ISSN 2394-739X
IJRSET SEPTEMBER Volume 9 Issue 9
International Journal for Research in Science Engineering & Technology (IJRSET)
https://www.doi.org/10.5281/zenodo.7049471

A SYSTEMATIC REVIEW OF PREDICTION OF HEART DISEASES USING SUPERVISED MACHINE LEARNING ALGORITHMS

[1] R. KARTHIKEYAN, [2] B. SELVANANDHINI
[1] Research Scholar, [2] AP, Research Superviser,
[1, 2] Department of Computer Science,
[1, 2] Pollachi College of Arts and Science,
[1, 2] Tamil Nadu India.

ABSTRACT - Machine learning algorithms are playing a major role in prediction of heart diseases. As compared to other diseases, heart disease is getting more attention today since it affects humans unexpectedly. Early detection of heart diseases will save human lives. Many researchers have published numerous articles in prediction analysis of heart disease with the support of machine learning algorithms. Especially, supervised learning algorithms are getting more attention with respect to heart diseases prediction. This paper provides a system review of heart diseases prediction with the help of supervised learning algorithms. The impact that heart diseases have on human lives is examined in this paper. It also discusses the various supervised learning algorithms. Tools and datasets used for heart disease prediction are also presented. Performance metrics that are used for heart disease prediction are discussed and examined. We have examined the comparison of various supervised machine learning algorithms with respect to heart diseases prediction. The review clearly identifies the most applied supervised machine learning algorithm by a significant number of research studies.

Keywords: [Heart Diseases, Prediction, Machine Learning Algorithms, Supervision learning algorithms.]

1. INTRODUCTION

Artificial intelligence, the Internet of Things (IoT), machine learning algorithms, medical image processing, data mining and others play a major role in healthcare. As a result, they have developed solutions to address the problems of the traditional healthcare system. Humans are impacted by various health issues due to a variety of reasons. Heart diseases are getting more attention due to the fact that they are unexpectedly life-threatening. In the world population, cardiovascular diseases or heart diseases are the most common causes of human death. This disease affects the entire function of the blood vessels and causes infections in the coronary arteries. The result is weakening of the body and death (A.L.Bui et al., 2011; J.L.Opez-Sendon, 2011).

Early diagnosis of heart diseases will save human lives. This prediction can be achieved by machine learning algorithms. Over the last few years, machine learning algorithms and related techniques have been applied to heart datasets for automatic prediction, diagnosis, and treatment of heart diseases. In recent years, databases called Electronic Health Records (EHR) have been built by many researchers and health-related organizations. By identifying risk factors of heart disease, such databases assist in improving the quality of life for heart disease patients, as well as providing precise and appropriate diagnoses. Using computer assisted tools and techniques, computational sciences are being applied to the healthcare sector. With modern tools and techniques, machine learning algorithms are playing a major role in prediction. Their approach was considered as the most relevant and emerging method for predicting and diagnosing heart diseases. Machine learning algorithms are programmed in such a way to process and analyse the input data and predict the output values.

In machine learning algorithms, Support Vector Machine (SVM), Decision Tree, Random Forest Algorithm, Na*ve Bayes, Kernel Support Vector Machine, K Nearest Neighbour Algorithm, Artificial Neural Networks, Linear Regression, Gradient Boosting, Lightgbm, Catboost, Adaboost, Xgbrf, Logistic Regression, K Means Clustering and more algorithms are used to predict and diagnose heart disease. (Wu et al., 2007 and FajrIbrahemAlarsan et al., 2019, J. P. Li et al., 2020, D. Shah et al., 2020, M. S. Amin et al., 2019, S. Mohan et al., 2019, Katarya, R, 2021). This paper only discusses supervised machine learning algorithms such as Logistic Regression, Support Vector Machine, Random Forest, Decision Tree, Kernel Support Vector Machine, Nave Bayes, and K-Nearest Neighbour. This paper analyses the performance of the above mentioned algorithm with respect to various performance metrics. Finally, this paper suggests suitable machine learning algorithms for heart diseases prediction and diagnosis.

The remainder of the paper is organized as following section discusses the various machine learning algorithms, heart diseases and its related impacts, the role of machine learning algorithms in heart diseases, data resources and data extraction, finally the proposed work reviews the existing machine learning algorithms for heart diseases.

2. BACKGROUND

2.1 HEART DISEASES

In the recent report by the World Health Organization (WHO), about 17.9 million people died of Cardiovascular Diseases (CVD) and they accounted for 32% of all global deaths. Besides, out of this 17.9 million, 38% of deaths result from CVD in people aged 70 and below. Cardiovascular diseases are commonly referred to as heart diseases. It is generally referred to as a variety of different types of heart diseases. Heart diseases include congenital, rheumatic, deep vein thrombosis and pulmonary embolism, cerebrovascular, coronary heart disease, and peripheral arterial disease.

Coronary Artery Disease (CAD) is considered the most common type of heart disease. It typically affects the blood flow to the heart, resulting in a heart attack. Heart disease is more harmful than other diseases because it affects people without their knowledge. Until a person experiences symptoms and signs of a heart attack and heart failure, it cannot be diagnosed. As a result of frequent use of tobacco and alcohol, physical inactivity, a poor diet, and being overweight, CVD risk factors can be raised (T. Leiner et al., 2019).

Heart disease symptoms include shortness of breath, discomfort in the throat, upper belly area, jaw and neck. Normally, he feels pain in the chest area, he also feels pressure and tightness in the chest region. In addition to, coldness and weakness in legs and arms. Other than this, diabetes and frequent use of hypertension medications can also lead to heart diseases. The above factors have to do with the physical body of humans. Other determinants of heart diseases include population ageing, globalization, urbanization, changes in culture and economics, stress, poverty, and genetics. These factors also have an impact on heart diseases.

Early detection of heart diseases and appropriate treatment prevent heart related issues and can also prevent premature death. Medical field uses information technologies like Image Processing, Data Mining, Machine Learning, Artificial Intelligence, Internet of Things, etc., to deal with heart-health issues. The above technologies can be used to treat heart related issues but also other diseases. This review article will discuss the role of machine learning algorithms in detection of heart diseases.

1.2 Machine Learning Algorithms

The following section discusses the various machine learning algorithms used for heart diseases prediction. Machine learning algorithms have a crucial role in the prediction of heart diseases almost universally. To replace the manual tasks, machine learning algorithms are used. All the human tasks are fully automated and the automation process ranges from individual chess games to automatic decision making in industries. It plays a major role in predicting what will happen in the future. As a result of technological progress, it happened. Machine Learning is an application of Artificial Intelligence to allow systems to learn automatically without human intervention. Moreover, the system is able to adapt to the environment without programming by analysing the input data and predicting the output value using the experience acquired. The training data are used to train the machine learning algorithms. Then, it makes decisions and predictions based on the past data when updated data feed into the system. Figure 2 depicts the classification of machine learning algorithms. There are four major types of machine learning algorithms: supervised learning algorithms, unsupervised learning algorithms, semi-supervised learning algorithms, and reinforcement learning algorithms (Rani, P et al., 2021).

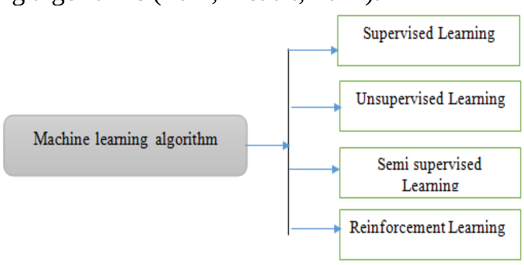

Figure 1. Classification of Machine Learning Algorithms

2.2.1 Supervised Learning Algorithms:

The supervised learning algorithm consists of a model that predicts the future with the help of labels. The label dataset is one where you already know the target answer, which is historical data or past data. Using known data as input, the machine learns and analyses the correlation of the fed data using features. Then, when new data is fed into the system without any label associated with those data, the machine is able to predict the new fed data accurately. This is done with the help of historical data or past data. The supervised learning algorithm is classified into two major categories: regression and classification.

The output value in regression is usually continuous and real. There is a relationship between two or more variables whereby changes in one variable are associated with corresponding changes in other variables. One variable is called a dependent variable and another variable is called an independent variable. When these two variables are fed into the machine, the machine tries to learn the relationship between these two variables. Afterward, the machine is trained, and finally it predicts the dependent variable based on the independent variable.

Classification involves classifying the output variables into two or more categories. The machine will identify the category of new observations based on the given training data. Generally, it will take labelled input data. The machine will learn the given trained dataset or observation, and based on this it will classify the new observation into a number of groups or classes. The following figure depicts the supervised learning algorithm.

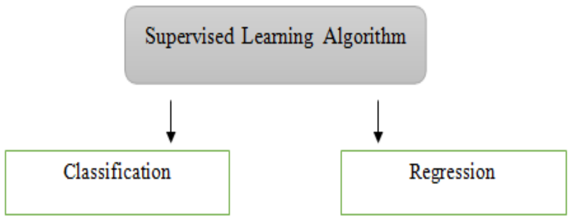

Figure 2. Classification of Supervised Machine Learning Algorithms

The classification model is again classified into two major groups: linear models and non-linear models.

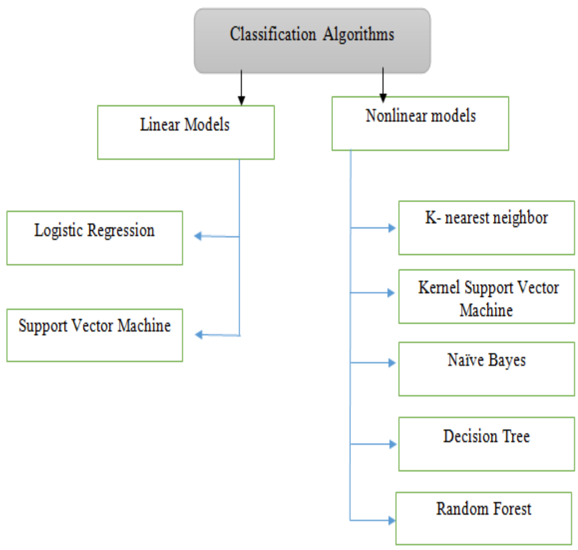

Figure 3. Organization of classification algorithms

The following figure depicts the prediction of heart diseases with respect to classification algorithms.

ISSN 2394-739X
IJRSET SEPTEMBER Volume 9 Issue 9

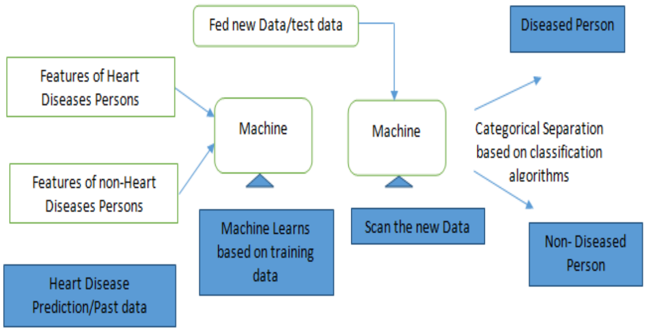

Figure 4. The prediction of heart diseases with respect to classification algorithms.

The following figure depicts the prediction of heart diseases with respect to regression algorithm.

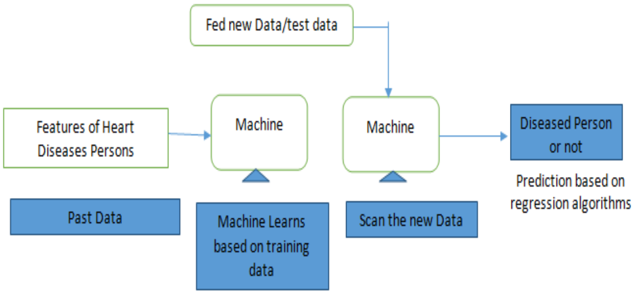

Figure 5. The prediction of heart diseases with respect to regression algorithm.

2.2.1.1 Logistic regression
As with linear regression (Sajja, G. S et al., 2021), the logistic regression is derived from the logit function. It is used to predict whether a binary value will occur, such as yes or no. In addition, it is also useful to define the relationship between the dependent variable and one or more independent variables. There are various types of logistic regression such as ordinal, multinomial and binary. Logistic regression is used to find the most relevant risk factors as well as overall risks of heart diseases.

2.2.1.2 Support vector machine
Support Vector Machine (SVM) (Faieq, A. K et al., 2022) is used for both classification and regression. This defines a boundary or boundary line that separates the n-dimensional spaces into several classes. The newly added data point will be added to the relevant category in the future. Thus, the ideal line boundary is called the hyperplane. SVM will choose the great vectors or points that are used to create a hyperplane. These extreme points are called a support vector, so that this algorithm is known as a support vector machine. SVM is classified into two major types: linear SVM and non-linear SVM.

2.2.1.3 K-nearest neighbour (KNN) algorithm
KNN (Chethana, C et al., 2021) algorithm is used in both regression and classification. However, it is mainly used for classification purposes. As the algorithm does not make any assumptions about the underlying data, it is called a non parametric algorithm. Additionally, it is also called the lazy learner algorithm as it does not learn anything from the trained data immediately. Instead, it will store the datasets and perform functions over a period of time based on the corresponding features. Based on the available data, it will try to find relevant similarity data points, and then classify those data points into well-suited categories. This is how KNN algorithm works.

2.2.1.4 Kernel support vector machine
The term kernel is used in conjunction with SVM because mathematical functions have been used along with SVM to manipulate the data. As a rule, it will take data as input, and then it will transform that data into the required form for processing the data. Numerous kernel functions are available, including Gaussian kernel, Gaussian Kernel Radial Basis Function (RBF), Polynomial kernel, and Sigmoid kernel (Rubini, P. E., et al., 2021).

2.2.1.5 Naïve Bayes
Nave and Bayesian algorithms are both used in this algorithm. This algorithm is mostly used in text classification with extreme datasets. Taking quick predictions from the dataset, this is considered a fast machine learning algorithm. The term Naïve refers to a state of total independence between the occurrence of one feature and those of other features. Another term is Bayesian, which is founded on Bayes' theorem. The Bayes theorem is composed of Bayes rules or Bayes laws that govern the likelihood of a hypothesis given prior knowledge. Theorem is based on conditional probability. In this model, the trained dataset is converted into a frequency table. Then, based on the probability of the given features, a likelihood table will be generated. Finally, the Bayes theorem will be used to calculate the posterior probability (Arumugam, K et al., 2021).

2.2.1.6Decision Tree
This model is called a decision tree because it resembles a tree. This structure starts with a root, then branches, and finally it is shaped like a tree. The algorithm mimics human reasoning and looks like a tree structure, making it easy to understand. In a decision tree, all the internal nodes correspond to the characteristics of the dataset. The decision rules are represented by branches. The outcome of a given dataset is represented by leaves. There are two key nodes in decision trees, namely the leaf node and the decision node. There is a distinction between decision nodes and leaf nodes. Decision nodes contain multiple branches, whereas leaf nodes are the output of the decision nodes and they do not contain any further branches. Adaptive Selection Measure (ASM) is used to select the best attribute that can be achieved by two metrics such as Gini index and information gain (Ali, M. M. et al., 2021).

2.2.1.7 Random Forest
Random forest algorithm (Katarya, R et al., 2021) is used for both regression and classification. A decision tree that combines more than one decision tree. The approach works based on the concept of ensemble learning, which combines multiple classifiers to solve a complex problem. The random forest consists of multiple decision trees with various subsets of dataset. Each decision tree produces an output, and the average output will be considered to be the overall accuracy of the prediction. The larger number of decision trees will lead to more accuracy. Additionally, it is used to avoid overfitting issues.

3. METHODOLOGY
Many researchers (M. Abdar et al., 2019, I. D. Mienye et al., 2020, Kumar, M.N et al., 2018, Tougui, I et al., 2020, Katarya, R et al., 2021, Ali, M. M et al., 2021) follow

the same methodologies with respect to prediction of heart diseases. Basically, six steps are involved in the detection or diagnosis of heart diseases. The steps include loading the dataset gathered from various medical databases, pre-processing the dataset, selecting the appropriate attributes, running machine learning algorithms, applying performance evaluation metrics and analyzing the results of the machine learning algorithm. The following figure depicts the common methodology used by researchers.

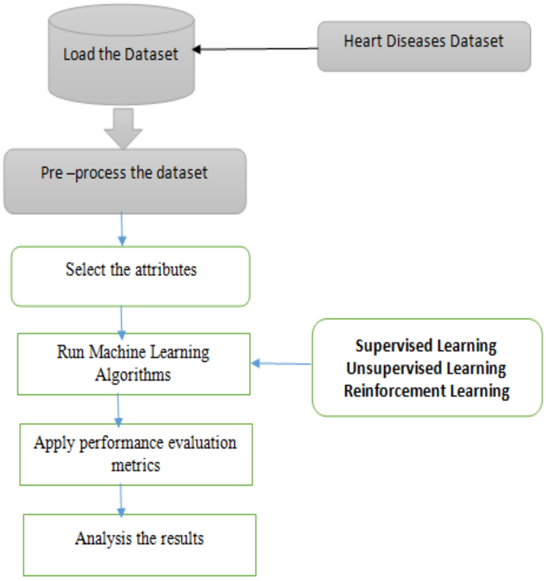

Figure 6. Common methodology used for heart diseases prediction

4. Tools that are used for analysing the machine learning algorithms.

The following section discusses the most relevant tools that are used for analysing heart diseases prediction. To train the models, create your own algorithms, identify new methods, and play with the data, these tools are used. The following table demonstrates the various tools and their purpose.

S.No	Name of the tool	Open source/Proprietor	Details
1	Scikit-Learn	Open source	Classification, Clustering and Regression
2	Knime	Open source	Support operations like manipulation and data mining
3	Tensor flow	Open source	All machine learning models and neural networks
4	Weka	Open source	Support to implement all the machine learning models
5	Rapid miner	Open source	Support to implement all the machine learning models and mainly used for non-programmer
6	Shogun	Open source	Most supported algorithms include Support Vector Machine, K-means, KNN
7	Numba	Open source	Python based tool support all the machine learning algorithms
8	Winpython	Open source	Python based tool
9	Caffe	Open source	It supports all the deep learning algorithms and it is based on C++
10	Open NN	Open source	Especially developed for neural network and it is based on C++.

Table 2. Commonly used tools for heart diseases prediction

5. Performance evaluation metric and commonly used dataset for heart diseases prediction

This section deals with the performance evaluation metrics and datasets used by most researchers with respect to heart diseases prediction. The prediction ability of heart diseases is largely determined by the Receiver Operating Characteristics (ROC) curve and confusion matrix. The confusion matrix is also known as a contingency matrix or error matrix. The graphical representation of ROC curve is shown below. It simply stated the performance of various classification models under various threshold classifications. Two parameters such as the true positive rate and the false positive rate can be used to draw the curve.

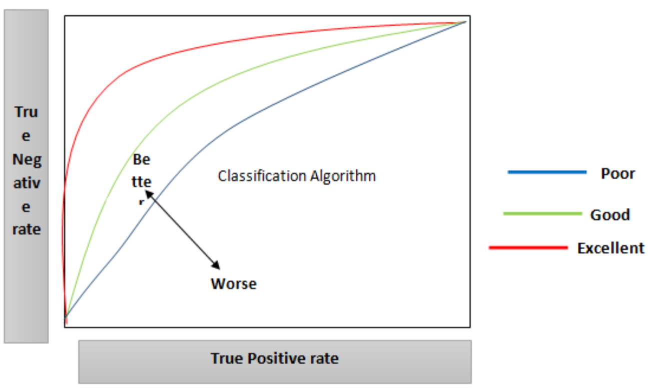

Figure 7. Illustration of RoC

ISSN 2394-739X
IJRSET SEPTEMBER Volume 9 Issue 9

Similarly, the confusion matrix provides four factors, such as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The graphical representation of the confusion matrix is shown in the figure. As a 2x2 matrix, it represents the data.

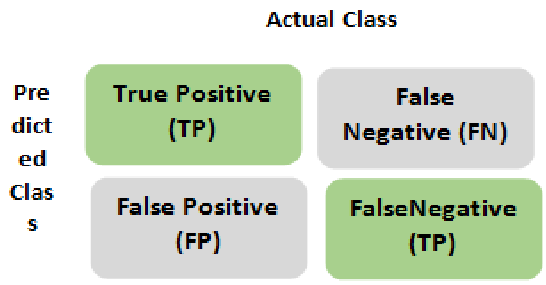

Figure 8. Illustration of Confusion matrix

True positive denotes positive cases identified by the classifier and similarly true negativite denotes negative cases identified by the classifier. This is also true for false positive and false negative cases. Besides, the following performance evaluation measurements which are based on confusion matrix are used to analyse the performance of various classification algorithms. The following figure shows the various performance evaluation measurements.

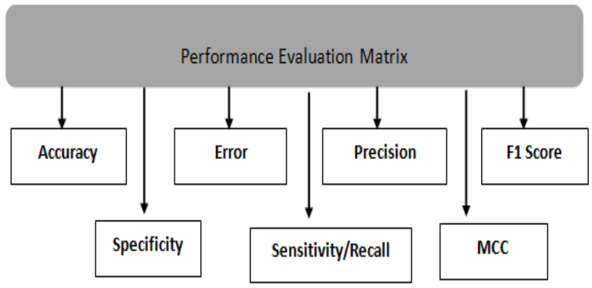

Figure 9. Performance evaluation metrics

The following performance metrics are widely used to evaluate machine learning algorithms.

5.1 Accuracy

A model's accuracy is determined based on the following equation and is defined as whether the model correctly predicts the data or not. Another way to define it is the portion of the estimate that is right.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

5.2 F1 score

F1 score defines the percentage of correct prediction that a model has made. It will be calculated based on the below equation.

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Presicion} * \text{Recall}}$$

5.3 Precision

Precision can be described as how a model is able to make a reliable prediction, and it can be described using the following equation.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

5.4 Sensitivity or recall

Sensitivity is also known as recall and can be defined as the percent of correct predictions found by a given model. It can be defined by the following equation.

$$\text{Sensitivity or Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

5.5 Specificity

A model's predictive ability is determined by the way in which it identifies a true negative of a model.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Error

It denotes the overall incorrect prediction made by the given model and it can be represented as

$$\text{Classification Error} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

5.6 Matthews Correlation Coefficient (MCC)

This predictor only generates a high score if it gains positive results in all four confusion matrices, such as true positive, false positive, true negative, and false negative.

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{\text{TP} + \text{FP}} * (\text{TP} + \text{FN}) * (\text{TN} + \text{FP}) * (\text{TN} + \text{FN})}$$

7. Dataset used for heart diseases prediction

The following table illustrates the most commonly used dataset by the research community for heart disease prediction. The number of attributes used in the dataset are 14.

S.No	Attribute name	Description	Values assigned
1	Age	Age of a person in years	Numbers
2	Sex	Denotes the gender of a person	Male or Female or Transgender
3	CP	Type of chest pain of a person	0 – Typical angina; it causes low blood supply to the heart 1 – Atypical angina; pain not related to heart 2 – Non – angina pain ; it represents esophageal spasms i.e not related to heart 3 – Asymptomatic; No sign of diseaes
4	trestbps	It represents the resting blood pressure	mm or HG; above 130mm-140 mm will be consider as serious concern
5	Fbs	It denotes the fasting blood sugar	>120mg/dl ; 1 – true;0-false
6	chol	It denotes the serum cholesterol	>200mg/dl will be consider as serious concern
7	thalach	It denotes the maximum heart rate	It ranges the value between 71 to 202

8	restecg	It denotes the resting electrocardiographic	O – nothing related to heart diseases 1-ST-T wave abnormally; possibility of mild symptoms to severe symptoms 2- possibility of left ventricular hypertrophy; possibility of enlarging heart's pumping chamber
9	exang	It denotes the exercise induced angina	1 – Possibility of Yes 0 – Possibility of No
10	slope	It denotes the slope of the peak exercise ST segment	It consists of three values; 0 – Upsloping 1 – flat 2 – downsloping
11	Ca	It denotes the number of major vessels that coloured by fluoroscopy	Values ranges from 0 to 3
12	thal	It denotes the thalasesemia	It consists of three values; 3 – normal 6 – fixed defect 7 - reversible defect
13	oldspeak	It denotes the ST depression include by exercise relevant to rest	
14	num	It denotes the number of diagnoses heart diseases.	It denotes 0 – absence 1,2,3 and 4 - present

Table 3. Commonly used dataset for heart diseases prediction

8. Review of machine learning algorithms

The following section discusses some of the notable supervised machine learning algorithms.

Amin UL Haq et al., 2018 developed a hybrid intelligence system to predict heart diseases. The proposed model consists of the following phases: pre-processing, feature selection, cross validation, machine learning classifiers and classifier performance evaluation methods. With respect to machine learning algorithms, the following machine learning algorithms were applied: NB, DT, SVM, KNN, ANN and logistic regression. Among the algorithms, logistic regression has shown excellent results with an accuracy of 89%.

A machine learning algorithm based on particle swarm optimization (PSO) and ant colony optimization (ACO) for the prediction of heart diseases has been proposed by YounessKhourdifi et al., 2018. The proposed method makes use of Fast Correlation based Feature Selection (FCBF) as well as different classification algorithms such as Support Vector Machine, Naive Bayes, Random Forest, Artificial Neural Network, and K-nearest neighbour. The combination of these two makes a hybrid approach to predicting heart disease. Performance metrics used in this proposal are precision, recall, accuracy, and F1 score. It shows the maximum accuracy of 99.65% with the optimized model proposed by PSO, ACO and FCBF. Fair IbrahemAlarsan et al., 2019 developed an ECG-based machine learning approach for predicting heart diseases using ECG features. With the help of gradient boosted tree and random forest algorithms, the proposed model has been implemented.

Mohd Ashraf et al., 2019 developed a deep neural network based heart disease prediction. Data selection, Data preprocessing, the use of deep neural networks in heart diseases prediction, and finally validation and result analysis have all been performed using R programming. The proposed work has also been compared with other machine learning algorithms in terms of performance metrics such as mean square error and accuracy.

Abdul Saboor et al., 2022 proposed a heart diseases prediction with various machine learning algorithms such as random forest, logistic regression, support vector machine and more. They have analysed the proposed method with various performance metrics such as f-measures, accuracy, sensitivity and specificity. The following table depicts the comparison of various machine learning algorithms.

Authors name	**Year of publication**	**Algorithm used**	**Dataset And type of data**	**Tools/ Languages**	**Prediction performance**	**Best algorithm with accuracy**
Amin UL HAQ et al	2018	NB, DT, SVM, KNN, ANN and Logistic Regression	Cleveland heart disease dataset 2016, Medical Data	Python	Accuracy, Specificity, Sensitivity, MCC, AUC and Processing Time	Logistic Regression, 89%
YounessKhourdifi et al.,	2018	Support Vector Machine, Naive Bayes, Random Forest, Artificial Neural Network, and K-nearest	UCI machine learning repository, Medical Dataset	Weka	TP Rate, FP Rate, Precision, Recall, F-Measusre	PSO, ACO and FCBF, 99.65%

		neighbour, particle swarm optimization (PSO) and ant colony optimization (ACO) and Fast Correlation based Feature Selection (FCBF)				
Fair Ibrahem Alarsan	2019	Gradient boosted tree and Random forest algorithms	MIT-BIH Arrhythmia and MIT-BIH Supraventricular Arrhythmia	ML-libs and Scala language	Accuracy, Specificity, Sensitivity, Correct classification	GDB Tree algorithm with 96.75%, random Forest for binary classification with 97.98%
Mohd Ashraf et al.,	2019	Deep Neural Network	UCI repository	R Programmi ng	Mean square error and Accuracy	Deep neural network with accuracy of 95%
Abdul Saboor et al.,	2022	Logistic regression, support vector machine, random forest and	Cleveland heart disease dataset, available from the University of California, Irvine (UCI) online repository	R Programmi ng	Precision, recall, f measure	Support vector machine with the accuracy of 96.72%
Rajkum arGanga ppaNad akinama ni et al.,	2021	Random forest, linear regression, naïve bayes, J48 and JRIP	Hungarian and Statlog (heart) dataset	Weka	Mean absolute error,	Random forest algorithm with maximum accuracy of 99.81%
Md. Asfi-Ar-Raihan Asif et al.,	2021	Random forest, Naïve Bayes, Adaboost, K-nearest neighbour, XGboost	The University of California, Irvine (UCI) data repository	Weka	Accuracy, Precision, Sensitivity, Specificity, F1-Score	Soft voting assemble classifer with accuracy of 92%
Xiao-Yan Gao et al.,	2021	K-nearest neighbour Support vector machine Random forest Naive bayes Ensemble algorithms Decision tree	Cleveland heart disease dataset	Python	accuracy, recall, precision, F-measure, and ROC.	Decision tree with accuracy of 98.6%
Rohit Bharti et al.,	2021	Decision tree, Logistic Regression, XGBoost, Support Vector Machine, k nearest neighbor	Public Health Dataset, Cleveland, Hungary, Switzerland, and Long Beach	R Programmi ng	Specificity, Sensitivity, Accuracy	K-nearest neighbor accuracy with 84.8%
RituAg grawal et al.,	2020	Random forest, Decision tree, Support Vector machine, KNN, LDA, GBC	Heart Failure Clinical Record Data Set 2020	Python	Confusion Matrix, Roc Curve, Precision, Recall Rate, accuracy, F1-Score	Random forest with accuracy of 86.67%
Vijeta Sharma et al.,	2020	Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree	Cleveland Heart Disease Dataset	Weka	Precision, Recall, Accuracy, ROC	Random forest with accuracy of 99%

Table 1. Comparative analysis of various supervised learning algorithms

CONCLUSION AND FUTURE DIRECTION

The use of machine learning algorithms, especially supervised machine learning algorithms, in healthcare services plays a major role in the prediction of heart diseases by analysing Medical Health Records (MHRs). Researchers and medical professionals are facing numerous problems with respect to heart diseases diagnosis and prediction. Though many researchers are addressing the problems with a variety of computer-aided solutions such as different machine learning algorithms, the need for solutions to the above said problem is still in demand as it has not been fully addressed. To overcome these issues, in this paper a systematic review has been conducted with the aid of supervised machine learning algorithms. The purpose of this review is to gain insight into how machine learning algorithms have been used for the prediction of heart disease. The following machine learning algorithms are discussed: decision trees, random forests, support vector machines, K nearest neighbors, kernel support vector machines, followed by an introduction to several prediction algorithms followed by different supervision methods. These algorithms work with respect to heart disease prediction. In addition, several factors related to heart prediction have been examined. Various performance evaluation metrics have also been discussed for analyzing the performance of the machine learning algorithms. In addition, various tools that are used for heart disease prediction have also been discussed.

In the last part of the paper, we have compared various machine learning algorithms in terms of their performance. As a result of the systematic review, we conclude that the random forest algorithm is an appropriate algorithm for heart disease prediction due to its high accuracy. Furthermore, the random forest algorithm provides better accuracy results. Further research will explore the various machine learning algorithms including unsupervised machine learning algorithms, semi supervision machine learning algorithms and reinforcement learning.

REFERENCES

[1]. Amin UlHaq, Jian Ping Li, Muhammad HammadMemon, Shah Nazir, and Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, 2018.

[2]. FajrIbrahemAlarsan and MamoonYounes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms" ,J Big Data (2019) 6:81

[3]. YounessKhourdifi, Mohamed Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", International Journal of Intelligent Engineering and Systems, Vol.12, No.1, 2019

[4]. Mohd Ashraf, M. A. Rizvi and Himanshu Sharma, "Improved Heart Disease Prediction Using Deep Neural Network" ,Asian Journal of Computer Science and Technology, ISSN: 2249-0701 Vol.8 No.2, 2019, pp. 49-54

[5]. Davari, D. A. et al. 2017. Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Computer Methods and Programs in Biomedicine. 138, (2017), 117--126.

[6]. Abdul Saboor, Muhammad Usman, SikandarAli,AliSamad, Muhmmad Faisal Abrar , and NajeebUllah , " A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, Volume 2022, Article ID 1410169, 9 pages

[7]. RajkumarGangappaNadakinamani, A. Reyana ,Sandeep Kautish , A. S. Vibith, Yogita Gupta, Sayed F. Abdelwahab , and Ali Wagdy Mohamed, "Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques", Computational Intelligence and Neuroscience, Volume 2022, Article ID 2973324, 13 pages

[8]. Md. Asfi-Ar-Raihan Asif, Mirza MuntasirNishat, Fahim Faisal, Rezuanur Rahman Dip, Mahmudul Hasan Udoy, Md. FahimShikder and Ragib Ahsan, "Performance Evaluationand Comparative Analysisof and Different Machine Learning Algorithms in Predicting CardiovascularDisease", Engineering Letters, 29:2,2021

[9]. Xiao-Yan Gao,Abdelmegeid Amin Ali, Hassan Shaban Hassan, and Eman M. Anwar , "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method", Complexity, Volume 2021, Article ID 6663455, 10 pages

[10]. RohitBharti,AdityaKhamparia ,Mohammad Shabaz ,Gaurav Dhiman ,SagarPande and Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Computational Intelligence and Neuroscience, Volume 2021, Article ID 8387680, 11 pages

[11]. RituAggrawal, Saurabh Pal, "Sequential Feature Selection and Machine Learning Algorithm-Based Patient's Death Events Prediction and Diagnosis in Heart Disease", SN Computer Science (2020) 1:344

[13]. Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta, "Heart Disease Prediction using Machine Learning Techniques" , ISBN: 978-1-7281-8337-4/20/$31.00 ©2020 IEEE, 2020 2nd International Conference on Advances in Computing,

[14]. Communication Control and Networking (ICACCCN) L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," Nature Reviews Cardiology, vol. 8, no. 1, pp. 30–41, 2011.

[15]. J. L´opez-Send´on, "The heart failure epidemic," Medicographia, vol. 33, pp. 363–369, 2011. X. Wu and V. Kumar, Top 10 Algorithms in Data Mining, Springer, Berlin, Germany, 2007.

[16]. FajrIbrahemAlarsan,andMamoonYounes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms", J Big Data (2019) 6:81

[17]. Abdul Saboor,Muhammad Usman ,Sikandar Ali , Ali Samad ,Muhmmad Faisal Abrar ,and NajeebUllah, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, 2022

[18]. Cardiovascular Diseases (Cvds), "World health organization," https://www.who.int/news-room/fact%20sheets/detail/cardio vascular-diseases-(cvds), 2022.

[19]. A Coronary, "Heart disease," Available from: https://www.aihw.gov.au/reports/australias-health/coronaryheart-disease, 2022.

[20]. 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.

[21]. D. Shah, S. Patel, and S. Kumar Bharti, Heart Disease Prediction Using Machine Learning Techniques, Springer Nature Singapore Pte Ltd, Berlin, Germany, 2020.

[22]. M. S. Amin, Y. K. Chiam, and K. D. Varathan, “Identification of significant features and data mining techniques in predicting heart disease,” Telematics and Informatics, vol. 36, pp. 82–93, S. Mohan, C.Irumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” IEEE Access, vol. 7, pp. 81542–81554, 2019.

[23]. M. Abdar, W. Ksia̧ ̇zek, U. R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak, “A new machine learningtechnique for an accurate diagnosis of coronary artery disease,” Computer Methods and Programs in Biomedicine, vol. 179, Article ID 104992, 2019.

[24]. Mienye, Y. Sun, and Z. Wang, “An improved ensemble learning approach for the prediction of heart disease risk,” Informat. Med. Unlocked, vol. 20, Article ID 100402, 2020.

[25]. Kumar, M.N.; Koushik, K.V.S.; Deepak, K. Prediction of heart diseases using data mining and machine learning algorithms and tools. Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. 2018, 3, 887–898.

[26]. Tougui, I.; Jilbab, A.; El Mhamdi, J. Heart disease classification using data mining tools and machine learning techniques. Health Technol. 2020, 10, 1137–1144.

[27]. T. Leiner, D. Rueckert, A. Suinesiaputra et al., “Machine learning in cardiovascular magnetic resonance: basic concepts and applications,” Journal of Cardiovascular Magnetic Resonance, vol. 21, no. 1, pp. 61–14, 2019.

[28]. Katarya, R., &Meena, S. K. (2021). Machine learning techniques for heart disease prediction: a comparative study and analysis. Health and Technology, 11(1), 87-97.

[29]. Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Computers in Biology and Medicine, 136, 104672.

[30]. Rani, P., Kumar, R., Ahmed, N. M., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, 7(3), 263-275.

[31]. Sajja, G. S., Mustafa, M., Phasinam, K., Kaliyaperumal, K., Ventayen, R. J. M., &Kassanuk, T. (2021, August). Towards application of machine learning in classification and prediction of heart disease. In 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1664-1669). IEEE.

[32]. Faieq, A. K., &Mijwil, M. M. (2022). Prediction of Heart Diseases Utilising Support Vector Machine and Artificial Neural Network. Indonesian Journal of Electrical Engineering and Computer Science, 26(1), 374-380.

[33]. Chethana, C. (2021, May). Prediction of heart disease using different KNN classifier. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1186-1194). IEEE.

[34]. Rubini, P. E., Subasini, C. A., Katharine, A. V., Kumaresan, V., Kumar, S. G., &Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. Annals of the Romanian Society for Cell Biology, 904-912.

[35]. Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2021). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings.

[36]. Cleveland Clinic Foundation, “Heart disease data set,”Available at http://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[37]. Statlog, “Heart disease data set,” Available at: https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29.

[38]. Z. Alizadeh Sani, “Heart disease data set,” Available at: https://archive.ics.uci.edu/ml/datasets/Z-Alizadeh+Sani.