SN Computer Science (2025) 6:334
https://doi.org/10.1007/s42979-025-03860-2

ORIGINAL RESEARCH

A Survey on Machine Learning Techniques for Heart Disease Prediction

Priti Shinde[1] · **Mahesh Sanghavi**[2] · **Tien Anh Tran**[3]

Received: 9 March 2024 / Accepted: 6 March 2025

Abstract
Machine learning is the field of data science that makes decisions according to the data. To process huge amounts of data, discover patterns, and find co-relations among data, this field has achieved remarkable success. It is very useful in the field of medicine, especially in cardiology. This application in the cardiology field has provided a great help to medical practitioners. Heart disease is now becoming a dangerous disease throughout the globe. Hence the main objective of this paper is to provide a review of the most frequently used ML techniques in cardiology, which ML techniques have gained maximum success in cardiology, and the overall performance of these techniques. We performed a systematic review from Jan 2018 to June 2023. We have selected 68 studies focusing on the prediction and classification of heart diseases, and heart failure too. Due to the limitations of image and signal data such as storing capacity, and noise, we have considered only text datasets for study. We have summarized why especially ML techniques are needed and their performance. The results obtained from this study show that RF, SVM, KNN, DT, LR, and NB have shown great performance in the prediction and classification of heart disease. Moreover, RF and LR have achieved great accuracy and proved more efficient than others.

Keywords Machine learning · Heart disease · Cardiology · Prediction · Classification

Introduction

The life of human beings has been changed in this 20th era. There are various reasons behind this change. It may be due to living style, use of technology in day-to-day life, etc. Hence everyone is busy in their daily schedule. But they don't even know that due of this lifestyle many problems will occur. Rather they have occurred. The non-communicable diseases come under this category. These diseases are diabetes, heart disease, cancer, TB, etc. These diseases increasingly cause the death of a person each year. Among these noncommunicable diseases, heart disease is the leading cause of death nowadays across the world [1].

Cardiology is the branch of medicine that deals with the study of the heart, its blood vessels, and major arteries of the heart that supply blood to the heart and other body parts. Nowadays, heart disease is becoming the leading cause of death worldwide. Early prediction of these chronic disorders is very important [2]. According to WHO, 32% of global deaths is due to cardiovascular disease across the globe. Among these 32, 85% of deaths are due to heart disease and stroke [2]. Hence heart disease is becoming the largest cause of death worldwide. There are various risk factors are present for the occurrence of this disease. They are stress, alcohol consumption, obesity, diet, zero exercise routine, high blood pressure, increased levels of cholesterol, etc. These factors have been noticed prominently in high-income countries. But particularly low/middle-income countries should be focused on these factors where the chances of the prevalence of disease have increased rapidly due to lifestyle change. That's why it has gained the attention of researchers to do a study on this heart disease.

To diagnose this disease medical practitioners are using their conventional methods which include angiography and certain investigations. These tests are costly and are done when there are high indicators of disease. One can use the non-conventional method to overcome the problems caused

✉ Tien Anh Tran
anhtt.mtb@vimaru.edu.vn

Priti Shinde
priti25.shinde@gmail.com

Mahesh Sanghavi
sanghavi.mahesh@gmail.com

[1] Department of Computer Engineering, MET's IOE, Bhujbal Knowledge City, Adgaon, Nashik, India

[2] SNJBs Late Sau KBJ CoE, Chandwad, India

[3] Vietnam Maritime University, Haiphong 180000, Vietnam

Published online: 01 April 2025

by conventional ones. Daily, there is a huge amount of data is generated through hospitals and medical research centers. Hence machine learning provides a base to extract, and analyze this data. Using machine learning one can extract and discover patterns, and correlations among huge amounts of data. Hence this SLR has been developed to find the machine learning techniques used in cardiology. In this study, we mainly focused on types of heart disease, their risk factors & occurrence. Also, which ML techniques are used prominently in this disease detection have been thoroughly given by surveying the selected studies. We have conducted a survey in which initially we considered 676 studies. Later on, by applying exclusion criteria we considered a total of 68 studies. To the best of our knowledge, there is no such SLR conducted that focuses on ML techniques, types of heart diseases, and their causes till now. Hence, we are motivated to do this survey of studying machine learning techniques in cardiology.

In "Research Methodology" the details of the research methodology are elaborated. The PRISMA technique has been used for this research survey. In "Results and Discussion", the results are discussed. There are implications and limitation of the results are present which are discussed in "Summary". Finally, the conclusion has given.

Research Methodology

The method used by academics that focuses on finding and evaluating all appropriate literature on a topic and making conclusions about research questions is called Systematic literature review. In such reviews, the objectives are given with predefined criteria. It contains inclusion /exclusion criteria, research questions, quality assessments, and implications of results with research limitations as well. A systematic review tries to organize altogether experiential evidence. This evidence matches with predefined eligibility criteria to satisfy the research question [3]. There are ways to conduct a systematic literature review process, one of them is Preferred Reporting Items for Systematic Reviews and Mata Analyses (PRISMA) [4]. For attempting evidence-based search PRISMA is used. In our review process, we have attempted the PRISMA approach for review. The earlier was PRISMA 2009 guidelines but it has been updated in PRISMA 2020. PRISMA has been accepted because of its 27 evidence-based checklists and phase-wise analysis. Hence it provides clarity and transparency for conducting a systematic literature review [1].

According to [1], a systematic literature review is defined as a review in which questions are framed and organized, clear methods are used to discover, relevant answers to conduct appropriate research. This research is done to test data from different studies on review evidence-based search, and meta-analysis.

Research Questions

The main aim of our study is to find and study different machine algorithms for heart disease. This study has been conducted from 2015 to 2023. The whole study has been divided into 5 research questions (RQs) for getting a review idea. Below are the research questions:

RQ1: What are the different types of heart disease, their causes, symptoms, and risk factors?

RQ2: Identify high-risk heart disease and identify the factors contributing to the development of this disease.

RQ3: Which ML techniques are used in cardiology?

RQ4: What are the advantages and disadvantages of ML techniques when used in cardiology?

RQ5: What is the overall performance of ML techniques in cardiology?

RQ6: Are there any ML techniques that perform better than other ML techniques?

Search Strategy

As the research questions are identified and determined, the next step becomes how to find the data for these questions. This step is a search strategy. The base databases have been identified to gather data in terms of search strings. Digital databases that are used to search strings are IEEE Explore, Scopus, Elsevier, Springer, PubMed, ACM digital library, Google Scholar and Scientific Reports. These databases contain a huge amount of research articles from computer science and medical backgrounds. Due to this, we get material of area of interest. These databases index papers from journals and conferences. Hence, they find a good resource for study by previous researchers also.

The complete search string is given below which contains a combination of Boolean operators such as AND, OR:

"CAD **OR** coronary heart **OR** coronary artery **OR** heart **AND** (disease **OR** failure **OR** diagnosis)

AND ("machine learning" **OR** "data mining" **OR** "disease prediction" **OR** classification **and** "deep learning)".

By applying such a search string in advanced search, we got many papers/articles as a result.

Study Selection

The candidate papers are selected by applying search strings in the digital databases. It helps to filter the papers to find the most relevant articles, and papers based on title, abstract, and keyword search.[5] Such a type of search is called advanced search, a facility provided in these giant databases. By observing the results given by this advanced

search, most papers are repeated or found the same. Hence the recent ones are chosen for further study. For repeated studies, only one study is chosen. The complete flowchart for study selection is given in Fig 1.

There are criteria applied for the selection of studies. To select which study should be included, inclusion and exclusion criteria are applied. These criteria are relevant to RQ Table 1. To select appropriate articles/papers/studies the inclusion/exclusion criteria are developed [5].

The following are the inclusion criteria:

IC1: Studies consisting of newly developed ML-based models in disease prediction

IC2: Studies validating existing ML-based models

IC3: Studies presenting an overview of existing approaches that apply ML techniques for disease prediction

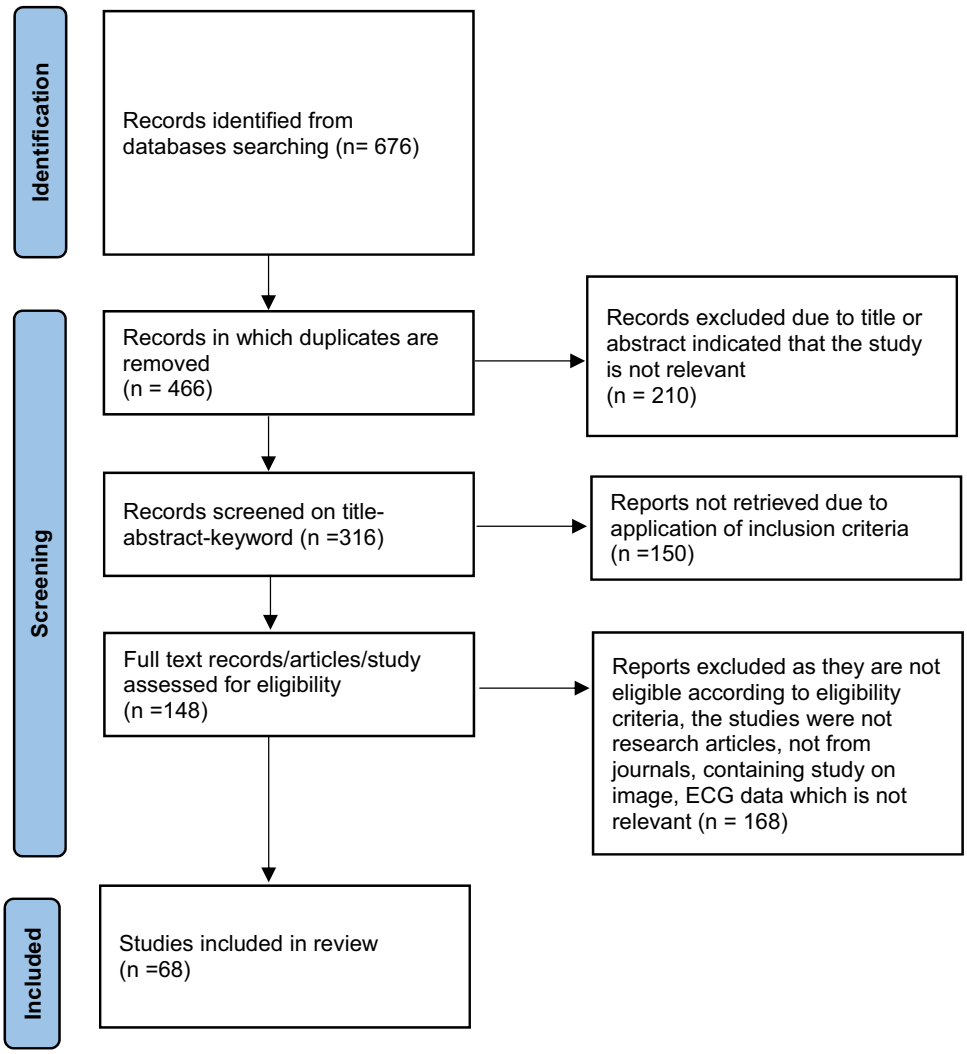


Fig. 1 PRISMA flowchart

Table 1 Strengths of classification ML techniques

ML technique	Strengths	Papers
KNN	Simple algorithm and easy-to-understand	[22, 19]
DT	Has the ability to handle both numerical and categorical data. Generates very simple and understandable rules	[23, 24]
RF	Robust in estimating missing data, gives evaluations about which features are more important	[21, 15]
SVM	Can be used for both classification and regression. Mostly used for binary classification	[17, 18]
Logistic regression	Widely used for predicting categorical dependent variables using independent variable	[19]
XGBoost and ADABoost	For decreasing biasing error, these methods are used efficiently	[25, 26]

IC4: Studies providing only comparisons between ML techniques in disease prediction

The studies are discarded if any of the following exclusion criteria are satisfied.

EC1: Studies on clinical data/categorical data are included only. Other such as images, signal are rejected.

Heart disease detection is crucial to avoid further problems or risks. Angiography is one of the techniques to diagnose heart disease. But it is invasive making it harmful for patients by dissecting arteries and is costly too. Image-based detection techniques are expensive and also not suitable for huge data processing in developing countries. Medical image contains vast amounts of data due to which computational challenges has increased. Also, they require more storage space and processing power. Often images require preprocessing power extensively in terms of normalization, noise reduction etc. This becomes time-consuming for preprocessing. Nowadays deep learning techniques are used for preprocessing of image data. But it acts like a black box in processing which becomes difficult for data interpretation.

ECG signals are noise-prone. If a patient is in movement or having poor electrode contact which complicates the analysis of the signal and ultimately degrades model performance.

Becoming a non-communicable disease, the medical history of patient is crucial towards diagnosis of disease. This contains past diagnosis treatments, family history, lifestyle, lab reports, image reports, ECG interpretation, smoking, alcohol consumption habits. We can collect and create this type of data is present only in the form of text. Textual data is easy for processing and also for store. It is stored in the form of an electronic health record (EHR). By considering all these, we have selected studies which contains text datasets only.

EC2: Studies before 2018 and after 2023 are not included in this review.

Data Extraction Strategy

The title and abstract of the studies are checked first. The same entries of articles/studies were removed by discussion with another author. Then the information that is relevant to our study was extracted such as publication source as is it from the journal, title, methodology/algorithms used, key findings, experimental det up/datasets used in each study, and performance measures. The extracted data is maintained in the form such as Excel and doc files. These files contain all the data which is relevant to inclusion criteria as well as the information which is relevant to study.

Threats to Validity

Following are the main threats to valid the study.

1. Bias in study selection: The study selection mainly depends upon source databases, search string, the criteria for selection. The search string hence is selected in such a way that it will extract the required study from the digital databases. Sometimes it may happen that in such automatic search, some studies may not be extracted by digital database. One has to select it manually. This causes study selection bias.
2. Bias in publication: It refers to the studies that give positive results of the studies/algorithms that are published. That means the studies which contain negative results or findings are not published sometimes. Hence the fourth inclusion criterion given in the study focuses on a comparative analysis of the published studies which tries to reduce this publication bias.

Bias in data extraction: Data extraction is the major critical task in this study. To overcome this task, the papers are read and analyzed by the authors independently. Later on, a discussion was held among the researchers involved in this study for data inclusion or exclusion.

Results and Discussion

In this section, the review of the selected study is presented. The main objective of this study was to find and investigate studies based on ML techniques in cardiology. For the same total 676 studies were selected, out of which 68 studies considered in to review. Initially, the overview of the study is given. Later on, the answers to each research question are given.

Overview of the Studies for the Selection Process

The review started with selecting study/papers/articles in eight digital electronic databases as shown in Fig 1. The total number of studies identified in depth for review was 676. The studies that is not relevant such studies removed from 676 and the total became 466. From these, the articles which do not support inclusion criteria were discarded and the total number of articles for further studies became 316. These 316 records were screened according to title-abstract-keyword. The further reports were excluded as they were not eligible according to eligibility criteria, which were not research articles, from conference papers, containing data from image modality, and ECG modality were discarded. These 141 full-text studies were read and finally total of 68 studies were included for review.

These articles/studies were based on machine learning techniques for heart disease prediction. The selected papers were studied in depth to identify relevant studies.

Figure 2 shows the distribution of the selected papers over the years. In Fig. 2, the number of papers has increased gradually over the years, but as the study limits up to June 2023, hence fewer papers are included in the year 2023.

RQ1: What are the Different Types of Heart Disease, Their Causes, Symptoms, and Risk Factors?

The term cardiovascular disease is also used for heart disease in a few studies. But it refers to the disease of arteries/veins that carry blood to other parts of the body. Heart disease is a specific term that means disease of heart-related muscles only.

The term heart disease refers to a range of conditions that affect the normal working of the heart. Heart disease is related to the disease of heart muscles, heart blood veins, and valves. It occurs by saturation of fatty substance i.e., a plaque inside veins, and arteries. Various parts of the body receive blood from the Heart. The blood flows through different blood vessels and veins. When those vessels have a problem carrying blood due to plaques or blood clots then that disease is called cardiovascular disease. Cardiovascular disease is a disease related to blood vessels that carry blood from the heart toward different body organs such as the brain, kidney, limbs, etc. All types of heart disease can be called as cardiovascular disease but not all cardiovascular diseases are heart disease.

Following are the details of types of heart disease which are given in the figure below (Fig. 3).

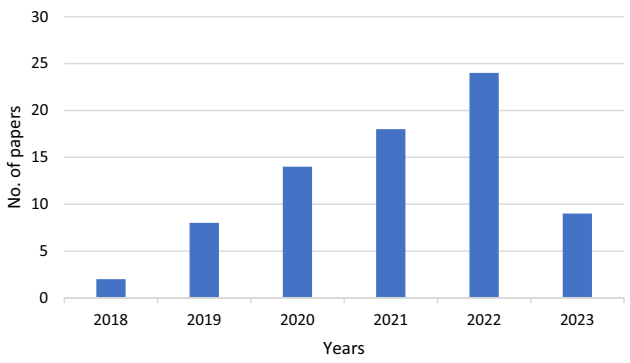


Fig. 2 Distribution of papers

Coronary Artery Disease

It is caused by the hardening of coronary arteries. The fatty substance called plaque is deposited at the wall of arteries. Due to this substance, the heart does not get oxygenated blood which ultimately in heart attack and heart failure. Nowadays the disease occurrence for this type has increased highly [6].

Arrhythmia

It is the problem with the rhythming of heartbeats. The problem is caused by due irregular beating of the heart which is the electrical activity of the heart to pump the blood. This causes brain stroke, heart attack, or even heart failure [7].

Congenital Heart Disease

This type of heart disease is present from the birth itself. These are heart defects from birth which are carried through the life of one.[8]

Cardiomyopathy

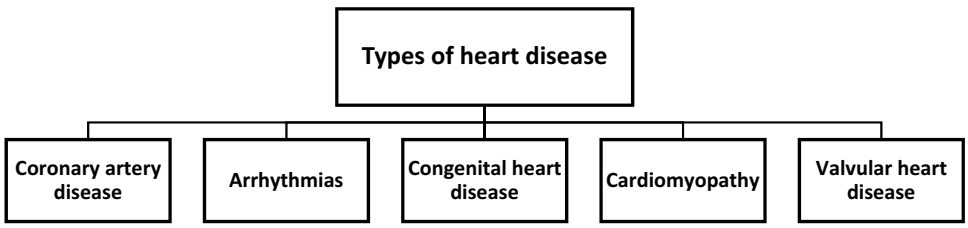
Cardiomyopathy is when heart muscles become harder to pump blood to the rest of the body. The muscles become hard due to a lack of certain supplements and nutrients.[9]

Valvular Heart Disease

In the human heart, a total of four valves are present in the heart. If any valve is having a problem or not working properly then this type of disease occurs.

The above types of diseases can be tested by using different test methods such as ECG which checks electric impulses, an echocardiogram uses ultrasound, chest CT scan, angiography, chest X-ray, stress test, cardiac catheterization, heart biopsy, pulse oximetry, angiogram [10].

Fig. 3 Types of heart disease



RQ2: Identify High-Risk Heart Disease and Identify the Factors Contributing to the Development of this Disease

It has been found in the papers that; coronary artery disease is becoming the most dangerous one of all [11]. This disease is also referred to as ischemic heart disease or coronary heart disease. It is also referred to as Atherosclerosis of coronary arteries. Atheros means gruel, a fatty substance, and sclerosis means hardening.

Below are some statistics about CAD disease across the globe:

- About 25,000 people under the age of 75 in the UK die per year because of CAD [11].
- CHD death rates are highest in Scotland and the north of England [11].
- In Western Europe, only Ireland, Germany, Sweden, and Luxembourg had a higher death rate than the UK in the same year [11].
- South Asian people living in the UK (people from India, Pakistan, Bangladesh, and Sri Lanka) have a higher premature death rate from CHD (46% higher for men; 51% higher for women) [11].
- With its rapidly increasing disease burden and mortality rate, coronary artery disease (CAD)—the most widely seen cardiovascular disease (CVD), is the growing disease in India [12].
- Coronary artery disease (CAD) accounts for approximately 610,000 deaths annually (estimated 1 in 4 deaths) and is the leading cause of mortality in the United States [13].

The risk factors associated with CAD are given by epidemiological research and these are given are smoking, diabetes, hypertension, hyperlipidemia, obesity, and stress [14]. Despite the stated risk factors other studies reported that systolic blood pressure ≥ 140 mmHg/diastolic blood pressure ≥ 90 mmHg, total cholesterol ≥ 240 mg/dL are also major risk factors [13].

RQ3: Which ML Techniques are Used in Cardiology?

Figure 4 presents the distribution of ML techniques that were used in the selected studies per objectives. Various techniques and models were developed to develop a decision support system for clinicians, and doctors and give a way for processing large medical datasets. These techniques are mainly from classification and prediction objectives. Very few techniques were belonging to clustering and association techniques namely 6% and 1% respectively. Nowadays many researchers have adopted a new way of combination of multiple classification techniques to increase the accuracy. These also include ensembles.

According to Fig. 4, the Random Forest model has been used widely by most of the researchers then SVM, KNN, DT, LR, and NB. In these 74 studies, the given ML techniques from Fig. 4 occur 283 times. This means if we count the occurrence of each technique,

and do the sum then it gives a total count of 283. The 16% of studies used RF as a classifier for disease. If the random forest is used with a feature selection technique, then instead of conventional RF, the selected combination has given the highest accuracy [15]. By using a smaller number of features from datasets, the RF provides good accuracy results from 90 to 95% [16]. Then next to RF, SVM was used in 13% of

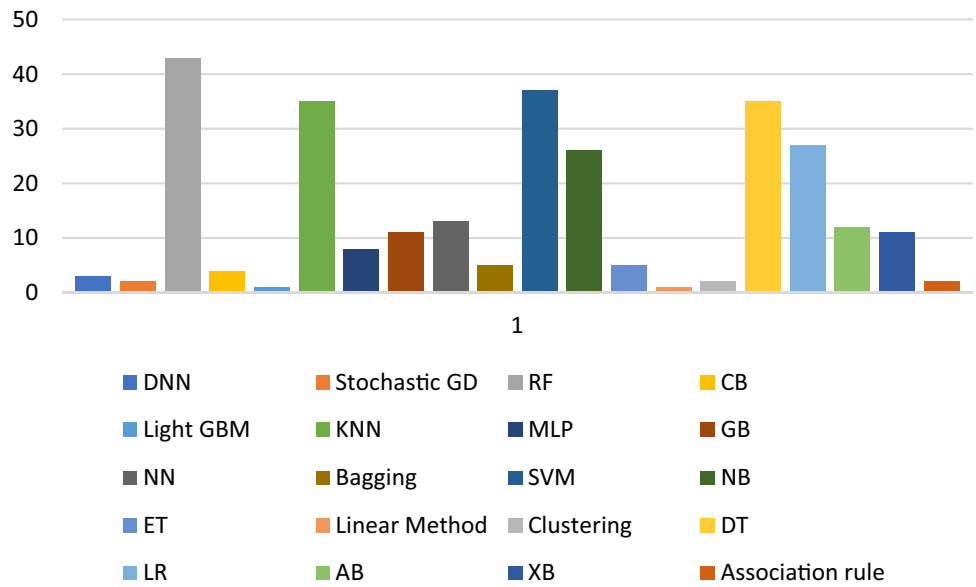


Fig. 4 Different ML algorithm used

studies. It has been observed that the focus of researchers has been changed from using a single technique for the prediction of disease to using ML technique + feature selection algorithm + cross-validation. Such types of studies have developed tremendously since 2022. Hence in this regard, if SVM has been used, then it has a classification accuracy of 92.37% [17]. SVM finds its advantage in prediction for binary classification. The KNN and DT have been used in 12% of studies. Many of the studies are now combining these ML techniques as an ensemble to gain more benefits of classification and achieving maximum accuracy. The KNN and DT are easy to implement and understand. Hence most of the researchers find them useful for implementation. Also, in many other studies for 10%, a logistic regressor has been used. The ensemble of LR + KNN with feature selection and fold cross-validation has achieved an accuracy of 99% [18].

RQ4: What Are the Advantages and Disadvantages of ML Techniques when Used in Cardiology?

The studies selected here present various models that were developed using ML techniques that are applied in cardiology. Every technique has advantages and disadvantages. Table 1 gives some common advantages of techniques that are given in the papers. These strengths and limitations are given for random forest, decision tree, SVM, KNN, and of other techniques. The research trend has changed. The previous studies used only one classification technique till some past years. But nowadays, many studies are using ensemble techniques to get maximum classification and prediction accuracy. Each ML technique has its advantages that motivate researchers to use it in their model development. Such as the SVM gives the classification accuracy as 92.37 and 92.30 due to its robustness. By combining the strengths of major classification techniques, the ensembles are developed by researchers. Hence due to these ensembles, the classification accuracy has increased [19, 20].

Despite all the strengths of techniques, there are limitations also present. Hence researchers have used other techniques due to the limitations of the previously used. Some studies are reporting small datasets and missing data challenges [21]. Due to the missing data challenge, researchers have used data preprocessing techniques excessively [17].

Researchers have used the above techniques effectively. However, these techniques have certain limitations. Other researchers overcome these limitations in their implementations by developing classification/prediction techniques. Using other ML algorithms as individual or an ensemble.

RQ5: What is the Overall Performance of ML Techniques in Cardiology?

As per the insights from studies, the random forest classifier has been used by most researchers to classify diseases. Random forest provides good accurate results for each of the datasets such as breast cancer, diabetes, heart disease, thyroid, surgery data, dermatology, and liver disorder [21]. If the Random Forest algorithm is implemented with a feature extraction algorithm, then its accuracy increases. The study reported that maximum prediction accuracy has been given by recursive feature elimination and random forest. 99.7 [27]. The SVM algorithm has given the maximum classification accuracy in many studies. The types of SVM such as nuSVM, LinSVM, and SVC were used in one of the studies, and when used with feature extraction techniques reported classification accuracy was 93.08 [28]. It has been observed in studies if one combines individual ML techniques with feature selection algorithms then they give high prediction/classification accuracy. If they apply as an individual without using any feature extraction technique then the performance of the algorithm is not satisfied.

The KNN algorithm has also proven to have maximum accuracy in many studies. Many researchers have used it with other techniques as an ensemble. Also, if it has been used on a single database then its accuracy is reported as 97.83 [29].

Summary

This section summarises the findings of each research question (RQs). The principal findings of this study are presented and discussed. The research question-wise summary is as below.

RQ1: What Are the Different Types of Heart Disease, Their Causes, Symptoms, and Risk Factors?

As shown in the Table 2, the symptoms and risk factors for types of disease have been summarised. There are common symptoms for every type such as chest pain, fatigue, shortness in breath, and weak feeling. Only some specific symptoms are there for every type which differentiates them into heart disease type. Also, if we observe the risk factor column, the common factors can be age, gender, high bp, high cholesterol, smoking habits, alcohol consumption, and lifestyle. Hence one can say that if any person is having above certain symptoms, then he/she can be prone to heart disease. Further investigation will decide the type of heart disease.

RQ2: Identify High-Risk Heart Disease and Identify the Factors Contributing to the Development of this Disease

As per the statistics of the CAD in "RQ1: What are the Different Types of Heart Disease, Their Causes, Symptoms,

Table 2 Heart disease types, causes, symptoms, risk factors

Types of heart disease	Causes	Symptoms	Risk factors
Coronary artery disease	Hardening of coronary arteries due to deposition of plaque	1. Angina/chest pain 2. Shortness in breathing 3. Fatigue 4. Heart attack	Age, gender, family history, smoking, high BP, high cholesterol, diabetes, obesity, lifestyle, unhealthy diet
Arrhythmia	Problem with rhythming of heartbeats	1. Feeling week 2. Irregular heartbeats 3. Chest pain 4. Shortness of breathing 5. Fatigue	Age, CAD, high BP, Imbalance in electrolytes, Thyroid problems, High alcohol consumption
Congenital heart disease	From childbirth	1. Skin color changes to blue 2. Breathing rapidly 3. Fatigue 4. Poor blood circulation	Genetic factors, Abnormal chromosomes
Cardiomyopathy	Heart muscles become harder	1. Shortness of breathing 2. Fatigue 3. Swelling 4. Feeling week	Family history, heart conditions, metabolic disease, deficiency in nutrition, alcohol and drug consumption, weak immune system, radiation and chemotherapy
Valvular heart disease	Valve problems	1. Shortness of breathing 2. Fatigue 3. Swelling 4. Feeling week 5. Chest pain 6. Irregular heart beats	Age, rheumatic fever history, infection, heart conditions, congenital heart problems, weak immune system, radiation therapy

and Risk Factors?", CAD can be called a dangerous heart disease among the remaining types. The statistics highlight its significance and regional variation towards its hazards. There is a substantial impact of the disease on young people in the UK, due to differences in healthcare access, socioeconomic factors, lifestyle, obesity, and smoking habits, it becomes dangerous in Scotland and the north of England [11]. There should be certain policy changes and healthcare improvements in such countries. The higher death rate in the South Asian population in the UK indicates that cultural dietary habits, lifestyle, and potential barriers to accessing healthcare cause CAD dangerous. In India, the increasing trend towards urbanization, changes in lifestyle, and increased occurrence of risk factors such as obesity, diabetes, sedentary lifestyle, hypertension, and less physical activity make CHD hazardous. In the USA, according to statistics, there is a comprehensive need for cardiovascular healthcare strategies. In view of the above discussion, there is a clear need of disease prevention strategies, public education, promotion of healthy life style, and routine health screenings. In high-burden population countries, there should be increased access for healthcare.

RQ3: Which ML Techniques are Used in Cardiology?

In 68 studies, it has been found that random forest has been used by many researchers as the classifier.

Then SVM, DT, KNN and etc. The top 10 or widely used ML algorithms are shown in the Fig. 5. The Fig. 5 indicates the top 10 algorithms and there count of usage in 68 studies. There is a growing trend among researchers that instead of using a single ML technique one can combine other techniques with ML also. These are feature selection algorithms, and cross-validation techniques to enhance the model's accuracy and reliability. The use of ensembles has been increased a little bit in some studies. These ensembles are combined with individual ML techniques. This combination takes advantage of strengths of different classifier and lead to better predictive performance. Also, the researchers have focused on feature selection. Feature selection is a crucial step as it identifies the most relevant features. By selecting features, the model accuracy is increased and computational complexity is reduced.

RQ4: What Are the Advantages and Disadvantages of ML Techniques When Used in Cardiology?

The application of ML techniques in heart disease prediction gives significant benefits including improved accuracy, strength, and ability to handle complex datasets. The ease of implementation and understanding of decision trees and logistic regression if combined with the predictive power of random forests and support vector machines gives a better toolkit for medical researchers. Ensemble methods further enhance the strengths of

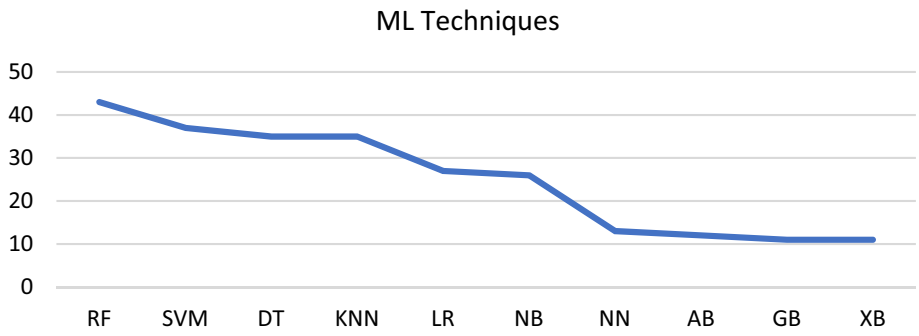


Fig. 5 Most used ML algorithms in cardiology

individual techniques that offer the best use of multiple algorithms. Hence in this way ML becomes an valuable asset in ongoing research of heart disease prediction, enabling more accurate, good treatments and ultimately better patient care.

RQ5: What is the Overall Performance of ML Techniques in Cardiology?

The summary of some of the ML techniques implemented by the authors and the accuracies of those techniques are given in the Table 3. If we use, feature selection, dimensionality reduction technique, and ensemble then the accuracy for disease prediction has increased. The same has been indicated in Table 3.

Implications of Research and Limitations of the Review

This section might be useful for research scholars in finding gaps in selected studies. Heart disease is becoming life life-threatening disease all over the globe, hence Table 2 provides information regarding the limitations of selected papers. The below gaps have been found while studying in literature.

- Size of dataset: Many studies rather than all of the studies have taken small sample sizes. ML techniques show slower performances when it comes to large data sizes. If the size of the dataset is adequately small then the same ML techniques give better performance. The size of the dataset has a significant impact on the accuracy of ML techniques.

Table 3 Highest accuracies obtained for ML techniques in review studies

Author	Proposed ML technique	Accuracy obtained
Ali, Md Mamun, et al. 2021	RF + feature importance technique	100
Azam Mehmood Qadr et al. 2023	DT + principal component technique	100
Y Khourdifi et al. 2019	Optimized technique + feature selection + KNN	99.65
Ghulab Nabi Ahmad et al. 2022	DT, RF + LDA	99.40
Pronab Ghosh, et. Al. 2021	Random forest bagging method	99.05
Velusamy, Durgadevi et al. 2021	WAVEn ensemble	98.97
XY Gao et al., 2021	Nu-SVC	94.66
D. Senitta et al. 2022	Bagging SMO, NB, SMO, NN, ICA-KNN + feature selection	98.38
G Wang et al. 2022	RF + dimensionality reduction with feature selection	98.3
M.A. Khan 2020	FS + DNN	98.1
J. Jasmine Gabriel et al., 2023	BORUTA Sharp algorithm for feature selection trained with RF, AB, CB, ET	97.70
Ayesha Noor et al. 2023	Stacking model using dimensionality reduction and data imbalancing approach combined with	97
Divya Krishnani et al. 2019	RF + preprocessing technique	96.8
Ashima Tyagi 2022	Ensemble of DT, RF	96.7
Norma Latif Fitriyani et al. 2020	DBScan, Smote ENN, KNN, Xgboost	95.90

Table 4 Comparison of machine learning techniques for heart disease prediction from selected studies

Sr. no.	Author and year	Work proposed	Comparison of ML techniques	Proposed technique	Accuracy (in %)	Dataset	Gaps/limitations/challenges addressed
1	Pavleen Kaur et al. 2018 [21]	IoT based healthcare network for heart disease prediction	K-NN, linear support vector machine, decision trees, Random Forest, and MLP	Random Forest	82.3	Cleveland	Large dataset should be considered
2	Ashir Javeed et al. 2019 [15]	Developed a framework by using feature selection and prediction algorithm for heart failure diagnosis	RSA, RF	RSA + RF	93.33	Cleveland	Problem of overfitting of data
3	Senthilkumar Mohan et al. 2019 [30]	Developed hybrid random forest model	NB, LR, DL,DT, RF, gradient boost, SVM	Hybrid random forest with linear model	88.7	Cleveland	Traditional algorithms are not giving accurate results for prediction and diagnosis
4	Divya Krishnani et al. 2019 [23]	Proposed preprocessing approach for detecting coronary heart disease	RF, DT, KNN	RF	96.8	Framingham Heart study	Real time and larger dataset required for prediction
5	Norma Latif Fitriyani et al. 2020 [25]	Proposed heart disease prediction model for clinical decision support system	DBScan, Smote ENN, KNN, Xgboost	Xgboost	95.90	Statglog and Cleveland	Different outlier detection method can be investigated
6	Jian Ping Li et al. 2020 [17]	Fast mutual information feature selection algorithm	SVM, LR, ANN, KNN, NB, DT	Feature selection algorithm + SVM	92.37	Cleveland	Focus on feature selection algorithms
7	Yar Muhammad et al. 2020 [31]	Feature selection algorithm in combination with classification technique	KNN, DT,ETC, RF, LR, NB, ANN, SVM, AB	ET	92.09	Cleveland and Hungarian	To use more optimization techniques, feature selection and classification algorithms
8	Devansh Shah 2020 [22]	Focused on probability of developing heart disease	NB, DT,KNN,RF	DT	80.26	UCI repository	To implement more complex and combination of models
9	Debabrata Swain et al. 2020 [24]	Analysis of classification algorithm for heart disease prediction	LR, SVM, KNN NB, DT, RF	LR	88.29	Framingham	Techniques should use parameter tuning and and ensembles for prediction
10	Rahul Katarya 2020 [32]	Studied ML techniques for the prediction of cardiac disease	LR, KNN, SVM, NB, DT, RF, ANN, MLP, DNN	SVM	92.30	Cleveland	Deep learning techniques can also be focused
11	Khan 2020 [18]	IoT Framework for effective prediction of heart disease	FS + DNN	DNN	98.1	Hungarian, Framingham, Public health	To use more feature selection and optimization techniques

Table 4 (continued)

Sr. no.	Author and year	Work proposed	Comparison of ML techniques	Proposed technique	Accuracy (in %)	Dataset	Gaps/limitations/challenges addressed
12	Aqsa Rahim et al. 2021 [33]	ML framework for CVD detection	Ensemble of KNN and LR	ENSEMBLE OF KNN & LR	95.50	Framingham	Robust need of a framework for CVD where data balancing, feature selection missing value handling, and classification is improved
13	Tsatsral Amarbayasgalan et al. 2021 [34]	CHD risk prediction framework, two datasets have desinged	FS + DNN	DNN	–	Own	Overfitting
14	Pronab Ghosh, et al. 2021 [35]	Developed new hybrid classifiers using existing classifiers	DT, RF, K nearest neighbors, adaboost, gradient boosting, bagging	Random Forest bagging method	99.05	Combination of the Cleveland, Hungary, Switzerland, VA Long Beach, Statlog heart disease dataset	The model should be robust against the dataset, and missing values. Focus deep learning
15	Swathy and Saruladha 2021 [36]	Comparison of data mining, machine learning, and deep learning models for CVD prediction	For heart disease prediction NN, ANN can work better	–	–	–	AI and data mining domains can be searched more for disease prediction
16	Bhanu Prakash Doppala et al. 2021 [37]	Using attribute reduction developed hybrid method for heart disease prediction	SVM, DT,LR, NB, RBF network	GA-RBF	94.20	Cleveland	Existing methods are not efficient and accurate for heart disease prediction
17	Saba Bashir et al. 2021 [38]	Four ensemble-based voting scheme is proposed	1.SVM + NB + AUTO NLP on Cleveland 2. SVM + NB + AUTO NLP on Hungarian 3.SVM + PERCEPYTON + NN on Switzerland 4.DT + NB + AUTO NLP on Long beach	SVM + NB + AUTO NLP ON Cleveland	83	Cleveland, Hungarian, Switzerland, Long beach	Other ensembles such as Adaboost, bagging boosting, stacking can be used
18	Chandan Pan, Arnab Poddar et al. 2022 [39]	Combination of categorical, numerical features to predict heart disease	GB, Xgboost, Adaboost, Catboost, ANN, SVM, RF, DT, LR	Ensemble of SVM, AdaBoost	–	Cleveland	–
18	Victor Chang et al. 2022 [40]	AI model using machine learning techniques used for heart disease prediction	KNN, DT, SVM, RF, LR	RF	83	Cleveland	Other ML techniques can be investigated using feature selection
19	Muhammad Salman Pathan et al. 2022 [27]	The most important risk factors for disease occurrence and disease classification	LR, DT, NB, RF, MLP, ET	SVM	75	CVD, Framingham	Enhancing prediction accuracy by using other combinations of ML

Table 4 (continued)

Sr. no.	Author and year	Work proposed	Comparison of ML techniques	Proposed technique	Accuracy (in %)	Dataset	Gaps/limitations/challenges addressed
20	Karthick Kanagarathinam et al. 2022 [41]	Hybrid dataset in identification of CVD	NB, XGBoost, KNN, MLP, SVM, CatBoost	CatBoost	94.34	Hungarian, Switzerland, Hungarian, Long beach	–
21	Ritesh Sonawane, Hitendra Patil 2022 [42]	Clustering model for heart disease prediction	Kmeans clustering	Kmeans clustering	–	ECG signal data	Other ML techniques can be explored
22	Ashima Tyagi 2022 [43]	Feature selection approach using ML techniques	DT,RF, SVM, KNN, ensemble	Ensemble	96.7	Z-Alizadeh Sani	Early diagnosis of CAD is more important
23	Nagaraj M et al. 2022 [44]	Genetic algorithms used for prediction for heart disease	RF, XGBoost, NN, Adaboost, bagged trees	–	Better accuracy	UCI	Deep learning models & association rule mining can be used
24	Abdul Aleem et al. 2022 [20]	Feature selection algorithm for improving prediction	GA, PSO, NB, SVM, J48, DT, LR, RF	–	–	Hungarian	New optimization techniques can be used
25	Priyank Gupta et al. 2022 [45]	New majority noting ensemble method	LR, NB, DT, RF, SVM, XGBM, LGBM, KNN	Ensemble of NB, RF, SVM, LGBM	88.33	Cleveland	Boosting algorithms can be used to improve prediction accuracy
26	Sfurti Sarah et al. 2022 [46]	Compared various ML models and give the best one	LR, DT, NB, SVM, KNN, RF	LR	85.25	Cleveland	Combinations of hybrid algorithms and feature selection techniques can be implemented
27	Ghulab Nabi Ahmad et al. 2022 [47]	Model that matches with real world	LDA, RF, GBC, DT, SVM, KNN,	DT, RF	99.40	Cleveland, Hungary, Switzerland, Statlog and Long Beach	Other feature selection techniques can be used
28	Nurul Absar et al. 2022 [48]	System that predicts heart disease using combination of datasets	RF, DT, ADB, KNN	RF, KNN	93.43,97.83	Cleveland, Hungary, Switzerland and Long Beach	Implement k fold cross validation, overfitting issue removal
29	Shakeel Ahmad 2023 [49]	DNN using CNN	CNN, BiLSTM	CNN, BiLSTM	94.50	UCI	Practical applications should be implemented in more
30	Pooja Rani et al. 2021 [50]	Hybrid system of ML techniques	SVM,NB, LR, RF, AdaBoost	RF	86.60	Cleveland	Other feature selection techniques can be used
31	XY Gao et al. 2021 [51]	Ensemble methods	PCA, LDA, KNN, DT, RF, NB, boosting, bagging	Bagging + PCA + DT	98.6	Cleveland	Other feature selection techniques can be used
32	Moloud Abdar et al. 2019 [52]	Novel nested ensemble nu-Support Vactor classification model for CAD	NB, SVC, RF, stacking, bagging, voting, stochastic gradient descent	Nu-SVC	94.66	Cleveland and Z Alizadeh Sani	Evolutionary algorithm is needed to select level automatically

Table 4 (continued)

Sr. no.	Author and year	Work proposed	Comparison of ML techniques	Proposed technique	Accuracy (in %)	Dataset	Gaps/limitations/challenges addressed
33	J. Jasmine Gabriel et al. 2023 [53]	Feature selection, hyperparameter tuning and optimization algorithm for CAD	BORUTA Sharp algorithm for feature selection trained with RF, AB, CB, ET	BORUTA Sharp algorithm for feature selection trained with RF, AB, CB, ET	97.70	Z Alizadeh Sani	Focus on feature selection, hyperparameter optimization
34	Huazhong Yang 2023 [54]	Optimizing the performance of LightGBM algorithms	DT, RF, CB, XGB, ADA, BG, GBM and HY_OptGBM algorithms	DT, RF, CB, XGB, ADA, BG, GBM and HY_OptGBM algorithms	93.0	Framingham Heart Institute	Performance improvement of systems
35	Jikuo Wang 2020 [55]	8 feature selection methods and 10 ML classification algorithms are evaluated	RF, ET, ADB, SVC, MLP, XGB, GPC, GNB, LR, NB with feature selection methods	GNB, GB, RF, ET, ADB, MLP, XBG	95.84	Z Alizadeh Sani	More connections with hospitals is needed
36	Azam Mehmood Qadr et al. 2023 [56]	Principal component heart failure feature engineering technique by creating new features	LR, RF, SVM, DT, XGB, NB, KNN, MLP, GB	DT	100	Heart failure dataset	Need to increase the dataset using data balancing
37	Xiaoming Yuan et al. 2022 [57]	Binary & multiclass prediction	Fuzzy GBDT, GBDT, DT, Bagging GBDT, Bagging-Fuzzy GBDT	Bagging Fuzzy GBDT	90	Cleveland, Hungary, Switzerland and VA Long Beach	Severity of disease should also be detected by models
38	Sushree Chinmayee Patra 2023 [58]	A two-step ensemble technique	KNN, SVM, DT, ET, RF, ADABoost, XGBoost	ET	89.14	Feamingham	Other datasets can be sued with other ML,DL technique
39	Ankur Gupta 2020 [59]	Framework that includes data imputation, partitioning, feature extraction, normalization, ML algorithms	LR, KNN, SVM, DT, RF	FAMD + RF	93.44	Cleveland	Need of multiclass classification for heart disease
40	Abid Ishaq et al. 2020 [60]	Class imbalance problem has been solved using SMOTE and 9 ML techniques used	DR, AdaBoost, LR, SGD, RF, GBM, ETC, GNB, SVM	ETC	92.62	Heart failure clinical record set from UCI	Combinations of ML techniques with feature selection can be used for better prediction of disease
41	SEA Ashri et al. 2020 [61]	Hybrid classifiers using ensemble technique with majority voting model	LR, SVM, KNN, DT, RF	RF	93	Cleveland, Statlog	Real time data prediction for heart disease
42	Abdellatif, Abdallah et al. 2022 [62]	The model that selects features performs disease prediction using parameter optimization	SVM, KNN, XgBoost, SMOTE-RF, GNB,LR	RF	95.5	Statlog, heart failure clinical records	General frameworks on ML ensembles including outliers detection and removal
43	Sara Ghorashi et al. 2022 [63]	A model DNN using long short term memory	DNN model with LSTM	DNN model with LSTM	71.5	Health data from UAE hospitals	–

Table 4 (continued)

Sr. no.	Author and year	Work proposed	Comparison of ML techniques	Proposed technique	Accuracy (in %)	Dataset	Gaps/limitations/challenges addressed
44	Abdallah Abdellatif et al. 2022 [64]	SMOTE with ML classifiers and HPO	SVM, SGD,KNN, ET, XGBoost, LR	SMOTE + ET	84.53	Cleveland, Statlog	Use of real time clinical data
45	Ayesha Noor et al. 2023 [63]	Stacking model using dimensionality reduction & data imbalancing approach combined with ML classifiers	Passive aggressive classifier (PAC), ridge classifier (RC), SGD, XGBoost, LogitBoost	LogitBoost	97	Kaggle	Clinical data
46	D. Senitta et al. 2022 [65]	Feature selection algorithm with ML classifiers	Bagging SMO, NB, SMO, NN, ICA-KNN	Own algorithm	98.38	Cleveland, Hungary, Switzerland anLong Beach	Convergence accuracy and speed can be studied
47	G Wang et al. 2022 [66]	Feature selection using dimensionality reduction with 6 classifiers	RF, LR, ANN, GB, KNN, SVM	RF	98.3	Cleveland and Hungarian	More large dataset
48	CM Bhatt et al. 2023 [67]	Heart disease classification with real dataset	DT, XGBoost, RF MLP	MLP	87.28	CVD	Need of more data
49	Y Khourdifi et al. 2019 [68]	Feature selection using 9 classifiers	KNN, SVM, NB, RF, MLP, ANN	Optimized technique + KNN	99.65	Cleveland	Other feature selection techniques need to use
50	Velusamy, Durgadevi et al. 2021 [69]	Heterogeneous ensemble	KNN, RF, SVM	WAVEn ensemble	98.97	Z Alizadeh sani	LAD, RCA, LCX Stenosis need to be predicted
51	Ali, Md Mamun, et al. 2021 [70]	Using feature importance technique classification of disease	KNN, RF, DT, AdaBoost M1, LR, MLP	RF	100	Kaggle	Real time data
52	Kibria et al. 2022 [71]	Fusion model to diagnose CVD	ANN, SVM, LR, DT, RF, AdaBoost	Fusion model	95.08	Cleveland	Hospital data

- Distribution of data: As there are no standard rules for data distribution, hence it gives different accuracies for ML techniques. This is due to random splitting. Some researchers have used 70–30%, some used 80–20% to train and test the data. This random sampling affects the accuracy of algorithms.
- Need for real-time data: As the available datasets have been developed for a long back, hence there is a need for real-time data for disease classification or prediction. Very few ML techniques have used real-time data or the latest datasets.
- Feature selection/extraction and hyperparameter optimization: It has been found that feature selection greatly affects disease prediction accuracy. Hence the trend of researchers has shifted towards using feature selection techniques from almost 2021. There are still other feature selection techniques exist which should be focused. The same observation is with hyperparameter tuning. Nowadays the trends of researchers have increased towards the optimization of hyperparameters. However, the technique or algorithm which is used for tuning has been kept unidentified.
- Use of ensembles: The use of ensembles for disease prediction should be increased, as it provides better disease prediction accuracy.
- Real-time clinical help from clinicians: It is the need for disease prediction to examine, and test clinical data. Due to this real-time decision-making will be more helpful for clinical practitioners/doctors.

In Table 4 the details of implemented papers by different researchers have been given. The limitations of the study are that the studies included are from 2018 to June 2023. They have selected as per inclusion criteria. Other authors can develop their inclusion criteria and different papers can be selected for studies.

As we see in the table that total of 52 studies are included out of 68. These are implemented studies and the remaining studies from 68 are review studies.

Conclusion

The systematic review done here investigates the studies that were implemented in cardiology using ML techniques. Initially, 676 papers were identified. After applying selection/inclusion criteria 68 studies were taken for this review. A total of 5 research questions were developed. To answer this research, question these 68 studies were studied. Different ML techniques, their advantages, and limitations were studied in this regard to answer the research questions.

As per Table 4, many researchers have used prediction and classification techniques to predict heart disease. Also, the Hybrids and ensembles have been used to improve the prediction accuracies. Many studies have used combined datasets as input to these ML techniques. RF, DT, SVM, and LR are the most used ML techniques in these studies. Also, this review provides research scholars with future direction such as using more datasets, investigating other datasets that are not used frequently in this study also collaborating with clinicians to work on real-time data.

Author Contributions Priti Shinde: writing the draft, methodologies, collecting results, investigation; Mahesh Sanghavi: investigation, editing, collecting the results; Tien Anh Tran: editing, supervision, conclusion.

Funding Open Access funding provided by the University of Malta. There is no funding to support this study.

Data Availability Statement The data availability is provided if any requirements.

Declarations

Conflict of interest The authors declare that there are no competing interests in this study.

Research involving human and/or animals Not applicable.

Informed consent Not applicable.

References

1. Ahsan MM, Siddique Z. Machine learning-based heart disease diagnosis: a systematic literature review. Artif Intell Med. 2022;128:102289.
2. https://www.ncbi.nlm.nih.gov/books/NBK535419
3. https://training.cochrane.org/handbook.
4. McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. J Clin Epidemiol. 2021.
5. Kadi I, Idri A, Fernandez-Aleman JL. Knowledge discovery in cardiology: a systematic literature review. Int J Med Inform. 2017;97:12–32.
6. https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/diagnosis-treatment/drc-20350619.

7. https://www.nhlbi.nih.gov/health/arrhythmias#:~:text=An%20arrhythmia%2C%20or%20irregular%20heartbeat,down%20while%20resting%20or%20sleeping.
8. https://www.mayoclinic.org/diseases-conditions/adult-congenital-heart-disease/diagnosis-treatment/drc-20355461.
9. https://www.nhlbi.nih.gov/health/cardiomyopathy.
10. https://www.nhlbi.nih.gov/health/heart-valve-diseases/diagnosis.
11. Sanchis-Gomar F, et al. Epidemiology of coronary heart disease and acute coronary syndrome. Ann Transl Med. 2016;4(13):256.
12. https://www.upmc.com/services/heart-vascular/conditions-treatments/coronary-artery-disease.
13. https://www.deccanchronicle.com/lifestyle/health-and-wellbeing/010723/the-rising-scourgecad-takes-heavy-toll-india.html.
14. Malakar AK, et al. A review on coronary artery disease, its risk factors, and therapeutics. J Cell Physiol. 2019;234(10):16812–23.
15. Javeed A, et al. An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection. IEEE Access. 2019;7:180235–43.
16. Manduchi E, et al. Genetic analysis of coronary artery disease using tree-based automated machine learning informed by biology-based feature selection. IEEE/ACM Trans Comput Biol Bioinform. 2021;19(3):1379–86.
17. Li JP, et al. Heart disease identification method using machine learning classification in e-healthcare. IEEE Access. 2020;8:107562–82.
18. Khan MA. An IoT framework for heart disease prediction based on MDCNN classifier. IEEE Access. 2020;8:34717–27.
19. Garg A, Mago V. Role of machine learning in medical research: a survey. Comput Sci Rev. 2021;40: 100370.
20. Aleem A, Prateek G, Kumar N. Improving heart disease prediction using feature selection through genetic algorithm. In: International conference on advanced network technologies and intelligent computing. Cham: Springer International Publishing; 2021.
21. Kaur P, Kumar R, Kumar M. A healthcare monitoring system using random forest and internet of things (IoT). Multimed Tools Appli. 2019;78:19905–16.
22. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Comput Sci. 2020;1:1–6.
23. Krishnani D, et al. Prediction of coronary heart disease using supervised machine learning algorithms. In: TENCON 2019–2019 IEEE region 10 conference (TENCON). IEEE; 2019.
24. Swain D, et al. An efficient heart disease prediction system using machine learning.In: Machine learning and information processing: Proceedings of ICMLIP 2019. Springer Singapore; 2020.
25. Fitriyani NL, et al. HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access. 2020;8:133034–50.
26. Abdeldjouad FZ, Brahami M, Matta N. A hybrid approach for heart disease diagnosis and prediction using machine learning techniques. In: The impact of digital technologies on public health in developed and developing countries: 18th international conference, ICOST 2020, Hammamet, Tunisia, June 24–26, 2020, Proceedings 18. Springer International Publishing; 2020.
27. Pathan MS, et al. Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthc Anal. 2022;2:100060.
28. Azmi J, et al. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Med Eng Phys. 2022;105:103825.
29. Javeed A, et al. Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: a systematic review and future directions. Comput Math Methods Med. 2022;2022:1–30.
30. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542–54.
31. Muhammad Y, et al. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep. 2020;10(1):19747.
32. Katarya R, Meena SK. Machine learning techniques for heart disease prediction: a comparative study and analysis. Health Technol. 2021;41(1):87–97.
33. Rahim A, et al. An integrated machine learning framework for effective prediction of cardiovascular diseases. IEEE Access. 2021;9:106575–88.
34. Amarbayasgalan T, et al. An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets. IEEE Access. 2021;9:135210–23.
35. Ghosh P, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access. 2021;9:19304–26.
36. Swathy M, Saruladha K. A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques. ICT Express. 2022;8(1):109–16.
37. Doppala BP, et al. A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset. Distrib Parallel Databases. 2021: 1–20.
38. Bashir S, et al. A knowledge-based clinical decision support system utilizing an intelligent ensemble voting scheme for improved cardiovascular disease prediction. IEEE Access. 2021;9:130805–22.
39. Pan C, et al. Impact of categorical and numerical features in ensemble machine learning frameworks for heart disease prediction. Biomed Signal Process Control. 2022;76:103666.
40. Chang V, et al. An artificial intelligence model for heart disease detection using machine learning algorithms. Healthc Anal. 2022;2:100016.
41. Kanagarathinam K, Sankaran D, Manikandan R. Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset. Data Knowl Eng. 2022;140:102042.
42. Sonawane R, Patil H. Automated heart disease prediction model by hybrid heuristic-based feature optimization and enhanced clustering. Biomed Signal Process Control. 2022;72: 103260.
43. Tyagi A, Singh VP, Gore MM. Improved detection of coronary artery disease using DT-RFE based feature selection and ensemble learning. In: International conference on advanced network technologies and intelligent computing. Cham: Springer International Publishing; 2021.
44. Lutimath NM, et al. Prediction of heart disease using genetic algorithm. In: Proceedings of second doctoral symposium on computational intelligence: DoSCI 2021. Springer Singapore; 2022.
45. Gupta P, et al. Heart disease detection scheme using a new ensemble classifier. In: Advances in data and information sciences: Proceedings of ICDIS 2021. Singapore: Springer Singapore; 2022. p. 99–110.
46. Sarah S, et al. Heart disease prediction using core machine learning techniques—a comparative study. In: Advances in data and information sciences: Proceedings of ICDIS 2021. Singapore: Springer Singapore; 2022. p. 247–260.
47. Ahmad GN, et al. Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection. IEEE Access. 2022;10:23808–28.
48. Absar N, et al. The efficacy of machine-learning-supported smart system for heart disease prediction. Healthcare. 2022;10(6):1137.
49. Ahmad S, et al. Diagnosis of cardiovascular disease using deep learning technique. Soft Comput. 2023;27(13):8971–90.

50. Rani P, et al. A decision support system for heart disease prediction based upon machine learning. J Reliab Intell Environ. 2021;7(3):263–75.
51. Gao X-Y, et al. Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. Complexity. 2021;2021:1–10.
52. Abdar M, et al. NE-nu-SVC: a new nested ensemble clinical decision support system for effective diagnosis of coronary artery disease. IEEE Access. 2019;7:167605–20.
53. Gabriel JJ, Anbarasi LJ. Optimizing coronary artery disease diagnosis: a heuristic approach using robust data preprocessing and automated hyperparameter tuning of eXtreme gradient boosting. IEEE Access. 2023;11:112988–3007.
54. Yang H, et al. Predicting coronary heart disease using an improved LightGBM model: performance analysis and comparison. IEEE Access. 2023;11:23366–80.
55. Wang J, et al. A stacking-based model for non-invasive detection of coronary heart disease. IEEE Access. 2020;8:37124–33.
56. Qadri AM, et al. Effective feature engineering technique for heart disease prediction with machine learning. IEEE Access. 2023;11:56214–24.
57. Yuan X, et al. A stable AI-based binary and multiple class heart disease prediction model for IoMT. IEEE Trans Ind Inform. 2021;18(3):2032–40.
58. Patra SC, Uma Maheswari B, Pati PB. Forecasting coronary heart disease risk with a 2-step hybrid ensemble learning method and forward feature selection algorithm. IEEE Access. 2023;11:136758–69.
59. Gupta A, et al. MIFH: a machine intelligence framework for heart disease diagnosis. IEEE access. 2019;8:14659–74.
60. Ishaq A, et al. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access. 2021;9:39707–16.
61. Ashri SEA, El-Gayar MM, El-Daydamony EM. HDPF: heart disease prediction framework based on hybrid classifiers and genetic algorithm. IEEE Access. 2021;9:146797–809.
62. Abdellatif A, et al. Improving the heart disease detection and patients' survival using supervised infinite feature selection and improved weighted random forest. IEEE Access. 2022;10:67363–72.
63. Ghorashi S, et al. Leveraging regression analysis to predict overlapping symptoms of cardiovascular diseases. IEEE Access. 2023;11:60254–6.
64. Abdellatif A, et al. An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods. IEEE Access. 2022;10:79974–85.
65. Cenitta D, VijayaArjunan R, Prema KV. Ischemic heart disease prediction using optimized squirrel search feature selection algorithm. IEEE Access. 2022;10:122995–3006.
66. Wang G, Lauri F, Hassani AHE. Feature selection by mRMR method for heart disease diagnosis. IEEE Access. 2022;10:100786–96.
67. Bhatt CM, et al. Effective heart disease prediction using machine learning techniques. Algorithms. 2023;16(2):88.
68. Khourdifi Y, Baha M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. Int J Intell Eng Syst. 2019;12(1):242–52.
69. Velusamy D, Ramasamy K. Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset. Comput Methods Programs Biomed. 2021;198: 105770.
70. Ali MM, et al. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Comput Biol Med. 2021;136:104672.
71. Kibria HB, Matin A. The severity prediction of the binary and multi-class cardiovascular disease—a machine learning-based fusion approach. Comput Biol Chem. 2022;98:107672.