STM JOURNALS

Research & Reviews: A Journal of
Bioinformatics

https://journals.stmjournals.com/rrjobi

ISSN: 2393-8722
Volume 12, Issue 3, 2025
September—December
DOI (Journal): 10.37591/RRJoBI

Review RRJoBI

Early Detection of Heart Disease Using Machine Learning Techniques

Dhananjay Yeola[1,*], Prashant Yawalkar[2], Ranjana Dahake[2]

Abstract

Coronary illness stays one of the main sources of death around the world. Exact expectations of coronary illness can altogether work on quiet results by empowering early intercession and customized treatment plans. Throughout the course of many recent years, AI (ML) methods have been extensively investigated for anticipating coronary illness, attributing to their remarkable capacity to analyze complex data patterns and generate precise predictions based on historical clinical records. With the continuous growth of healthcare datasets and advancements in computational power, machine learning has emerged as a powerful tool for assisting medical professionals in identifying individuals at risk of heart disease at an early stage. This literature review aims to provide a detailed examination of the current body of research surrounding coronary illness prediction through ML techniques, emphasizing a range of approaches, algorithms, and methodologies. Furthermore, it highlights key models, comparative findings, and emerging trends, thereby demonstrating the potential of ML-driven decision-making in improving diagnostic accuracy and enhancing patient outcomes.

Keywords: Artificial intelligence (AI), machine learning (ML), coronary illness prediction, heart disease risk assessment, clinical data analysis

INTRODUCTION

Now a days health related issued are faced by many of us and heart disease is the most common reason of death. Machine learning (ML) and related techniques have a huge role. By examining vast datasets to find trends, correlations, and risk factors that conventional approaches might miss, machine learning is being used more to forecast heart diseases. Numerous machine learning models and techniques are used to improve clinical decision-making, increase early detection, and predict heart disease.

Medical data, such as patient demographics, medical history, lifestyle factors, lab results, electrocardiograms (ECG), imaging data (such as CT scans or echocardiograms), and other clinical test results, are all used by machine learning models.

***Author for Correspondence**
Dhananjay Yeola
E-mail: dhananjayy_ioe@bkc.met.edu

[1]Teaching Assistant, Department of Computer Engineering, Mumbai Educational Trust's Institute of Engineering, Bhujbal Knowledge City, Nashik, Maharashtra, India
[2,3]Student, Department of Computer Engineering, Mumbai Educational Trust's Institute of Engineering, Bhujbal Knowledge City, Nashik, Maharashtra, India

Received Date: March 26, 2025
Accepted Date: September 01, 2025
Published Date: November 11, 2025

Citation: Dhananjay Yeola, Prashant Yawalkar, Ranjana Dahake. Early Detection of Heart Disease Using Machine Learning Techniques. Research & Reviews: A Journal of Bioinformatics. 2025; 12(3): 34–45p.

Data preprocessing is essential before being used to develop models for machine learning, the data frequently needs to be cleaned and preprocessed (e.g., addressing missing values and standardizing numerical data, encoding categorical features).

OUTLINE OF THE USE OF MACHINE LEARNING

Various processing steps are involved in the prediction of heart disease using machine learning such as data preprocessing, modeling, evaluation, feature importance, etc.

- *Data Preprocessing:* The datasets should be cleaned and preprocessed by handling missing values, normalizing numerical features, and encoding categorical variables.

- *Modeling:* Train various machine learning models (such as logistic regression, random forests, SVM, and neural networks) using the dataset's features to classify or predict the course of heart disease [1].
- *Evaluation:* Use performance metrics, such as ROC-AUC, F1-score, recall, accuracy, and precision, to evaluate the model's effectiveness.
- *Feature Importance:* Determine the key characteristics that influence the prediction of heart disease, as this can offer important information about patient care and illness prevention.

Benefits of Machine Learning for Predicting Heart Disease

Machine learning offers several advantages in the prediction of heart and brain disease like early detection, improved accuracy, personalized risk assessment, and scalability [2].

- *Early Detection:* Even in patients who are asymptomatic, machine learning models can analyze big datasets and spot patterns early on, assisting in the prompt diagnosis of heart disease.
- *Improved Accuracy:* Machine learning can identify intricate correlations between variables that conventional approaches might overlook, thanks to sophisticated algorithms.
- *Personalized Risk Assessment:* Based on information about each patient, machine learning can generate customized risk scores, which facilitates the customization of interventions and therapies.
- *Scalability:* ML models can be used at scale in clinical settings because they can process and analyze vast amounts of data faster than manual method [3, 4].

Common Types of Heart Diseases

Heart diseases refer to a range of states that affect the heart. Among these illnesses are coronary artery disease (CAD), Heart Attack (Myocardial Infarction), Heart Failure, Arrhythmia, Valvular Heart Disease, Congenital Heart Disease, Cardiomyopathy, Cardiomyopathy, Pericarditis, Aortic Disease, Hypertension (High Blood Pressure), etc. (Table 1).

- *Coronary Artery Disease (CAD):* The most prevalent kind of heart disease is this one. It happens when plaque (atherosclerosis) builds up in the coronary arteries, which supply blood to the heart muscle, narrowing or blocking them [5, 6].
- *Heart Attack (Myocardial Infarction):* When oxygen-rich blood cannot reach a portion of the heart muscle, a heart attack occurs. The heart muscle may be harmed by this [7, 8].
- *Heart Failure:* A condition in which the heart's ability to pump blood to meet the body's needs is inadequate. Damage to heart muscles or other heart conditions may cause this [9, 10].
- *Arrhythmia:* Heartbeats that are erratic, too fast (tachycardia), or too slow (bradycardia) are referred to as arrhythmias. The heart's capacity to adequately pump blood may be impacted by this [11, 12].
- *Valvular Heart Disease:* This happens when one or more heart valves malfunction, causing the blood to flow backward or interfering with the blood's normal flow. Aortic stenosis, valve regurgitation, and mitral valve prolapse are common types [13, 14].
- *Congenital Heart Disease:* These are congenital cardiac abnormalities. They can have a variety of symptoms and impact on the structure or function of the heart [15].
- *Cardiomyopathy:* A disorder that affects the heart muscles and makes it more difficult for the heart to pump blood. This may result in heart failure. Dilated, hypertrophic, and restrictive cardiomyopathy are among the common types [16].
- *Pericarditis:* Inflammation of the heart's surrounding lining, or pericardium. It may arise due to an infection, trauma, or other circumstances.
- *Aortic Disease:* This includes diseases where the aorta, the big artery that transports blood from the heart to the other parts of the body, is weak or damaged, such as aortic aneurysms or aortic dissection [25].
- *Hypertension (High Blood Pressure):* Chronic hypertension can harm the heart and arteries, resulting in heart failure, coronary artery disease, and stroke, among other heart conditions [20–26].

Table 1. Study of different heart diseases.

Type of Heart Disease	Research Paper Title	Research Details
Coronary Artery Disease (CAD) [17]	"Coronary Artery Disease Classification with Different Lesion Degree Ranges Based on Deep Learning" (2024)	• This work presents a classification methodology for coronary artery disease using invasive coronary angiography images. • *Accuracy:* F-measure and AUC of 92.7% and 98.1%, respectively. • One of the limitations of this work is that the used dataset only contains information about 42 patients from one hospital.
Heart Failure (HF) [18]	"An Enhanced Random Forests Approach to Predict Heart Failure from Small Imbalanced Gene Expression Data" (2021)	• The study investigates a publicly available dataset of gene expressions of patients having STEMI. Random Forests variant method and our results about heart failure genes can have a strong clinical significance. • Limitation is that the small size of patients having heart failure (9 individuals) restricts the generalizability of this approach.
Myocardial Infarction (MI) [19]	"Lightweight Method of Myocardial Infarction Detection and Localization from Single Lead ECG Features Using Machine Learning Approach" (2024)	• This research focuses on binary classification or implements complex classifiers for localization to achieve good accuracy. • K-NN classifier has been used to achieve accuracy and F1-score of 99.74% and 99.20%, respectively. • Improvisation of the classifiers may result in improved performance.
Hypertension (HTN) [20]	"Machine Learning-Enabled Hypertension Screening Through Acoustical Speech Analysis: Model Development and Validation" (2024)	• This study proposes a novel framework for detecting hypertension through acoustic analysis of speech. By recording speech across multiple sessions and analyzing its temporal and spectral characteristics, author aims to identify indicators of hypertension. • For the first threshold, the balanced *accuracy* achieved was 84% for females and 77% for males. For the second threshold, the corresponding balanced accuracies were 63% for females and 86% for males. • This study had several limitations. First, the number of hypertensive cases in both sexes was limited, and the ethnicity of the participants was primarily Indian.
Valvular Heart Disease [21]	"Exploiting Data-Efficient Image Transformer-Based Transfer Learning for Valvular Heart Diseases Detection" (2024)	• This paper proposes a transfer learning methodology using the DeiT (Data-Efficient Image Transformer) model pre-trained on image datasets for VHD classification. • DeiT-based transfer learning approach achieved an overall *accuracy* of 97.44%. Moreover, our Conv–DeiT method outperformed the DeiT-based transfer learning with an impressive overall accuracy of 99.44%.
Arrhythmias [22]	"Morphological Arrhythmia Classification Based on Inter-Patient and Two Leads ECG Using Machine Learning" (2024)	• This paper explores inter-patient-based arrhythmias classification using combined two ECG leads automatically by employing machine learning methods, specifically ensemble learning. • Ensemble Learning achieves performance *accuracy* 87%, recall 87.4%, precision 88.4%, and F1-score 87%. • In this study, there are several challenges like data quality and variability, limited data availability, class imbalance, generalization to different populations, computational complexity, integration with clinical workflows.
Congenital Heart Disease [23]	"Precision Diagnosis: An Automated Method for Detecting Congenital Heart Diseases in Children from Phonocardiogram Signals Employing Deep Neural Network" (2023)	• This research study focuses on developing a strong binary classification system for congenital heart diseases (CHDs) utilizing deep neural networks. • *Accuracy:* 98.57%, due to constraints in the dataset and the limited timeframe of the study, this analysis specifically focused on a small range of CHDs, which resulted in intrinsic limitations. Initially, the study was limited to binary classification, namely distinguishing between normal and abnormal cases.
Cardiomyopathy [24]	"HCM-Echo-VAR-Ensemble: Deep Ensemble Fusion to Detect Hypertrophic Cardiomyopathy in Echocardiograms" (2024)	• This research proposes *HCM-Echo-VAR-Ensemble*, a novel framework that performs binary classification (HCM vs. no HCM) of echocardiogram videos directly using an ensemble of state-of-the-art deep VAR architectures models (SlowFast and I3D) and fuses their predictions using majority averaging ensembling. • *HCM-Echo-VAR-Ensemble* achieved *accuracy* of 95.28%. • This work has a few limitations. It explored only echocardiogram views such as A2C, A4C, PLAX, and PSAX for HCM diagnoses.

Aortic Disease [25]	"Classification of Aortic Stenosis Using Time–Frequency Features from Chest Cardio-Mechanical Signals" (2020)	• This paper introduces a novel method for the detection and classification of aortic stenosis (AS) using the time-frequency features of chest cardio-mechanical signals collected from wearable sensors, namely seismo-cardiogram (SCG) and gyro-cardiogram (GCG) signals. Such a method could potentially monitor high-risk patients out of the clinic. • The average *accuracies* achieved are 96.25% from decision tree, 97.43% from random forest, and 95.56% from neural network.

These conditions can often be prevented or managed with proper treatment, lifestyle changes, and medical care.

RELATED WORK

One of the main causes of death in the world is heart disease. Predicting its likelihood accurately can aid in preventing it. It has been demonstrated that ML algorithms can accurately forecast heart conditions based on a variety of medical data parameters. An overview of recent and earlier studies that have used machine learning algorithms to forecast heart diseases is provided in this section. ML algorithms, such as SVM, ANN, DT, LR, and RF [5], have been used in several studies to evaluate medical data and forecast heart conditions.

A recent investigation by [6] predicted the risk of heart disease in a multiethnic population using machine learning models. To stratify cardiovascular disease *(*CVD) risks, the authors used a sizable dataset from electronic health records and connected it with socio demographic data. The multiethnic population's CVD risk was accurately predicted by the models.

Similarly, a deep learning (DL) algorithm was used in another study by [7] to predict coronary artery disease (CAD). To train the DL model, the researchers used coronary computed tomography angiography (CCTA) images and clinical data. High accuracy in predicting the presence of CAD was attained by the model that was presented.

Several models for machine learning were used in a study by [8] to predict CVD based on clinical data. The researchers employed DTs, RFs, and K-nearest neighbor (KNN) models. The authors found that these models were highly accurate in predicting CVD.

Similarly, a study by [9] employed machine learning methods to identify the risk factors for heart disease. To identify risk factors for coronary heart disease, the authors accustomed data from the National Health and Nutrition Examination Survey (NHANES). According to the authors, risk factors could be effectively identified by the suggested machine learning algorithm.

The ability of various machine learning algorithms to predict heart diseases was examined in another study by [10]. The authors employed several models, such as ANN, DT, and LR. The models' high precision in predicting heart diseases was reported by the authors.

ML algorithms have demonstrated high accuracy in numerous studies and are now commonly used to predict heart diseases. ML algorithms have been used to predict various heart diseases, including CAD and CVD, by considering medical data parameters such as clinical data, sociodemographic information, and medical images. The reviewed studies have demonstrated the effectiveness of models, such as DTs, DL, ANN, RF, and KNN, in predicting heart diseases. It is anticipated that more suitable models and features will be created for precise heart disease prediction as machine learning algorithms continue to advance.

According to earlier research on HD prediction, machine learning techniques may be able to identify characteristics associated with the illness and create reliable prediction models. To fill in these gaps in the corpus of existing knowledge, more effort is necessary. Here are a few gaps and how the suggested method closes them.

One ML algorithm, such as DT, LR, RF, or SVM, has been used in HD prediction research. Although each of these algorithms has demonstrated potential, no thorough evaluation or comparison of ML techniques exists. This limits generalizability and complicates the search for the optimal HD predictor. This gap is filled by the proposed study. Ten machine learning classifiers – Naive Bayes, SVM, voting, XGBoost, AdaBoost, bagging, DT, KNN, RF, and LR – are compared and assessed. The article assesses the optimal algorithm for HD prediction using performance metrics such as accuracy, sensitivity, precision, specificity, F1-score, and AUC.

Unbalanced classes in HD prediction datasets make it difficult to make correct predictions for the minority class (HD-positive patients). Although some studies have attempted to address this issue by using oversampling or under sampling, a thorough analysis of the techniques and their effects on prediction accuracy is required. The proposed article closes this gap by addressing the issue of unequal classes as well. SMOTE is used to guarantee that the dataset is balanced. This work investigates how well SMOTE improves HD prediction accuracy and how it affects the performance of various ML algorithms [11].

AN OVERVIEW OF PREDICTING HEART DISEASE

Machine Learning Techniques

Several machine learning models are used to predict heart disease are Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Neural Networks, Gradient Boosting Machines (GBM) [12], etc.

- *Logistic Regression:* This statistical model is commonly used to problems containing binary classification such as forecasting a patient's likelihood of developing heart disease. It uses input features, like blood pressure, age, and cholesterol levels, to estimate the likelihood of an event (like the existence or absence of heart disease).
- *Decision Tree:* These models are helpful for identifying high-risk patients because they simplify a complex decision-making process. To estimate the chance of developing heart disease, the model learns to divide data into branches according to feature thresholds (such as age or cholesterol).
- *Random Forest:* A model composed of several decision trees. By decreasing over fitting, it increases prediction robustness and accuracy. Additionally, random forests can be used to determine which characteristics – such as diabetes or smoking – are most crucial for predicting heart disease.
- *Support Vector Machines (SVM):* Tasks involving classification are handled by SVMs. By optimizing the margin between classes, they can establish a decision boundary (e.g., heart disease vs. no heart disease). Both linear and non-linear data can be handled by SVMs.
- *K-Nearest Neighbors (KNN):* KNN is a straightforward algorithm that uses the "k" most similar patients in the dataset to classify a patient. It performs best when there are distinct patterns in the data and is especially helpful for grouping patients according to characteristics like age, cholesterol, or ECG readings [3].
- *Neural Networks:* For intricate pattern recognition tasks, deep learning models in particular, artificial neural networks, or ANNs are employed. They can extract features from large datasets (such as those from medical imaging and ECG) that may not be visible using more conventional techniques.
- *Gradient Boosting Machines (GBM):* LightGBM and XGBoost are examples of models that combine numerous weak learners (simple models) to create strong predictors. Because of their resilience and capacity to manage incomplete or unbalanced data, they have demonstrated a high degree of efficacy in the prediction of heart disease and other cardiovascular disorders.

Heart Disease Prediction Workflow Using Machine Learning

There are various steps for heart disease prediction using machine learning which include data collection, feature engineering, model training, model validation, prediction, etc.

- *Data Collection:* Collect imaging data, ECG readings, clinical data (such as lab results and demographics), and other pertinent medical records.
- *Feature Engineering:* Determine the key risk factors for heart disease, including age, BMI, smoking, cholesterol, and family history.
- *Model Training:* Utilizing labeled data (such as patients with and without heart disease), set a machine learning model (such as random forest or SVM) on the dataset.
- *Model Validation:* Utilize methods, like cross-validation, to evaluate the model's performance and modify hyper parameters to increase prediction accuracy.
- *Prediction:* Based on their data, utilize the learned model to determine if a new patient is at risk for heart disease (Figure 1).

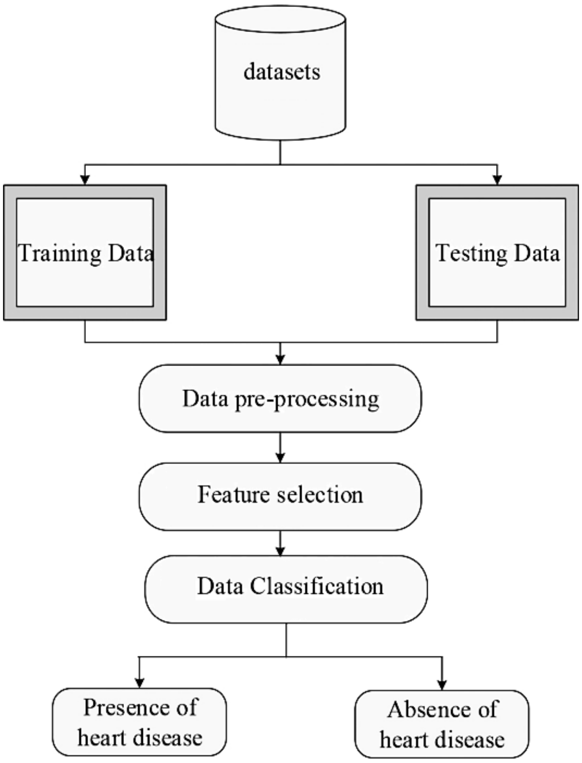

Figure 1. Workflow of heart disease prediction.

Heart Disease Prediction Using Various Data Types

Heart Disease Prediction uses various types of data like clinical data, ECG data, medical imaging, genomic data.

- *Clinical Data:* To determine the possibility of heart disease, machine learning methods can utilize structured data such as age, gender, blood pressure, cholesterol, and smoking habits.
- *ECG Data:* Raw ECG signals can be analyzed by deep learning algorithms, especially convolutional neural networks (CNNs), which can detect irregular heart rhythms and forecast possible cardiac issues.
- *Medical Imaging:* Medical images, like CT scans, MRIs, and echocardiograms, can be analyzed by machine learning (ML) methods, specifically convolutional neural networks (CNNs), to find heart disease symptoms like abnormal heart valves or coronary artery blockages.
- *Genomic Data:* Heart disease risk is influenced by genetic factors, and ML models can be trained to examine genomic data and forecast a person's vulnerability to cardiovascular diseases.

Cleveland Heart Disease Dataset (UCI Repository)

This is one of the most popular datasets used for heart disease classification. It contains 303 instances with 14 attributes, including patient demographics, cholesterol levels, resting blood pressure, age, and the existence of heart disease. Features of the dataset include age, sex, resting blood pressure, chest pain type, cholesterol, electrocardiographic results, fasting blood sugar, maximum heart rate, exercise-induced angina, ST depression, and more.

- *Access:* Available at the UCI Machine Learning Repository.

Framingham Heart Study Dataset

This dataset comes from the long-running Framingham Heart Study, which aims to identify the common factors contributing to cardiovascular disease. It contains data on over 4,000 patients and includes a wide range of demographic, medical, and lifestyle variables [12].

Features are age, gender, hypertension, diabetes, cholesterol, smoking habits, physical activity, family history, and more.

- *Access:* This dataset is available from multiple sources, including PhysioNet and Kaggle.

Heart Disease UCI (Cleveland) Dataset on Kaggle

A variation of the Cleveland Heart Disease dataset available on Kaggle, often used for the task of binary classification to predict whether heart disease is present or not.
Features are like the Cleveland dataset, it includes features like age, sex, cholesterol, ECG results, and maximum heart rate.

- *Access:* Available on Kaggle.

Heart Failure Clinical Records Dataset

This dataset contains clinical data related to heart failure patients, including features like age, gender, smoking habits, diabetes, and levels of various biomarkers.
Features are age, ejection fraction (a measure of heart function), creatinine kinase levels, serum sodium, smoking, and diabetes status.

- *Access:* Available on Kaggle.

Cardiovascular Disease (CVD) Dataset

This dataset is used for predicting cardiovascular disease risk based on patient characteristics such as age, gender, and medical history.

Features are sex, age, cholesterol levels, blood pressure, ECG results, lifestyle factors and family history.

- *Access:* Available on Kaggle.

MIMIC-III (Medical Information Mart for Intensive Care) Database

This dataset is a large and diverse dataset from patients admitted to intensive care units (ICUs), which includes heart disease-related information, among other conditions. It contains detailed clinical records, including lab results, vitals, diagnoses, medications, and more.

Features are patient demographics, lab results, clinical notes, heart disease diagnoses, and co-morbidities.

- *Access:* Available from PhysioNet (requires registration for access).

ST-T Wave Dataset

This dataset contains ECG data and is used to predict the occurrence of myocardial infarctions or heart attacks based on ST-segment and T-wave abnormalities.

Features are ECG signal features such as heart rate, ST-segment changes, and T-wave abnormalities.

- *Access:* Available from UCI Machine Learning Repository.

Cardiotocography (CTG) Dataset

Though primarily used for studying fetal heart rate patterns, this dataset includes cardiotocography features that can also help in predicting heart conditions.

Features are fetal heart rate, uterine contraction measurements, and other indicators related to fetal well-being.

- *Access:* Available on UCI Machine Learning Repository.

Cincinnati Heart Disease Dataset

This dataset, from the University of Cincinnati, contains clinical data of patients who were identified with heart disease, including laboratory results and health history.

Features are age, cholesterol, ECG readings, smoking status, and diabetes.

- *Access:* Available from Kaggle.

Pima Indians Diabetes Dataset (Related to Heart Disease Risk)

While primarily focused on diabetes, this dataset includes determinants, such as blood pressure, age, and BMI, which are also relevant to heart disease prediction, given the strong link between diabetes and cardiovascular health.

Features are age, glucose levels, BMI, blood pressure, insulin levels, and diabetes status.

- *Access:* Available on Kaggle.

The above datasets provide a strong foundation for anyone looking to explore heart disease prediction using machine learning, helping healthcare professionals and researchers develop tools for early diagnosis and personalized treatment.

RESULT AND DISCUSSION

Challenges and Limitations

Predicting heart diseases comes with several challenges that impact the accuracy, reliability, and overall success of the prediction models. These challenges can stem from data-related issues, model limitations, and real-world complexities. The availability and quality of the data are among the primary obstacles. Feature Selection and Engineering, Non-linear Relationships, Data Privacy and Ethics, Model Overfitting and Underfitting, Interpretability and Explainability, Temporal and Long-term Prediction, External Factors, Generalization Across Populations, Model Evaluation, Real-time Prediction and Integration [4].

Data Quality and Availability

- *Missing Data:* Missing values for specific features in incomplete datasets can cause bias or inaccurate predictions and weaken prediction models.
- *Imbalanced Datasets:* Because there are frequently far more healthy people than heart disease patients in these datasets, biased models may not be able to accurately predict the presence of heart disease in underrepresented classes.
- *Noisy Data:* The quality of the model's predictions may be impacted by noisy medical data, which can be caused by measurement errors, inaccurate data entry, or inconsistent reporting.

Feature Selection and Engineering

- *Irrelevant Features:* It can be difficult to determine which characteristics are most pertinent to heart disease prediction. The accuracy and interpretability of the model may be diminished by adding superfluous or irrelevant features.
- *Complex Interactions:* Numerous factors that affect the possibility of heart disease, such as genetics, lifestyle, and environmental factors, interact in intricate ways. Effectively capturing these interactions can be challenging.

Non-linear Relationships

- *Complexity of Heart Disease:* Numerous factors, such as genetic predisposition, lifestyle decisions (diet, exercise, and smoking), and environmental factors, all have an impact on heart disease. Traditional linear models may have trouble capturing the non-linear effects of these factors.

Data Privacy and Ethics

- *Sensitive Nature of Medical Data:* Predicting heart disease requires using sensitive patient data, which raises questions about data security and privacy as well as the morality of doing so.
- *Bias in Data:* Models that perform poorly for groups may result from bias in the data caused by the underrepresentation of populations (e.g., women, older people, or minority groups).

Model Overfitting and Underfitting

- *Overfitting:* Models may become overly complicated and fit the training data too closely when dealing with small or noisy datasets, making it difficult for them to generalize to new data.
- *Underfitting:* However, an overly simplistic model might perform poorly because it misses significant patterns in the data.

Interpretability and Explainability

- *Black-box Nature of Models:* Doctors may find it challenging to comprehend how predictions are made by complex machine learning models (like deep learning), which could impede the model's clinical application.
- *Trust and Adoption:* Even if a predictive model has good statistical performance, medical professionals may be hesitant to use it if they do not fully comprehend the logic behind its predictions.

Temporal and Long-Term Prediction

- *Data from Different Time Points:* Predicting heart disease frequently involves taking long-term factors into account (e.g., changes in lifestyle, progression of symptoms). Such longitudinal data is difficult to gather and analyze.
- *Changes in Risk Over Time:* The risk of heart disease varies over time because of interventions, aging, and lifestyle choices. Effectively capturing these temporal shifts in risk is challenging.

External Factors

- *Socioeconomic Factors:* Although they may be difficult to measure or may not be included in datasets, socioeconomic status, healthcare access, and environmental factors, all have a great impact on heart disease.
- *Lifestyle Changes:* The risk of heart disease can be greatly impacted by variables such as smoking, alcohol use, diet, and exercise. Accurately incorporating this subjective data can be difficult.

Generalization Across Populations

- *Generalizability of Models:* Due to variations in demographics, lifestyle, and access to healthcare, a model developed on one population or dataset might not generalize well to other populations. It can be difficult to guarantee that a model works as well in different populations.

Model Evaluation

- *Evaluating Performance:* It can be challenging to evaluate how well heart disease prediction models perform, particularly when dealing with unbalanced data or when there is no clear diagnostic benchmark in the dataset.
- *Metrics and Thresholding:* The model's efficacy can be greatly impacted by selecting the right evaluation metrics (such as precision, accuracy, F1-score, recall, ROC AUC, and prediction threshold).

Real-Time Prediction and Integration

- *Real-time Monitoring:* Predicting heart disease in clinical settings requires seamless data collection and analysis to be integrated with real-time health monitoring systems.
- *Clinical Decision Support:* Since prediction models need to support current healthcare workflows, it can be difficult to incorporate them into clinical decision support systems that help medical professionals without overburdening them.

Strategies to Address Challenges

There are many strategies to address challenges, some of them are discussed over here. They are data preprocessing, feature selection, explainability models, ensemble methods, transfer learning.

- *Data Preprocessing:* Addressing class imbalance by using sophisticated methods, like SMOTE, imputation, or resampling, to handle missing data.
- *Feature Selection:* Using methods, like PCA or feature importance analysis, to reduce dimensionality and choose pertinent features by applying domain knowledge.
- *Explainability Models:* Employing models with greater transparency (such as decision trees, LIME, and SHAP) or, in the case of black-box models, post-hoc explanation methods.
- *Ensemble Methods:* Combining several models (such as XGBoost and Random Forest) to increase robustness and accuracy.
- *Transfer Learning:* Enhancing prediction in various populations by applying pre-trained models to comparable medical data.

Despite these challenges, advances in machine learning, medical technology, and data collection methods continue to improve the potential of heart disease prediction methods.

FUTURE DIRECTIONS

As the recent study shows various new techniques are combined and can be worked together for accurate results, AI and real time systems are also important in healthcare applications. Various future directions include,

- *Integration of Multimodal Data:* To improve prediction accuracy, future studies will probably concentrate on merging data from multiple sources, including imaging, clinical, and genetic data.
- *Explainable AI (XAI):* Developing models that are not only accurate but also interpretable and explicable is becoming increasingly important as the healthcare sector depends more on AI for crucial decision-making.
- *Real-Time Prediction:* Real-time heart disease prediction using machine learning models may become a significant trend with the introduction of wearable technology that continuously monitors heart health indicators.
- *Personalized Healthcare:* By considering each person’s distinct risk factors and genetic predispositions, machine learning may be able to develop personalized health risk models that go beyond generalized prediction models.

CONCLUSIONS

Through the analysis of intricate datasets and the discovery of insights that may result in earlier detection and more individualized treatment regimens, machine learning holds the potential to revolutionize the prediction and management of heart disease. To guarantee its efficacy and safety, however, cautious data handling, model selection, and clinical validation are needed.

The use of machine learning techniques for heart disease prediction has advanced significantly in recent years, with multiple studies showcasing the models’ capacity to precisely identify patients at risk. While more recent techniques, like deep learning and hybrid models, show promise for further increasing prediction accuracy, supervised learning techniques, like decision trees, logistic regression, and support vector machines, continue to be the most widely used. With developments in multimodal data integration and explainable AI, the future of machine learning in heart disease prediction appears bright, despite obstacles like data imbalance, model interpretability, and generalization.

REFERENCES

1. Dalvi JJ, Khole SM, Kudale B. A survey on heart disease prediction using machine learning techniques. Int Res J Modern Eng Technol Sci. 2023.
2. Ramalingam VV, Dandapath A, Karthik Raja M. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol. 2018.
3. Chaithra CS, Siddesha S, Aradhya VNM, Niranjan SK. A review of machine learning techniques used in the prediction of heart disease. Rev Intell Artif. 2024.
4. Nashif S, Raihan MR, Islam MR, Imam MH. Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World J Eng Technol. 2018.
5. Williams R, Shongwe T, Hasan AN, Rameshar V. Heart disease prediction using machine learning techniques. In: International Conference on Data Analytics for Business and Industry. 2021.
6. Brendel JM, Walterspiel J, Hagen F, Kübler J, Brendlin AS, Afat S, et al. Coronary artery disease detection using deep learning and ultrahigh-resolution photon-counting coronary CT angiography. Diagn Interv Imaging. 2024.
7. Shadmi R, Mazo V, Bregman-Amitai O, Elnekave E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. ResearchGate. 2018.
8. Uddin KMM, Ripa R, Yeasmin N, Biswas N, Dey SK. Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset. Intell Based Med. 2023.
9. Oh T, Kim D, Lee S, Won C, Kim S, Yang J, et al. Machine learning-based diagnosis and risk factor analysis of cardiocerebrovascular disease based on KNHANES. Sci Rep. 2022.
10. Limbitote M, Mahajan D, Damkondwar K, Patil P. A survey on prediction techniques of heart disease using machine learning. Int J Eng Res Technol. 2020.
11. El-Sofany HF. Predicting heart diseases using machine learning and different data classification techniques. IEEE Access. 2024.
12. Ali MM, Al-Doori VS, Mirzah N, Hemu AA, Mahmud I, Azam S, et al. A machine learning approach for risk factors analysis and survival prediction of heart failure patients. Healthc Anal. 2023.
13. Huang AA, Huang SY. Use of machine learning to identify risk factors for coronary artery disease. PLoS One. 2023.
14. Al-Janabi MI, Qutqut MH, Hijjawi M. Machine learning classification techniques for heart disease prediction: A review. Int J Eng Technol. 2018.
15. Niranjana G, Elizabeth Shanthi I. A survey on heart disease prediction techniques. Int J Trend Sci Res Dev. 2021.
16. Rahim A, Rasheed Y, Azam F, Anwar MW, Rahim MA, Muzaffar AW. An integrated machine learning framework for effective prediction of cardiovascular diseases. IEEE Access. 2021.
17. Jiménez-Partinen A, Thurnhofer-Hemsi K, Rodríguez Capitán J, Molina-Ramos AI, Palomo EJ. Coronary artery disease classification with different lesion degree ranges based on deep learning. IEEE Access. 2024.
18. Chicco D, Oneto L. An enhanced random forests approach to predict heart failure from small imbalanced gene expression data. IEEE/ACM Trans Comput Biol Bioinform. 2021;18(6).
19. Anwar SMS, Pal D, Mukhopadhyay S, Gupta R. A lightweight method of myocardial infarction detection and localization from single lead ECG features using machine learning approach. IEEE Sensors Council. 2024.
20. Taghibeyglou B, Kaufman JM, Fossat Y. Machine learning-enabled hypertension screening through acoustical speech analysis: Model development and validation. IEEE Access. 2024.
21. Jumphoo T, Phapatanaburi K, Pathonsuwan W, Anchuen P, Uthansakul M, Uthansakul P. Exploiting data-efficient image transformer-based transfer learning for valvular heart diseases detection. IEEE Access. 2024.
22. Zakaria H, Nurdiniyah ESH, Kurniawati AM, Naufal D, Sutisna N. Morphological arrhythmia classification based on inter-patient and two leads ECG using machine learning. IEEE Access. 2024.
23. Alkahtani HK, Ul Haq I, Ghadi YY, Innab N, Alajmi M, Nurbapa M. Precision diagnosis: An automated method for detecting congenital heart diseases in children from phonocardiogram signals employing deep neural network. IEEE Access. 2024.

24. Almadani A, Sarwar A, Agu E, Ahluwalia M, Kpodonu J. HCM-Echo-VAR-ensemble: Deep ensemble fusion to detect hypertrophic cardiomyopathy in echocardiograms. IEEE Open J Eng Med Biol. 2025;6.
25. Yang C, Aranoff ND, Green P, Tavassolian N. Classification of aortic stenosis using time–frequency features from chest cardio-mechanical signals. IEEE Trans Biomed Eng. 2020;67(6).
26. Sadad T, Rehman A, Munir A, Saba T, Tariq U, Ayesha N, Abbasi R. Brain tumor detection and multi-classification using advanced deep learning techniques. Microscopy research and technique. 2021 Jun;84(6):1296-308.

