GSJ: Volume 12, Issue 1, January 2024, Online: ISSN 2320-9186

www.globalscientificjournal.com

A Review on Algorithm Based Techniques used for Heart disease Prediction.

Anazodo William, Okengwu U. A. , Odia K. M.

CENTER FOR INFORMATION AND TELECOMMUNICATIONS ENGINEERING.

UNIVERSITY OF PORT HARCOURT

Abstract

Reviewing Algorithm-based Techniques is crucial for predicting heart disease, given its global significance as one of the most fatal issues that remains invisible until it abruptly manifests upon reaching its limits. Accurate and timely diagnosis is imperative in addressing this condition. Although the healthcare industry generates vast amounts of data daily, its efficient utilization by researchers and practitioners is lacking. Despite the wealth of data, the healthcare industry is characterized by a knowledge deficit. Various data mining and machine learning techniques and tools are available to extract valuable insights from databases, facilitating more precise diagnosis and decision-making. As research on predicting heart disease advances, there is a growing need to consolidate the incomplete body of work in this field. This research paper aims to summarize recent studies, presenting comparative results on proposed heart disease prediction systems and drawing analytical conclusions. The investigation reveals that Naive Bayes with Genetic Algorithm, Decision Trees, and Artificial Neural Networks demonstrate effectiveness. However, the enhanced Naïve Bayes Algorithm developed in this study stands out for improving the accuracy of the heart disease prediction system across different scenarios. This paper provides a comprehensive summary of commonly used data mining and machine learning techniques, along with their complexities in the context of heart disease prediction.

Keywords: Data mining, Machine learning, Heart disease, Classification, Naive Bayes, Artificial Neural Networks, Decision Trees, Associative Rule

1.0 INTRODUCTION

There are many studies that have explored the prediction of diseases through various machine learning techniques and algorithms applicable in medical settings. This paper conducts a comprehensive review of research papers, examining the methodologies and outcomes employed in these studies. The primary focus of this research is the performance evaluation of disease prediction approaches, specifically exploring different variations of supervised machine learning algorithms. In recent years, there has been a notable surge in interest within the data science research community regarding disease prediction and, more broadly, medical informatics.

This heightened interest can be attributed to the widespread integration of computer-based technologies into the healthcare sector, manifesting in various forms such as electronic health records and administrative data. The subsequent availability of extensive health databases has provided researchers with substantial resources for investigation. This research particularly centers on disease risk prediction models utilizing machine learning algorithms, including support vector machines, logistic regression, and artificial neural networks. Specifically, the emphasis is on supervised learning algorithms, where models leverage labeled training data from patients for their training.

Recognizing the importance of prior knowledge in comprehending and analyzing any educational field, this research underscores the need for a foundational understanding of the basic concepts related to the paper before delving into its substantive content. This preliminary understanding serves as the cornerstone for successfully grasping and interpreting the paper's insights.

Classification: Classification is a supervised data mining and machine learning technique. It is a two step process, first step is called learning step where the model is constructed and trained by a predetermined dataset with class labels (training set) and second step is classification (testing) step where the model is used to predict class labels for given data (test data) to estimate the accuracy of classifier model.

Associative rule: Associative rule miming is a data mining technique which is used to find associative rules or patterns in data. In association rule mining, a pattern is discovered based on a relationship of a particular item to other items in the same transaction. It finds frequent item sets in data by using predefined support and confidence values. The association rule technique is used for heart disease diagnosis to discover the relationship of different attributes used for analysis and sort out the patient with all the risk factor which are required for prediction of disease.

Clustering: Clustering is basically an unsupervised machine learning technique, It is the task of dividing the dataset or population into a number of groups such that records or objects in the same groups are more similar to each other and dissimilar to the records or objects in other groups. Clustering helps to understand natural grouping or structure in a dataset and has no predefined classes K-means algorithm is cluster based algorithm.

Decision making: Decision tree is a technique that is used as a decision support tool that uses a tree-like graph or model of decisions. It takes as input a record or object described by a set of attributes and returns a "decision with predicted output value for the input". The input attributes can be discrete or continuous. After performing a sequence of tests decision tree reaches its decision. Each non leaf node of a decision tree corresponds to a test for the relevant attribute value, and the branches from the node are labeled with the possible outcomes of the test. Each leaf node in the tree specifies the value (decision) to be returned if that leaf is reached. J48, Random Forest (RF) and Logistic Tree Model (LTM) are Decision tree implementation algorithms.

Artificial Neural Networks: Artificial neural networks are machine learning algorithms having non linear data processing ability. Artificial neural network, sometimes just called a neural network, is a mathematical model or computational model based on biological neural network. In other words, it is an emulation of biological neural system. Neural networks consist of input and output layers, as well as (in most cases) a number of hidden layers. They are excellent tools for finding complex patterns in data and improve performance continuously from past experiences.

Genetic Algorithm: Genetic algorithm is a method for solving optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm iteratively modifies a population of individual solutions. At each step individuals are selected randomly as parents from the current population by genetic algorithm to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution In genetic algorithm solutions are represents by chromosomes. Chromosomes are made up of genes, which are individual elements that represent the problem. The collection of all chromosomes is called population. The genetic algorithm uses three main types of rules (operators) at each step to create the next generation from the current population: a) Selection is used in selecting individuals for reproduction. b) Crossover is used to combine two parents to form children for the next generation. C) Mutation is used to alter the new solutions in the search for better solution. Mutation prevents the GA to be trapped in a local minimum.

Cross Validation: Cross-validation serves as a technique to assess predictive models by partitioning the original dataset into a training set for model training and a test set for evaluation. In k-fold cross-validation, the original sample is randomly divided into k subsets of equal size. One subset is designated as the validation data for testing the model, while the remaining k-1 subsets contribute to training the model. This cross-validation process is iterated k times (the folds), with each subset used exactly once as the validation data. The average accuracy of the k-folds is then computed as the final accuracy. The 10-fold cross-validation technique is commonly employed in most experiments. In this approach, all instances of the dataset are utilized, divided into 10 disjoint groups where nine groups are dedicated to training, and the remaining one is employed for testing. The algorithm undergoes 10 iterations, and the average accuracy across all folds is calculated.

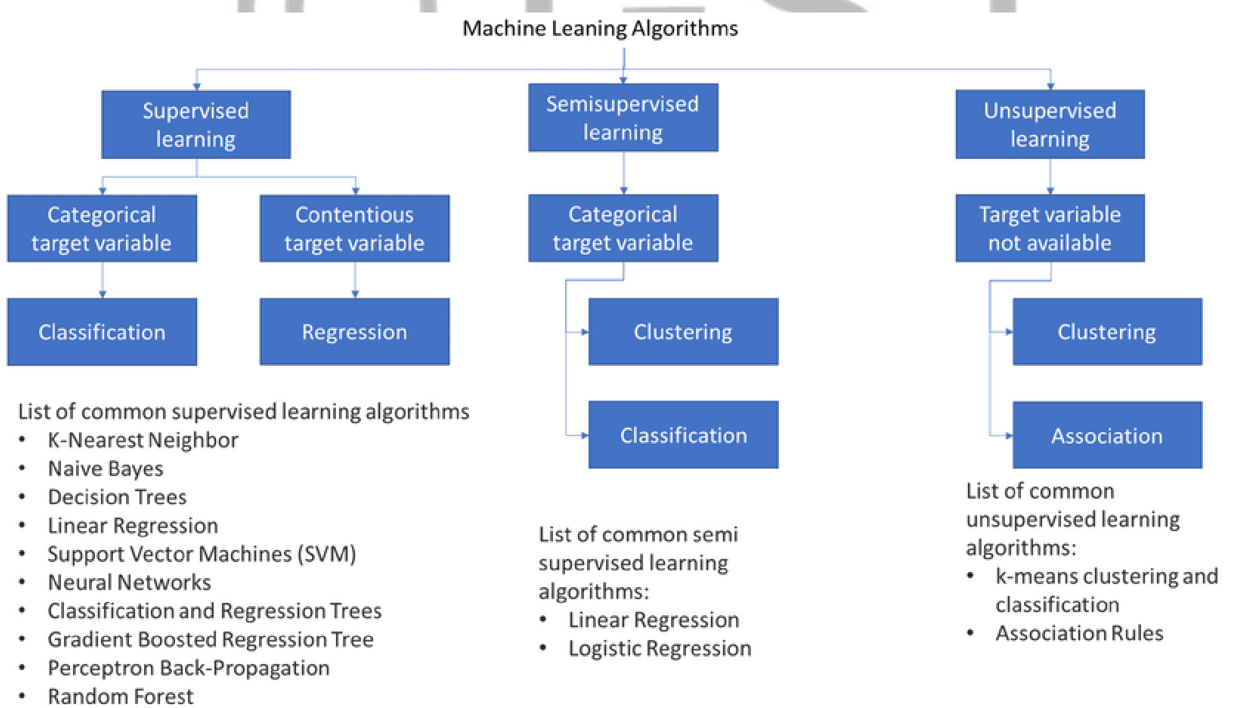


Figure 1.0 Category of Machine Learning Algorithms. (Amani Aldahiri et al, 2021)

2.0 LITERATURE / TECHNOLOGY REVIEW

Many researchers try to use machine learning and data mining Algorithm to predict complex psychological problems of individuals, such as predicting diseases, stress disorders and anxiety disorders.

CLASSIFICATION

i. (Priyanga et al, **2017)** Implemented the Naïve Bayes technique for the classification task in data mining, this system requires patients to input attribute values to obtain precise results. The UCI dataset was employed for training the data, and the trained data was then compared with user-input values. While traditional data mining techniques often fall short in accuracy, Naïve Bayes demonstrated a close-to-accurate result. For classification purposes, Naïve Bayes was utilized, presenting results categorized as low, average, high, and very high. The system performed both classification and prediction tasks. The accuracy of the system hinges on the algorithm and database employed, with the Naïve Bayes algorithm achieving a remarkable 96% accuracy—surpassing the performance of all other traditional data mining techniques.

ii. (Aman et al **2021)** Developed was a heart disease prediction system utilizing the WEKA tool and employing a 10-Fold cross-validation approach. The author utilized SMO, J48, KStar, and Bayes Net classification methods to forecast accuracy. According to the research findings, SMO demonstrated an accuracy of 89%, Bayes Net achieved 87% accuracy, while Multilayer Perceptron, J48, and KStar yielded accuracies of 86%, 86%, and 75%, respectively. However, these accuracy rates were deemed unsatisfactory in comparison to others. Mohan et al. proposed a distinctive strategy for identifying key characteristics through classification techniques, aiming to enhance the accuracy of cardiovascular disease forecasting. The suggested method for heart disease, incorporating a hybrid random forest with a linear model (HRFLM), showcased improved effectiveness with an overall accuracy of 88.7%, as indicated by the research results.

iii. (Joo et al, **2018)** Explored in this study was the reliability of machine learning techniques in predicting cardiovascular disease risks. The authors conducted a longitudinal cohort study involving 3.6 million patients seeking hospital admission in England, evaluating the discrimination and calibration performance of 19 predictive models. For instance, the random forest tree prediction score varied from 2.9 to 9.2 percent, and the neural network prediction score ranged from 2.4 to 7.2 percent. The study recommended avoiding the use of logistic models for predicting long-term risks and emphasized the importance of regularly assessing the levels considered among various models. Machine learning serves as a valuable tool in addressing diverse challenges in data science, with the utilization of existing data aiding in outcome prediction. The authors delved into the application of ensemble classification, a robust machine learning technique, to enhance the performance of multiple classifiers. Although ensemble classification improved prediction classification by 7%, the study underscored its potential benefits. The Cleveland heart dataset was employed for training and testing, revealing that random forest and MP5 achieved an 85.48% accuracy in heart disease prediction.The extraction of information from various facets of human life is known as data mining, with healthcare mining being a prevalent application. The study utilized the random forest algorithm to predict the occurrence of heart disease in patients, considering 303 samples from the Kaggle dataset. Performance metrics such as accuracy, sensitivity, and specificity were employed, and the algorithm demonstrated a notable prediction rate of 93.3% in the classification of heart disease.

iv. (Lambodar Jena et al, **2015)** Concentrating on the prediction of risks associated with chronic diseases, this study leveraged distributed machine learning classifiers, employing techniques such as Naive Bayes and Multilayer Perceptron. Specifically, the research aimed to forecast Chronic Kidney Disease, achieving impressive accuracy rates with Naïve Bayes at 95% and Multilayer Perceptron at 99.7%.

v. (Dhomse Kanchan B. et al, **2016)** studied special disease prediction utilizing principal component analysis using machine learning algorithms involving techniques like Naive Bayes classification, Decision Tree, and Support Vector Machine. The accuracy of this system is 34.89% for Diabetes and 53% for Heart disease.

vi. (Bardhwaj et al, 2017) Explored was a comprehensive overview of machine learning techniques applied in the healthcare domain to address diverse diseases. The study shed light on the potential value inherent in medical big data, emphasizing its utility in clinical decision support, diagnostics, treatment decisions, fraud detection, and prevention. The researchers succinctly outlined the nine-step data mining process while underscoring the imperative for efficient decision support within the healthcare system. Notably, their experiment yielded results demonstrating the viability of machine learning models for early disease diagnosis. While their findings offer relevance to this project to some extent, it is pertinent to note that their research is not primarily centered on the diagnosis of heart diseases. Consequently, our focus shifts to reviewing literature that aligns more closely with the project's objective, specifically exploring how machine learning algorithms can be effectively employed in the diagnosis of heart disease.

vii. Tripoliti et al, **(2017)** focused on machine learning methodologies evaluating heart failure. They researched severity estimation of heart failure and the prediction of re-hospitalization, mortality, and destabilizations. They performed an extensive study on related works of heart failure.

ARTIFICIAL NEURAL NETWORK

i. (Aryal et al. **2016)** Introduced was a system utilizing machine learning algorithms for the screening of microbiome-based cardiovascular disease. The analysis involved the examination of fecal ribosomal RNA of 16S from both cardiovascular and non-cardiovascular patients, with the samples sourced from the American Gut Project. Five distinct machine learning algorithms, namely decision trees, random forests, neural networks, elastic nets, and support vector machines, were trained for the study. Varied bacterial Vaux were discerned through the process. Notably, the random forest algorithm exhibited an improved characteristics curve of 0.70. Recognizing the demonstrated potential of random forest in cardiovascular disease prediction, it was incorporated along with one other machine learning algorithm in the ongoing study.

ii. (Chen et al, **2022)** used ANN with multiple features to diagnose cardiac disease. The results showed that ANN achieved the best accuracy of 80%.

iii. (Okengwu et al, **2023**) carried out a review on technologies applied to classification of tomato leaf virus diseases, A mitigation system was designed to automatically determine and categorize tomato illnesses using digital imaging sensors. The motor controlled image capturing device recognizes any illness on the plant. This was achieved with the convolutional Neural Networks (CNN) and also other image classification models were used like SVM and K-Means algorithm. The CNN Model was applied in the most of the works reviewed with accuracy as high as 98.13% better than K-Means.

iv. (Kannan et al**, 2015)** used four machine learning algorithms: LR, RF, SVM, and stochastic gradient boosting (SGB) to predict heart diseases. The model prediction showed that LR has a best accuracy of 86.5%.

v. (Raza, 2016) applied an ensemble learning model, multilayer perceptron, LR, and NB to classify heart diseases. The result shows that ensemble learning has improved the prediction performance of cardiac disease compared to other algorithms. (Winn 2017) used feature subset selection (CFS) with sequential minimal optimization (SMO) to predict heart diseases. The result shows that the CFS-SMO algorithm has achieved the best accuracy 86.96%.

vi. (Nalluri et al, **2018)** used two techniques (XGBoost and LR) to improve heart disease prediction. The result showed that LR with an accuracy of 85.68% was better than XGBoost, which achieved an accuracy of 84.46%.

vii. (Bhatet al**, 2019)** proposed a model that is a combination of multilayer perceptron network (MLP) with a back propagation algorithm to diagnose heart disease. The result shows that the proposed model has reduced error and an improved accuracy of 80.99%.

viii. (Abushariah et al, **2020)** utilized ANN and adaptive neuro-fuzzy inference system (ANFIS) to predict cardiac disease. ANN has an obtained optimal accuracy of 87.04%, but ANFIS has achieved the lowest accuracy of 75.93%.

ASSOCIATIVE RULE

i. (Aakash Chauhan et al, **2013)** proposed a disease prediction model for heart disease by utilizing evolutionary rule learning. Association Rule is used in this proposed system. This system is not very efficient because it has an accuracy of 53%.
ii. Furthermore, (Riyaz et al, **2022**) proposed a model that predicts heart disease using SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree classification techniques. SVM had the best accuracy of 98.43%
iii. (Purushottamet al, **2015**) devised a system employing the decision tree algorithm to formulate rules for the prediction of heart disease. The decision tree algorithm, versatile in addressing both regression and classification challenges, is employed to generate a training model capable of predicting classes. This algorithm constructs a tree representation where internal nodes correspond to attributes, and leaf nodes correspond to class labels. The system exhibited a commendable accuracy of 90%, showcasing its strong performance. While a single data mining technique can yield reasonable accuracy in predictive systems, the integration of hybrid data mining techniques is recommended to further enhance accuracy levels.
iv. (Malav et al, **2017**) carried out predictive analysis which was done on UCI dataset by using K-means and ANN
v. (Han et al, **2017)** Evaluated was the efficacy of various machine learning algorithms in predicting the risk of rapid progression of coronary atherosclerosis. The study focused on analyzing the qualitative and quantitative computed tomography angiography plaque features of 983 patients. The model's score was subsequently compared to the cardiovascular atherosclerosis risk score, and crucial clinical variables were scrutinized. However, the authors underscored the

ongoing challenge of effectively assessing unnoticed biases in the dataset using machine learning techniques.

CLUSTER ANALYSIS

i. (Jyothi Soni et.al,**2013**) Suggested is a heart disease prediction method utilizing the association rule mining technique, generating an extensive set of rules when applied to the dataset. Frequent item set mining is employed to identify all frequent item sets, with commonly used association rule mining methods such as Apriori and FPgrowth being applied. The inclusion of a genetic algorithm is proposed to optimize the subset of attributes necessary for heart disease prediction by reducing the actual data size.Classification, a supervised learning method, is introduced to extract models that describe crucial data classes. Three classifiers Decision Tree, Naïve Bayes, and Classification via clustering are employed to diagnose the presence of heart disease in patients.

ii. (Ashish Chhabbi et al, **2016**) used a dataset collected from UCI repository to perform different data mining techniques to predict heart disease. They applied K-means algorithm and Naïve Bayes and their results revealed that tuning the number of clusters of the k-means algorithm gave better results of 96.3%

iii. (Jyothi Soni et.al,**2013**) proposed for predicting the heart disease using association rule mining technique, they have generated a large number of rules when association rules are applied to dataset .Frequent Item set Mining is used to find all frequent item sets. Association rule mining methods like Apriority and FPgrowth are used frequently. Genetic algorithm have been used in to reduce the actual data size to get the optimal subset of attributed sufficient for heart disease prediction. Classification is one of the supervised learning methods to extract models describing important classes of data. Three classifiers e.g. Decision Tree, Naïve Bayes and Classification via clustering have been used to diagnose the Presence of heart disease in patients.

DECISION MAKING

i. (Krittanawong et al, **2018)** evaluated machine learning algorithms' overall predictive ability of predicting cardiovascular disease. The strategy was created using various databases published in March 2019. The ability of predicting diseases such as coronary artery disease, cardiac arrhythmias, heart failure, and stroke was observed. The area under the curve metric was used in the prediction analysis. However, because of the heterogeneity of machine learning algorithms, identifying an optimal algorithm for the cardiovascular disease remains a challengeC4.5 had the best accuracy of 86.32%

ii. **(**Duan et al, **2014)** Explored was the correlation between concentrations of heavy metals in blood and urine and the mortality rates associated with cardiovascular disease and cancer. The study utilized datasets obtained from the National Health and Nutrition Examination Survey. Poisson's regression was employed to investigate both single and multi-metal exposure. Participants in the study spanned an age range of twenty-five to eighty-five years. The analysis included the examination of factors such as age, gender, education, body mass index, serum cotinine, and medical comorbidities. While the study identified a connection between metal concentrations in both blood and urine and cancer mortality, the authors emphasize the study's origin in the quest for further research on cardiovascular disease.

iii. (Devansh Shah et al, **2012)** Introduced is a method for predicting the likelihood of a patient developing heart disease. The paper employs various data mining techniques to enhance the effectiveness of heart disease prediction. The primary goal is to anticipate the potential occurrence of heart disease through computerized prediction, offering valuable insights for medical practitioners and patients. Additionally, the paper identifies which attributes play a more significant role in enhancing prediction precision. This information can potentially save costs associated with multiple patient trials, as not all attributes contribute equally to predicting the outcome.The research utilizes the Cleveland dataset from the UCL repository and employs different classifier algorithms, conducting tests through the WEKA data mining tool in ARFF format. The author specifically utilizes four data mining classification techniques k-nearest neighbor, naive bayes, decision tree, and random forest.

iv. (Noor Basha et al, **2013)** proposed an early detection of heart syndrome using machine learning techniques, to predict and analyze the heart related syndrome in patients. The model for heart disease patients with various machine learning algorithms such as, NB, SVM, DT, RF, and KNN algorithms and techniques to achieve accuracy and efficiency on analysis and prediction of heart syndromes. An exploratory data analysis (EDA) heart disease dataset to summarize their main characteristics with visual methods. Out of which KNN algorithm found to be that very effective and efficient performance on accuracy score on heart disease prediction with the accuracy of 97.33%

v. (Chaitrali S. Dangare et al, **2014)** propose a method to predict heart disease more accurately. This paper has analyzed prediction systems for heart disease using more number of input attributes. This research paper added two more input attributes i.e. obesity and smokings are used to get more accurate result. After pre-processing, data mining classification techniques such as, Decision Tree, Neural Network, and Naïve Bayes are used. The result analysis shows that the accuracy of Neural Network is more accurate than Decision Trees, and Naïve Bayes.

vi. (Lakshmanarao et al, **2015)** Proposed a selection of machine learning techniques employed in heart disease detection, incorporating sampling techniques to address unbalanced datasets. The researcher utilized a Kaggle dataset containing numerous patient records encompassing approximately 15 features, including demographic, behavioral factors, and medical risk factors. The machine learning model predicts, based on these attributes, whether a patient faces a 10-year risk of future coronary heart disease (CHD). Various data mining methods are applied to predict the overall risk. To address the challenge of unbalanced datasets, oversampling and undersampling prove beneficial in balancing samples between the two different classes.

vii. (Krittanawong et al, **2019**) investigated the overall predictive performance of machine learning systems in predicting cardiovascular illness. The plan was designed with the use of several databases that were released in March 2019. There was evidence of the ability to anticipate disorders such as coronary artery disease, cardiac arrhythmias, heart failure, and stroke. The prediction analysis employed the area under the curve measure. Identifying an optimal machine learning method for cardiovascular illness, however, remains a difficulty due to the variety of machine learning algorithms.

viii. (Joo et al, **2018)** investigated the consistency of machine learning techniques for predicting the risks of cardiovascular disease. The authors conducted the longitudinal cohort study on 3.6 million patients seeking admission to hospitals in England. The discrimination and calibration performance of the 19 predictive models were evaluated. For example, the random forest tree prediction score ranged from 2.9 to 9.2 percent, while the neural network prediction score ranged from 2.4 to 7.2 percent. It was suggested that when considering various models avoid using logistic models to predict long-term risks and that the levels considered between models be evaluated regularly. Machine learning is used to solve many problems in data science. Existing data aids in the prediction of outcomes in machine learning. As a powerful machine learning technique, the authors investigated ensemble classification to improve multiple classifiers. The ensemble classification improves the prediction classification, but only by 7%. For training and testing, the Cleveland heart dataset was used. According to the authors in random forest and MP5 produced 85.48% in heart disease prediction. The process of extracting information from all aspects of human life is known as data mining. The most common data mining application is healthcare mining. The random forest algorithm was used in the study to predict the occurrence of heart disease in patients. A total of 303 samples from the Kaggle dataset were considered. The metrics used to evaluate performance were accuracy, sensitivity, and specificity. In the classification of heart disease, the algorithm achieved a prediction rate of 93.3%.

ix. **(**Han et al**, 2019) Evaluated** The effectiveness of multiple machine learning algorithms to predict the risk of rapid progression of coronary atherosclerosis and looked at the qualitative and quantitative computed tomography angiography plaque features of 983 patients. The score of the model was compared to a risk score for cardiovascular atherosclerosis. They compared the most essential clinical variables. However, the authors pointed out that employing machine learning techniques to evaluate undiscovered biases in the dataset remains a challenge. In data science, machine learning is used to solve a variety of challenges. In machine learning, existing data contributes in the prediction of outcomes. The authors used ensemble classification as a potent machine learning technique to improve several classifiers. Ensemble classification increases prediction classification by just 7%. The Cleveland heart dataset was utilized for training and testing.

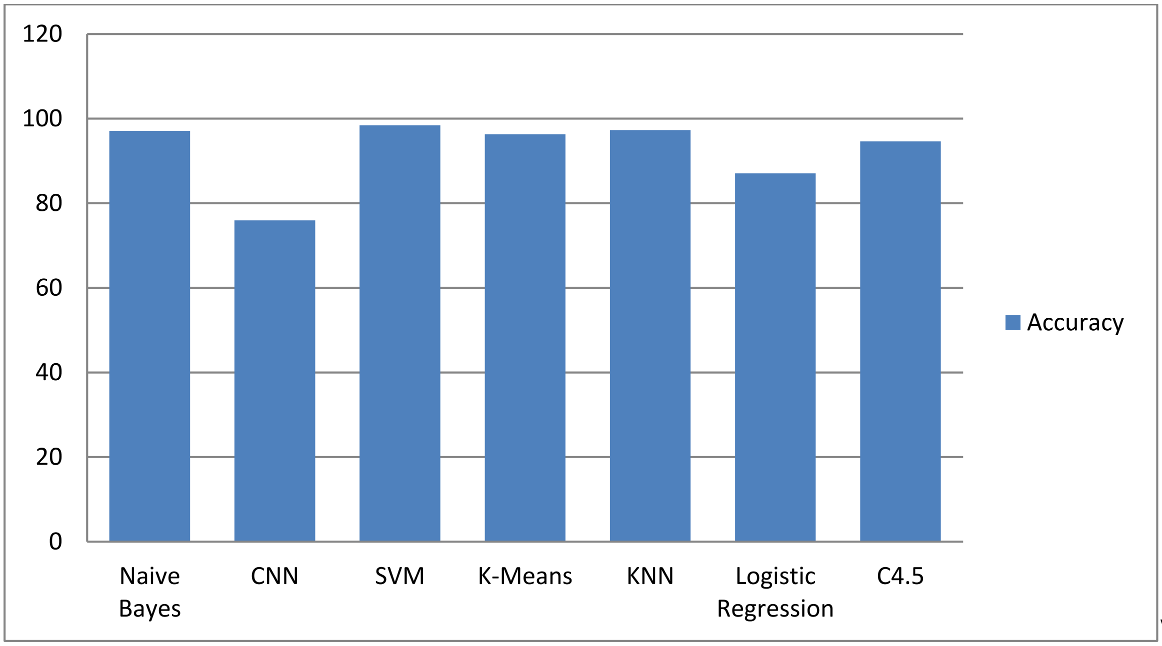
3.0 DISCUSSION

Different datasets of heart disease patients are used in different experiments. In most experiments dataset used is taken from online Cleveland database of UCI repository. The dataset consists of 303 records with 14 essential attributes (total attributes 75) with some missing values also. Fewer experiments have been done on different datasets. From the study we also find that Neural Networks with same datasets provide 100% accuracy in one experiment whereas in another experiment gives 76.55% accuracy with 8 attributes. Naive Bayes also gives high accuracy above (96%) in most experiments with different number of attributes. Decision lists (C4.5) also performs very well in accuracy goes up to 94.62 % in a case. So, different techniques used indicate the different accuracies depend upon number of attributes taken and tool used for implementation. From this study we come up with following observations that should be taken in consideration in future research work for high accuracy and more accurate diagnosis of heart disease by using intelligent prediction systems. In most experiments Small and same dataset has been used to train prediction models. So, we have to take real data in a large quantity of heart disease patients from reputed medical institutes of our country and use that data to train and test our prediction models. Then we have to examine the accuracy of our prediction models on large datasets. We have to consult highly experienced experts of cardiology to prioritize the attributes according to their effect on patient's health and also if necessary add more essential attributes of heart disease for more accurate diagnosis and high accuracy. However looking at different experiments the table below shows the different accuracy rates for some Algorithms that was researched on.

Authors	Techniques	Algorithms	Accuracy
Priyanga et al, 2017	Classification	Naïve Bayes	97.1%
Abushariah et al, 2020	Artificial Neural Networks	Convolutional Neural Network	75.93%
Riyaz et al, 2022	Classification	Support vector Machine	98.43%
Ashish Chhabbi et al, 2016	Clustering	K-Means	96.3%
Noor Basha et al, 2013	Classification	K-Nearest Neighbour	97.33%
Sid Ahmed et al, 2021	Regression	Logistic Regression	87.07
Krittanawong et al, 2018	Decision Tree	C4.5	94.62%

Table 1.0 Algorithm Accuracy Table

The visual representation of table 1.0 shown in the chart below displays the visual comparison and Data Representation of the different algorithms that have been reviewed.



\\\

Figure 2.0 Accuracy for different Techniques reviewd

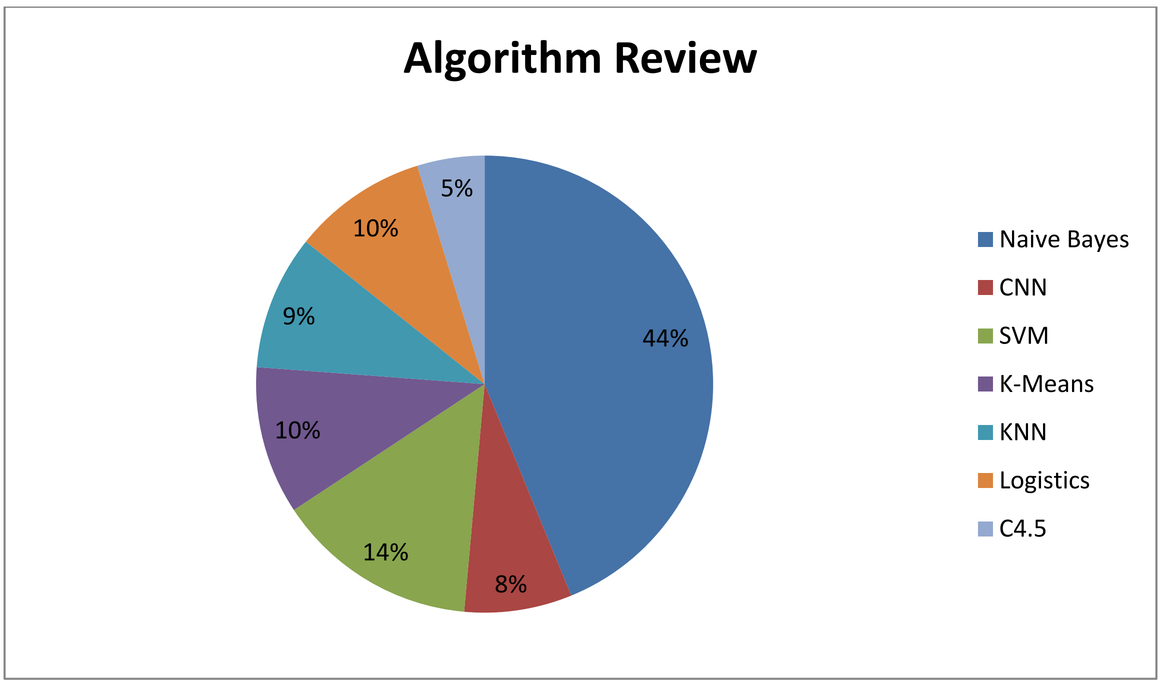


Figure 3.0 Categories of technologies mostly used in this Review

4.0 CONCLUSIONS

Upon reviewing various recent research papers addressing heart disease prediction through diverse data mining and machine learning techniques, it becomes evident that these studies employ a range of methodologies and algorithms. The exploration of heart disease prediction involves the utilization of different experimental tools in conjunction with various data mining and machine learning techniques. There is a discernible need to advance the field by developing more intricate hybrid models that integrate diverse data mining and machine learning techniques. Furthermore, the inclusion of text mining for extracting insights from unstructured medical data, abundantly available in medical institutes, is identified as a crucial enhancement.

The study highlights the effectiveness of incorporating Genetic algorithms for optimization and feature selection, significantly improving the overall performance of intelligent prediction models. Notably, the research places a greater emphasis on classification techniques compared to regression and association rule approaches. To ensure more robust and comparable results in future research, it is imperative to give due consideration to these aspects.

The accuracy of research outcomes is intricately linked to the selection of research tools and methodologies. Consequently, the choice of an appropriate experimental tool for implementing techniques emerges as a pivotal parameter influencing the accuracy of predictions in this field.

5.0 ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Centre that has contributed to the successful completion of this research paper.

Center for Information and Telecommunication Engineering (CITE)

Your Platform has been truly instrumental, and we are sincerely grateful.

6.0 CONFLICT OF INTEREST

The authors declare that they do not have any conflict of interest.

References

1) Abhishek Taneja, “ Heart Disease Prediction System Using Data Mining Techniques”, Oriental Journal Of Computer Science and Technology, Vol. 6, pp. 457-466, Dec.2013.

2) Asha Rajkumar, and Mrs. G. SophiaReena, “Diagnosis of Heart Disease Using Data Mining Algorithm”, Global Journal of Computer Science and Technology, Vol. 10, pp. 38-43, Sept. 2010.
3) B.Venkatalakshmi, and M.V Shivsankar, “Heart Disease Diagnosis Using Predictive Data mining”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Special Issue 3, March-2014.
4) Cardiovascular disease webpage on WHO [Online]. Available: https://www.who.int/cardiovascular_diseases/en/ , 2019.
5) Chaitrali S.Dangare, and Sulabha S.Apte, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications (0975 – 888), Vol. 47, No.10, June.2102.
6) Dr. T. Karthikeyan, and V.A.Kanimozhi, “Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network” , International Journal of Advanced Research in Science, Engineering and Technology, Vol. 4, Issue 1, January 2017.
7) G.Subbalakshmi, K. Ramesh, and M.C. Rao, “Decision Support in Heart Disease Prediction System using Naive Bayes”, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2, No. 2, Apr-May 2011.
8) Nidhi Bhatla, and Kiran Jyoti, “An Analysis of Heart Disease Prediction using Different Data Mining Techniques”, International Journal of Engineering Research & Technology (IJERT), Vol. 1, Oct.2012.
9) Heart disease webpage on MAYO CLINIC [Online]. Available: https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118 , 2019
10) Hlaudi Daniel Masethe, and Mosima Anna Masethe, “Prediction of Heart Disease Using Classification Algorithms”, in Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. II WCECS 2014, 22-24 Oct. 2014, San Francisco, USA.
11) Sarangam Kodati, and Dr. R Vivekanandam, “A Comparative Study on Open Source Data Mining Tool for Heart Disease” , International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, March-2018.
12) Jaymin Patel, Prof.Tejal Upadhyay, and Dr. Samir Patel, “ Heart Disease Prediction Using Machine Learning and Data Mining Technique”, IJCSC, Vol. 7, No. 1, pp.129-137, September-2015. [12] K.Gomath, Dr. Shanmugapriyaa, “Heart Disease Prediction Using Data Mining

Classification", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol.4, Issue 2, February-2016.

13) MA.Jabbar, Dr. Priti Chandra, and B.L. Deekshatulu, "Cluster Based Association Rule Mining For Heart Attack Prediction", Journal of Theoretical and Applied Information Technology, Vol. 32 No.2, October-2011.

14) Nikita Shirwalkar, and Tushar Tak, " Human Heart Disease Prediction System Using Data Mining Techniques", International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, Mar.2018.

15) Navdeep Singh and Sonika Jindal, " Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms", International Journal of Advance Research, Ideas and Innovations in Technology, Vol.4, Issue 2, 2018.

16) Okengwu U.A , Hillard A.Akpughe, Eyinanabo Odogu, and Taiye Ojetunmibi, ""Review on Technologies Applied to Classification of Tomato Leaf Virus Diseases", European Journal of Artificial Intelligence and Machine Learning, Vol. 2, Issue 4, October 2023