Heart Disease Prediction Using Machine Learning

Nkosiyazi Myaka
School of Mathematics, Statistics, and Computer Science
University of KwaZulu-Natal

Abstract

The number of heart disease cases globally is rising at an unprecedented and exponential rate every day. Early detection and accurate heart disease prediction can significantly improve patient outcomes and reduce mortality rates. Machine learning (ML) techniques have shown promising results in various medical applications, including heart disease prediction. In this paper, we developed an ML model for the prediction of heart disease based on a set of clinical and demographic features. The project utilized a dataset comprising patient data collected from Kaggle. The Behavioural Risk Factor Surveillance System (BRFSS) dataset incorporates features such as age, gender, race, body mass index, diabetes, alcohol drinking, asthma, and other relevant clinical parameters. The decision tree classifier, k-nearest neighbors classifier (KNN), random forests classifier, and neural network classifier are the four machine learning classifier models that are used to construct this prediction model. In the first portion, several clinical variables that are crucial in determining whether a patient has heart disease are collected. Secondly, various machine learning classifiers are defined on the provided dataset and their accuracy is computed. These models were explored and compared based on their performance in predicting heart disease. The developed ML model underwent rigorous evaluation using appropriate performance metrics, such as accuracy, precision, recall, and F1 score. The model's performance was validated using cross-validation techniques to ensure its generalizability and robustness. This project aimed to create an accurate and reliable machine-learning model that could predict the likelihood of heart disease in patients based on their clinical and demographic information. It can be inferred from the experimental results that neural networks yield the best outcomes. It is indisputable that random forest frequently yields results that are almost as good. Thus, it stands to reason that the neural network and random forest classification algorithms should be applied to

obtain the most accurate heart disease predictions.

Keywords: heart disease, prediction, machine learning, clinical features, neural network, random forest, decision tree, k-nearest neighbors, rigorous evaluation

1 Introduction

Heart disease is a serious health problem that affects millions of people worldwide. It is the leading cause of death in many countries and poses a significant challenge to healthcare providers. The early detection of heart disease can help prevent serious health problems and reduce the risk of mortality. The risk of heart disease has been predicted in recent years using machine learning algorithms and data mining approaches. These methods forecast the chance of acquiring heart disease based on a number of characteristics, including age, gender, blood pressure, cholesterol, Body Mass Index (BMI), diabetes status, and other variables. To increase the precision of heart disease prediction models, it is necessary to determine the best feature selection strategies and data mining tactics.

Using machine learning (ML), one can manipulate and extract implicit, known or previously unknown, and possibly helpful information from data. [1]. The area of machine learning is extremely broad and diversified, and its application and breadth are growing daily. ML uses a variety of classifiers from ensemble learning, supervised learning, and unsupervised learning to find the accuracy of a given dataset and make predictions. Since it will benefit many, we can apply that knowledge to our machine-learning project for heart disease prediction. [2].

Heart disease is one of the many serious illnesses that has attracted a lot of attention in medical research. Although diagnosing heart illness is a difficult undertaking, it can provide an automated estimate of a patient's heart state so that subsequent treatments can be tailored to the patient's needs. The signs, symptoms, and physical examination of the patient are typically used to make the diagnosis of cardiac disease. Heart disease is more common in people who smoke, have high blood pressure, have high cholesterol, have a family history of the condition, are obese, have high blood pressure, and don't exercise. [3]. The provision of high-quality services at reasonable prices is a significant problem for healthcare organizations, including hospitals and medical facilities [3]. Correct patient diagnosis and efficient therapy delivery are prerequisites for providing quality care. There are both numerical and categorical data in the heart disease database that is currently accessible. Prior to additional processing, these records undergo cleaning and filtering to remove unnecessary information from the database [4].

These days, cardiovascular diseases (CVDs) are a broad category that includes many disorders that may have an impact on your heart. According to WHO estimates, CVDs cause 17.9 million deaths worldwide. According to a World Health Organization (WHO) estimate, cardiovascular diseases (CVDs) accounted for 17.9 million fatalities in 2019, or 32% of all deaths worldwide [5], with an annual mortality rate exceeding 17.7 million [6]. According to the Australian Institute of Health and Welfare (AIHW), 42% of fatalities in Australia are attributed to CVD, making it the country's biggest cause of death in 2018 [7]. It is very difficult to diagnose and treat cardiac disease when cutting-edge technologies and medical professionals are not available [8]. A sound diagnosis and treatment can save the lives of a great number of people [9]. Heart disorders are diagnosed by a doctor based on an assessment of the patient's medical history, the results of the physical examination, and an analysis of any worrisome symptoms. Nevertheless, the results of this diagnostic approach are not enough to identify patients with cardiac disease. Moreover, its analysis is computationally demanding and expensive [10]. Thus, we use machine learning classifiers to construct a non-invasive prediction system to address these problems. Using artificial fuzzy logic and machine learning classifiers, an expert decision system effectively diagnoses heart disorders. As a result, the death ratio starts to decrease [11, 12].

A knowledge-finding method for analysing data and condensing it into meaningful information is data mining (DM) [13]. The goal of the current study is to forecast the likelihood of developing cardiac disease based on the patient data set [14]. In reality, the main objectives of DM are descriptions and predictions [15]. Aspects or variables in the data set are used in prediction in DM to determine the values of other qualities that are unknown or will change in the future [16]. The focus of description is on finding patterns that help humans understand the data. Predictions in DM are meant to assist in identifying patterns in patient data so that the patients' health can be improved [13].

This research aimed to investigate the application of ML models to heart disease prediction and critically compare them in terms of efficiency, accuracy, etc.

First, the data is evaluated using various Exploratory Data Analysis (EDA) approaches to describe the nature of our data. Next, standardization is applied to fix the data in case of errors including empty data cells. Once more, various types and samples of data are found by analyzing the data utilizing EDA techniques. The models are then trained and tested on two separate portions of the data. The heatmap in Figure 1 makes it evident which of the main risk factors for heart disease are smoking, diabetes, and diffwalking. A heatmap is a color-coded, system-based graphical depiction of data that shows several values in relation to one another on a single graph. They are commonly utilized in analytics, but researchers also regularly use them to characterize the behavior of human values in their studies. The link and influence of several parameters in cardiovascular disease are eloquently depicted in Figure 1's heat map. The primary causes of heart disease are advanced age, physical health issues, asthma, kidney disease, and other diseases.

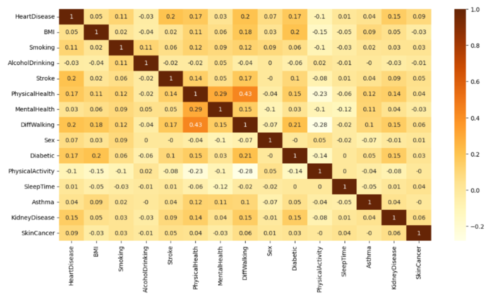

Figure 1: Fig. 1. Heatmap of the Various CVD Parameters

2 Literature Review

Heart disease is a serious public health issue, and early detection and prevention depend heavily on the ability to forecast its incidence. Numerous studies have been carried out utilizing different ML techniques to create predictive models for heart disease.

Medical data information resources are analyzed and predicted using machine learning approaches. In medicine, making the diagnosis of heart disease is an important and time-consuming task. All conditions that impact the heart are included under the umbrella term "heart disease." Heart disease exposure due to a variety of causes or symptoms is a problem that is not helped by incorrect assumptions that frequently have unanticipated consequences. A Supervised ML algorithm underpins the data classification, improving accuracy.

The researchers [17] employed data mining techniques and pattern recognition to create prediction models for cardiovascular diagnosis. The Naïve Bayes technique outperformed other employed techniques in the studies, which employed the classification algorithms Naïve Bayes, Decision Tree, K-NN, and Neural Network [17]. In order to extract information on heart illness from a heart disease warehouse, the researchers [18] utilize a K-means clustering technique. They then use the MAFIA (Maximal Frequent Item Set technique) approach to determine the weightage of the frequent patterns that are important for heart attack predictions. A layered neuro-fuzzy method was presented by the researchers [13] to forecast the onset of coronary heart disease using a MATLAB tool simulation. When the neuro-fuzzy integrated approach was used to do analysis for occurrences of coronary heart disease, the error rate was extremely low and the work efficiency was great [13]. Additionally, a novel method for association rule mining based on sequence number and grouping transactional data sets for heart disease prediction was put out by the researchers [14]. The suggested strategy was implemented in the C programming language and was thought to be scalable and effective because it only required a tiny cluster at a time [14].

To predict heart disease risk factors in the Taiwanese population, such as hypertension, diabetes, and excessive cholesterol, researchers employed a random forest algorithm. The study discovered that the random forest model has an accuracy of 89.9% in predicting heart disease risk variables. The application of an SVMs model for heart disease prediction was examined in a different study by Razzak et al. (2018). The study made use of a patient dataset that contained details about age, sex, blood pressure, and cholesterol levels. The SVMs model's accuracy of 84.47% shows the potential of this method for predicting cardiac disease [15]. In order to predict heart disease risk, the researchers utilized a range of machine-learning methods, such as decision trees, random forests, and SVMs. The study made use of a patient database that contained data on age, sex, smoking habits, and family history of heart disease. The SVMs model showed the promise of ML in this field by achieving the greatest accuracy of 76% [14].

The study's objective was to develop a machine learning-based model to predict the risk of heart disease using patient electronic health records (EHRs). For the study, a deep neural network (DNN) model was created, and it was trained on a substantial EHR dataset. The model's 87.2% accuracy in predicting the risk of heart disease illustrates the potential of machine learning in this area.In [16] Researchers employed a deep learning technique to predict heart disease from electrocardiogram (ECG) data. In order to predict cardiac disease, the authors first used a convolutional neural network (CNN) to extract information from ECG signals. Afterward, they employed an LSTM network. The study's 90.1% accuracy rate in predicting heart illness was achieved by using ECG signals [17].

A decision tree algorithm was utilized by researchers to forecast cardiac disease in a sizable Iranian population. The study has a 91.1% accuracy rate in predicting heart disease using demographic, clinical, and laboratory data. Researchers employed a hybrid strategy to predict cardiac disease by fusing statistical methods with ML algorithms. The study's prediction of heart disease had a 93% accuracy rate using logistic regression, Artificial neural networks (ANNs), and SVMs [15].

To forecast cardiac illness in a Chinese population, researchers employed machine learning. The accuracy of the study's DNN prediction of heart disease using EHRs was 94%. Heart disease prediction has made extensive use of ML techniques such as decision trees, KNN, random forests, SVMs, and neural networks [7]. These methods have demonstrated promising results in accurately classifying individuals as being at high or low risk for heart disease and identifying the most significant risk factors for heart disease. However, further research is required to verify how well these methods perform in various populations and to make the models easier to understand. Evaluating the various classification methods, including decision trees, KNN, random forests, and neural networks, is the major goal. Subsequently, the performance is assessed using metrics for specificity, sensitivity, accuracy, and precision [14].

3 Methodology

The research allowed the use of the heart disease dataset data that was available at the time to achieve the objectives described above. The steps that were attempted are highlighted below:

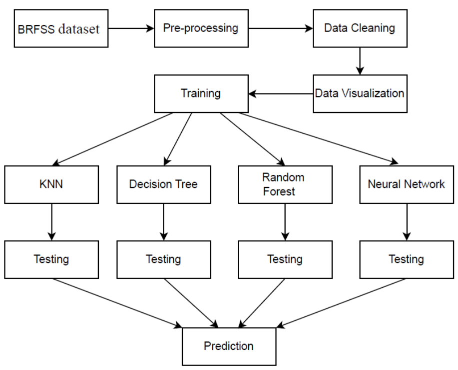

Figure 2: Model Flowchart

Clinical features	Description
Heart disease	Patients who have ever reported having coronary heart disease (CHD) or myocardial infarction (MI).
BMI	Body Mass Index (BMI).
Smoking	Have you ever smoked a hundred cigarettes or more in your lifetime? (Yes or No as the response).
Alcohol drinking	Adolescent males who consume over 14 drinks per week and adult females who consume over 7 drinks per week are considered heavy drinkers.
Stroke	Ever been told you had a stroke?
Physical health	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good? (0-30 days).
Mental Health	Considering your mental state of mind, how many days in the last thirty days did you feel that way? (0 to 30 days).
DiffWalking	Do you find it difficult to climb stairs or walk?
Sex	Are you male or female?
AgeCategory	Fourteen-level age category.
Race	ethnicity value.
Diabetic	(Ever told) (you had) diabetes?
PhysicalActivity	adults who said they engaged in exercise or physical activity outside of their regular jobs within the last 30 days.
Gen Health	In general, how would you describe your health?
Sleep time	How many hours of sleep do you get on average in a day?
Asthma	Have you been informed you have asthma?
KidneyDisease	Were you ever told you had kidney illness, excluding kidney stones, bladder infection, or incontinence?
SkinCancer	Have you been informed you have skin cancer?

Table 1: Clinical features

3.1 Behavioural Risk Factor Surveillance System (BRFSS) dataset

The Behavioural Risk Factor Surveillance System (BRFSS) dataset is used in this analysis, and it incorporates features such as age, gender, race, body mass index, diabetes, alcohol drinking, asthma, and other relevant clinical parameters. Contains 18 columns and 319795 entries. There are four category features and fourteen numerical features, with no null values. The dataset has both beneficial knowledge and unhelpful qualities. So, in pre-processing the useful data is selected and attributes are replaced with corresponding numerical equivalents. Diabetic column is converted to integers, facilitating numerical analysis and processing.

3.2 Visualization

3.2.1 Exploratory Data Analysis

The graphical visualization of the data makes it easier to understand its trend. This method uses a graphical representation of the data for data representation. Bar charts are used in this research to show the cleaned data. Recall that on the x-axis, men are represented by number one and women by number zero. If you examine the graphs below closely, you will notice that women make up the majority of instances without heart disease and men make up the majority of cases with heart disease. Smokers make up the greatest group of people with heart disease. Even though they don't smoke, there are in-

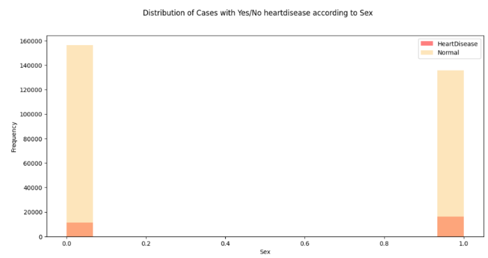

Figure 3: Distribution according to sex

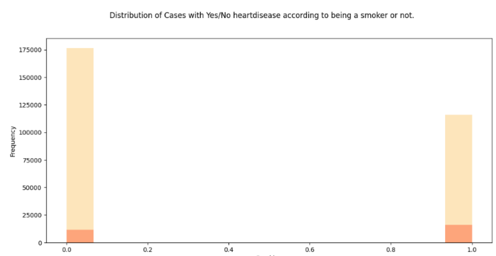

Figure 4: Distribution according to being a smoker or not

stances of heart disease, but there are other contributing variables. Additionally, we observe that heart disease is more common in those 80 years of age and older.

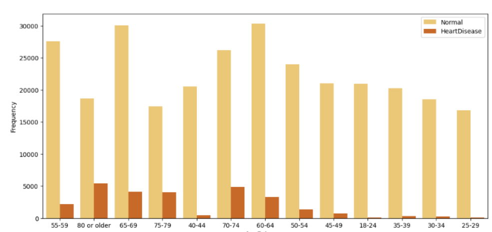

Figure 5: Distribution according to age category

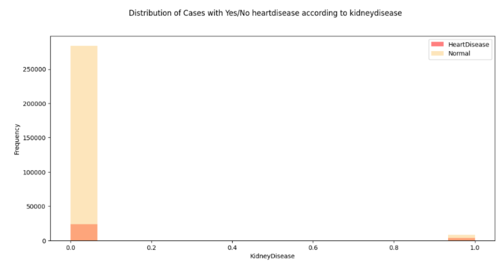

Figure 6: Distribution according to kidney disease

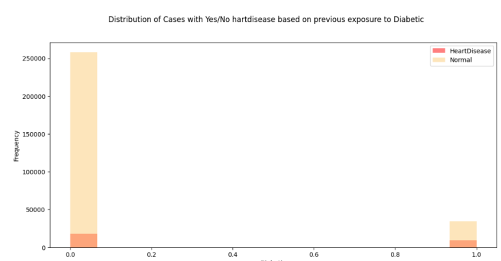

Figure 7: Distribution based on previous exposure to diabetic

3.3 Data pre-processing

This is a crucial step in the pre-processing stage, when data is taken from the heart disease dataset. This stage is necessary since the raw data is incomplete and unreliable, necessitating further pre-processing operations to produce ready raw data. A ColumnTransformer was created to encode specific categorical features in the dataset, namely AgeCategory, Race, and GenHealth, using one-hot encoding. The transformer was applied to both the training data X_{train} and the test data X_{test} . After transforming the data, the resulting one-hot encoded features were concatenated with the original dataset, effectively replacing the categorical columns with their numerical representations. Subsequently, the original categorical columns, namely AgeCategory, Race, and GenHealth, were removed from the dataset, leaving only the one-hot encoded representations for further modeling. This transformation was essential as machine learning algorithms often require numerical inputs, and one-hot encoding categorical variables enables these algorithms to interpret and learn from the data effectively. The dataset's features were standardized using a StandardScaler object, ensuring they follow a standard normal distribution with a mean of 0 and a standard deviation of 1. This transformation, applied to both training X_{train} and test data X_{test}, is vital for machine learning models relying on distance-based computations. Standardization guarantees equal contribution from all features, preventing any single feature from overpowering the model due to its scale variation.

3.4 Model training and optimization

The dataset is prepared for machine learning analysis. The features, excluding the HeartDisease column, are stored in the variable features, while the HeartDisease column is stored in the variable target, representing the prediction goal. The data is then split into training and testing sets using the `train_test_split` function. 80% of the data is allocated for training X_{train} and y_{train}, and 20% is reserved for testing X_{test} and y_{test}. The shuffle=True parameter ensures the randomization of data before splitting, and a specific random seed (random_state=44) is set for reproducibility. The printed shapes provide an overview of the sizes of the training and testing datasets, a fundamental step in machine learning workflows to enable model training and evaluation on distinct datasets.

3.5 Modelling

This approach assesses the performance of a machine-learning model using various metrics. The function takes three parameters: the trained model, the test data X_{test}, and the corresponding true labels y_{train}. First, the function predicts the labels using the test data and calculates metrics such as accuracy, precision, recall, and F1-score, providing insights into the model's classification performance. It computes Cohen's Kappa score, a statistical measure indicating the agreement between predicted and actual classifications. The function also generates a Receiver Operating Characteristic (ROC) curve and calculates the area under the curve (AUC), which measures the model's ability to distinguish between classes. Furthermore, it constructs a confusion matrix, displaying the counts of true positive, true negative, false positive, and false negative predictions. The function returns a dictionary containing these metrics, offering a comprehensive evaluation of the model's classification accuracy and reliability.

3.6 Model Training and Evaluation

Data is the input and output of any machine learning algorithm. A training dataset is the term for the input data, which is examined for knowledge to confirm the model's functionality. Train and exam databases make up the two halves of the Heart dataset. Models are created by running algorithms on the training dataset to learn, and they are subsequently assessed using the testing dataset. The entire process of building machine learning and data science models depends on algorithms. supervised, semi-supervised, unsupervised, and reinforcement algorithms are a few examples of these algorithms. These categories are further subdivided into many algorithms, of which this study uses four: Random Forest, Decision Tree, K-Nearest Neighbor, and Neural Networks, all of which are Supervised Algorithms [19].

4 Models

4.1 K-Nearest Neighbour

KNN estimates the closest output by taking the data points out of the dataset. It has exceptionally high prediction accuracy. The heart disease dataset has multiple features, which makes this technique a good fit for heart disease prediction. Along with most of KNN, KNN extracts knowledge and logic based on the Euclidean distance Samples function d(xi, xj). In terms of math [20].

4.2 Random Forest

The tree-based classifier method used in machine learning is called the random forest. The Random Forest Classifier builds many trees depending on various parameters, and the algorithm's success is measured by averaging the trees' projected outcomes. To get the best outcomes, it constructs a few decision trees and makes use of them. [20].

4.3 Decision Tree

Although decision trees are a supervised learning technique, they are mostly employed to solve classification problems. They can also be used to solve regression problems. They operate by recursively splitting the dataset into subsets, optimizing the decision-making process to accurately predict class labels or numerical values. While easy to interpret and suitable for various data types, careful management of tree depth is crucial

Paper	Dataset	Year	Technique	Accuracy	Error rate
An Analysis of Heart Disease Prediction using Different Data Mining Technique [21]	BRFSS dataset	2012	Naïve Bayes	52.33%	47.67%
			Decision Tree	52%	48%
			KNN	45.67%	54.33%
Heart Disease Prediction using Machine Learning Algorithms	BRFSS dataset		Logistic Regression	87.88%	12.12%
			KNN	74.19%	25.81%
			Random Forest	86.81%	13.19%
Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers [22]	BRFSS dataset	2011	WAC	57.75%	42.25%
			CBA	58.28%	21.72%
			CMAR	53.64%	46.36%
			CPAR	52.32%	47.68%
Improving Heart Disease Prediction Using Feature Selection Approaches [23]	BRFSS dataset	2019	Logistic Regression	82.56%	17.44%
			Random Forest	84.17%	15.83%
			Naïve Bayes	84.24%	15.76%
			LR-SVM	84.85%	15.15%
			Decision Tree	82.22%	17.78%

Table 2: Comparative Analysis with previous studies

Paper	Dataset	Year	Technique	Accuracy	Error Rate	Precision	Recall	F1	Kappa	AUC
Heart Disease Prediction Using Machine Learning	BRFSS	2023	Decision trees classifier	86.4%	13.6%	23.2%	25.5%	24.3%	16.9%	58.8%
			K-nearest neighbors' classifier (KNN)	90.5%	9.5%	35.7%	14.4%	20.5%	16.4%	56%
			Random forests classifier	90.6%	9.4%	35.5%	11.8%	17.7%	14.1%	54.9%
			Neural networks classifier	91.6%	8.4%	54.2%	9.8%	16.7%	14.4%	84.1%

Table 3: Comparative Analysis

to prevent overfitting and enhance their predictive power.

4.4 Neural Networks

Neural networks are essentially built on neurons, which are just brain cells. A biological neuron takes in information from several sources, integrates it in some fashion, applies a nonlinear operation to the outcome, and produces the ultimate result as an output. A detailed comparison between the models employed in this article and the models from earlier research is provided in Table 2.

5 Results

The decision trees, random forests, neural networks, and k-nearest neighbors (KNN) classifiers were all used with the datasets. The findings vary with respect to the algorithm and the test proportion of test data. The neural network's classifier may attain a maximum accuracy of 91% when the fraction of test data is 0.2 percent. Various classifiers produce varying outcomes and degrees of accuracy. The dataset has been used over time because it is readily available and simple to use. A detailed comparative study of the models employed in this paper is provided in Table 3.

6 Conclusion

This study presents a very comprehensive, practical, and highly preferred machine learning based model that aids in the early diagnosis of cardiac disorders by medical professionals, allowing patients to take preventative actions during a window for rectification. Based on the results displayed in Table 3 above, it can be concluded that a classification model's usage of four (04) distinct classifiers greatly influences the ratio of test to training data. These experimental results suggest that a neural network classifier with a 0.2 test data size yields the best results.

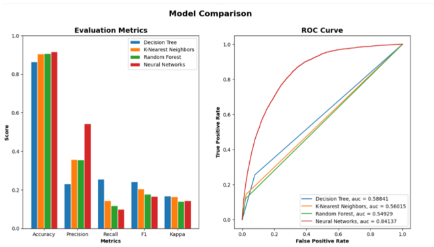

Figure 8: Visual Comparison between accuracies of algorithms

But it's impossible to overlook the fact that Random Forest frequently yields results that are almost as good. Thus, it stands to reason that the best heart attack predictions can be obtained by using the random forest and neural network classification techniques.

References

1. Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), 43-48.

2. Jindal, H., et al. (2021). Heart disease prediction using machine learning algorithms. *IOP Conf. Ser.: Mater. Sci. Eng.*, 1022, 012072, 1-4.

3. Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. *Int J Comput Sci Net Secur*, 8, 343-350.

4. World Health Organization. (n.d.). Cardiovascular Diseases(Cvds). Retrieved from `https://www.who.int/newsroom/fact%20sheets/detail/cardiovascular-diseases-(cvds)`.

5. Dwivedi, A. K., Imtiaz, S. A., & Villegas, E. R. (2019). Algorithms for automatic analysis and classification of heart sounds - a systematic review. *IEEE Access*, 7.

6. Coronary. (2020). Heart disease. Retrieved from https://www.aihw.gov.au/reports/australias health/coronaryheart-disease.

7. Allen, L. A., Stevenson, L. W., Grady, K. L., et al. (2012). Decision making in advanced heart failure: a scientific statement from the American heart association. *Circulation*, 125(15), 1928-1952.

8. Ghwanmeh, S., Mohammad, A., & Al-Ibrahim, A. (2013). Innovative artificial neural networks-based decision support system for heart diseases diagnosis. *Journal of Intelligent Learning Systems and Applications*, 5(3), Article ID 35396.

9. Al-Shayea, Q. K. (2011). Artificial neural networks in medical diagnosis. *Int. J. Comput. Sci. Issues*, 8(2), 150-154.

10. Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of The Royal Society Interface*, 8(59), 842-855.

11. Ansarullah, S. I., & Kumar, P. (2019). A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method. *International Journal of Recent Technology and Engineering*, 7(6S), 1009-1015.

12. Nazir, S., Shahzad, S., Mahfooz, S., & Nazir, M. (2018). Fuzzy logic-based decision support system for component security evaluation. *The International Arab Journal of Information Technology*, 15(2), 224-231.

13. Sen, A. K., Patel, S. B., & Shukla, D. P. (2013). A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level. *International Journal of Engineering and Computer Science*, 2(9), 1663-1671.

14. Jabbar, M., Chandra, P., & Deekshatulu, B. (2011). CLUSTER BASED ASSOCIATION RULE MINING FOR. *Journal of Theoretical & Applied Information Technology*, 32(2), 196-201.

15. Rao, R. (2011). SURVEY ON PREDICTION OF HEART MORBIDITY USING DATA MINING TECHNIQUES. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 1(3), 14-34.

16. Vijiyarani, S., & Sudha, S. (2013). Disease Prediction in Data Mining Technique – A Survey. *International Journal of Computer Applications & Information Technology*, II(1), 17-21.

17. Peter, T. J., & Somasundaram, K. (2012). AN EMPIRICAL STUDY ON PREDICTION OF HEART DISEASE USING CLASSIFICATION DATA MINING TECHNIQUES.

18. Patil, S. B., & Kumaraswamy, Y. S. (2009). Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction. *International Journal of Computer Science and Network Security (IJCSNS)*, 9(2), 228-235.

19. Ishtake, S. H., & Sharma, P. S. (2013). Intelligent Heart Disease Prediction System Using Data Mining Techniques. *International Journal of Healthcare & Biomedical Research*, 94-101.

20. Masethe, H. D., & et al. (2014). Prediction of Heart Disease using Classification Algorithms. *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA.

21. Bhatla, N., & Jindal, K. (2012). An Analysis of Heart Disease Prediction using Different Data Mining Techniques. *International Journal of Engineering Research & Technology*, 1-4.

22. Bashir, S., & Shahzad, Z. S. (2019). Improving Heart Disease Prediction Using Feature Selection Approach. *16th International Bhurban Conference on Applied Sciences & Technology (IBCAST)*, 619-623. Islamabad: IEEE.