Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

A Review of Heart Disease Classification Base on Machine Learning Algorithms

Mayaf Zakwa Hasan[1], Adnan Mohsin Abdulazeez[2]
mayaf.zakawa@dpu.edu.krd, adnan.mohsin@dpu.edu.krd
[1]Technical College of Zakho, Duhok Polytechnic University, Kurdistan Region, Iraq
[2]Technical College of Engineering, Duhok Polytechnic University, Kurdistan Region, Iraq

Article Information

Submitted : 5 Apr 2024
Reviewed: 7 Apr 2024
Accepted : 20 Apr 2024

Keywords

Heart Disease, Machine Learning(ML), Classification Algorithm.

Abstract

Heart disease is currently the leading cause of death. This problem is acute in developing countries. Predicting heart disease helps patients avoid it in its early stages and can also help medical practitioners find out the main causes. Machine learning has proven over time to play an important role in decision making and forecasting through massive data sets created by the healthcare sector. This review provides an overview of heart disease prediction using applied machine learning algorithms such as Naïve Bayes, Random Forest, Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbour (KNN). And these differences in the techniques are a reflection of many strategies for predicting heart disease. We present a synopsis of classification techniques that are primarily used in the predicted of heart disease. Additionally, we review several previous studies that conducted over the past four years, that used machine learning algorithms to predict cardiovascular.

A. Introduction

After the brain, the heart is one of the most important organs in the human body. The heart's main job is to pump blood to every portion of the body [1]. If the heart isn't beating correctly, other organs may cease working properly as well. Thus, maintaining the health of the heart and other organs becomes difficult [2], in any condition that may impair the heart's ability to operate is called heart disease [1]. Heart disease claims the lives of 17.5 million people annually. These days [3].

The most prevalent illness in the world that poses a risk to human life also known as cardiovascular disease [4]. Heart disease encompasses a variety of heart-related disorders, such as blood vessel disease, heart failure, heart attack, arrhythmia, stroke, etc. [5]. Therefore, heart disease diagnosis or prognosis is necessary. Heart disease can be diagnosed using a variety of techniques. Although angiography is a widely used technique for identifying cardiac illness, doctors find it challenging to use it because to its high cost and intricate analysis [1], particularly in developing nations [4]. Reducing the risk of heart disease and stopping the rising death toll are largely dependent on early detection of heart disease and the use of suitable treatment [6]. As a result, non-invasive techniques for predicting cardiac disease have been developed; however, they are time-consuming and may produce inaccurate results [1]. The provision of high-quality services and accurate and efficient prediction is currently the main problem in the field of medical sciences [7]. We require an automated method in order to prevent these mistakes and produce better and quicker results. Researchers have discovered in recent years that machine learning algorithms are highly effective at analysing medical data sets [1].

Machine learning is an area that can help predict disease prognosis. Without special programming, machines can now learn thanks to machine learning [8] Machine learning has applications in medicine that include diagnosing, detecting, and predicting diseases [7]. Predicting cardiovascular disease has become a new approach thanks to the use of machine learning. Using patient datasets, several machine learning methods are used to extract important information from the health club [9]. Many classifiers and algorithms, including K-nearest, Decision Tree, Random Forest, Support Vector Machine (SVM), Naïve Bayes, and others, can provide a solution to this situation thanks to advances in machine learning and artificial intelligence [8]. The field of machine learning is becoming more important for early diagnosis of diseases [6].

In medical applications, Classification is the most widely used machine learning technique in medical contexts since it is applicable to common real-world problems. Classification algorithms use training data to build a model first, which is then used on test data to produce predictions. Many categorization techniques have been used to diagnose diseases, with very positive results. These methods have the potential to produce quick results and lower diagnostic mistake rates [9].

The aim of this review is to provide a comprehensive overview of machine learning techniques in heart disease prediction. The structure of this paper is as follows. Section II discusses about Theoretical background ML and Algorithm of Heart Disease. Section III presents Related Work. Section IV Discussions. Finally, Section V presents Conclusion and Future Directions.

B. Research Method

1. Machine Learning

It is a subfield of artificial intelligence (AI) [10], [11], and is a large field of learning that deals with machines that mimic human capabilities [11]. Machine learning is frequently used in many different fields to address difficult problems that cannot be easily solved by computer-based methods [12]. Machine learning is one of the effective testing techniques that relies on machine learning training and testing [11]. It includes algorithms designed to perform various tasks including prediction, classification, and decision making. To facilitate the learning process, these algorithms require training data. After this learning phase, the algorithms create a model that serves as the output. This model is later tested and validated using unseen real-time test datasets. The final accuracy of the model is then compared with the actual values, proving the overall validity of the predicted results [13], [14].

Three different kinds of machine learning algorithms exist [11]:

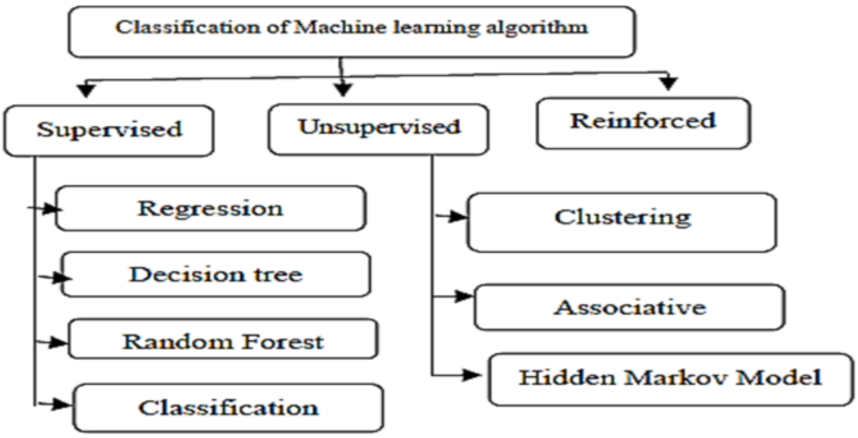

Figure 1 Classification of Machine Learning [11]

Techniques for machine learning are categorised as [14]:

a) Supervised Learning

A type of learning that occurs under supervision or with a mentor present is known as supervised learning. Here, a training dataset acts as the teacher for making predictions about a particular dataset. In other words, if testing data is used, there is always a matching training dataset that offers the direction needed to learn [11]. However, a drawback of supervised machine learning algorithms is that their ability to identify interesting patterns within the data is limited [15].

b) Unsupervised learning

is defined as self-contained learning that takes place in the absence of direction or instruction. The procedure is not being guided by an outside teacher. Unsupervised learning involves an algorithm that, given a dataset, analyses it on its own to find patterns and relationships. It then classifies and arranges newly introduced data inside the established patterns, based on these inferred associations [11]. Furthermore, it is possible to view unsupervised learning as more sovereign or agnostic [16] .

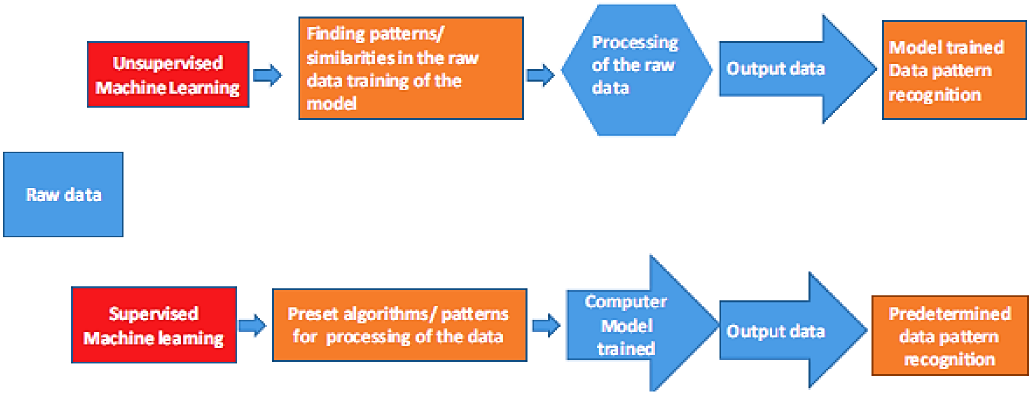

Figure 2. Unsupervised and Supervised Learning [17]

1.1 Machine Learning Algorithms

This section offers a comprehensive review of the most commonly used machine learning algorithms for disease diagnosis [18].

1.1.1 Logical regression

Logistic regression is a supervised learning technique used to address binary classification issues [7], [8], [18]. Essentially, it works as a regression model that predicts the probability that an item or data input belongs to a particular category [8]. The logistic function is particularly valuable because it produces values between zero and one [4], [18], making it interpretable as a probability [4]. Logistic regression uses a sigmoidal function to model the data, as shown in Figure 3 [8]. Notably, logistic regression offers several key advantages including simplicity of implementation, computational efficiency, effectiveness based on training data, and ease of interpretation. In addition, input features do not require scaling [8].

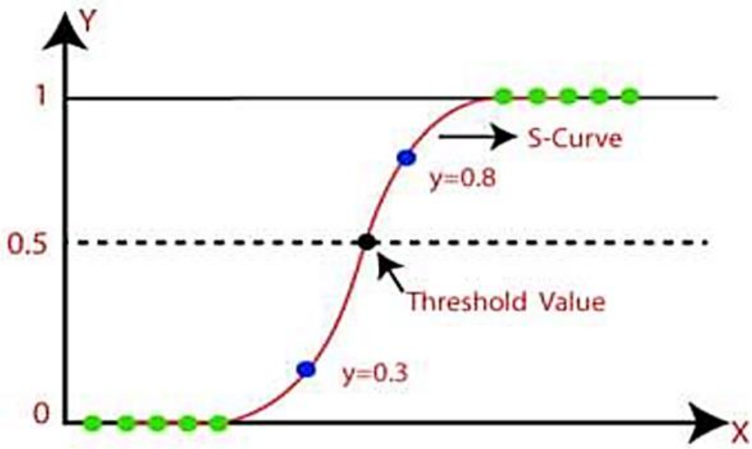

Figure 3. Logistic Function [19]

1.1.2 Decision Tree

The Decision Tree (DT) serves as a supervised machine learning algorithm utilized for addressing regression and classification tasks [8], [13], accomplishing this by repeatedly segmenting data according to particular variables [8].This model, resembling a tree structure, forms the basis of classification [7], [13]. Comprising root nodes, branches, and leaf nodes, the decision tree, analyses data by following the path from the root to the leaf node [8], [13]. Both numerical and categorical data are managed by the decision tree [7], [8]. Its primary objective is to construct a model capable of predicting a target variable by learning straightforward decision rules from training data [8]. The decision tree algorithm simplifies the interpretation of results, offering higher accuracy compared to alternative methods by organizing the dataset into a tree-like graph [7], as illustrated in Figure 4 [20]. However, the possible drawback is when working with large datasets and making decisions based on a single attribute at a time [7].

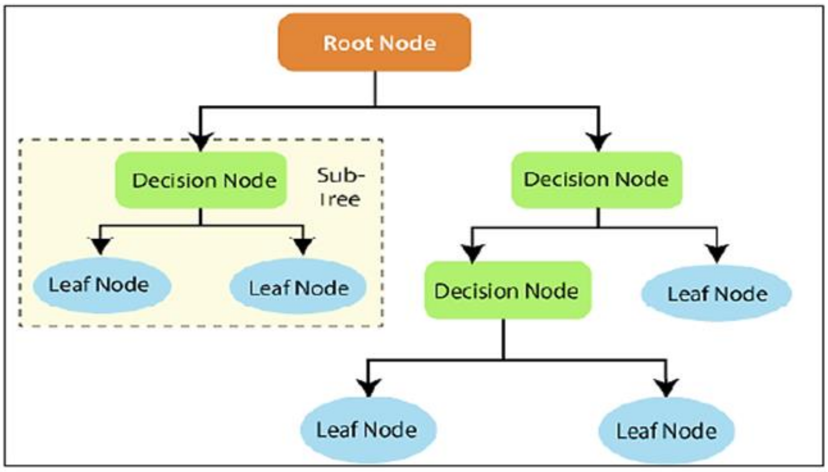

Figure 4. Decision tree [20]

1.1.3 Support Vector Machine

The Support Vector Machine (SVM) stands as a machine learning algorithm employed in supervised learning scenarios [13], [21], [22], aimed at scrutinizing data and uncovering patterns in both classification and regression analyses [13], [22], [23]. Generally, SVM is considered when the data is classified as a two-class problem. The ideal hyper plane that isolates every data point from one class to the other class is what defines data in this strategy. The model is deemed better the greater the difference or edge between the two classes. Support vectors are the data points that are located near the margin's edge [13]. The primary objective of a support vector machine is to identify the optimal highest-margin separating hyperplane between the two classes [7], [23]. Owing to its computational effectiveness, SVM has emerged as a proficient approach in addressing high-dimensional space problems, particularly with large datasets [23].

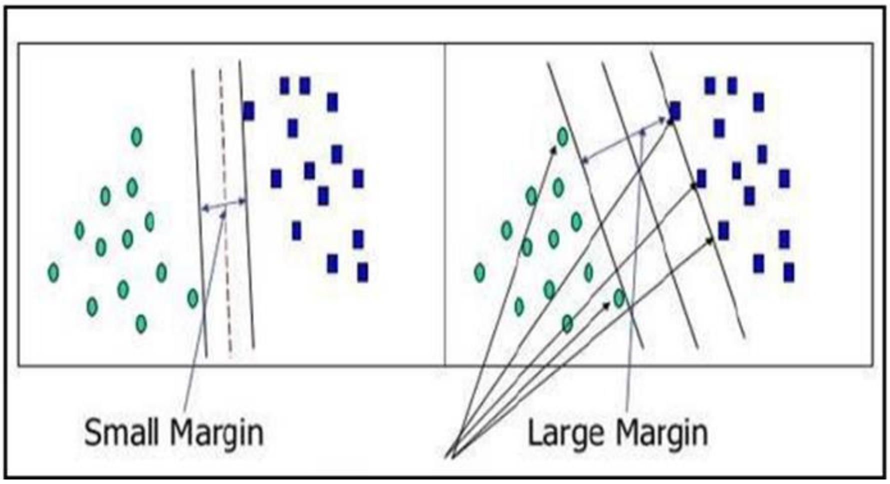

Figure 5. Margin and Support Vectors. Left: small margin between 2 classes Right: large margin between 2 classes [24]

1.1.4 Naive Bayes

Naïve Bayes (NB) serves as a classification algorithm rooted in probability and statistics [8], [14], operates under a supervised learning paradigm [4], [14]. The basis of this approach is the Bayes theorem [14], [25]. And it has become a staple in machine learning applications owing to its straightforwardness in assigning equal importance to all features in the final decision-making process [8]. The reason for its extensive use is that it produces excellent results, outperforming even more complex categorization techniques [25]. The Naïve Bayes approach proves to be both versatile and apt across various domains. This technique boasts several advantages, such as its suitability for extensive datasets and its overall utility, making it applicable for both binary and multiclass classification challenges [8].

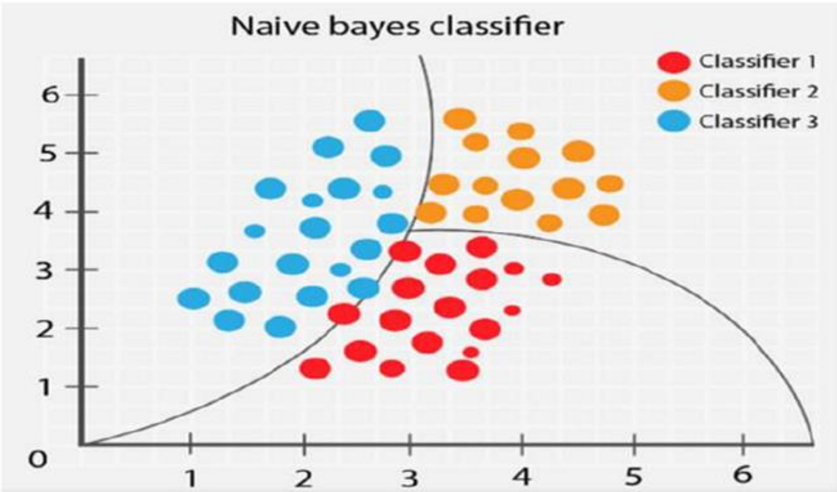

Figure 5. Naive Bayes algorithm

1.1.5 Random Forest

Random Forest is a supervised machine learning framework that may be used for both regression and classification applications [2],[26],[27]. It comprises a collection of trees [2],[28], with each tree constructed through repeated sampling of the dataset. Prediction is made by aggregating the outputs of all individual decision trees [2]. Increasing the number of trees is believed to enhance the accuracy and reliability of the model's predictions [2],[26]. Random Forest is particularly effective for large datasets and can also handle data with high dimensionality [2]. By considering the predictions of each tree, Random Forest produces a final output by selecting the class with the highest probability, determined through majority voting [27],[28]. Figure 7 provides an illustration of selecting procedure [28].

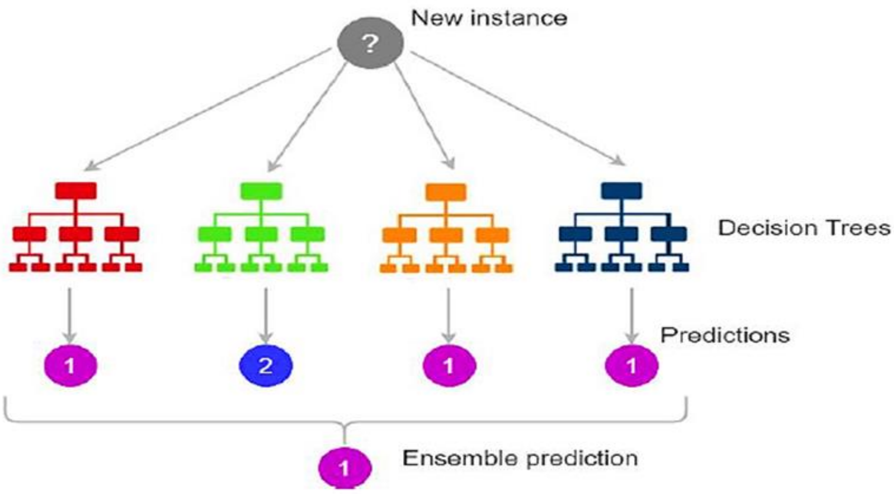

Figure 6. Random Forest prediction [28]

1.1.6 K-nearest neighbour Algorithm (KNN)

The KNN algorithm stands out as one of the simplest algorithms to grasp [4],[29],[30]. It is used for pattern recognition utilising a non-parametric approach for classification and inference [4]. In order to classify data points according to their proximity to one another, KNN measures the distance between each one [11]. Large datasets are especially well-suited for KNN classification, which requires more processing time during testing than training [4],[31]. However, it has strong computing powers [27]. KNN uses training datasets to forecast where K-nearest neighbours will be, and it uses Euclidean distance to measure how close the training dataset is to the target [29]. Nevertheless, a shortcoming in the KNN method is its susceptibility to local data structures [4].

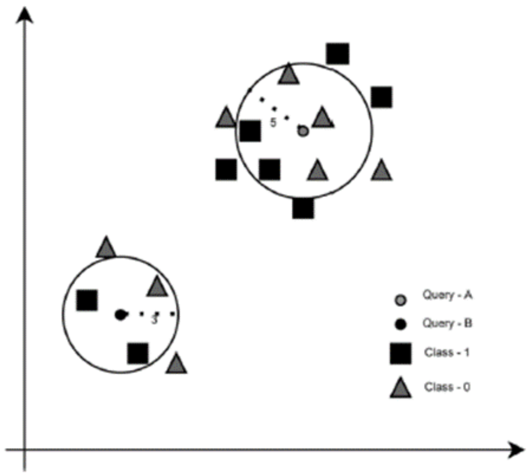

Figure 7. K-nearest neighbor Algorithm [32]

1.1.7 Gradient Boosting

Gradient boosting is a machine learning technique that solves regression and classification problems [33],[34], by building a predictive model from a collection of weak predictive models [33]. An ensemble of decision trees is called gradient boosting. The approach relies on learning decision trees iteratively in order to minimise the loss function. Gradient boosting can handle nonlinearities and deal with categorical information because of the properties of decision trees [33]. The algorithm's strength is its capacity to train and enhance the classifiers iteratively, leading to an ensemble model that capitalises on each classifier's unique strengths and ultimately improves overall classification accuracy [34]. Boosting is a technique used to elevate underperforming models to a higher level of proficiency. In boosting, every subsequent tree is trained on a revised rendition of the initial dataset [33].

1.1.8 AdaBoost (Adaptive Boosting)

AdaBoost is an iterative process developed by to increase the performance of weak classifiers, sometimes known as base classifiers. Through repeated correction of these base classifiers' mistakes, AdaBoost improves the data's classification performance. This repeated refining helps to reduce variance and bias. The algorithm's efficacy stems from its ongoing iteration of classifiers, which leads to an ensemble model that optimises each classifier's distinct strengths and improves overall classification accuracy [34].

1.1.9 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are a component of neural networks within machine learning. ANNs operate in a manner akin to the human brain, resembling the structure and function of human neuron cells [1],[2]. They find applications across various domains of medicine [35] and serve as a non-linear statistical architecture for tackling complex problem-solving tasks [2]. This model comprises numerous interconnected elements, or neurons, collaborating to execute a task. A neural network with a single layer is termed a perceptron, producing a singular output [1], as depicted in Fig. 10.

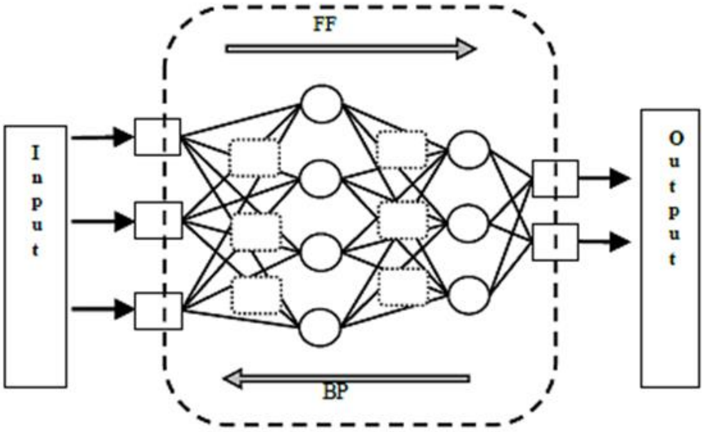

Figure 8. Artificial Neural Network [36]

C. Literature Review

In [37], the researchers have focused on how machine learning (ML) can be used to predict and understand heart disease symptoms. They revealed several notable factors, including heart rate, cholesterol, chest pain, features associated with ST depression, and vascular characteristics of the heart. According to their study, combining chi-square and PCA enhances performance for most classifiers, but applying PCA directly from raw data leads to less-than-ideal results.

In [38], the researchers have proposed a systematic approach to machine learning-based heart disease detection. To increase accuracy and efficiency, they used several feature selection strategies and classification algorithms. Cross-validation was used to evaluate the performance and a new feature selection approach was presented. They discovered that the proposed technical results are successful and applicable, superior to previous methods and have the potential for use in healthcare applications.

In [39], the authors proposed using data mining and machine learning techniques to predict heart disease using a variety of algorithms. Emphasis was placed on the importance of early detection and the function of machine learning in medical diagnosis. Positive results from one model suggested that it may be useful in predicting heart disease. She indicated that alternative models could be studied in further research to increase the accuracy of diagnosis.

In [40], the study focused on applying machine learning algorithms in classifying cardiovascular data sets for the purpose of disease prediction. They emphasized how important the dimensions of the data set are to the performance of the algorithm and plan to improve disease prediction models in the future by reducing their dimensions. Their main goal was to provide information on early diagnosis and management of cardiovascular diseases.

In [41], the authors proposed using several feature selection strategies and classification algorithms to examine how well machine learning models diagnose heart conditions. Principal components analysis and chi-square testing were used to identify separate feature sets from the four heart disease datasets analyzed. They found that the best accuracy is achieved when combining the BayesNet method and chi-square feature selection. They emphasized the importance of feature selection in enhancing heart disease prediction algorithms.

In [42], the researchers focused on highlighting the importance of data analytics, especially when it comes to applying machine learning algorithms to predict heart disease. They used heart disease datasets to evaluate different supervised learning techniques. They recommend future investigations using larger data sets and other machine learning methods to create more accurate real-time prediction models for heart disease.

In [14], this study explored different data mining techniques for predicting heart disease, using supervised learning algorithms such as Naïve Bayes, decision tree, K-nearest neighbor, and random forest. Using a dataset from the Cleveland Database, the study evaluated 14 features to evaluate the performance of the algorithm. K-nearest neighbor showed the highest accuracy. They suggest that future research should consider additional data mining techniques to enhance predictive accuracy for early detection of heart disease.

In [43], the study proposes a clinical support system that predicts heart disease using machine learning algorithms, to help doctors make more accurate diagnoses and decisions. Naïve Bayes emerged as the best-performing algorithm, showing notable accuracy in both split and cross-training tests. It has been recognized that it can be applied to smaller data sets. The results highlighted the potential of the predictive model to enhance heart disease risk prediction, thereby improving clinical diagnosis and treatment decisions.

In [2], this research aims to help medical professionals diagnose and treat patients by examining the importance of using machine learning methods to predict heart disease, given the vast amount of medical data available. Through comparative analysis, random forest emerged as the most effective algorithm for predicting heart disease, as determined by evaluation and accuracy metrics. They suggested that future studies should focus on implementing deep learning algorithms to enhance predictive capabilities.

In [44], the researchers worked on automatic detection of cardiovascular diseases using different machine learning algorithms. They found that the Gradient Boosting Classifier performed best across all features, and Random Forest was the best classifier. Which helped doctors diagnose and make decisions faster and more efficiently in cardiovascular diseases.

In [45] discussed the study on the effectiveness of machine learning algorithms in predicting heart disease in order to overcome difficulties in clinical decision-making and diagnostic accuracy. They found the possibilities of accurate disease prediction, using supervised algorithms such as KNN, DT and RF. They found key predictors using feature importance ratings, demonstrating the value of machine learning in enhancing patient care and outcomes.

In [46], the authors addressed the issue of major cardiac anomalies. They proposed a new method for real-time feature extraction (less than 30 milliseconds) and ECG type prediction using machine learning and digital ECG data. They implemented the XGBoost algorithm. They created a standard for detecting heart abnormalities. It has made great progress and shown excellent performance in hospitals and countries.

In [47], the study worked on using deep learning and machine learning methods to predict heart disease. Using the 1988 data set, in order to improve accuracy, the researchers compared several classifiers and methodologies. He highlighted pre-processing, normalization, and outlier data identification. Which led to the development of computational intelligence and neuroscience.

In [48], the researchers have proposed a variety of feature selection techniques to examine how well classification models predict heart disease. The primary goal was to evaluate the effects of 10 feature selection strategies and six classification algorithms on the accuracy of heart disease prediction. They highlighted how important feature selection is in improving the success of prediction algorithms in diagnosing heart disease.

In [49], the study addressed the topic of machine learning (ML) in heart disease diagnosis by using machine learning algorithms, including supervised techniques such as Random Forest and K-Nearest Neighbor (KNN), to classify people based on characteristics such as age, cholesterol and weight. Health status. And chest pain. The results highlight the effectiveness of machine learning in diagnosing cardiovascular disorders.

In [50], the researchers presented a comprehensive model designed to accurately predict heart disease. By using LASSO and Relief methods to select features, perform pre-processing and collect data in an efficient manner. Moreover, they developed state-of-the-art hybrid classifiers that combine traditional classifiers with bagging and boosting techniques, such as DTBM, RFBM, KNNBM, ABBM, and GBBM. They evaluated typical performance indicators using machine learning methods.

In [51], the authors presented the development of a model that uses three machine learning classification techniques to identify cardiovascular disease. Their model has proven to be highly robust and accurate, correctly identify people at risk of heart disease, and reducing pressure on healthcare systems. It has reduced expenses while improving medical treatment. Their study, conducted in Pynb format, provided insight into predicting patient heart disease.

In [52], the authors proposed developing a diagnostic framework using several machine learning techniques to classify heart diseases. Its goal is to improve prediction accuracy by optimizing model parameters for effective detection of heart disease. They pointed out the possibility of an effective health care monitoring system for early detection of heart disease. Data mining classification techniques have accelerated the learning process, simplified models, and improved prediction accuracy, especially when combined with optimization methods.

In [53], the study proposes a hybrid decision support system that can identify heart problems early and is particularly useful in settings where access to specialized health care is scarce. Among the pre-processing techniques, mixed feature selection and multivariate imputation. Used a variety of classifiers, their system aimed to reduce delays in diagnosing heart disease while improving accuracy compared to current techniques.

In [54], the research discussed the predictive accuracy of eleven machine learning classifiers for heart disease risk. To select the best features, they used three feature evaluators, and adjusted hyperparameters to improve accuracy. Prediction accuracy has been greatly increased by this approach. The researchers have advised future research on different feature selection algorithms and methods. He also intended to merge several data sets in order to collect more information and conduct more research in order to increase the accuracy of predictions.

In [55], the researchers worked to creating a machine learning model that used the DiScRi dataset to predict the incidence of diabetes and cardiovascular disease. They identified algorithm risk variables that affect the chance of developing the condition, such as family history, HbA1c, and HDL. The model provided the possibility of early identification and treatment in prevention.

In [56], the researchers used a variety of algorithms combined with sequential feature selection, using K-fold cross-validation for validation purposes. Their analysis, using datasets from a variety of sources, showed that Random Forest and Decision Tree classifiers achieved remarkable levels of accuracy. Their future efforts seek to expand the model's reach and utility, and they highlighted advances in feature selection methods and considering incorporating a random forest classifier to improve predictive performance

In [57], the authors describe how to prediction of cardiovascular disease using logistic regression. They stressed the importance of early detection to receive timely treatment. They prepared the data after cleaning, filling in missing values, and selecting strongly associated characteristics. They have used the classification process for many courses and test sections. They recommended further research using multiple datasets.

In [58], the researchers focused on applying machine learning algorithms to predict heart disease. In order to solve the problem of early diagnosis and lower mortality rates. They used a variety of prediction algorithms and collected data sets. The goal of their study was to improve the accuracy of predicting the rate of heart disease by examining a large amount of medical data. They have advanced the field of medicine by using data science to enhance the prediction of heart disease.

In [59], the study proposed using machine learning algorithms to predict heart disease, to improving accuracy and reliability in diagnosis. They used classifiers and evaluated metrics such as sensitivity and specificity. Preprocessing and hyperparameter tuning improved prediction accuracy, they highlighted the potential of machine learning in cardiovascular healthcare.

In [60], the researchers discussed how to using the capabilities of the diagnostic system and applying heuristic algorithms to classify heart disorders. They relied on a reduced set of features and revealed several different classifications to predict the presence of heart disorders. They found that the results achieved the highest accuracy among the algorithms tested, which confirmed their usefulness in increasing prediction accuracy for diagnosing heart diseases.

In [61], the authors have proposed using machine learning techniques to improve the accuracy of heart disease prediction. Using GridSearchCV, a five-fold mutual improvement of hyperparameters was achieved, introducing innovation. Real-time heart disease prediction has been made possible through the potential integration of the model with useful applications such as online platforms or mobile devices, which have leveraged patients' electronic health records to improve accuracy.

In [62], the researchers presented the development of a machine learning model for predicting CVD risk. Different machine learning models were compared, which are useful for early diagnosis and more effective treatment outcomes. Their goal was to address the problem of cardiovascular diseases (CVDs), and they also emphasized how machine learning algorithms can improve the effectiveness of identifying and treating diseases.

In [63], the research discussed machine learning algorithms to predict and diagnose cardiovascular problems early. The aim of their research was to improve the diagnostic and therapeutic skills of medical professionals. The performance of several algorithms has been evaluated and they used in global healthcare. To increase accuracy, they advised further research into machine learning methods and patient data.

In [6], the authors worked on a supervised machine learning approach called Classification and Regression Tree (CART) algorithm for heart disease. They demonstrated that the model is reliable in predicting heart disease. Furthermore, the model's decision criteria simplify its implementation in clinical settings and eliminate the need for additional knowledge. The CART model has made it easier to identify and manage heart disease, and is a vital tool for patients and healthcare professionals alike.

Table 1. Summary of studies conducted on heart disease using ML

Authors and year of pub	Dataset	Method	Pros	Limit	Result
Escamila et al 2020 [37]	Cleveland-Hungarian	DT, Gradient-Boosted Tree, LR, Multilayer Perceptron, Naïve Bayes, and RF	Improved prediction accuracy, Dimensionality reduction, Relevance of features	Small sample size, Limited extension to other diseases	98.7% Cleveland, 99.0% for Hungarian, 99.4% Cleveland-Hungarian
Li et al 2020 [38]	Cleveland Heart Disease dataset	Sequential Backward Selection, KNN, ANN, SVM, Naïve Bayes, DT	high accuracy, used various algorithms, Information feature selection algorithm.	Research gap in improving prediction accuracy and reducing calculation time.	80.4% to 92.37%

Obiri and Sarku 2020 [39]	Cleveland Heart Diseases	Linear Regression, Decision Tree, Gaussian Naïve Bayes	improved accuracy in diagnosis and the potential for reducing the number of tests needed for diagnosis.	only three classification models, a tiny sample size, a lack of thorough information on feature selection, and little investigation into the interpretability of the models.	82.75% 79.31% 76%
Princy et al 2020 [40]	cardiovascular disease dataset from Kaggle	Decision Tree, Logistic Regression, Random Forest, Support Vector Machine with RBF Kernel, Naive Bayes, and K-Nearest Neighbors	comparison of pre and post dimensionality reduction accuracy of algorithms, and the implementation of various classification techniques for predicting.	The model requires larger dataset testing, interprets Neural Network processes more complexly, and dataset dimensionality significantly impacts algorithm performance.	0.73
Spencer et al 2020 [41]	Cleveland, Long-Beach-VA-Dataset, Hungarian, Switzerland	BayesNet, LR, Stochastic Gradient Descent, KNN IBK with K, Adaboost M1 with Decision Stump, Adaboost M1 with Logistic, RF	Evaluate machine learning models, compare different feature selection methods to enhance accuracy and identify influential features.	a few machines learning and feature selection methods.	80% to 95.2% 85%
Sujatha and Mahalakshmi 2020 [42]	Heart disease dataset obtained from the Kaggle	Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine, K-Nearest Neighbor, and Logistic Regression.	High accuracy was achieved using the Rapid Miner tool Determine the feature selection technique.	Limited discussion of the specific difficulties encountered and no comparison with other studies	83.52%
Shah et al 2020 [14]	UCI	Naïve Bayes Classifier, Decision Tree, K-Nearest Neighbour, Random Forest Algorithm	high accuracy, data mining techniques complex medical, insights into attributes.	need to additional data mining techniques, limited discussion, need for more advanced models	83% 80.26% 90.78% 86.84%
EL Hamdaoui et al 2020 [43]	Cleveland heart disease dataset	Naïve Bayes, K-Nearest Neighbour, Support Vector Machine, Random Forest, and Decision Tree	Accurate Prediction, Clinical Decision Support, Model Validation	Decrease in accuracy, dataset Size small, overfitting by using cross-validation	82.17% 76.56% 79.20% 69.30% 75.57%
Katarya and Meena 2020 [2]	UCI	Logistic Regression, Naïve Bayes, SVM, KNN, Decision Tree, Random Forest, Artificial Neural Network, Deep Neural Network, Multilayer Perceptron	thorough evaluation algorithms, providing insights into their effectiveness and enabling comparison of these algorithms.	lack a more thorough examination of the computational resources needed to put the methods into practice.	95.60%
Alalawi and Alsuwat 2021 [44]	Cardiovascular Disease heart diseases	RF, SVM, DT Voting, LR, KNN, Naïve Bayes	Comprehensive Approach High Accuracy Utilization of Diverse Datasets Potential Clinical Impact	Limited Dataset Size Feature Engineering Challenges Comparison with Existing Methods	94% 65% 89% 67% 89% 71% 73%

Ali et al 2021 [45]	heart disease	RF, Neural Network, Naïve Bayes, Classification via regression analysis, SVM, DT, Genetic Algorithm, LR, Majority vote with MP, RF, BN, NB Cascaded Neural Network	Improved clinical services, lower costs, and early diagnosis of cardiovascular illnesses	Specific dataset small sample size lack of detailed information	100%
Bertsimas et al 2021 [46]	(The 2017 Challenge dataset from Physionet), (Tianchi Hefei High-Tech Cup ECG), (University & Shaoxing People's Hospital)	XGBoost	The models outperformed original papers in predictive performance, enabling real-time detection of heart anomalies from ECG signals with predictions in under 30 milliseconds.	The study's proprietary datasets, random data split between training and testing sets, and lack of time complexity information for cross-dataset models limit objective comparisons	0.99 0.99 0.93
Bharti et al 2021 [47]	Cleveland, Hungary, Switzerland, Long Beach V.	LR, KNN, Decision Tree, RF, SVM, XGBoost	various algorithms, Comparison of different classifiers and technique, comprehensive dataset	not capture the most current trends or data, focuses on a specific set of attributes and features,	83.31% 84.86% 84.86% 80.3% 83.29% 71.4% 94.2%
Dissanayake & Johar 2021 [48]	Hungarian, Cleveland, and Hungarian-Cleveland	decision tree, random forest, support vector machine, K-nearest neighbor, logistic regression, and Gaussian naive Bayes.	using feature selection techniques to improve classification accuracy and reduce model complexity.	effectiveness of feature selection methods may vary depending on the dataset and the combination of feature selection techniques and learning algorithms used.	88.52%
Garg et al 2021 [49]	dataset used from a website Kaggle	K-Nearest Neighbor Random Forest	Effective Prediction, Healthcare Improvement, Balanced Data	Specific Algorithms, Dataset Dependency, Scope of Analysis	86.88% 81.96%
Ghosh et al 2021 [50]	Cleveland, Hungary, Switzerland, VA Long Beach, and Statlog	Decision Tree, Random Forest, K-Nearest Neighbors, AdaBoost, and Gradient Boosting	improved prediction accuracy, providing a tightly correlated feature set for use with ML algorithms.	need for larger datasets, Dependency on specific feature selection techniques, may limit the generalizability of the model.	99.05%
Jindal et al 2021 [51]	UCI	K Nearest Neighbors Logistic Regression Random Forest Classifier	ability to predict and classify patients with heart disease accurately, and the reduction in medical costs	need for further research to enhance prediction accuracy.in Intelligent Heart Disease Prediction System (IHDPS) model	87.5%

Patro et al 2021 [52]	Cleveland, datasets from Beijing and China's environmental monitoring stations, and the Heart Stalog	Salp Swarm, Salp Swarm Optimized Neural Network, Bayesian Optimization, KNN, Naïve Bayes, SVM, RapidMiner, Particle Swarm Optimization, Ant Colony Optimization, and Fast Correlation-Based Feature Selection	improved accuracy, convergence, speed, and reliability	less control over multimodal strategies	SSA93.3% NB86.7% KNN93.3% RF99.6% FCBF 99.65%
Rani et al 2021 [53]	University of California, Irvine (UCI)	hybrid system, support vector machine, naive bayes, logistic regression, random forest, and adaboost classifiers	better accuracy compared to existing systems, It can assist doctors in making quick decisions	only diagnose the presence or absence of heart disease, not the severity.	86.6%,
Reddy et al 2021 [54]	Cleveland	Naïve Bayes, LR, Sequential Minimal Optimization, Instance-Based Classifier, AdaBoostM1 with (DS), AdaBoostM1 with LR, Bagging with REP Tree, Bagging with LR, Jrip	Improved Forecasting Performance: hyperparameter tweaking and attribute selection strategies.	Small Dataset Size, Limited Feature Selection Methods	86.47%
Abdalrad a et al. 2022 [55]	DiScRi	Logistic Regression, Evimp Functions, Generalised Cross-Validation, correlation matrix, Classification and Regression Trees	This prediction model can help with early diagnosis, preventative management, and lessening the cost of DM and CVD-related healthcare.	Small sample size, Data completeness: The absence of these data points affects the predictive accuracy and completeness of the model, Evaluation of model performance: The running time of the machine learning models developed was not recorded	94.09%
AHMAD et al 2022[56]	Heart Statlog Cleveland Hungary, Heart Disease Clinical Record Data Set 2020	Linear Discriminant Analysis, Random Forest, Gradient Boosting Classifier, Decision Tree, Support Vector Machine, K-Nearest Neighbours	comparative study of ML techniques, introduces a hybrid approach, feature selection methods.	Lacks detailed data set information and discussions of model biases. Further validation of the data sets, interpretability of the model and its clinical relevance.	100% 84.09% 100% 95.78% 85.71%
Ambrish G et al 2022 [57]	UCI	LR, SVM, RF, Stochastic Gradient Boosting, J48, and Multilayer Perceptron.	high accuracy achieved compared to previous research work	limited dataset	87.10%
Bora et al 2022 [58]	UCI heart disease heart disease	Logistic Regression, SVM, RF, Naïve Bayes, KNN, Gradient Boost, XGBoost	comprehensive analysis and, potential for early diagnosis	limitations involve data quality dependency and scalability challenges.	73% to 96%

Saboor et al 2022 [59]	UCI, Z-Alizadeh Sanitation Heart, Hungarian and Long Beach VA, Cleveland, Hungary, Switzerland	KNN, Neural Networks, SVM, Genetic Algorithms, J48 (Decision Trees), RF, Naive Bayes	dataset standardization and refinement, high accuracy	complexity levels, noting insufficient risk factors or attributes from detailed clinical data	83.66% to 96.72%
Ullah et al 2022 [60]	Cleveland	DT, neural network, SVM, KNN, LR.	high accuracy, proving prediction accuracy through cooperative techniques	need for further research, potential obstacles in implementing these algorithms in real-world	97.90%
Chandrasekhar and Peddakrishna 2023 [61]	UC Irvine ML Repository-Cleveland IEEE Dataport heart disease	RF, KNN, LR, Naïve Bayes, Gradient Boosting, AdaBoost	Enhanced Prediction Accuracy, novel approaches such as GridsearchCV, Comparative Analysis	a limited amount of patient data, limited data set size	Cleveland 93.44% IEEE 95.00%
Dalal et al 2023 [62]	UCI	QUEST, random forest, neural network, Bayesian network	use of advanced machine learning approaches, and early identification and therapy for at-risk patients	potential biases in data, limited generalizability, and variations in algorithm performance	99.10%
Khan et al 2023 [63]	Khyber Teaching Hospital and Lady Reading Hospital in Pakistan .	DT, RF, LR, Naïve Bayes, SVM	potential for effective prediction and diagnosis in hospital settings	need for detailed patient data and clinical datasets	0.8372 0.8501 0.8308 0.7474 0.8308
Ozcan and Peker 2023 [6]	heart disease dataset .	Classification and Regression Tree	Provides a high accuracy rate, the model is easily integrated into clinical decision support systems	Unstructured data was not taken. Further research is conducted to incorporate clinical, demographic, and social determinants to improve the quality and validity of the model.	87%

D. Result and Discussion

The table gives a summing up view of articles concentrating on different years from 2020 to 2023. They are based on multiple ways of predicting heart diseases. The findings present valuable insights, serving as a foundation for potential future research, showing the accuracy of the algorithms such as support vector machines and random forests. Logistic regression, decision trees and neural networks. It is interesting to note that in analysis of these very complex medical data sets, the capability to learn by machines is also emphasized. First, we will involve datasets like the Cleveland Heart Disease Dataset and those provided by sources like Kaggle, UCI, and hospitals as inputs.

Accuracy rates reported in these studies show wide variation, ranging from approximately 68% to 100%. This diversity illuminated how algorithm or dataset qualities vary and also affects the predictive performance. Increased prediction accuracy, reduced dimensionality and extraction of essential attributes are some of the advantages reported in researches.

These advantages highlight how machine learning can be used to increase the effectiveness and efficiency of heart disease prediction algorithms. Although research demonstrates that there are some positive results, the limitations, which include small sample sizes, the inability to generalize to other conditions, and inadequate increase in predictive accuracy as well as taken time, clearly suggest further research work.

Second, in later research, it should be investigated how the details of data collection and model performance relate to the results. Routine medical practice will be extremely shaped by the data from these tests. Machine learning-based models for prediction can substantially decrease the overall costs of cardiovascular disease and intensify early identification and early intervention, thereby lowering personal and population outcomes. Despite that, but for a careful consideration of these factors, the models could hardly be used to measure efficiency and outcome of such care.

E. Conclusion and Future Directions

In conclusion, the main findings emerged that machine learning techniques were used across different datasets, methodologies and algorithms to achieve high accuracy rates in predicting heart disease. Even despite the notable progress in predictive accuracy and identification of relevant features, there are still many hurdles to overcome, such as small sample sizes, limited application to other diseases, and the need for further research to increase prediction accuracy and shorten calculation times. Moreover, in future research endeavors, issues such as data completeness, model performance evaluation, and scalability limitations must be carefully considered. Despite these obstacles, the results of these studies demonstrate the potential of machine learning for early diagnosis of cardiovascular disease, preventive care, and cost-saving measures. The main goals of future research efforts should focus on addressing these obstacles and leveraging the advantages of machine learning to improve cardiovascular disease management and prediction.

Subsequent research efforts ought to focus on addressing the identified challenges and using machine learning's potential in predicting and managing cardiovascular illness. This means exploring new techniques, improving the caliber and quantity of datasets, enhancing the interpretability of models, and carrying out exhaustive validation tests in a range of clinical settings.

F. References

[1] Katarya., *Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) : 02-04, July 2020*. 2020.

[2] R. Katarya and S. K. Meena, "Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis," *Health Technol (Berl)*, vol. 11, no. 1, pp. 87–97, Jan. 2021, doi: 10.1007/s12553-020-00505-7.

[3] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, P. Singh, and N. Kumar, "Latest trends on heart disease prediction using machine learning and image fusion," in *Materials Today: Proceedings*, Elsevier Ltd, 2020, pp. 3213–3218. doi: 10.1016/j.matpr.2020.09.078.

[4] Yadav, *2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020) : conference proceedings : 5-7 March, 2020*. 2020.

[5] I. Tougui, A. Jilbab, and J. El Mhamdi, "Heart disease classification using data mining tools and machine learning techniques," *Health Technol (Berl)*, vol. 10, no. 5, pp. 1137–1144, Sep. 2020, doi: 10.1007/s12553-020-00438-1.

[6] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2022.100130.

[7] P. Anbuselvan, "Heart Disease Prediction using Machine Learning Techniques," 2020. [Online]. Available: www.ijert.org

[8] I. Ibrahim and A. M. Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 10–19, Mar. 2021, doi: 10.38094/jastt20179.

[9] M. Balakrishnan, A. B. Arockia Christopher, P. Ramprakash, and A. Logeswari, "Prediction of Cardiovascular Disease using Machine Learning," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1767/1/012013.

[10] C. Krittanawong *et al.*, "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Sci Rep*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-72685-1.

[11] K. Singh, Institute of Electrical and Electronics Engineers. Uttar Pradesh Section, IEEE Industry Applications Society, and Institute of Electrical and Electronics Engineers, *ICE3-2020 : International Conference on Electrical and Electronics Engineering : February 14-15, 2020*. 2020.

[12] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.

[13] V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," in *Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 177–181. doi: 10.1109/ICACCCN51052.2020.9362842.

[14] D. Shah, S. Patel, and S. K. Bharti, "Heart Disease Prediction using Machine Learning Techniques," *SN Comput Sci*, vol. 1, no. 6, Nov. 2020, doi: 10.1007/s42979-020-00365-y.

[15] G. Quer, R. Arnaout, M. Henne, and R. Arnaout, "Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review," *Journal of the American College of Cardiology*, vol. 77, no. 3. Elsevier Inc., pp. 300–313, Jan. 26, 2021. doi: 10.1016/j.jacc.2020.11.030.

[16] K. Seetharam *et al.*, "Applications of Machine Learning in Cardiology," *Cardiology and Therapy*, vol. 11, no. 3. Adis, pp. 355–368, Sep. 01, 2022. doi: 10.1007/s40119-022-00273-7.

[17] P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, “Artificial Intelligence, Machine Learning, and Cardiovascular Disease,” *Clinical Medicine Insights: Cardiology*, vol. 14. SAGE Publications Ltd, 2020. doi: 10.1177/1179546820927404.

[18] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-Learning-Based Disease Diagnosis: A Comprehensive Review,” *Healthcare (Switzerland)*, vol. 10, no. 3. MDPI, Mar. 01, 2022. doi: 10.3390/healthcare10030541.

[19] A. Arista, “Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19,” *Sinkron*, vol. 7, no. 1, pp. 59–65, Jan. 2022, doi: 10.33395/sinkron.v7i1.11243.

[20] B. Jijo and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[21] C. U. Kumari, A. S. D. Murthy, B. L. Prasanna, M. P. P. Reddy, and A. K. Panigrahy, “An automated detection of heart arrhythmias using machine learning technique: SVM,” in *Materials Today: Proceedings*, Elsevier Ltd, 2021, pp. 1393–1398. doi: 10.1016/j.matpr.2020.07.088.

[22] M. N. R. Chowdhury, E. Ahmed, M. A. D. Siddik, and A. U. Zaman, “Heart Disease Prognosis Using Machine Learning Classification Techniques,” in *2021 6th International Conference for Convergence in Technology, I2CT 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021. doi: 10.1109/I2CT51068.2021.9418181.

[23] H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, and M. Maaroufi, “A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques,” in *2020 International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020. doi: 10.1109/ATSIP49331.2020.9231760.

[24] D. Abdullah and A. M. Abdulazeez, “Machine Learning Applications based on SVM Classification: A Review,” *Qubahan Academic Journal*, pp. 1–10, Nov. 2021, doi: 10.48161/Issn.2709-8206.

[25] A. Batuhan Yilmaz, Y. Selim Taspinar, and M. Koklu, “Classification of Malicious Android Applications Using Naive Bayes and Support Vector Machine Algorithms,” *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2022, no. 2, pp. 1–6, 2022, doi: 10.1039/b000000x.

[26] R. YILMAZ and F. H. YAĞIN, “Early Detection of Coronary Heart Disease Based on Machine Learning Methods,” *Medical Records*, vol. 4, no. 1, pp. 1–6, Jan. 2022, doi: 10.37990/medr.1011924.

[27] G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima, and S. M. H. Akhter, “Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique with and Without Sequential Feature Selection,” *IEEE Access*, vol. 10, pp. 23808–23828, 2022, doi: 10.1109/ACCESS.2022.3153047.

[28] H. Aly Elzeheiry, S. Barakat, and A. Rezk, “Different Scales of Medical Data Classification Based on Machine Learning Techniques: A Comparative Study,” 2022, doi: 10.3390/app.

[29] M. S. Raja, M. Anurag, C. P. Reddy, and N. R. Sirisala, “Machine Learning Based Heart Disease Prediction System,” in *2021 International Conference on Computer*

Communication and Informatics, ICCCI 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402653.

[30] S. Farhad Khorshid and A. M. Abdulazeez, "BREAST CANCER DIAGNOSIS BASED ON K-NEAREST NEIGHBORS: A REVIEW PJAEE, 18 (4) (2021) BREAST CANCER DIAGNOSIS BASED ON K-NEAREST NEIGHBORS: A REVIEW," 2021.

[31] R. Nanmaran *et al.*, "Investigating the Role of Image Fusion in Brain Tumor Classification Models Based on Machine Learning Algorithm for Personalized Medicine," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/7137524.

[32] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10358-x.

[33] D. Chumachenko, I. Meniailov, K. Bazilevych, T. Chumachenko, and S. Yakovlev, "Investigation of Statistical Machine Learning Models for COVID-19 Epidemic Process Simulation: Random Forest, K-Nearest Neighbors, Gradient Boosting," *Computation*, vol. 10, no. 6, Jun. 2022, doi: 10.3390/computation10060086.

[34] M. M. Chowdhury, R. S. Ayon, and M. S. Hossain, "An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset," *Healthcare Analytics*, vol. 5, Jun. 2024, doi: 10.1016/j.health.2023.100297.

[35] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques," *ICT Express*, vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.icte.2021.08.021.

[36] A. G. Farizawani, M. Puteh, Y. Marina, and A. Rivaie, "A review of artificial neural network learning rule based on multiple variant of conjugate gradient approaches," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1529/2/022040.

[37] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Inform Med Unlocked*, vol. 19, Jan. 2020, doi: 10.1016/j.imu.2020.100330.

[38] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[39] D. Ananey-Obiri and E. Sarku, "Predicting the Presence of Heart Diseases using Comparative Data Mining and Machine Learning Algorithms," *Int J Comput Appl*, vol. 176, no. 11, pp. 17–21, Apr. 2020, doi: 10.5120/ijca2020920034.

[40] Princy, *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) : 13-15 May, 2020*. 2020.

[41] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Digit Health*, vol. 6, 2020, doi: 10.1177/2055207620914777.

[42] P. Sujatha and K. Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease," in *2020 IEEE International Conference for Innovation in Technology, INOCON 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/INOCON50539.2020.9298354.

[43] H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, and M. Maaroufi, "A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques," in *2020 International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020. doi: 10.1109/ATSIP49331.2020.9231760.
[44] H. H. Alalawi and M. S. Alsuwat, "Detection of Cardiovascular Disease using Machine Learning Classification Models," 2021. [Online]. Available: www.ijert.org
[45] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput Biol Med*, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104672.
[46] D. Bertsimas, L. Mingardi, and B. Stellato, "Machine Learning for Real-Time Heart Disease Prediction," *IEEE J Biomed Health Inform*, vol. 25, no. 9, pp. 3627–3637, Sep. 2021, doi: 10.1109/JBHI.2021.3066347.
[47] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," *Comput Intell Neurosci*, vol. 2021, 2021, doi: 10.1155/2021/8387680.
[48] K. Dissanayake and M. G. M. Johar, "Comparative study on heart disease prediction using feature selection techniques on classification algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2021, 2021, doi: 10.1155/2021/5581806.
[49] A. Garg, B. Sharma, and R. Khan, "Heart disease prediction using machine learning techniques," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1757-899X/1022/1/012046.
[50] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
[51] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1757-899X/1022/1/012072.
[52] S. P. Patro, G. S. Nayak, and N. Padhy, "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective," *Informatics in Medicine Unlocked*, vol. 26. Elsevier Ltd, Jan. 01, 2021. doi: 10.1016/j.imu.2021.100696.
[53] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *J Reliab Intell Environ*, vol. 7, no. 3, pp. 263–275, Sep. 2021, doi: 10.1007/s40860-021-00133-6.
[54] K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam, H. N. Chua, and S. Pranavanand, "Heart disease risk prediction using machine learning classifiers with attribute evaluators," *Applied Sciences (Switzerland)*, vol. 11, no. 18, Sep. 2021, doi: 10.3390/app11188352.
[55] A. S. Abdalrada, J. Abawajy, T. Al-Quraishi, and S. M. S. Islam, "Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: a retrospective cohort study," *J Diabetes Metab Disord*, vol. 21, no. 1, pp. 251–261, Jun. 2022, doi: 10.1007/s40200-021-00968-z.

[56] G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima, and S. M. H. Akhter, "Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique with and Without Sequential Feature Selection," *IEEE Access*, vol. 10, pp. 23808–23828, 2022, doi: 10.1109/ACCESS.2022.3153047.
[57] A. G, B. Ganesh, A. Ganesh, C. Srinivas, Dhanraj, and K. Mensinkal, "Logistic regression technique for prediction of cardiovascular disease," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 127–130, Jun. 2022, doi: 10.1016/j.gltp.2022.04.008.
[58] N. Bora, S. Gutta, and A. Hadaegh, "Using machine learning to Predict Heart Disease," *WSEAS Transactions on Biology and Biomedicine*, vol. 19, pp. 1–9, 2022, doi: 10.37394/23208.2022.19.1.
[59] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/1410169.
[60] A. Ullah, S. A. Khan, T. Alam, S. Luma-Osmani, and M. Sadie, "Heart disease classification using various heuristic algorithms," *International Journal of Advances in Applied Sciences*, vol. 11, no. 2, pp. 158–167, Jun. 2022, doi: 10.11591/ijaas.v11.i2.pp158-167.
[61] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes*, vol. 11, no. 4, Apr. 2023, doi: 10.3390/pr11041210.
[62] S. Dalal *et al.*, "Application of Machine Learning for Cardiovascular Disease Risk Prediction," *Comput Intell Neurosci*, vol. 2023, pp. 1–12, Mar. 2023, doi: 10.1155/2023/9418666.
[63] A. Khan, M. Qureshi, M. Daniyal, and K. Tawiah, "A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction," *Health Soc Care Community*, vol. 2023, pp. 1–10, Feb. 2023, doi: 10.1155/2023/1406060.