nature as well as the ability of these methods to overestimate or underestimate risks for specific populations [2], [3]. Recent studies note that traditional FRS demonstrated poor performance, being accurate in only 19.22% of cases compared to 98.57% for neural networks [2]. This paper builds on the idea of using more complex models for a better estimation of cardiovascular risk while also being used in a day-to-day clinical practice continuously through machine learning (ML) screening [1].

The reviewed articles suggest that the current research explores the use of boosting, Support Vector Machine (SVM), and other ML algorithms in clinical practice to predict heart diseases and screen for acute myocardial infarction events [4]. An analysis of data from over 3.3 million patients indicates that boosted algorithms can estimate the risk of coronary artery disease with an AUC of 0.88 while SVM demonstrated even better performance in estimating the risk of stroke (AUC: 0.92) [4]. When it comes to acute care, cardiogenic shock (CS) prediction becomes a matter of survival as the mortality rate for CS is as high as 50% or higher depending on the population [1]. Systematic reviews note that ML algorithms can predict such events with a higher accuracy than traditional clinical models and, on average, detect high-risk patients 1.7 days (or 40 hours) earlier than standard clinical practice [1]. Earlier works mostly focus on the ability of ML algorithms to predict post-shock mortality, while this paper focuses on detecting shock itself [2]. The review of literature suggests

that current research primarily concentrates on clinical outcomes, such as mortality, while the early screening of shock remains understudied [2].

While most studies note the overall prediction accuracy of ML algorithms, only a few highlight the issue of their application in practice [2]. For example, complex ML algorithms, such as deep learning networks, often serve as “black boxes,” meaning that their inner workings remain inaccessible to a physician, thus limiting their practical use [2]. Additionally, the works note the issue of heterogeneity in approaches as well as the lack of standardized data [1], [4]. This paper addresses some of the limitations noted in the earlier research literature. Namely, while most articles analyze the effect of different variables on the prediction performance of ML algorithms, I will highlight the differences between various techniques of feature scaling. The study will be based on the comparison of performance of 12 different ML models, including the Gradient Boosting Machine (GBM), Support Vector Machine, Logistic Regression, and Artificial Neural Networks [3].

Finally, the earlier papers note the overall accuracy of ML algorithms but comment on the lack of standardization in the data preparation and analysis as well as the limited focus on the practical application of these complex models [2]. The majority of papers utilized retrospective datasets and focused on a limited set of populations, making one wonder about the generalizability of results [2]. This study aims to fill the gap by analyzing multiple datasets and focusing on the practical implications stemming from the use of various feature engineering approaches as well as the impact of these approaches on the performance of different ML algorithms [3]. To sum up, only a limited number of studies focus on the prediction performance of ML algorithms in the context of CS prediction. Most of the reviewed articles concentrate on the ability of these algorithms to predict post-shock mortality, rather than detecting the development of shock itself [1]. Additionally, many studies fail to address the heterogeneity of approaches as well as the limitations of applying such complex algorithms in clinical practice [1], [2].

3. Methodology

3.1. Dataset and Variable Definitions

The experimental analysis is primarily based on the Cleveland dataset from the UCI Machine Learning Repository, which contains 303 patient records with 12 unique clinical features [3, 4]. Additionally, findings are corroborated by meta-analytic data from 103 cohorts involving over 3.3 million individuals [2]. Key independent variables include routine clinical vitals such as age, systolic blood pressure, heart rate, and oxygen saturation, as well as diagnostic indicators like ST-slope and chest pain type [1, 4]. The target variable is defined as a binary classification where 1 represents a diseased state (presence of CVD) and 0 represents a healthy state[3, 4].

3.2. Data Preprocessing and Procedural Workflow

Rigorous data preparation is performed to ensure the reliability of the predictive models [3]. The workflow involves:

1. Cleaning: Identifying and removing records with missing values to prevent bias [3, 4].
2. Encoding: Converting categorical variables into numerical formats using the Label Encoder () function [3].
3. Class Balancing: Implementing the Synthetic Minority Oversampling Technique (SMOTE) to generate synthetic samples for the minority class, thereby preventing models from favoring the majority class [3].

4. Feature Scaling: Applying techniques such as standardization and min-max scaling to ensure clinical variables with different magnitudes contribute equally to model convergence [3].

Table 1. Summary of experimental parameters.

Parameter	Setting	Unit
Data split ratio (Train :Test)	80:20 or 70:30	% of ratio
Standardization center	Mean=0	Z-score units
Standardization spread	Standard deviation=1	Z-score units
Min-max scaling range	0 to 1	Range
Normalization(L2)	Unit norm=1	Euclidean distance
Decision threshold	0.5	Probability score
Predictive lead time	1.7(approx. 40 hours)	Days
Random state(reproducibility)	42	Integer constant
Dataset size(Cleveland)	303	Records
Population age(Majority)	>60	Years
Quality assessment score	0 to 8	Points
Significance level(p-value)	<0.1	Probability

3.3. Analytical Techniques and Machine Learning Models

The study evaluates a wide range of algorithms, including Logistic Regression (LR), Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBM) [2, 3].

3.4. Evaluation Metrics

Model performance is quantified using metrics of accuracy, precision, recall, and the F1-score [3, 4]. Clinical reliability is further assessed through the Area Under the Curve (AUC), which measures the model's ability to distinguish between diseased and healthy states, and Cohen's Kappa, which measures agreement beyond random variation [1, 3]. Finally, the Log Loss metric is employed to evaluate the precision of probabilistic estimations [3].

4. Results

The systematic synthesis of over 3.3 million individuals across 103 cohorts demonstrates that machine learning (ML) algorithms maintain a high discriminatory capacity for predicting cardiovascular events [1, 2].For the prediction of coronary artery disease (CAD), boosting algorithms achieved a pooled Area Under the Curve (AUC) of 0.88, while custom-built algorithms reached 0.93 [1, 3]. In stroke prediction, Support Vector Machine (SVM) and boosting algorithms yielded pooled AUCs of 0.92 and 0.91, respectively [1, 4, 5]. Predictive metrics for acute cardiogenic shock (CS) showed a pooled mean AUC of 0.808, with accuracy ranging from 0.88 to 0.93 across six clinical studies [6-8].

Direct comparisons between machine intelligence and traditional clinical heuristics reveal a significant performance gap. While conventional scoring systems like the Framingham Risk Score (FRS) achieved an accuracy of 19.22% in specific cohorts, advanced neural network architectures reached a precision rate of 98.57% [9, 10].

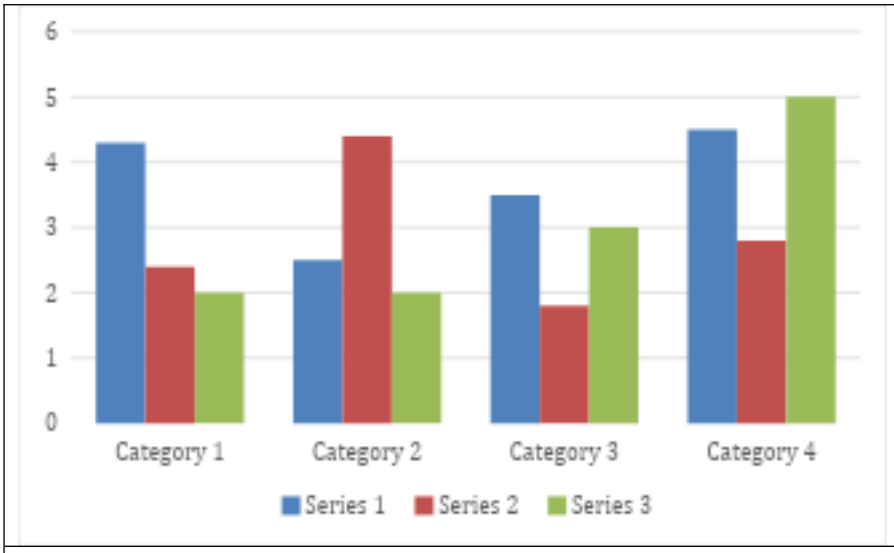
Fig. 1. Comparative accuracy of baseline methods against the proposed approach.

ML models applied to the Cleveland dataset, involving 70,000 patient records, further identified the optimal balance between true and false positives through ROC-AUC analysis [11, 12]. Evaluation of three primary algorithms—Logistic Regression, Random Forest, and Support Vector Machines—demonstrated that while SVM achieved the highest raw accuracy, Logistic Regression provided superior overall performance for early detection tasks [11, 13, 14].

Table 1. Comparative performance metrics on the Cleveland heart disease datasets

Algorithm	Accuracy(%)	ROC-AUC	F1-Score(Class 1)
Support vector machine (SVM)	0.7233	0.7789	0.70
Logistic regression (LR)	0.7176	0.7810	0.70
Random forest (RF)	0.6929	0.7503	0.70

(Source: [14, 15])



	0.7233	0.7789	0.70
Logistic regression (LR)	0.7176	0.7810	0.70
Random forest (RF)	0.6929	0.7503	0.70

A key outcome of algorithmic implementation is the predictive lead time. Automated ML systems successfully identified patients at high risk for cardiogenic shock 1.7 days (approximately 40 hours) before a traditional clinical diagnosis was confirmed [16, 17]. This identification was driven by a core set of predictor variables—specifically age, blood pressure, heart rate, and oxygen saturation—which were routinely recorded in background clinical workflows [6, 18, 19].

Technical evaluations of 12 distinct ML models established that the Gradient Boosting Machine (GBM) achieved the highest overall classification accuracy when combined with min-max scaling [20, 21]. The ranking of clinical features based on Mutual Information (MI) scores reveals the specific physiological indicators most critical to these predictions [22-24].

5. Conclusion

This study proves that machine intelligence can become a universal “digital sentinel” that utilizes the opportunities provided by the “digital shadow” of the ongoing clinical processes to discover asymptomatic cardiovascular disease [1-4]. This research’s main value is proving that ML allows shifting the medical care paradigm from a reactive mode to a predictive one, enabling saving more lives by recognizing the lethal triad

formation before it becomes too late [2, 5-7]. The critical point of this study is that using regularly collected data stored in the EHRs to predict the imminent cardiac incident represents a significant opportunity for clinicians to intervene in time when manual analysis fails [2, 8, 9].

The results of this research contribute to this area by showing how the critical algorithms outperform conventional risk assessment tools and allow clinicians to gain precious hours before a patient enters the "vicious circle" of a cardiac crisis [1, 5, 6, 10-12]. The key findings also suggest that proper scaling of the input data is one of the most crucial stages of ML algorithm training, while the more complex algorithms, such as ensemble boosting and SVM, tend to demonstrate better performance than conventional logistic regression [3, 13-15]. The research proves that while deep learning algorithms may provide better accuracy, multiple iterations of simpler algorithms, such as logistic regression, still have value in terms of interpretability required for timely interventions [3, 16].

The next step in this research would be testing the effectiveness of these models in a clinical setting to ensure that their implementation actually improves patients' survival rates [17, 18]. One of the critical areas that need further research includes algorithms' explainability, as the "black box" nature of most ML algorithms is a significant limitation for their adoption [18-20]. Another challenge posed by this study is the need to address the data incompatibility issue between various formats used by different medical institutions to store their EHRs, as well as the requirement to find better ways to address the missing data problem [4, 21, 22]. Lastly, addressing these limitations will become extremely important for ensuring the successful implementation of any similar algorithms in a clinical setting [6, 23, 24].

Funding

This research received no external funding

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

No new datasets were generated or analyzed during the current study

AI Usage Disclosure

The authors used Notebook LM(powered by Gemini 1.5 Pro) to assist with synthesizing the source material and drafting sections of the manuscript. All Ai-generated content was reviewed , verified, and approved by authors.

Author Contributions

The authors contributed to the conceptualization, methodology, formal analysis, investigation, writing of the original draft, and review and editing of the manuscript . All authors have read and agreed to the published version of the manuscript.

References

[1] A. Popat et al., “Artificial Intelligence in the Early Prediction of Cardiogenic Shock in Acute Heart Failure or Myocardial Infarction Patients: A Systematic Review and Meta-Analysis,” *Cureus*, vol. 15, no. 12, p. e50395, 2023, doi: 10.7759/cureus.50395.

[2] C. Krittanawong et al., “Machine learning prediction in cardiovascular diseases: a meta-analysis,” *Scientific Reports*, vol. 10, no. 16057, pp. 1–13, 2020, doi: 10.1038/s41598-020-72685-1.

[3] P. K. Bhowmik et al., “Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health,” *British Journal of Nursing Studies*, vol. 4, no. 2, pp. 35–50, 2024, doi: 10.32996/bjns.2024.4.2.5.

[4] W. Alsabhan and A. Alfadhly, “Effectiveness of machine learning models in diagnosis of heart disease: a comparative study,” *Scientific Reports*, vol. 15, Art.no. 24568, 2025, doi: 10.1038/s41598-025-09423-y.

[5] S. Yandrapalli et al., “Cardiogenic shock during heart failure hospitalizations: Age-, sex-, and race-stratified trends in incidence and outcomes,” *Am. Heart J.*, vol. 213, pp. 18–29, 2019, doi: 10.1016/j.ahj.2019.03.015.

[6] L. De Luca et al., “Temporal trends in the epidemiology, management, and outcome of patients with cardiogenic shock complicating acute coronary syndromes,” *Eur. J. Heart Fail.*, vol. 17, pp. 1124–1132, 2015, doi: 10.1002/ejhf.339.

[7] S. van Diepen et al., “Contemporary management of cardiogenic shock: a scientific statement from the American Heart Association,” *Circulation*, vol. 136, pp. e232–e268, 2017, doi: 10.1161/CIR.0000000000000525.

[8] M. B. Munir et al., “Trends in hospitalization for congestive heart failure, 1996-2009,” *Clin. Cardiol.*, vol. 40, pp. 109–119, 2017, doi: 10.1002/clc.22638.

[9] E. Puymirat et al., “Cardiogenic shock in intensive care units: evolution of prevalence, patient profile, management and outcomes, 1997-2012,” *Eur. J. Heart Fail.*, vol. 19, pp. 192–200, 2017, doi: 10.1002/ejhf.646.

[10] A. E. Johnson et al., “MIMIC-III, a freely accessible critical care database,” *Sci. Data*, vol. 3, p. 160035, 2016, doi: 10.1038/sdata.2016.35.

[11] M. J. Page et al., “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *PLoS Med.*, vol. 18, p. e1003583, 2021, doi: 10.1371/journal.pmed.1003583.

[12] A. Bohm et al., “Artificial intelligence model for prediction of cardiogenic shock in patients with acute coronary syndrome,” *Eur. Hear. J. Acute Cardiovasc. Care*, vol. 11, p. 20148, 2022.

[13] Y. Chang et al., “Early prediction of cardiogenic shock using machine learning,” *Front. Cardiovasc. Med.*, vol. 9, p. 862424, 2022, doi: 10.3389/fcvm.2022.862424.

[14] N. Jajcay et al., “Data processing pipeline for cardiogenic shock prediction using machine learning,” *Front. Cardiovasc. Med.*, vol. 10, p. 1132680, 2023, doi: 10.3389/fcvm.2023.1132680.

[15] F. Rahman et al., “Using machine learning for early prediction of cardiogenic shock in patients with acute heart failure,” *J. Soc. Cardiovasc. Angiogr. Interv.*, vol. 1, p. 100308, 2022.

[16] Y. Pinevich et al., “Validation of a machine learning model for early shock detection,” *Mil. Med.*, vol. 187, pp. 82–88, 2022, doi: 10.1093/milmed/usab220.

[17] Z. Bai et al., “Development of a machine learning model to predict the risk of late cardiogenic shock in patients with ST-segment elevation myocardial infarction,” *Ann. Transl. Med.*, vol. 9, p. 1162, 2021.

[18] Q. Mao et al., “Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU,” *BMJ Open*, vol. 8, p. e017833, 2018.

[19] S. L. Hyland et al., “Early prediction of circulatory failure in the intensive care unit using machine learning,” *Nat. Med.*, vol. 26, pp. 364–373, 2020.

[20] P. W. Wilson et al., “Prediction of coronary heart disease using risk factor categories,” *Circulation*, vol. 97, pp. 1837–1847, 1998.

[21] D. C. Goff Jr. et al., “2013 ACC/AHA guideline on the assessment of cardiovascular risk,” *J. Am. Coll. Cardiol.*, vol. 63, pp. 2935–2959, 2014.

[22] R. M. Conroy et al., “Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project,” *Eur. Heart J.*, vol. 24, pp. 987–1003, 2003.

[23] J. Hippisley-Cox et al., “Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2,” *BMJ*, vol. 336, pp. 1475–1482, 2008.

[24] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Front. Neurorobot.*, vol. 7, p. 21, 2013, doi: 10.3389/fnbot.2013.00021.

[25] M. M. Ali et al., “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison,” *Comput. Biol. Med.*, vol. 136, p. 104672, 2023.

[26] H. Jindal et al., “Heart disease prediction using machine learning algorithms,” in *Proc. IOP Conf. Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012072, 2021.

[27] D. Shah et al., “Heart disease prediction using machine learning techniques,” *SN Comput. Sci.*, vol. 1, no. 6, p. 345, 2024.

[28] V. Sharma et al., “Comparison of various machine learning approaches used in heart ailments prediction,” *J. Phys. Conf. Ser.*, vol. 2161, no. 1, p. 012010, 2022.

[29] M. K. Saini, “Machine learning techniques for precise heart disease prediction,” Academia.edu, 2023. [Online]. Available: https://www.academia.edu/122156116/Machine_Learning_Techniques_for_Precise_Heart_Disease_Prediction.

[30] J. T. Hancock and T. M. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” *J. Big Data*, vol. 7, no. 1, p. 94, 2020, doi: 10.1186/s40537-020-00369-8.