International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I4P228

Volume 70 Issue 4, 316-331, April 2022

Review Article

Decision Support System for Reliable Prediction of Heart Disease using Machine Learning Techniques: An Exhaustive Survey and Future Directions

Deepali Yewale[1], S. P. Vijayaragavan[2], Mousami Munot[3]

[1]*Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Bharath University, Chennai, India*
Department of Electronics & Telecommunication, AISSMS Institute of Information Technology, Pune, India
[2]*Department of Electrical and Electronics Engineering, Bharath Institute of Higher Education and Research, Bharath University, Chennai, India*
[3]*Department of Electronics and Telecommunication, SCTR'S Pune Institute of Computer Technology, Pune, India*

[1]deepali_yewale@yahoo.co.in

Received: 11 March 2022 Revised: 19 April 2022 Accepted: 22 April 2022 Published: 26 April 2022

Abstract *- The Centers for Disease Control and Prevention statistics say 17.9 million people died from cardiovascular diseases (CVD), representing 32% of global deaths. This will increase and may reach 50% in 2050. CVD continue to be the prominent cause of mortality globally, making early detection of heart disease critical. Previously, knowledge-centred clinical decision support systems were created, which applied medical professionals' expertise and manually transferred data into computer systems. This procedure is time-consuming and is highly reliant on the judgment of a medical professional, which may be subjective. Machine learning (ML) algorithms have been applied to solve this problem by automatically gaining information from raw data. This study aims to thoroughly review the decision support system (DSS) using the ML approach for the CVD prediction for the University of California Irvin (UCI) dataset. Firstly, the exhaustive survey is carried out to understand and study the approaches adopted by different researchers. In the preceding sections, a few important aspects of heart disease study are discussed, including Risk factors of heart disease, Types of heart disease, ML approaches in the design of prediction systems, and optimization techniques for performance improvement. The surveyed papers are evaluated using different performance matrices. After that, I discovered the literature gaps and presented them in the comparative analysis section. This survey will assist investigators who wish to use ML or data mining approach in Heart disease projection.*

Keywords *- Cardiovascular disease, Heart disease prediction, Machine Learning, Decision support system*

1. Introduction

For the last two decades, heart disease has been the leading cause of mortality. According to The World Health Organization (WHO), heart disease has surpassed cancer as India's leading cause of death.[1] Concerning the 2016 Global Burden of Disease study, issued on September 18, 2017, 1.7 million Indians died in 2016 because of heart disease. Heart disease has become more prevalent, resulting in a reduction in human efficiency. According to the World Health Organization (WHO), India lost up to $237 billion due to heart and cardiovascular diseases between 2005 and 2015. [2] For the last two decades, CVD has continued to be the prominent cause of mortality globally, making early detection of heart disease critical. The heart is a crucial tissue in the human body that pumps blood to all the body's organs. If the heart fails to operate properly, the brain and the other organs will cease working, and people will die within minutes. Alterations in work-associated stress, lifestyle, and bad dietary habits lead to a growth in heart illnesses.

Heart disease is responsible for the majority of fatalities worldwide. [3] According to the study conducted by the WHO, approximately 56.7 million fatalities were reported globally in 2019, with around 17.9 million deaths due to cardiovascular disease, as illustrated in Fig. 1.

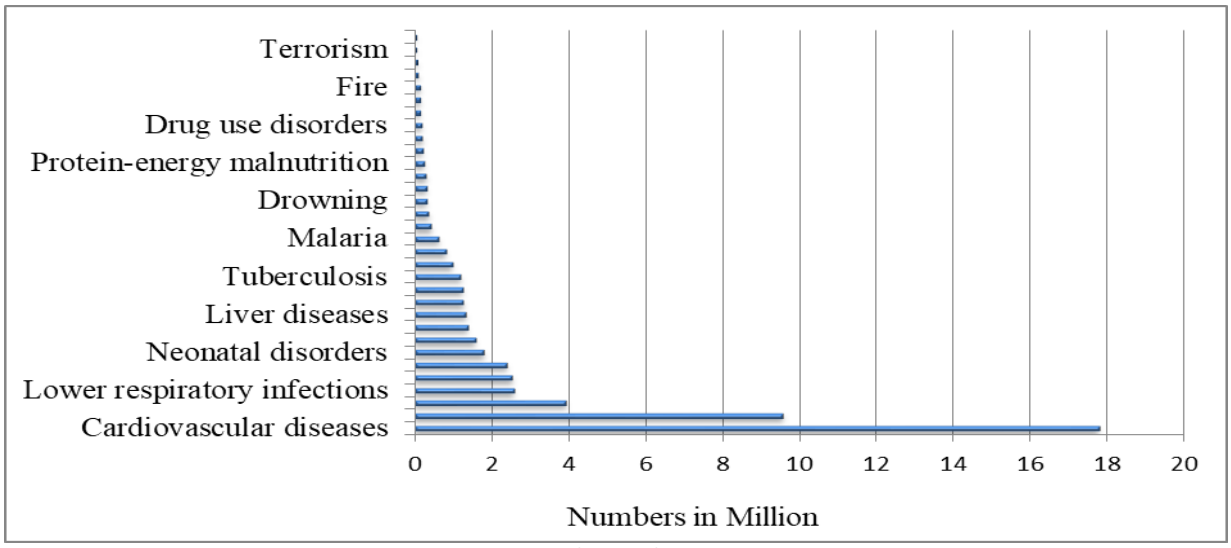

Fig. 1 Annual number of deaths for various reasons [4, 5]

Recently, computer technology and machine learning methods have been used to develop software to help doctors provide early-stage decisions about cardiac disease. Clinical and pathological data are used to diagnose cardiac disease. Medical practitioners may use the prediction system to forecast the patient's heart disease level established on their clinical information. In the biomedical sector, data mining is critical for illness prediction. When it comes to biological diagnosis, patients' information may contain repetitive and linked symptoms and indications, particularly if they have more than one kind of illness in the same category. It is possible that the doctors would not be able to identify it properly. Also, this procedure is time-consuming and highly reliant on a medical professional's judgment, which may be subjective.

Consequently, the ability to predict heart-related diseases reliably and accurately is essential.ML algorithms have been applied to solve this problem by automatically gaining information from raw data. When heart expert physicians are not accessible, an accurate automated system may predict a significant part in the initial phase of detection of heart disease in impoverished nations.

ML integrates predictive analysis, statistical analysis, and database technology to uncover connections and hidden patterns from large datasets. [6] ML methods have been applied in various medical settings, containing the calculation of surgical effectiveness, medical testing, medications, and the discovery of connections between diagnostic and clinical data. [7] The Medical diagnosis is a difficult procedure that requires accurate data from the patient, philosophical knowledge earned through the clinical training, and a thorough knowledge of the medical information. The healthcare industry gathers a lot of data, which is not mined enough to uncover concealed perceptions for improved decision-making. [8]

Clinical decisions often depend on physician views and skills more readily than data-rich truths hidden in records. [9] Regrettably, not every doctor is knowledgeable in the sub-speciality, and it is a limited resource in many areas. Individuals' information may contain duplicate and linked signs and symptoms in diagnosing medical issues, particularly when a person has more than one kind of illness in the same group. It is conceivable that physicians won't be able to identify it properly. [10] Due to complex interdependencies on various variables, correct illness detection at an early phase is difficult. [11] Medical organizations all around the globe gather data on a variety of health-related issues. Using a number of machine learning methods, this data may be utilized to obtain useful insights. The quantity of data collected is massive, and it is often noisy. ML techniques can rapidly examine datasets too big for human brains to comprehend. Consequently, these techniques are very useful in accurately predicting the presence or absence of heart-related diseases. [12]

2. Background

2.1 Heart Disease

Heart diseases are illnesses that should be considered a global health priority. Furthermore, cardiac disorders significantly strain patients, caregivers, and healthcare systems. Heart disease affects nearly 30 million people worldwide, with survival rates poorer than any other disease. Patients with heart disease are affected by high blood sugar, poor hygiene, cholesterol deposits, physical inactivity, SmokingSmoking, unhealthy diet, hand change smoking, high blood pressure, overweight, and viral infection. Despite the worldwide existence rate of 48.9 million, 50 individuals are born with cardiac abnormalities.

2.2 Types of Heart Diseases

Heart diseases are broadly categorized as coronary artery disease, myocardial infarction, heart failure, arrhythmia, heart valve disease and Cardiomyopathy. [13]

2.2.1. Coronary artery disease

Coronary artery disease is malaise introduced by the diminished circulation of blood. The improper blood supply in arteries may damage the vein and produce a problem with the heart's regular systolic and diastolic function.

2.2.2. Acute myocardial infarction

Acute myocardial infarction is a medical term for cardiac arrest. A cardiac arrest occurs when a lipid material in the blood alters the flow pace, causing tissue damage in the arteries. The blockage in the arteries may not be capable of supplying the oxygenated blood supply to the body. This may further result in the malfunction of other organs.

2.2.3. Heart failure

Heart failure can occur when the heart cannot pump adequate blood to the body organs due to stiffness in the heart muscles. The common symptoms of heart failure are shortness of breath and swelling.

2.2.4. Arrhythmia

Arrhythmia, also called dysrhythmia, is due to irregular rhythm of the heart caused when Heart's Electrical system does not function properly. There can be a Fast Heart rhythm or slow heart rhythm. Arrhythmia may be without symptoms, and if symptoms occur, they may include chest pain and dizziness.

2.2.5. Heart valve disease

Heart valve disease occurs when any of the four valves in the heart doesn't function properly. It may interrupt the flow and circulation of blood. Some of the symptoms of heart valve disease include the abnormal sound of the heartbeats and chest pain.

2.2.6. Cardiomyopathy

Cardiomyopathy is also recognized as heart muscle disease. People with this disease; have an enlarged heart that cannot pump blood away from the heart. Symptoms include breathing problems, swollen feet and a bloated stomach.

Fig. 2 illustrates the categorization of the heart diseases in one of the chosen surveys where an ML-based prediction system is implied to detect different types of Heart disease discussed in section 2.2. [14]

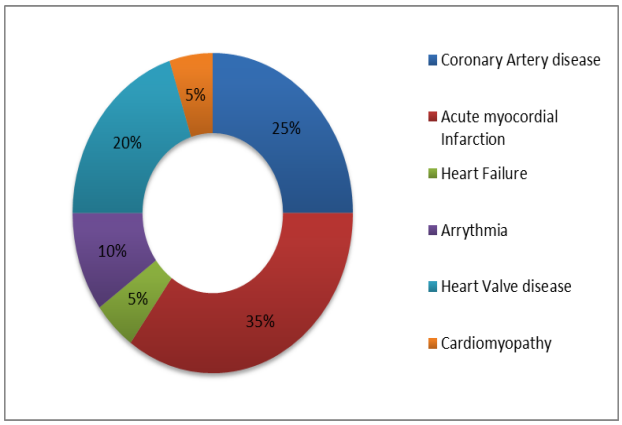

Fig. 2 Distribution of heart disease for the selected studies [14]

2.3 Risk Factors of Heart Diseases

Numerous considerations are intensifying the risk of acquits CVDs. Four primary CVDs incorporate coronary heart disease, transitory ischemic attack, peripheral arterial disease and aortic disease. There are certain features which affect the function or structure of the heart.

Fig. 3 shows the major controllable endanger factors of cardiac discomfort.

Fig. 3 Risk Factors of Heart Disease [15]

2.3.1. High Blood Pressure or hypertension

High Blood Pressure (BP) arises when the blood pressure escalates to alarmingly excessive levels. The quantity of blood flowing through the blood vessels and the level of the resistance blood meets when the heart is beating are considered when measuring BP. Narrowed arteries exacerbate atherosclerosis. BP will be greater if arteries are tightened.

2.3.2. High cholesterol

Cholesterol is a kind of lipid present in the body. It is a waxy, fat-like substance generated through the liver naturally. It is needed for cell membrane formation, hormone synthesis, and vitamin D absorption. Cholesterol cannot travel through circulation on its own since it does not dissolve in water. The liver produces lipoproteins to assist in the transport of cholesterol. Lipoproteins are particles made up of fat and protein. They transport cholesterol and triglycerides (another kind of fat) through your circulatory system. Lipoproteins are categorized into 2 categories: low-density lipoprotein and high-density lipoprotein. If ignored, high cholesterol may cause many fitness problems,

involving heart attack and stroke. The vast majority of the time, elevated cholesterol is asymptomatic. As a result, people should get their cholesterol levels checked regularly.

2.3.3. stress

A sense of mental or bodily strain is described as stress. Any incident or concept that creates people irritated, furious, or frightened may trigger it. The body's response to a challenge or demand is termed stress. Stress might be helpful in small doses, for example, when it facilitates your escape from danger or fulfilling a deadline. Stress may be harmful to health if it lasts for a long period**.**

- Acute stress-This is a temporary stressor that will pass soon. It helps in trading through potentially hazardous circumstances. It may also occur if you try something different or interesting. At some point in their lives, everyone experiences severe stress.

- Chronic stress is a kind of tension that lasts a long time. If people have money difficulties, an unpleasant marriage, or job problems, they may be suffering from chronic stress. Chronic stress is defined as stress that lasts for weeks or months.

2.3.4. Smoking Smoking

Smoking is the act of breathing a substance that has been scorched and then tasted and submerged into the circulation. The extremely frequent substance applied is the tobacco leaves enclosed into a small rectangle of rolling paper to form a tiny, spherical chamber called a cigarette. The burning of dried plant leaves conveys and disintegrates dynamic synthetic compounds into the lungs, where they are quickly immersed in the circulation and reach body tissue. Smoking is a major risk factor for cardiac arrest since it causes the patient's blood enzymes to coagulate and raise the chance of heart failure. Synthetic compounds in tobacco smoke make the blood thicken and shape clumps inside veins and supply routes. Blockage from coagulation can prompt cardiovascular failure and unexpected demise.

2.3.5. Physical inactivity

Physical inactivity is a phrase used to describe individuals who do not regularly engage in the required amount of physical exercise. Physically inactive refers to a lack of physical exercise. On the other hand, sedentary refers to extended sitting or lying down periods.

2.3.6. Diabetes Mellitus

Diabetes is a situation in which blood glucose, usually recognized as blood sugar, is too high. The primary energy supplier is blood glucose, which occurs through the nutrition you consume. Insulin, a hormone generated through the pancreas, aids glucose assimilation into cells for usage as strength. Sometimes your body does not generate enough or any insulin, or it does not utilize it correctly. Glucose continues in your bloodstream and does not reach your cells as a consequence. Having more glucose in the blood might reason to health problems. Although there is no cure for diabetes, people take measures to manage it and continue strong. Diabetes is frequently described as "a touch of sugar" or "borderline diabetes". These phrases indicate that someone does not have diabetes or a weaker form of the disease, although diabetes disturbs everyone.

2.3.7. Obesity

Obesity is a prevalent and preventable illness that affects clinical and public health. It is often a substantial risk factor for the onset of various non-communicable illnesses, severe disability, and early mortality. Obesity is now a worldwide pandemic affecting people of all ages. Drug treatment is designated for fat or overweight individuals with obesity-related risk factors or illnesses. Compared to clinic-based weight-loss treatments, population-wide preventive initiatives have a better chance of halting the obesity epidemic and are more cost-effective.

2.4 ML Techniques

ML is a multidisciplinary field with origins in statistics, algebra, data processing, skill analytics, and other fields, making it difficult to develop a new definition.ML is an artificial intelligence technique that gathers information from training data. Types of Machine learning algorithms are discussed below.

2.4.1. Supervised Machine Learning

The development of algorithms that can generate common trends and assumptions by applying superficially supplied cases to predict the outcome of future cases is recognized as supervised ML. The goal of supervised ML classification algorithms is to classify information centred on the data that came before it. This information is used to train the algorithm, which creates a practice that assigns inputs to associated outputs. A classification issue is a popular form of the supervised learning task. In the presence of uncertainty, the goal of supervised ML is to produce a prototype that creates projections centred on proof. A supervised learning algorithm instructs a prototype to create realistic projections to react to additional information by applying a recognized set of input data and recognized replies to the data.

Fig. 4 below shows the projection of heart disease utilizing the supervised ML approach. [16]

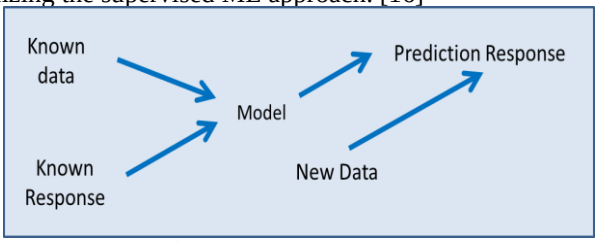

Fig. 4 Framework for heart disease prediction using supervised Machine Learning

2.4.2. Semi-Supervised Machine Learning

It is a method for determining the best classifier from a set of uncategorized and categorized data. It transfers high classification performance by using unlabeled data. The method's success is entirely dependent on a few underlying assumptions. Semi-supervised learning may be a viable option for resolving the problem of cancer gene identification in recurrence.

2.4.3. Unsupervised machine learning

Unsupervised machine learning techniques are provided with unlabeled data. This algorithm will learn and discover hidden patterns from data. [17] The ability to explore the out-of-the-box abilities of created unsupervised learning approaches through slight human involvement throughout the procedure is enabled by minimal data processing upfront. The methods' output is applied to forecast survival time, with the prediction's efficacy serving as a proxy for the classification's usefulness.

2.4.4. Reinforcement learning

A computer program grants admittance to a dynamic environment to accomplish a particular goal in this kind of learning. Because it navigates its drawback, the programme receives feedback on rewards and punishments. [18] Tumour localization is an example of reinforcement learning in diagnosis and treatment. In particular, as a critical step in analyzing heart disease, accurate localization of the infected area can improve the surgical procedure and lower the recurrence rate.

2.5 Different Types of Classifiers in ML

2.5.1. Support Vector Machine (SVM)

SVM is the greatest widely used and recognized supervised algorithm for classification problems. The SVM is based on a simple line equation, y = mx + c, which allows the linear division domain manipulation. [19] SVM algorithms are split into two categories: linear and nonlinear models. [20] The linear SVM simply divides the data linearly to segregate the original data. The nonlinear SVM is used when the data is non-linearly separable and converted into a feature space where the data domain can be broken linearly to differentiate the groups. The basic diagram of the SVM algorithms is shown in Fig. 5, in which the plane is divided into two hyperplanes, the positive hyperplane and the negative hyperplane. SVM was applied to analyze gene expression to diagnose prostate cancer in its initial stages and choose genes and classify them. A predictive model is used for tasks requiring the projection of individual values applying more values in the dataset.

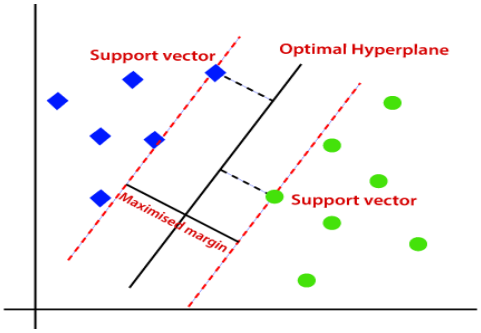

Fig. 5 Support Vector Machine

2.5.2. K nearest neighbour (KNN)

The KNN technique assumes that the existing and new cases are similar. It places the new case in the category with the greatest similarity to the remaining classifications. [21] The K-NN algorithm stores the accessible information and categorizes a new data point centred on its relationship. As new data becomes available, the K-Nearest Neighbor algorithm can be used to group them into a proper suite group. The K-NN algorithm could be applied in conjunction with Classification for Degeneration; however, it is primarily applied to solve categorization challenges. The K-NN algorithm is a non-parametric method, implying it does not consider the underlying data. It is known as the lazy learner technique because it doesn't store the data set and instead learns with the help of the training data set. It only acts on the data set when it's time to classify it. A new input data set is presented to the KNN algorithm, kept in a remarkably comparable classification to the new data. KNN is a popular and simple multi-variable classification technique that has been efficiently employed in various health functions, including cancer detection. KNN allocates an undetermined issue to be classified based on the K determining issues which are closest to them.

2.5.3. Naïve Bayes

The NB classifier is the straightforward probabilistic classifier that counts on the Bayes theorem. Each attribute variable is treated as an autonomous variable by Naïve Bayes. This classifier could be trained in supervised learning and used in complicated real-world circumstances. The main benefit of the Naïve Bayes is that it only involves a bit of instruction information, which is critical for characterization and classification. The naïve Bayes algorithm is the popular data mining system for classification. It depends on the probability of some class, assuming that the attributes are not dependent on each other of the given class. This assumption inspires the requirement to assess the multivariate probabilities by the training data. In practice, most attribute value combinations are either not present or not present in adequate amounts in the training results. [22] Consequently, it would not be accurate to estimate each related multi-variate likelihood explicitly. By its conditional freedom assumption, Naïve Bayes circumvents this predicament. Despite this strict presumption of freedom, in many real-world functions,

naïve Bayes is a very skilled classifier. The basic equation of the Naïve Bayes algorithm is demonstrated under **Equation 1**.

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)} \quad (1)$$

Cancer is classified using the Naive Bayes algorithm. The recommended algorithm's evaluation results showed that it was 98% accurate in predicting gastric cancer and 90% accurate in predicting heart disease. [23] The Nave Bayes algorithm is applied for cancer prediction, centred on a Gaussian distribution.

2.5.4. Random forest (RF)

RF is a widely used ML technique applied in ML algorithms to solve classification and regression problems. It is established on the principle of collective learning, the technique for merging multiple classifiers to resolve various complicated problems and enhance the model's accuracy. A basic diagram of the Random Forest algorithms is shown in Fig. 6. The random forest is the classifier that consists of different decision trees on subsets of datasets. To enhance the efficiency, it assumes the average of the different accuracies of the dataset. Random forest is a big data classification algorithm that uses one of many classification techniques. Cancer microarray data is subjected to random forest classification to achieve a more precise and solid classification performance. The random forest has a 100% accuracy rate in cancer classification. [24]

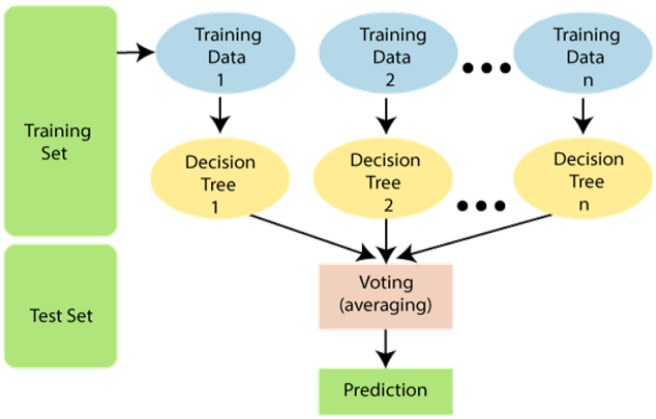

Fig. 6 Random Forest Algorithms

2.5.5. Neural Network (NN)

An NN is a collection of techniques that work on a mechanism that simulates how the human brain works, intending to detect hidden relationships in a data set. The NN stands for "naturally occurring artificial neurons". Fig. 7 depicts the Neural Network's basic structure: the hidden layer, the output layer, and the input layer. The number of hidden layers in a complex structure is increasing. [25] Because they eliminate the need for invasive procedures and the interpretation of imaging results, neural networking tools might be applied to diagnose cancer more effortlessly and essentially than conventional techniques.

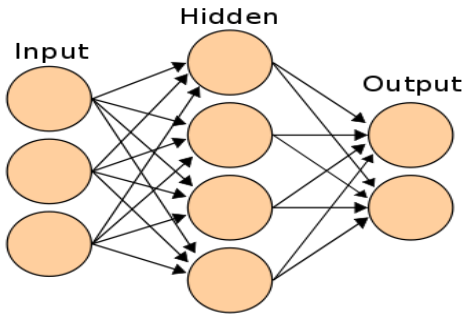

Fig. 7 Simple Neural Network

2.5.6. Decision tree (DT)

DT is a classification approach comprising an interior node and one terminal node through the class label. The root nodes of DT are the nodes at the very top. The decision tree is widespread as it is easy to construct and does not require any parameters. [26] A decision tree aids disease examination by presenting a doctor's perception of health problems or revealing background information about specific patients. It also assists in detecting a patient's condition. It provides a model that advises the patient on when to utilize the appropriate medication at a particular time, thanks to a web-centred organization linked to a computerized patient/therapeutic record.

As shown in Fig. 8, the Distribution of ML techniques and their corresponding heart disease.

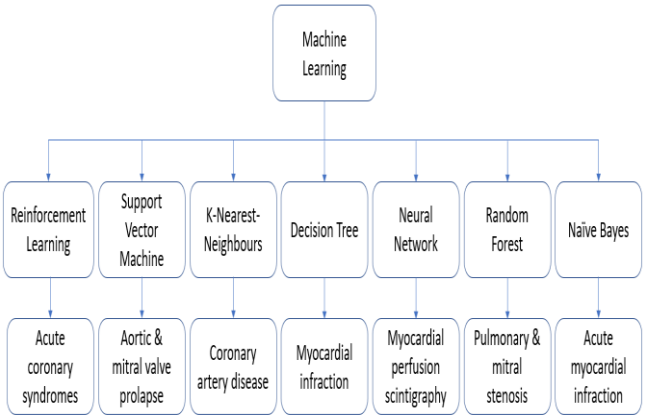

Fig. 8 Distribution of ML fields corresponding to heart diseases [13]

2.6 Categorization of ML Techniques According to Studies

This subsection presented the categorization of ML-based DSS for heart disease prediction in the selected study of the review paper. [14] In supervised algorithms, output datasets are provided to instruct the machine and obtain the required output. In contrast, unsupervised algorithms do not have output datasets, and the data is instead grouped into various classes. Evolutionary techniques, which may be employed both supervised and unsupervised, fall into a third

group. Evolutionary algorithms (EA) are inspired by biological mechanisms and implemented for optimization in ML. In 59 per cent of the research, supervised machine learning algorithms were utilized, while unsupervised ML algorithms were applied in 35 per cent, and evolutionary algorithms were used in 6 per cent. To address the issue of heart disease prediction DSS, most of the researchers used supervised and unsupervised ML Techniques. Fig. 9 illustrates the categorization of ML methods according to scholarly work. [14] There is scope to work on EA and verify the results in the design of a heart disease prediction system. With further research, it is anticipated to actualize EA to optimize the prediction model.

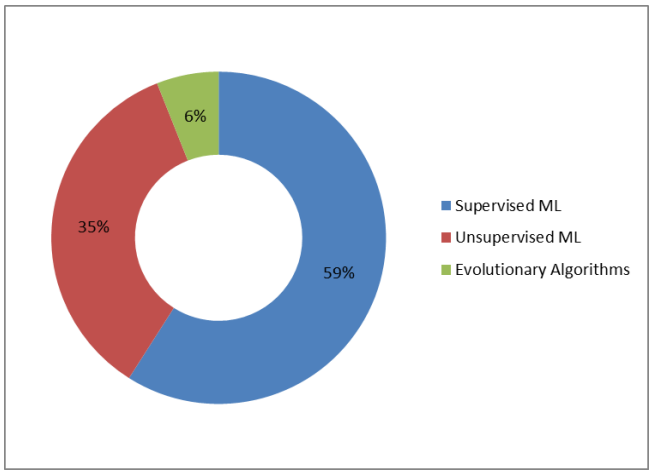

Fig. 9 Major classification of ML algorithms [14]

2.7 Basic Framework Applied for Heart Disease Prediction

This segment explains heart disease datasets, pre-processing methods, element decline/assortment strategies, information division and authentication procedures, categorization methods, and execution assessment measures in short. The researchers practised various datasets such as Framingham, UCI, Statlog, z-alizadehsani, and MIT-BIH Arrhythmia. For cardiac disease prediction utilizing electrocardiogram (ECG) signals, several researchers utilized the MIT-BIH Arrhythmia dataset from Physio Net. [27] But sometimes, a short period of ECG signal recording failed to slow up symptoms of heart disease. In such a scenario, a longer recording and monitoring period of more than 24 hrs may help in the prognosis of abnormality in the heart. Cardiologists' examination of ECG graphs is complex, time-consuming, and more error-prone. Only a few researchers have utilized data from private hospitals in their regions to conduct their studies. But the focus of this review paper is on the UCI dataset as the dataset is rich in information and is an easily available open source dataset. The UCI dataset is utilized in most trials since it has fewer missing values and is a stable dataset.

2.7.1. Datasets

The majority of the researchers apply the UCI database from the ML repository. This directory includes 4 databases through the Cleveland Clinic, the Hungarian Institute of Cardiology, the VA Medical Centre, and the University Hospital of Zurich in Switzerland. Table 1 summarises the number of occurrences found across all databases. [28]

Table 1. Review of database

Sr. No.	Database	Number of instances
1	Hungarian	294
2	Cleveland	303
3	Long Beach VA	200
4	Switzerland	123

The above-stated entire database has 76 raw attributes, but only 14 attributes are applied by most of the investigators relevant to predicting heart disease. These attributes comprise almost all clinical as well as medical data required for heart disease prediction and are summarised in Table 2 . [28]

Only 6 values are missing out of 303 instances from the Cleveland dataset, leaving 297 examples for experimentation with Cardiac disease. Practitioners used either 303 instances by imputing missing values or 297 instances by removing missing values from the dataset in their experiment.

Table 2. List of 14 Attributes

Sr. No	Name	Number in the Actual database
1.	Diagnosis value	58
2.	Thalassemia	51
3.	Number of vessels coloured by fluoroscopy	44
4.	The slope of the ST segment	41
5.	ST depression	40
6.	Exercise-induced angina	38
7.	Maximum Heart Rate	32
8.	Resting ECG	19
9.	Fasting blood sugar	16
10.	Serum Cholesterol	12
11.	Resting Blood Pressure	10
12.	Chest pain type	9
13.	Sex	4
14.	Age in years	3

2.7.2. Basic Framework

This section explains the model to forecast heart disease. The model's architecture is shown in Fig. 10. The following are the stages required in the design of the heart disease prediction model:

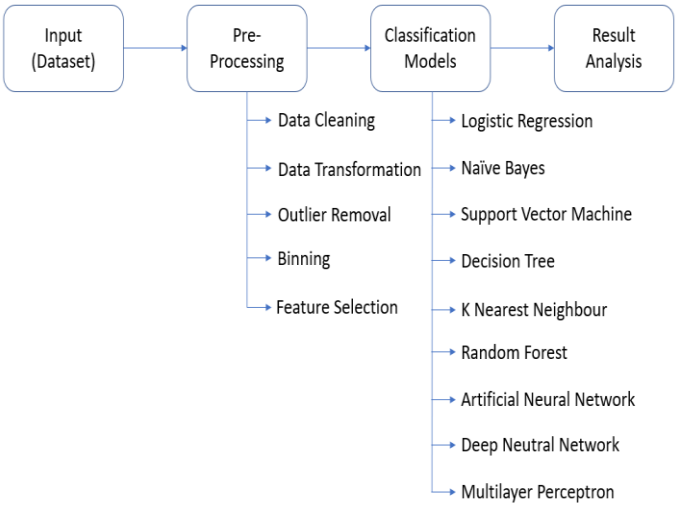

Fig. 10 Basic Framework

Input- Extract the necessary data in the first step from the UCI repository, a well-known repository or website.
Pre-processing- The primary step after obtaining the dataset is to do data pre-processing. Perform different pre-processing methods in this phase, such as data cleaning, Transformation, managing missing values, attribute selection, etc.

Classification models- The dataset is subjected to several ML methods following the feature extraction.
Result analysis- Compare the findings of the different algorithms in this phase. Using the accuracy metrics produced, determine which technique is the greatest for forecasting heart disease.

3. Literature Review

Cardiac disease can be detected via early-stage symptoms, challenging in today's culture. If not identified in a timely way, this may cause death. This segment recommends a literature review of computational techniques utilized by previous researchers in the domain under discussion. Discussed the effects and benefits of the various methods researchers employ in this area.

The researchers recommended a hybrid decision support system to aid in the early recognition of heart illness based on the patient's clinical indications. Rani et al. [29] opted for a multivariate imputation method to cope with missing data. A feature selection hybrid method linking recursive feature removal and the evolutionary algorithm is used to pick acceptable features from the given dataset. Data pre-processing techniques included the Synthetic Minority Oversampling Technique to treat data imbalance. The random forest classifier was found to give the best accurate results for the system compared to SVM, NB, LR and adaboost in a Python-based simulation environment. It was tested on the UCI's heart disease dataset, freely available in the UCI ML repository. The model has an accuracy of 86.6 per cent and a sensitivity of 84.14 per cent, which is better for predicting cardiac disease. Still, in future, any advanced method can be used to enhance the performance of the prediction system.

A combination of Image fusion and ML has been recommended in the medical sector, specifically for predicting heart disease. More sophisticated feature selection methods enhance algorithm accuracy and provide consistent results. Essentially, Diwakar et al. [30] conclude that correct and suitable datasets are utilized to build a heart disease projection prototype. The UCI dataset should be pre-processed, as this is the highly essential stage in formulating the dataset when utilizing the ML method and achieving better outcomes. A suitable method must be used to build a prediction model that yields trustworthy results. The authors discovered that Artificial Neural networks (ANN) had excellent effects on heart disease projection. Utilizing image fusion and ML to identify cardiac disease proved to be an essential activity that may help patients and healthcare professionals.

Katarya et al. [31] addressed LR, KNN, NB, SVM, DT, RF, ANN, Deep Neural Network (DNN) and Multi-Layer Perceptron (MLP) to predict heart disease. They offered a relative evaluation of the methods used in the forecast test. ML approach was efficient for data projection and management of huge data collected from the medical sector. Data normalization and Transformation are the initial stages applied for data management. NAN in the python library replaces the missing values present in the dataset. The random forest has the highest accuracy of 95.60% and sensitivity of 97.68 % for the UCI dataset. In the future, another advanced method can be implemented to outperform the model.

Recent research has used a feature selection approach to improve the performance of DSS in the prediction of heart disease. Prakash et al. [32] introduced the Optimality Criterion Feature Selection (OCFS) technique for extrapolating heart disease and successfully diagnosing heart disease. The researchers proposed to use only selected features using the OCFS technique on the UCI dataset. The relation between different attributes and their contribution toward disease prediction is determined to prepare a decision table. OCFS eliminates redundant features and obtains the minimum subset of features. The author discovered that the OCFS method takes the minimum time to implement than information entropy-based feature selection using rough set theory (RFS-IE) and Multivariate

Reconstructed Phase Space (MRPS). According to experimental results, for 10 patient data, OCFS took 8.31 ms, while RFS-IE and MRPS took excess time of 1.58 ms and 3.34 ms individually to compute disease prediction.

Ali et al. [33] presented a feature fusion and deep learning-based healthcare system. The feature fusion technique is used to produce relevant healthcare information. The information acquisition technique eliminates inappropriate and superfluous components while concentrating only on the most important features. This helps in decreasing computational load and increasing system performance. Additionally, the conditional probability method calculates a distinctive feature weight for every class, significantly improving system execution. Eventually, the ensemble deep learning model is prepared to forecast heart disease. The recommended approach is compared to conventional classifiers centred on feature selection, feature fusion, and allowance methods. The recommended system's accuracy for the UCI dataset (with 70% training dataset and 30% test dataset) is 98.5 per cent and sensitivity 96.6 per cent, which is greater than that of current techniques. This finding indicates that, compared to other cutting-edge technology, the approach is more successful at predicting cardiac disease.

Fitriyani et al. [34] proposed using outlier removal and data sampling techniques to balance the training dataset. The outliers are different from other observations in the dataset and may cause statistical analysis problems. Hence outliers are required to be removed. Imbalanced data must also be handled for a classification model to get a higher accuracy rate. The authors discovered that the suggested model obtains an enhanced accuracy of approximately 98.40 per cent and sensitivity of 98.33 per cent with XGBoost – based ML classifier on the UCI dataset. The presented model is compared with other basic classification models and prior studies for evaluation. It is found that the proposed approach outperformed all six-performance metrics accuracy, precision, sensitivity, F-measure, MCC, and AUC.

Garate-Escamila et al. [35] recommended the dimensionality reduction technique and utilized a feature selection plan to discover heart disease characteristics. This research proposal implements two diverse dimensionality reduction methods, Chi-square (CHI) for selecting the important features and principal component analysis (PCA) for extracting the features. Six ML classifiers, DT, LR, NB, RF, MLP, and Gradient Boosted Tree (GBT), are proposed with default values of their hyperparameters. Using CHI-PCA with the random forests, the maximum accuracy achieved was 98.7 per cent for the Cleveland dataset (with 70% training dataset and 30% test dataset). Experiment findings indicate that integrating chi-square with PCA enhances the performance of most classifiers. The proposed approach is validated for three Cleveland, Hungarian, and Cleveland-Hungarian datasets.

Before building a predictive model, selecting relevant features from the dataset is essential; various researchers have implemented filtering, wrapper and hybrid methods to select a set of features. Filter methods are fast to compute, but they fail to consider feature dependency with other features. Wrapper methods are classifier dependent and not reliable with a large set of features. Embedded methods combine the advantages of both the filter and wrapper methods, but the reduced set of features obtained may be problematic for model building. Hasan et al. [36] proposed a multi-stage feature selection technique combining all the three feature selection approaches and a feature importance algorithm. Finally, the feature sub-set is extricated to fulfil the "True" condition criteria. The clinical data (with 70% training dataset and 30% test dataset) of 70,000 patients with 12 features available at the Kaggle data repository is used for experimentation. Data pre-processing and outlier removal is performed. The new 13th feature, Body Mass Index, is calculated and added to the dataset. RF, SVM, NB, KNN and XGBoost create the Classification model. It is revealed that XGBoost with k fold cross-validation gives the best accuracy compared to other classifiers.

Various ensembles techniques are implemented in the state of the art research to improve the classification accuracy of weak classifiers. Latha et al. [37] proposed using ensemble techniques bagging, boosting, stacking, majority voting, and feature selection to amplify the classification results. Cleveland dataset from UCI is used for the experiment.NB, Bayes net, DT, MLP, and NN are classifiers used in the ensemble. The results of the experiment prove that ensemble is a good technique to improve the performance of the weak classifier. Majority voting proved to be the best ensemble approach with the highest accuracy. Further enhancement in accuracy is achieved with feature selection. The proposed method has provided an accuracy of 85.48 %.

Motivated from the previous research with ensemble technology, an improved randomized tree-based ensemble is developed by Mienye et al. [38] for the prediction model using Cleveland and Framingham heart disease dataset (with 70% training dataset and 30% test dataset). The whole dataset is randomly partitioned into smaller subsets. Each subset is treated by CART to create a model. Depending upon the accuracy of each CART model, a weightage based homogeneous ensemble classifier is formed. Test results on Cleveland and Framingham datasets accomplished 93% and 91% accuracy, outflanking other ML calculations and comparative learning work. Further ROC characteristic is added to validate the performance of the proposed approach.

Reliable data is the need of any medical diagnosis system. To get reliable data, Rahim et al. [39] designed an integrated four-phase framework for heart disease prediction. The first phase of the framework is handling missing values using the mean replacement approach. The second phase is to deal with imbalanced data using SMOTE. In the next phase, feature importance is used to select features. Finally, the Ensemble of LR and KNN classifiers is proposed as a prediction model. The proposed approach is validated on three datasets, Framingham, Heart Disease and Cleveland, with accuracies of 99.1%, 98.0% and 95.5 %, respectively.

To fit the ML classifier for any prediction system, the hyperparameters of the classifier need to be tuned appropriately. [40] Selecting accurate hyper_parametrs of the classifier is a challenging task as it will significantly impact the performance of the developed system. Emrana et al. [41] opted for a grid search approach for hyperparameter tuning and found a near 7 per cent rise in accuracy compared to traditional systems. The highest accuracy obtained with the proposed system is 91.80 per cent for Cleveland dataset. The author recommended using a feature selection approach with hyperparameter tuning to enhance the performance of the proposed system. Javeed et al. [42] proposed a feature selection learning system with a random search algorithm and RF optimized by grid search. The author claimed the model to be less complex than the traditional RF model. Only 7 subsets of 13 features are selected from the Cleveland dataset using the proposed approach. The proposed hybrid diagnostic system improved the accuracy of the RF model by 3.3 per cent and achieved 93.33 per cent accuracy. The problem of overfitting is also well handled by the said approach with good performance on both training and testing data.

Table 3 describes the summary of the review of the literature.

Table 3. Brief of Literature review

Author(s)	Algorithm	Description
Rani et al., 2021[29]	Random forest	A hybrid feature selection method combining recursive feature removal and a genetic algorithm is employed in the pre-processing phase.
Diwakar et al., 2021[30]	ANN	More clinics and hospitals are being inspired to disclose datasets that generate excellent outcomes and enhance their models that assist patients in curing heart disease in its initial phases.
Katarya and Meena 2020[31]	SVM	This research aims to utilize ML to predict heart illness and examine the findings.
Prakash et al., 2019 [32]	Optimality Criterion Feature selection algorithm	To extract the important features from the set of features for predicting heart disease.
Ali et al., 2020[33]	Ensemble Deep Learning Model	Suggested a health care structure centred on feature fusion and ensemble deep learning.
Fitriyani et al., 2020 [34]	XGBoost	Presented a successful heart disease prediction prototype for a DSS by detecting and removing outliers, added algorithm to balance training data distribution and the XGBoost to forecast heart disease.
Garate-Escamila et al., 2020 [35]	CHI-PCA with the Random Forest	Discovered that combining the Chi-square with the Principal Component Analysis (CHI-PCA) with the Random Forest yields more accuracy than previous methods.

4. Comparative Study

This research review paper focuses on the UCI dataset, so here in this section, a comparative analysis is done considering scholarly work for the same dataset.

The usage of feature selection methods boosted the forecasting of heart disease. Dun et al. [43] investigated the incidence of heart disease using deep learning methods, LR, SVM, and RF with feature selection and hyperparameter. The most accurate algorithm was NN, which had a score of 78.3 per cent. Medhekar et al. [44] used the NB algorithm to predict the heart disease and observed that the model achieved an accuracy of 89.58 per cent when considering 240 records in the testing dataset. The authors used 14 features of the UCI dataset. Wiharto et al. [45] applied different types of the SVM techniques like Binary Tree Support Vector Machine (BTSVM), Exhaustive Output Error Correction Code (EOECC), and Decision Direct Acyclic Graph (DDAG) to the UCI dataset.

Furthermore, the author first pre-processed the dataset with the help of the min-max scalar. The author observed that the BTSVM gives better results than other techniques, with an overall efficiency of 61.86 per cent. To the UCI dataset, Khateeb and Muhammad [46] applied various ML classification techniques like NB, bagging, DT, and KNN. The authors considered six different cases; in the first case, the author applied algorithms without feature reduction. In the second case, the author used the feature reduction techniques with considering the seven different attributes etc. The accuracy of the proposed approach is near 80 per cent. Chen et al. [47] applied the ANN algorithm to the 13 attributes of the UCI dataset and achieved an accuracy of 80 per cent.

Previous research focused on a 13-feature heart disease subgroup. Among the remarkable study, Sen [48] proposed implementing four classifiers, NB SVM, DT and KNN, to the UCI dataset with 13 attributes and achieved an accuracy of 84.51 per cent with SVM. SVM algorithm is proved to be efficient compared to the other three algorithms. Pouriyeh et al. [49] implemented an ensemble approach using NB, DT, SVM, MLP, KNN, Radial Basis Function (RBF), Single Conjunctive Rule Learner (SCRL) classifiers. Compared to the seven techniques mentioned above, SVM with Boosting is found to be more efficient. SVM with Boosting provided an accuracy of 84.81 per cent. Singh et al. [50] implemented supervised machine learning techniques on labelled Cleveland datasets, namely linear regression, logistic regression, SVM, DT and RF. Random forest works well with the nonlinear behaviour of the dataset. A manual search is carried out to select the parameters of the RF classifier. The highest accuracy is 85.81 per cent with an RF classifier when 20 numbers of split and 75 trees are selected from various trial and error sequences. Das et al. [51] implemented an ensemble-based neural network for the Cleveland dataset in SAS based software. Many neural networks are combined in an ensemble neural network (NN) by considering the previous model's probable target or predicted target. The author proposed to use three neural networks in the ensemble NN model. The proposed approach achieved a classification accuracy of 89.01 per cent.

Table 4 depicts the comparative analysis of various traditional ML techniques. Significant technique still needs to be implemented to reach a performance comparable to a medical professional.

Table 4. Comparative analysis of ML techniques

Author(s)	Algorithms	Dataset	Accuracy
Dun et al. 2016 [43]	Neural Network	UCI	78.3%
Medhekar et al. 2013 [44]	Naïve Bayes	UCI	89.58%
Wiharto et al.2015[45]	BTSVM	UCI	61.86%
Khateeb and Muhammad 2017 [46]	KNN	UCI	80%
Chen et al.2011 [47]	ANN	UCI	80%
Sen 2017 [48]	SVM	UCI	84.51%
Pouriyeh et al. 2017 [49]	Boosting	UCI	84.81%
Singh et al. 2017 [50]	Random Forest	UCI	85.81%
Das et al. 2009 [51]	Ensemble ANN	UCI	89.01%

Furthermore, some authors developed a hybrid approach to improve the model's accuracy. Ali et al. [33] proposed a medical structure centred on feature fusion and ensemble deep learning methods, with 98.5 per cent accuracy on the dataset of UCI. Fitriyani et al. [34] presented a model for the projection of heart disease that incorporates Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to eliminate and recognize anomalies, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to counterbalance the training data distribution, and utilized XGBoost to foresee heart disease. On the UCI dataset, the authors accomplished an efficiency of 98.4 per

cent. Garate-Escamila et al. [35] discovered that combining the Chi-square and Principal Component Analysis (CHI-PCA) by RF yields better results than other methods. The model has an efficiency of 98.7 per cent on the UCI dataset, which is the greatest accuracy when contrasted to other methods. Amma [52] proposed to use the most impressive hybrid model, i.e., the combination of a genetic algorithm and the neural network. The model achieves an accuracy of 94.2 per cent. Santhanam et al. [53] implemented feature selection using PCA component analysis and regression. With 92.0 per cent and 95.2 per cent accuracy of PCA regression and PCA-NN produced the most accurate PCA models. The comparison of the different hybrid models is depicted in Table 5. It is observed that the hybrid techniques give more accurate results than the traditional techniques.

Table 5. Comparative analysis based on the hybrid approach

Author(s)	Algorithms	Dataset	Accuracy
Ali et al. [33]	Feature fusion + deep ensemble learning	UCI	98.5%
Fitrivani et al. [34]	SMOTE-ENN + XGBoost	UCI	98.4%
Garate-Escamila et al.[35]	CHI-PCA + Random Forest	UCI	98.7%
Amma [52]	Genetic algorithm + Neural Network	UCI	94.2%
Santhanam et al. [53]	PCA-NN	UCI	95.2%

In the new era of the ML learning approach, researchers focused on tuning the hyperparameter of ML models. [40] Hyperparameters are the controlling knobs of the ML classifier. Manual search requires knowledge and expertise in the domain. If hyper-parameters are complex, they will be time-consuming and may cause a human error in interpreting correct hyper-parameters. To overcome the imperfection in manual search, automated hyper-parameter tuning algorithms such as Grid search and Random search, Bayesian optimization, and Genetic algorithm have been proposed. Kartik Budholiya et al. [54] proposed to optimize XGBoost hyperparameters, implementing the Bayesian optimization technique. XGBoost optimized using BO revealed an accuracy of 91.8% in predicting heart disease using the Cleveland dataset. Patro et al. [55] proposed DSS to predict heart disease using Bayesian optimized SVM. The results reveal that BO SVM outperformed with 93.3 per cent accuracy. R. Valamathi et al. [56] designed a framework for heart disease prediction with grid search, Randomized search and genetic optimization for RF and XGBoost classifiers. The prediction accuracy of the RF classifier is enhanced to 97.52 per cent with genetic (TPOT classifier).

5. Result Analysis

This segment contains the portion of result analysis, in which the information is evaluated on various algorithms depending on accuracy. Overall models are run on the dataset of UCI.

Table 6 presents the accuracy results of the hyper-parameter tuning algorithm for the classifier in predicting heart disease as per the survey carried out for this review paper.

Table 6. Accuracy of Hyper parameter tuning algorithm in the literature

Hyper-Parameter optimization Technique with Base Classifier	Accuracy
Grid Search with KNN [41]	91.80%
Grid Search + RF [42] (Features selection with RSA)	93.33%
Bayesian Optimization + XGBoost [54]	91.80%
Bayesian Optimization + SVM [55]	93.3 %
TPOT +RF	97.52 %

Table 7 presents the accuracy results of the 8 approaches of the various models. The ANN, Logistic regression and SVM all projected similar accuracy. Moreover, Random Forests outperforms other methods in terms of accuracy, but the fusion methodology CHI-PCA by Random Forest approach outperforms them all.

Table 7. Comparison of ML algorithms based on accuracy [28]

Algorithm	Accuracy
KNN	71.42%
DNN	76.92%
Decision Tree	81.31%
Naïve Bayes	90.10%
SVM	92.30%
ANN	92.30%
Logistic Regression	93.40%
Random forest	95.60%
CHI-PCA with Random Forest	98.7%

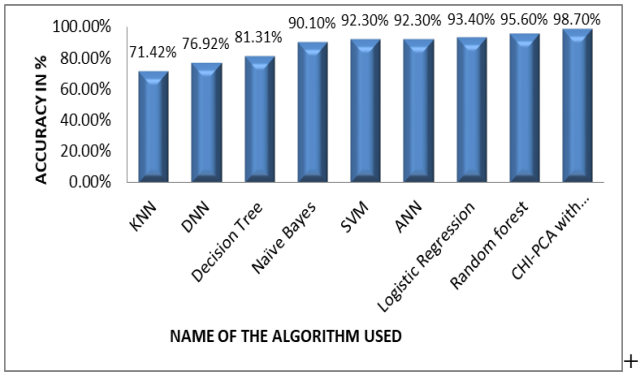

Fig. 11 Evaluation of various algorithms concerning the accuracy

The hybrid algorithm (CHI-PCA with Random Forest) has the highest accuracy of any algorithm at 98.70 per cent. In terms of accuracy, the hybrid algorithm beats all other algorithms, as shown in Fig. 11. The hybrid algorithm's superior performance is due to its capacity to generate many decision trees and work well on high-dimensional data.

Table 8 highlights the numerous advantages and disadvantages of the various approaches used to address the issue of heart disease projection.

Table 8. Comparison of ML algorithms from the perspective of heart disease prediction

Algorithms	Advantages	Disadvantages
Naïve Bayes	The conditional probabilities are easy to evaluate	It works better if all the features are independent. But in real life, many attributes are dependent on each other.
Logistic regression	Precise, Quick and less computation time.	Exceptionally superior computing energy utilization.
SVM	Great inaccuracy	For big datasets, it is quite slow; it consumes a lot of computation energy.
Decision Tree	Good accuracy	Require high calculation energy slow for larger datasets.
KNN	Quick and less computation time	Data requires a significant quantity of resulting memory.
ANN	Great accuracy	The datasets and epoch selected by the user affect the computation speed and accuracy.
Random forest	Great accuracy	Gradual involves higher computation energy for large datasets
DNN	Comparably admirable at accuracy	The datasets and epoch selected by the user affect the computation speed and accuracy.

6. Discussion and Conclusion

Despite meticulous efforts to construct DSS for heart disease prediction, it falls short of the performance and outcomes offered by a medical expert. Appropriate usage of expert knowledge and expertise to increase decision-making skills is one of the probable areas for further expansion for improved performance. Enormous efforts to develop an automated heart disease prediction system have been made. DSS with Machine Learning techniques is an attempt to avoid or eliminate the human effort in the process of disease prediction. From a clinical acceptability standpoint, the framework must have both subject expert knowledge and a trained, experienced workforce. This makes the development of DSS challenging, necessitating extra research. The following are some of the most notable findings, challenges and future directions in the development of DSS for Cardiac disease classification:

- Data pre-processing is beneficial in predicting cardiac disease for high quality, analysis-ready output. However, the complexity of the algorithms involves time intricacies, which slows down the process. It is necessary to develop and investigate the utility of data pre-processing techniques, which resolves the issue of computational intricacy and still conveys a comparable performance index. This issue needs further investigations and detailed analysis from the perspective of DSS.
- Feature selection is a prominent component in the supervised machine learning approach for better model learning. Selecting a subset of features simplifies the model and makes it easy to interpret for researchers. But most of the feature selection methods arbitrarily seek to identify only one solution to the problem. Several features are taken together can predict events equally. In that case, it would be misleading to return only one of them and claim that the rest are unnecessary. In this case, the researcher can try several feature selection algorithms and collect common

features from them for model training. The ultimate aim is to select the feature set with all the important risk factors.

- ML classifiers are used in the training and testing phase in heart disease prediction model building. Researchers have carried out extensive research using different ML algorithms. The performance of the heart disease prediction model with a single algorithm is found to be very poor. The performance of such systems can be improved with an ensemble approach. The ensemble approach helps to reduce variance. But Ensemble methods are found to be difficult to interpret. There is scope to find the best ensemble approach with a proper selection of ML classifiers involved in the Ensemble model. The hybrid model also improves the accuracy of the Classification model. Hyperparameter tuning of Machine Learning classifiers was found to be very effective in selecting appropriate classifier parameters and enhancing the performance of DSS. Moreover, a Heart disease prediction system, an application in medical spaces, needs to have very high classification accuracy. There is a genuine need to increase the reported classification accuracy by developing a better system.
- The popular Artificial neural networks machine, unfortunately, suffers from Catastrophic Forgetting. A deep neural network requires a large amount of data to perform better than other techniques. It is extremely expensive to train neural networks due to complex data models and unpredicted operation time. Neural network models are nonlinear and have a high variance, which can be disappointing when preparing a final model for making predictions. Ensemble learning combines the predictions from multiple neural network models to reduce the variance of predictions and reduce generalization error.
- There is a strong need for real-world clinical factors that are easily approachable and computed in real-time for the future of clinical cardiac disease detection via ML centred DSS. Despite the enormous quantity of patient data accessible in clinics and hospitals, little of it is published. Most investigators obtained their information from a similar resource: the UCI repository. Provided that the dataset's contents affect prediction accuracy, more hospitals should be motivated to spread high-quality datasets so that investigators may generate excellent findings and enhance the models that will benefit patients and treat cardiac disease in its early stages. There is a need for a practically deployable system for heart disease prediction.

References

[1] Ks. Reddy, B. Shah, C. Varghese, and A. Ramadoss, Responding to the Threat of Chronic Diseases in India, the Lancet. 366(9498) (2005) 1744-1749.

[2] S. Mendis, P. Puska, and B. Norrving, Global Atlas on Cardiovascular Disease Prevention and Control, World Health Organization. 15(1) (2011) 1-163.

[3] (2020) The Who Website. [Online]. Available: Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/the-Top-10-Causes-of-Death.

[4] (2017) The Wikipedia Website. [Online]. Available: Https://En.Wikipedia.Org/Wiki/List_of_Causes_of_Death_by_Rate#Cite_Note-3.

[5] (2019) Centers for Disease Control and Prevention. [Online]. Available: Https://Www.Cdc.Gov/Heartdisease/Statistics_Maps.Htm.

[6] P.K. Anooj, Clinical Decision Support System: Risk Level Prediction of Heart Disease Using Weighted Fuzzy Rules, Journal of King Saud University-Computer and Information Sciences.24(1) (2012) 27-40.

[7] M. C. Tu, D. Shin, and D. Shin, A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms, Eighth Ieee International Conference on Dependable, Autonomic and Secure Computing,(2009)183-187.

[8] G. Subbalakshmi, K. Ramesh, and M.C. Rao, Decision Support in Heart Disease Prediction System Using Naive Baye, Indian Journal of Computer Science and Engineering (Ijcse). 2(2) (2011) 170-176.

[9] L. Parthiban, and R. Subramanian, Intelligent Heart Disease Prediction System using Canfis and Genetic Algorithm, International Journal of Biological, Biomedical and Medical Sciences. 3(3) (2008) 157-160.

[10] D. Shanthi, G. Sahoo, and N. Saravanan, Input Feature Selection Using the Hybrid Neuro-Genetic Approach in Diagnosing Stroke Disease, Ijcsns. 8(12) (2008) 99-107.

[11] H. Yan, J. Zheng, Y. Jiang, C. Peng, and Q. Li, Development of A Decision Support System for Heart Disease Diagnosis using Multilayer Perceptron, Proceedings of International Symposium on Circuits and Systems, Iscas'03, 5 (2003) 709-712.

[12] M. Kukar, I. Kononenko, and C. Grošelj, Modern Parameterization and Explanation Techniques in the Diagnostic Decision Support System: A Case Study in Diagnostics of Coronary Artery Disease, Artificial Intelligence in Medicine. 52(2) (2011) 77-90.

[13] (2021) Medical News Today. [Online]. Available: Https://Www.Medicalnewstoday.Com/Articles/323724.

[14] Safdar, S. Zafar, N. Zafar, and N. F. Khan, Machine Learning-Based Decision Support Systems (Dss) for Heart Disease Diagnosis: A Review, Artificial Intelligence Review. 50(4) (2018) 597-623.

[15] (2019) Centers for Disease Control and Prevention. [Online]. Available: Https://Www.Cdc.Gov/Heartdisease/Risk_Factors.Htm.

[16] R. Chitra, and V. Seenivasagam, Heart Disease Prediction System using Supervised Learning Classifier, Bonfring International Journal of Software Engineering and Soft Computing. 3(1) 2013) 1-7.
[17] A. Mathur, and G.P. Moschis, Socialization Influences on Preparation for Later Life, Journal of Marketing Practice: Applied Marketing Science, 5 (6/7/8) (1999) 163 -176.
[18] L. P. Kaelbling, M. L. Littman, and A. W. Moore, Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research. 4 (1996) 237-285.
[19] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Schoellkopf, Support Vector Machines, Ieee Intelligent Systems and their Applications. 13(4), (1998) 18-28.
[20] T. Hastie, R. Tibshirani, and J. Friedman, Additive Models, Trees, and Related Methods, the Elements of Statistical Learning Springer Series in Statistics, Second Edition, New York,(2009) 295-336.
[21] O. Kramer, K-Nearest Neighbors. in: Dimensionality Reduction with Unsupervised Nearest Neighbors. Intelligent Systems Reference Library, Springer, Berlin, 51(1) (2013) 13-23.
[22] S. Chen, G.I. Webb, L. Liu, and X. Ma, A Novel Selective Naïve Bayes Algorithm, Knowledge-Based Systems, 192(105361) (2020) 1-12.
[23] G. Subbalakshmi, K. Ramesh, and M.C. Rao, Decision Support in Heart Disease Prediction System Using Naive Bayes, Indian Journal of Computer Science and Engineering (Ijcse). 2(2) (2011) 170-176.
[24] Y. Liu, Y. Wang, J. Zing, the New Machine Learning Algorithm: Random Forest, Ser. Lecture Notes in Computer Science. Springer, Berlin. 7473(1) (2012) 246-252.
[25] (2021) Investopedia. [Online]. J. Chen, Available: Https://Www.Investopedia.Com/Terms/N/Neuralnetwork.Asp.
[26] P. Sharma, and R. Bhartiya, Implementation of Decision Tree Algorithm to Analysis the Performance, International Journal of Advanced Research in Computer and Communication Engineering. 1(10) (2012) 861-864.
[27] (1999) Mit-Bih Polysomnographic Database.[Online]. Available: Https://Archive.Physionet.Org/Physiobank/Database/Slpdb/.
[28] (1988) Heart Disease Dataset Uci. [Online]. Available:Https://Archive.Ics.Uci.Edu/Ml/Datasets/Heart+Disease,
[29] P. Rani, R. Kumar, N.S. Ahmed, and A. Jain, A Decision Support System for Heart Disease Prediction Based upon Machine Learning, Journal of Reliable Intelligent Environments. 1(13) (2021) 1-13.
[30] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, and P Singh, Latest Trends on Heart Disease Prediction using Machine Learning and Image Fusion, Materials Today: Proceedings. 37 (2021) 3213-3218.
[31] R. Katarya, and Sk. Meena, Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis, Health and Technology. 11(1) (2021) 87-97.
[32] S. Prakash, K. Sangeetha, and N. Ramkumar, An Optimal Criterion Features Selection Method for Prediction and Effective Analysis of Heart Disease, Cluster Computing. 22(5) (2019) 11957-11963.
[33] F. Ali, S. El-Sappagh, S.M. Islam., D. Kwak, A. Ali, M. Imran, and K.S. Kwak, A Smart Healthcare Monitoring System for Heart Disease Prediction Based on Deep Ensemble Learning and Feature Fusion, Information Fusion.63 (2020) 208-222.
[34] Nl Fitriyani, M Syafrudin, G Alfian, J Rhee. Hdpm: An Effective Heart Disease Prediction Model for a Clinical Decision Support System. Ieee Access, 8, (2020) 133034-133050.
[35] A.K. Gárate-Escamila, A.H. El Hassani, and E. Andrès, Classification Models for Heart Disease Prediction Using Feature Selection and Pca, Informatics in Medicine Unlocked 19(100330) (2020) 1-11.
[36] N. Hasan and Y. Bao, Comparing Different Feature Selection Algorithms for Cardiovascular Disease Prediction, Health and Technology Springer Journal. 10(1) (2020) 1-14.
[37] C. Latha and S. Jeeva, Improving the Accuracy of Prediction of Heart Disease Risk Based on Ensemble Classification Techniques, Informatics in Medicines Unlocked. 16(100203) (2019)1-9.
[38] I. D. Mienye, Y. Sun, Z. Wang, An Improved Ensemble Learning Approach for the Prediction of Heart Disease Risk, Informatics in Medicines Unlocked. 20 (100402) (2020) 1-5.
[39] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, A. W. Muzaffar, An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases, Ieee Open Access. 9 (2021) 106575-106588.
[40] Li Yang. A.Shami, on Hyper Parameter Optimization of Machine Learning Algorithms: Theory and Practice, Neurocomputing Elsevier Journal. 415 (2020) 295-316.
[41] Ek Hashi, Md Zaman, Developing A Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction, Journal of Applied Science and Process Engineering.7(2) (2020) 631-647.
[42] A. Javeed, S. Zhou, An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection, Ieee Access. 7 (2019) 180235-180243.
[43] B. Dun, E. Wang, and S. Majumder, Heart Disease Diagnosis on Medical Data Using Ensemble Learning,Stanford.Edu, 1(1) (2016) 1-5.
[44] Ds.Medhekar, M.P. Bote, and S.D. Deshmukh, Heart Disease Prediction System Using Naive Bayes, Int. J. Enhanced Res. Sci. Technol. Engineering. 2 (3) (2013)1-5.
[45] W. Wiharto, H. Kusnanto, and H. Herianto, Performance Analysis of Multiclass Support Vector Machine Classification for Diagnosis of Coronary Heart Diseases, International Journal on Computational Science & Applications (Ijcsa).5 (5) (2015) 27-37.
[46] N. Khateeb, and M. Usman, Efficient Heart Disease Prediction System Using K-Nearest Neighbour Classification Technique, Proceedings of the International Conference on Big Data and Internet of Thing, (2017) 21-26.
[47] A.H. Chen, S.Y. Huang, P.S. Hong, C.H. Cheng, and E.J. Lin, Hdps: Heart Disease Prediction System, Computing in Cardiology. (2011) 557-560.
[48] Sk. Sen, Predicting and Diagnosing Heart Disease Using Machine Learning Algorithms, International Journal of Engineering and Computer Science. 6(6) (2017) 21623-21631.

[49] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, A Comprehensive Investigation and Comparison of Machine Learning Techniques in Heart Disease, Ieee Symposium on Computers and Communications (Iscc), (2017) 204-207.
[50] Yk. Singh, N. Sinha, and S.K. Singh, Heart Disease Prediction System Using Random Forest, International Conference on Advances in Computing and Data Sciences Springer, (2017) 613-623.
[51] R. Das, I. Turkoglu, and A. Sengur, Effective Diagnosis of Heart Disease Through Neural Networks Ensembles, Expert Systems With Applications. 36(4) (2009) 7675-7680.
[52] Nb. Amma, Cardiovascular Disease Prediction System Using Genetic Algorithm and Neural Network, International Conference on Computing, Communication and Applications, (2012) 1-5..
[53] T. Santhanam, E.P. Ephzibah, Heart Disease Classification Using Pca and Feed-Forward Neural Networks, Mining Intelligence and Knowledge Exploration, Springer, Switzerland, (2013) 90-99.
[54] In Press, K. Budholiya, S. Shrivastava, V. Sharma, An Optimized Xgboost Based Diagnostic System for Effective Prediction of Heart Disease, Journal of King Saud University –Computer and Information Sciences.
[55] S. P. Patro, G.S. Nayak, N. Padhy, Heart Disease Prediction by Using Novel Optimization Algorithm: A Supervised Learning Prospective, Informatics in Medicine Unlocked. 26 (100696) (2021) 1-17.
[56] R. Valarmathi, T. Sheela, Heart Disease Prediction Using Hyperparameter Optimization (Hpo) Tuning, Biomedical Signal Processing and Control, Elsevier. 103033 (2021) 1-10.