BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

AI-BASED EARLY DETECTION OF HEART DISEASE: A REVIEW

[1]Triveni Gaikwad, [2]Prof. Poonam Pawar
[1]Student, Professor
Department of Artificial Intelligence & Data Science
Ajeenkya DY Patil School of Engineering, Pune, Maharashtra, India
[1]trivenigaikwad02@gmail.com, [2]pawar.punam08@gmail.com

Abstract

Cardiovascular diseases have become one of the major sources of morbidity in the world, therefore early detection of the people at risk would allow medical professionals more time to act. One of the approaches to the problem through ML is the use of the information that includes such variables as age, blood pressure, total cholesterol, blood glucose, heart rate, symptoms of the patient, and test results. In this paper, we give a brief description of the most popular algorithms for predicting the risk of heart diseases using ML: Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbors, and Neural Networks. Deep learning, Explainable AI (XAI), small data, class imbalance, bad data quality, privacy, and hospital validation will also be discussed. This narrative literature review is based on the synopsis of the project, some pieces of literature, and dataset documentation. There was no training of any ML model nor conducting of any experiments. As it follows from the papers under study, ML allows to separate patients into risky and non-risky categories; however, the results strongly depend on the selected dataset, goal of prediction, validation method, and patients themselves. Results from small benchmark datasets tend to be highly optimistic and cannot be used to predict clinical relevance of the models. We also present a process flow of how the data should be prepared, modeled, evaluated, interpreted and externally validated and hope that such models will help physicians more than replace them.

Index Terms: Artificial intelligence, heart disease, prediction of heart diseases, machine learning, deep learning, explainable artificial intelligence, clinical decision support, risk assessment.

I. INTRODUCTION

Heart diseases constitute a severe health issue that concerns a significant part of the population. Age, high blood pressure, cholesterol levels, diabetes, smoking, lifestyle, genetic predispositions, symptoms and results of diagnostic procedures are among the factors which affect development and progression of the disease. In case when patients requiring more attention can be identified in advance, they can be examined and treated by physicians at an earlier stage. Nevertheless, a high score does not mean that the person suffers from the disease. In order to diagnose heart disease, a patient should undergo a physical examination, tests, ECG imaging of his/her heart and get an evaluation by a physician.

Risk score estimation methods based on patients' medical history, clinical guidelines, and test results have always been quite common. Nevertheless, health data contain numerous interrelated variables, and there are some complicated relationships that cannot be easily defined by using rules. The algorithms for machine

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

learning use the data of previous cases and predict the probability of an outcome, considering the interactions not represented in rules. Still, there are some requirements for making predictions properly. Cardiovascular diseases risk prediction is gradually switching from classic classifiers towards deep and interpretable machine learning algorithms which give information about the decision process. In the recent systematic review on the subject of AI-based cardiovascular risk prediction models, the need for validation of the performance of such models and publication of their results is noted [1]. Also, in a systematic review and meta-analysis of machine learning models which used electronic health records data, the good pooled performance of some machine learning models was found despite a lot of variance between the analysed studies [2].

In this paper, the use of AI for the early diagnosis of heart disease will be discussed. The algorithms are reviewed, challenges regarding the datasets and assessment are highlighted, current research gaps are identified, and a potential workflow is suggested. What is provided is a structured analysis and a suggested approach to the problem; there is no solution or implementation provided in this study. The paper is just an analysis based solely on the project synopsis, dataset documentation, and literature review.

The structure of the paper continues with the following sections. The objectives and scope are outlined in Section II. Clinical data, machine learning, and explainability are presented in Section III and Section IV provides a review of the algorithms, followed by their comparison in Section V. The workflow that may be considered in the future development is outlined in Section VI. The literature review and research gaps are summarized in Sections VII and VIII.

II. OBJECTIVES AND REVIEW SCOPE

There are four aims of this paper and these are:

(1) To analyze the use of AI and Machine Learning methods for heart disease prediction,

(2) To discuss different algorithms and their fit into clinical datasets,

(3) To suggest future research directions and datasets,

(4) To state the limitations and gaps that cause less accuracy in early detection.

We are going to analyze structured data concerning such features as age, blood pressure, cholesterol, heart rate, symptoms and other test results. We will use such algorithms as Logistic Regression, SVM, Decision Tree, Random Forest, Naïve Bayes, KNN and Neural Networks including deep learning and XAI. This paper is a narrative review according to the outline and chosen papers. It is not a systematic review because there was no search protocol created before.

Since this paper is just an analysis of previous researches, it is useful to present the boundaries of this paper. The following table shows the comparison of our paper with an experimental one.

TABLE I. COMPARISON OF THE SCOPE OF THIS PAPER TO AN EXPERIMENTAL

Aspect	Literature Review	Experimental Study
Main activity	Synthesizes existing research	Trains and tests models
Patient data	Not used	Obtained by researchers

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Software	No code executed	Code developed and run
Results	No new performance metrics	Reports performance metrics
Outcome	Understanding and proposed workflow	Evidence of model performance

RESEARCH

There are a number of problems with the methodology. To begin with, the literature was collected manually and not through any systematic search. As a result, there is a possibility that some useful studies were missed. Furthermore, the datasets used in each paper vary in nature. This means that their findings cannot be compared and, as a result, the algorithms are not ranked according to their accuracy.

III. BACKGROUND

A. Clinical Data for Risk Prediction

Predictive dataset includes input attributes and the output attribute (label). Input attributes may include demographics, clinical attributes, symptoms/test categories. The labels could represent the presence or absence of a certain disease in question (as in most benchmark datasets) or cardiovascular events in the future. However, these two tasks are different because the presence of existing disease and predicting the future events are two distinct things; therefore, each research should state what they are predicting and when labels were created.

UCI Heart Disease repository is a common benchmark that comprises datasets from four different databases. Specifically, the Cleveland subset contains 303 observations and is mostly considered together with 14 attributes, one of which serves as the target attribute indicating the presence of heart disease [3]. It is easily available and reproducible, which makes it a good choice for training and comparisons of methods. Nonetheless, the dataset is rather small and its data reflect the practices of many years ago, thus, the results based on this dataset are hardly generalizable.

Table II shows 14 attributes of the Cleveland subset according to the repository documentation.

TABLE II. ATTRIBUTES OF THE CLEVELAND SUBSET [3]

Attribute	**Description**
age	Age in years
sex	Sex of the patient
cp	Chest pain type
trestbps	Resting blood pressure
chol	Serum cholesterol
fbs	Fasting blood sugar above 120 mg/dl (yes or no)
restecg	Resting electrocardiographic result
thalach	Maximum heart rate achieved
exang	Exercise-induced angina (yes or no)
oldpeak	ST depression induced by exercise relative to rest
slope	Slope of the peak exercise ST segment

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

ca	Number of major vessels (0–3) colored by fluoroscopy
thal	Thallium stress-test result (normal, fixed defect, reversible defect)
num	Diagnosis of heart disease (the target)

All of these characteristics are regular numbers or categorical codes and hence the data is considered structured data. ECG waves and medical images do not fall in this category. They are not represented in columns and usually managed using some other models.

B. Machine Learning and Deep Learning

TABLE III. DATA TYPES AND MODEL FAMILIES OFTEN CONSIDERED

Data type	Examples	Model families
Structured clinical records	Age, blood pressure, cholesterol, test results	Logistic Regression, SVM, tree-based methods, Naïve Bayes, KNN
Physiological signals	ECG recordings	Deep learning, such as convolutional or recurrent networks
Medical images	Echocardiograms, CT or MRI scans	Deep learning, such as convolutional networks
Combined data	Records together with signals or images	Fusion of several model types

This is merely a general guideline based on best practices. The ultimate decision will have to be made based on the data at hand and comparing different models.

C. Explainable AI

Explainable AI aims at explaining the factors responsible for generating the output of the model. SHAP gives a contribution value to each feature in a single prediction from a single framework [5]. LIME creates a simplified model for one particular prediction to explain how the real model behaves in its vicinity [6]. Both techniques help in understanding the role of features in getting either a higher or lower score in one particular case. Nevertheless, the explanation does not explain causality or prove the clinical validity of the model.

For example, the explanation of a heart disease model might tell that a high resting blood pressure raised the risk score, while youth reduced the score. A doctor would check if this is in line with medical knowledge.

D. Prediction Tasks in Cardiovascular Care

The notion of prediction comprises a variety of approaches to data analysis, yet the existing literature does not distinguish between them. Diagnosis implies the identification of the presence or absence of some problem at the moment, risk stratification – the identification of the patients who are more likely to need treatment compared to other patients, and prognosis – the prediction of what will happen in the future during the specified period of time. Each approach needs different terms, different samples of patients and criteria

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

for success.

Different levels of significance need to be assigned to the two kinds of errors depending on the specific task. In screening, the cost of the false-negative error is usually higher than that of the false-positive one; hence, a higher level of sensitivity needs to be used. In contrast, for the purposes of diagnostic confirmation, high specificity is necessary. The costs of the two kinds of errors being context-dependent, the estimate of accuracy in such conditions turns out to be meaningless.

IV. REVIEW OF COMMON ALGORITHMS

A. Logistic Regression

The Logistic Regression model determines the probability of whether the output will be Yes/No based on calculating the weighted sum of the input variables and then passing the output through the sigmoid function. It makes for a good baseline model since the coefficients of the model are easy to interpret, and the predictions can also be calibrated easily. The problem with it is that it fails with nonlinear relationships.

Support Vector Machine

SVM algorithm looks for a maximum margin boundary separator among classes and through the use of kernels, it becomes possible for the algorithm to create a non-linear boundary [7]. It generally works well with properly scaled features; however, its efficiency depends on choosing the right kernel and hyperparameters. Probability scores generated by the algorithm need to be calibrated to interpret them as risks and the algorithm is difficult to interpret.

B. Decision Tree and Random Forest

The predictions of the Decision Tree are made by posing questions and answering them according to the values of the variables. This has a straightforward structure which makes it easy to examine. The drawback of a very deep Decision Tree is the problem of overfitting. The Random Forest builds many trees based on randomly selected variables and data, and makes a vote of all of them [8]. This lowers the variance; however, this model handles nonlinear interactions but is hard to interpret compared to the Decision Tree.

C. Naive Bayes

Naïve Bayes follows the Bayes' theorem and assumes that the predictors are independent given the class variable. The classifier is very fast in computing the solution and therefore can be used as a benchmark for comparison. Predictors are correlated in clinical settings, and therefore this assumption may not necessarily always hold.

D. K-Nearest Neighbors

New data points are classified using the labels of the closest training data points. This classifier has been investigated earlier [9]. It is simple to implement but is sensitive to attribute scaling, distance metric and the value of K. The time to make predictions increases with an increase in the size of the data set, while the presence of irrelevant attributes impacts its performance negatively.

E. Neural Networks and Deep Learning

Complex patterns can be captured using the links between multiple units in neural networks. The approach

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

of deep learning can be used when dealing with big amounts of data, ECG signals, medical imaging data, or the combination of structured and unstructured data. The disadvantages are in the need for having a sufficient amount of data for learning, substantial computing power, as well as tuning and the ability to interpret results. More complicated models can be used only when they produce better results and not because it is assumed that they will be more accurate.

Preprocessing techniques also differ for different algorithms. KNN and SVM depend on distance values and margins, respectively. Thus, they are sensitive to feature scaling. However, logistic regression is often subject to regularization to prevent overfitting. Tree-based methods are less sensitive to feature scaling. It is important to consider the preprocessing techniques for each method when comparing them.

V. COMPARATIVE VIEW

The difference depends upon the degree of interpretability, computational power, and nonlinearity. None of the techniques is necessarily better than the others. To conduct an objective comparison between them, the algorithms must have the same data partitioning, the same pre-processing on the basis of the training data set alone, the same target definition, and the same evaluation process.

TABLE IV. COMPARISON OF COMMON ALGORITHMS

Algorithm	Potential strengths	Key cautions
Logistic Regression	Readable baseline; probability modelling	May miss complex nonlinear patterns
SVM	Flexible decision boundaries	Scaling and tuning; probability calibration
Decision Tree	Easy to visualize and explain	Can overfit; unstable splits
Random Forest	Ensemble robustness; nonlinear patterns	Less transparent; importance can mislead
Naïve Bayes	Fast and simple baseline	Conditional-independence assumption
KNN	Simple, non-parametric method	Scaling, distance choice, slow inference
Neural Network / DL	Can learn complex representations	Data-hungry; harder to explain and validate

Table IV includes only those strengths and precautions that have been mentioned in literature; there is no ranking of the accuracy of the mentioned algorithms. An intuitive approach will necessarily be less flexible than a very flexible approach which will necessarily be unintuitive; therefore, the balance depends on the use of the algorithm model. The interpretability will always be as significant in the clinical setting as the score, as the physician should confirm the result.

VI. PROPOSED CONCEPTUAL WORKFLOW

In the light of papers considered, the following approach could be developed for workflow. It is rather a future project than something that we have done already. For example, there is a guideline for the reporting of

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

predictive models called TRIPOD checklist including 22 items [10].

A. Dataset Selection and Target Definition

This will demand the data set to possess qualities that match the patient population, variables, and the outcome. An illustration of a data set suitable for a simple learning activity is the UCI Heart Disease Data set [3]. Whatever the case, it is crucial to know the source of the data, the population within the data set, what each variable means, how the outcome is coded and the limitations of the data set. If it involves prediction of risks in future, the data set should possess the required follow-up period because diagnosis of a disease cannot be considered a future event.

B. Data Preparation

First of all, we need to identify missing values, duplication, inconsistencies in units, outliers, and class distributions. Secondly, we need to select an encoding scheme for the categorical variables and the need for scaling of numeric variables. Imputation, scaling, feature selection, and resampling should all be done on training data only within each cross-validation fold. However, if we use the test data, then there will be information leakage to the training data and results will look better than what they really are.

C. Model Training and Comparison

Logistic Regression is a base algorithm that needs to be built before all others. One may compare it with SVM, Decision Tree, Random Forest, Naïve Bayes, KNN, and even with a neural network if the amount of data is enough. The method of stratified cross-validation looks to be the right one to use when choosing a model. It should be clear that a test dataset should not be touched until the decision will be made, while the tuning of hyperparameters should be done only on the training and validation datasets.

The baseline is a good example of how more complicated algorithms give an additional value. Small value means that the model is better to be simple.

D. Evaluation

Nevertheless, mere accuracy is not a proper criterion of evaluation. Recall (sensitivity of the positive class) stands for the share of true positives, while specificity denotes the share of true negatives. Precision, F1-score, ROC-AUC, and confusion matrix provide completely different information, and it is also important to test the calibration if there are probabilistic estimations generated by the model. It is not equally dangerous for the patients and practitioners to have a miss and a false positive, and the threshold of decision making should be chosen together with the experts of clinical field, as well as to provide uncertainty in the results when it is possible. Table V provides the main metrics, where TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively.

It is necessary to point out that the confusion matrix represents the foundation for the mentioned metrics, because it demonstrates the quantity of correct and incorrect predictions for each class. The change in the threshold of decision-making changes the numbers in the matrix cells, and therefore there is no ideal threshold to choose.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

TABLE V. COMMON EVALUATION MEASURES

Measure	Meaning	Definition
Accuracy	Share of all predictions that are correct	(TP + TN) / (TP + TN + FP + FN)
Sensitivity (recall)	Share of actual positives that are detected	TP / (TP + FN)
Specificity	Share of actual negatives correctly identified	TN / (TN + FP)
Precision	Share of positive predictions that are correct	TP / (TP + FP)
F1-score	Harmonic mean of precision and recall	2PR / (P + R)
ROC-AUC	Ability to rank positive cases above negative ones	Area under the ROC curve
Calibration	Agreement between predicted and observed risk	Assessed with calibration plots

Explainability and Validation

The accuracy figure would be misleading when one category vastly outnumbers the other because the system will exhibit a high degree of accuracy despite not predicting the minority category most of the time. This explains why the sensitivity, specificity, and all other figures in Table V must be presented.

TABLE VI. SUMMARY OF THE PROPOSED WORKFLOW

Stage	Main activity	Risk to avoid
A. Dataset	Select data and define the target	Mismatch between data and task
B. Preparation	Clean, encode, and scale the data	Leakage from test data
C. Training	Compare baseline and complex models	Tuning on the test set
D. Evaluation	Report several measures	Relying on accuracy alone
E. Explanation and validation	SHAP, LIME, external and subgroup testing	Reading attribution as causation; skipping external validation

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

VII. LITERATURE INSIGHTS

Based on the literature, the following features appear clearly. First of all, ML methods can find patterns in EHRs and clinical variables; nevertheless, the most productive algorithm depends on cases and even definitions of the issue. According to the findings of a systematic review and meta-analysis in 2024, there was a good performance result of some ML algorithms with different results of the other studies at the same time. Thus, although the results of pooled studies are quite impressive, they do not necessarily guarantee a similar result in another sample [2]. Regarding supervised machine learning techniques, there is evidence of varying results based on datasets and studies [11].

First of all, benchmark datasets provide researchers with an opportunity to reproduce and compare their results; nevertheless, in some cases, they may be too small or biased to make conclusions about the applicability of these techniques. For example, Cleveland's subset of UCI repository dataset consists of only 303 samples and is extensively used in scientific and educational purposes [3]. Although it is very popular, its popularity does not imply representativeness for the whole population. Any ML model built on such dataset will detect only local dependencies and perform badly on others.

Thirdly, interpretability and validation of the models are key factors of ensuring the safety of the models. Tools like SHAP and LIME can be used to interpret the model [5], [6], but they cannot resolve problems with the poor quality of the dataset, incorrect labeling and bias and change in the distribution of data through time. According to Cai et al., it is necessary to pay attention to independent validation and reproducibility of the reports for AI cardiovascular prediction models [1]. In general, literature proposes improving reporting, external validation and involvement of the clinicians, but not comparing of the models.

Table VII illustrates the role of each source in the present paper.

TABLE VII. SOURCES AND THEIR ROLE IN THIS PAPER

Ref.	Type	Role in this paper
[1]	Systematic review	Independent validation and transparent reporting
[2]	Systematic review and meta-analysis	Pooled discrimination and heterogeneity
[3]	Dataset documentation	Benchmark data and its limits
[4]	Review article	Background on deep learning
[5]	Conference paper	SHAP explanations
[6]	Conference paper	LIME explanations
[7]	Journal article	Support vector machines
[8]	Journal article	Random forests
[9]	Journal article	Nearest-neighbor rule
[10]	Reporting guideline	TRIPOD checklist

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[11]	Systematic review	Variation across studies and datasets

The similarity that can be noticed in the selected sources is the opinion about the fact that good performance on a specific data set is not a good indicator at all. The type of data, the target used, and validation method must be considered before the papers are evaluated, since all these aspects form different results.

The quality of reporting is the next problem that needs to be highlighted. In case when there is no information concerning the source of the data, outcomes, treatment of the missing values and the validation method used, the research cannot be replicated and even analyzed in spite of great statistics mentioned in the paper.

VIII. CHALLENGES AND RESEARCH GAPS

There are a number of things that should be resolved prior to implementing AI risk prediction models into practice. The following list of problems and possible ways to address them is provided in Table VIII.

Availability and quality of data: clinical data can be incomplete, contain errors and inaccuracies, as well as data collected from various hospitals can differ in terms of equipment, units of measurements, coding standards, and patient selection procedures.

Imbalanced data: some datasets lack observations in relation to the target variable, which cannot be reflected in the accuracy measures and necessitates the implementation of such techniques as resampling and class weights; however, resampling should not happen within the train folds.

Generalizability: small-size data collected from one single site with internal validation does not necessarily reflect the future patient population; therefore, external validation is required.

Trust and interpretability: The clinicians need something they can trust. However, feature attribution techniques are not reliable and can be misinterpreted very easily. It is necessary to test interpretability techniques with the end-users and it is not acceptable to use feature attribution techniques as the replacement for the clinical decision-making process.

Fairness, privacy, and security: Models might have a reflection of any historical bias present in data. The personal data of patients needs to be managed in the proper way, be available to the right people, to be protected and stored properly according to the data protection legislation.

Clinical integration: a model which is very good in terms of metrics will not help healthcare to be improved if it interferes with the workflow, creates lots of false alarms, or does not fit the target population. Prospective evaluation and monitoring after deployment are crucial and the models get outdated quickly.

TABLE VIII. CHALLENGES AND POSSIBLE MITIGATIONS

Challenge	**Possible Mitigation**
Data quality	Systematic checking, documentation, and cleaning
Class imbalance	Class weighting or resampling inside training folds, with suitable metrics

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Limited generalizability	External validation across sites and periods
Low trust	Explanations evaluated together with clinicians
Bias and privacy	Subgroup analysis, governance, and secure data handling
Poor workflow fit	Prospective evaluation and monitoring after deployment

User is also a vital entity to think about in this case. A characteristic which is very helpful for a researcher might be not so convenient for a clinician. Thus, the interface design and user training should be taken into account from the very beginning as well.

Another important issue here is time. Patient population, clinical parameters, and guidelines all change in time; thus, a model which was valid during its development process may become invalid in the future.

IX. FUTURE SCOPE

For future research, one might think about performing multiple algorithms comparisons based on bigger and more diverse clinical data sets with clearly stated outcome criteria and clear data pre-processing procedure. The future research has to demonstrate discrimination, calibration, sensitivity, specificity, and subgroup analysis along with the simple accuracy instead.

Another potentially useful approach is combining clinical databases with the ECG signal or imaging data, but this is contingent on the availability of appropriate data and clinical expertise. Machine learning models explanation techniques need to be assessed in terms of stability, validity, and concordance with clinical knowledge. It might be worth exploring the idea of privacy-preserving federated learning, where each hospital will train its piece of a shared model using its own clinical database, although there will be some extra problems.

It would include an assessment of the risk, the description of the major factors involved, some form of accounting for uncertainty, and the claim that the output is decision-supportive. The evaluation will be done by clinicians and it will center around its contribution to making the right decisions safely. No predictive models should be employed for the purposes of diagnosis or treatment without valid clinical evidence.

Table IX includes the proposed recommendations and justifications for each of them.

TABLE IX. FUTURE DIRECTIONS

Direction	**Reason**
Larger, more diverse datasets	Reduce dependence on small benchmarks
Clear outcomes and follow-up	Separate present disease from future events
Planned external validation	Check performance beyond the development site
Fusion with ECG or imaging	Use information that tables alone do not hold
Federated learning	Collaboration without sharing raw records
Clinician-involved evaluation	Test real usefulness and safety

What can be noticed in these recommendations is that there is one common thing: The proof of the effectiveness of any method must come only from testing, and not from an extremely high score. A person

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

who is interested in starting his practice after conducting the research can adopt the procedure mentioned in section VI and wait for independent verification of the results obtained.

X. CONCLUSION

Machine learning approaches to early detection of heart diseases have been examined in this paper, and they are connected with the following aspects: most widespread machine learning algorithms, deep learning, explainability, data preprocessing, evaluation and validation. Every approach described above (Logistic Regression, SVM, Decision Tree, Random Forest, Naïve Bayes, KNN and Neural Networks) has both pros and cons, and since the efficiency of all those algorithms is strongly dependent on the dataset and particular task, none of them can be regarded as the best. Existing studies point out the potential of ML for cardiovascular risks prediction; however, such issues as small size of the datasets, discrepancies between the studies, lack of generalizability, poor interpretability and clinical validation were also pointed out. The workflow that was developed in this paper can serve as a foundation for further research and comparison. Since no models have been trained and no clinical experiments conducted, new findings regarding the effectiveness of the approach cannot be provided. Machine learning can be treated as an auxiliary technology for health care specialists.

REFERENCES

[1] "Machine Learning Techniques for Heart Disease Prediction: A Systematic Literature Review," Springer, 2025. [Online]. Available: https://doi.org/10.1007/s42979-025-03860-2

[2] "Heart Disease Prediction Using Machine Learning Algorithms: A Comparative Study," arXiv, 2025. [Online]. Available: https://arxiv.org/abs/2505.09969

[3] "Artificial Intelligence-Based Methods for Cardiovascular Disease Detection: A Review," 2025. [Online]. Available: https://scholar.google.com

[4] "Systematic Review of AI-Based Heart Disease Prediction Models: Challenges and Future Directions," 2025. [Online]. Available: https://scholar.google.com

[5] "Artificial Intelligence for Heart Disease Prediction: A Review of Deep Learning and Hybrid Approaches," *Artificial Intelligence Review*, Springer, 2024. [Online]. Available: https://doi.org/10.1007/s10462-024-10899-9

[6] "Heart Disease Prediction Using Machine Learning: A Comparative Analysis," arXiv, 2024. [Online]. Available: https://arxiv.org/abs/2410.14738

[7] "Heart Disease Prediction Based on Machine Learning Using Health Indicators," arXiv, 2023. [Online]. Available: https://arxiv.org/abs/2306.00023

[8] "Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review," arXiv, 2021. [Online]. Available: https://arxiv.org/abs/2112.06459