IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 1, January 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001009 |

Effective Analysis and Prediction of Cardiovascular Disease Using Machine Learning Algorithm

Shubha Soni[1], Dr. Neeraj Sharma[2], Dr. Harsh Pratap Singh[3]

Research Scholar, Computer Science and Engineering Sri Satya Sai University of Technology and Medical Sciences, Sehore, India[1]

Associate Professor, Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore, India [2]

Assistant Professor, Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore, India [3]

ABSTRACT: Heart is considered as the second most important organ after brain. Any disturbance in Heart leads to Whole body disturbed. Diseases including heart disease as major are the result of the changes happening to us on daily basis and major that heat disease is among the among the top five killer disease worldwide. Therefore, predicting the disease at the right time and right moment is the important. Data mining is a basic and primary process in defining and discovering useful information and finding hidden patterns from large databases. Data mining and Machine Learning techniques find its use in medical sciences in resolving real health related issues by prediction and diagnosis of various diseases. This paper aims at analysing three machine learning algorithms comparatively for heart disease prediction viz., support vector machine, decision tree and random forest.

KEYWORDS: Machine learning, Random Forest, Logistic Regression, Prediction, Heart disease, Classification

I. INTRODUCTION

"Machine Learning" [1] is defined as "a method of manipulating and extracting implicit, previously unknown/known, and potentially relevant information from data." Machine Learning is a huge and complex field, and its scope and application are expanding all the time. Machine learning includes supervised, unsupervised, and ensemble learning classifiers that are used to predict and find the accuracy of a dataset. We can use that information to our HDPS initiative, as it will benefit a large number of individuals.

Cardiovascular diseases are a broad term that refers to a variety of problems that can affect your heart. According to the World Health Organization, 17.9 million people die each year from CVDs (cardiovascular diseases)[2].
It is the leading cause of death among adults. With the use of their medical history, our project can help forecast who is likely to be diagnosed with heart disease [3]. It recognises who has any heart disease symptoms, such as chest pain or high blood pressure, and can assist in detecting the disease with fewer medical tests and more effective therapies, allowing them to be treated appropriately.

This study presents an advanced hybrid intelligent approach (AHIA) for distinguishing persons with heart disease from healthy people. Additionally, each classifier's receiver optimistic curves and area under the curves are generated.In this paper, we have covered all of the classifiers, feature selection techniques, pre-processing methods, validation methods, and metrics for evaluating classifier performance. The suggested system's performance has been verified using both full features and a reduced set of features. The decrease of features has an effect on classifier performance in terms of accuracy and execution time. The suggested machine-learning-based decision support system will aid clinicians in accurately diagnosing heart patients.

The objectives of this project is to see if a patient's medical characteristics, such as gender, age, chest pain, fasting sugar level, and so on, indicate that they are likely to be diagnosed with cardiovascular heart disease. A dataset with the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001009 |

patient's medical history and attributes is chosen from the UCI machine-learning repository. We may use this dataset to forecast whether or not the patient will develop heart disease. Two datasets were obtained to predict this, one with 1026 instances and 14 attributes and the other with 303 instances and 14 attributes. When the two datasets are combined, the outcome is a dataset with 1329 instances and 14 attributes. Support vector machine (SVM), decision tree (DT), and random forest (RF) classifier techniques are used to train these medical attributes [10]. Random forest is the most efficient of these algorithms, with a 99.39 percent accuracy. Finally, we classify people as being at risk of developing heart disease or not, and this procedure is completely cost effective.

II. RELATED WORK

Senthilkumar Mohan et al. (2019) introduced a novel strategy for identifying key features using machine learning techniques, which improves the accuracy of cardiovascular disease prediction. Different combinations of features and numerous well-known classification algorithms are used to introduce the prediction model. Through the prediction model for heart disease with the hybrid random forest with a linear model, we achieve an improved performance level with an accuracy level of 88:7%. (HRFLM) [4].*Akram Ahmed Mohammedet al. (2018)*developed an approach for predicting cardiac disease using ensemble learning. To evaluate the stages of heart disease, the proposed methodology combined the Flask Web framework with the Random Forest machine learning algorithm. Heart disease is indicated by the presence of artery blockage. The more blocked arteries there are, the more advanced the stage of heart disease. Stages 1 and 2 indicate the presence of heart disease, whereas Stages 3 and 4 are referred to as chronic heart disease, and patients in these stages are at a very high risk of having a heart attack at any time. Parameters such as age, sex, blood pressure, and sugar levels are obtained from the Kaggle website and used in the forecast. According to the findings of the experiments, predictions made utilising the proposed method are consistently better than those obtained using the other methods [5].*Harshit Jindal et al. (2018)*developed a heart disease prediction system that uses the patient's medical history to determine whether the patient is going to be diagnosed with heart disease or not. To predict and classify patients with heart disease, we applied various machine learning methods such as logistic regression and KNN. To govern how the model may be utilised to increase the accuracy of prediction of Heart Attack in any individual, a very helpful technique was applied. The proposed model's strength was quite gratifying, as it was able to predict evidence of having a heart condition in a specific individual using KNN and Logistic Regression, which demonstrated a high level of accuracy when compared to previously utilised classifiers such as Naive Bayes and others. So, by utilising the specified model to determine the probability of the classifier correctly and accurately identifying cardiac illness, a considerable amount of strain has been relieved. Given's heart disease prediction technology improves medical care while lowering costs. This investigation has provided us with valuable information that can aid in the prediction of heart disease patients. It's based on the.pynb file format [6].*RiddhiKasabe and RiddhiKasabe (2018)*goal of this study is to compare and contrast various classification algorithms used in heart diagnostics. First, the numeric dataset for the heart is retrieved and preprocessed. The features that are condition to be found and categorised by machine learning are then extracted. When compared to previous methods, machine learning outperforms them. Following categorization, performance criteria such as accuracy and precision, as well as the F-measure, must be determined. Machine learning outperforms human intelligence. On the existing dataset, the comparative measure reveals that Random Forest is the best classifier for the diagnosis of heart disease [7].*InduYekkala et al. (2018)*employed the Random Forest algorithm and Feature Selection utilising rough sets to accurately forecast the occurrence of heart illness using the heart Stalog dataset from the UCI repository [8].*Amin UlHaq et al. (2018)* Using a heart illness dataset, researchers built a machine-learning-based diagnosis method for heart disease prediction. We used seven popular machine learning algorithms, three feature selection algorithms, the cross-validation method, and performance evaluation metrics like classification accuracy, specificity, sensitivity, Matthews' correlation coefficient, and execution time to evaluate the performance of seven classifiers. The proposed technology can quickly distinguish between those with heart disease and those who are healthy. Receiver optimistic curves and area under the curves were also computed for each classifier. All of the classifiers, feature selection techniques, pre-processing methods, validation methods, and classifier performance evaluation metrics that were employed in this paper have been discussed. The proposed system's performance has been verified using both full features and a reduced set of features. The decrease of features has an effect on classifier performance in terms of accuracy and execution time. The suggested machine-learning-based decision support system will aid clinicians in accurately diagnosing heart patients [9].

III. PROPOSED WORK

Heart disease is one of the most critical human diseases in the world and affects human life very badly. In heart disease, the heart is unable to push the required amount of blood to other parts of the body. Accurate and on time diagnosis of heart disease is important for heart failure prevention and treatment. 0e diagnosis of heart disease through traditional

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001009 |

medical history has been considered as not reliable in many aspects. To classify the healthy people and people with heart disease, non-invasive-based methods such as machine learning are reliable and efficient. In the proposed study, we developed a machine-learning-based diagnosis system for heart disease prediction by using heart disease dataset. We used seven popular machine learning algorithms, three feature selection algorithms, the cross-validation method, and seven classifiers performance evaluation metrics such as classification accuracy, specificity, sensitivity, Matthews' correlation coefficient, and execution time. The proposed system can easily identify and classify people with heart disease from healthy people. Additionally, receiver optimistic curves and area under the curves for each classifier was computed. We have discussed all of the classifiers, feature selection algorithms, pre-processing methods, validation method, and classifiers performance evaluation metrics used in this paper. 0e performance of the proposed system has been validated on full features and on a reduced set of features. 0e features reduction has an impact on classifiers performance in terms of accuracy and execution time of classifiers. The proposed machine-learning-based decision support system would assist the doctors to diagnosis heart patients efficiently.

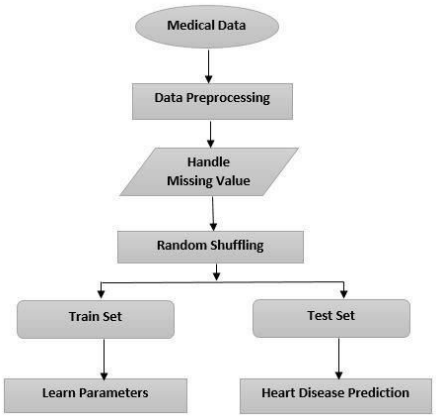

Fig. 1: Proposed Model

Working of Proposed Model

Step-1 :Data Set
The Cleveland data set is used for designing for machine learning based system for heart disease diagenesis. It has a sample size 303 patients,76 features and some missing values in these analyses six samples were removed due to missing value in the features column and left-over sample size is 297 with 13 more appropriate independent input features

Step-2 :Data Preprocessing:
The pre-processing of data is necessary for efficient representation of data and machine learning classifier which should we trained and tested and effective manner. Pre-processing technique have been applied to data set for effective use in classifier.

Step-3: Feature Selection
It is necessary for the machine learning process because some time irrelevant feature affect the classification performance of the machine learning classifier. It improves the classification accuracy and reduces the model execution time.

Step-4: Cross Validation
Cross validation are used to evaluate the performance of classifier. In order to evaluate the performance of classifier various performance evaluation matrix such as classification accuracy, classification error.

Step-5: Heart Diseases Prediction
In this step, we predict the heart diseases of person in near future by our machine learning approach.
Step-6: Parameter Calculation
In this step, we calculate the parameter, which helpful to predict the heart disease like accuracy, F1 Score, true positive rate false positive rate.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001009 |

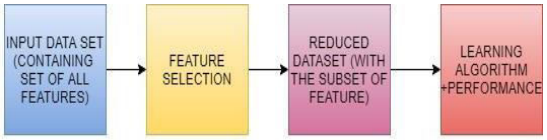

Fig. 2 Working Process of Proposed Work

IV. EXPERIMENTAL SETUP AND RESULT ANALYSIS

Dataset

The dataset has been obtained from the application of UCI machine learning repository [11]. Two datasets have been obtained one containing 1026 instances and 14 attributes and the second one containing 303 instances and 14 attributes. On combining the datasets, the resulting dataset contains 1329 instances and 14 attributes. The description of the attributes are being described in Table 1.

Table 1 Description of dataset used

S. No	Attribute Name	Description
01	Age	Age in years
02	Sex	Mal e/female
03	Cp	Constructive pericarditis
04	Trestbps	Resting blood pressure in mmHg on admission to hospital
05	Chol	Serum cholesterol in mg/dl
06	Fbs	Fasting blood sugar (greater than 120mg/dl). Values :1=true, 0=false.
07	Restecg	Resting electrocardiographic results. Values :0=normal, 1=having ST-T wave abnormality.
08	Thalch	Maximum heart rate achieved.
09	Exang	Exercise including angina.value:1=yes,0=no
10	Oldpeak	St depression induced by exercise relative to rest
11	Slope	The slope of peak exercise ST segment.value:1=up sloping, 2=flat, 3=down sloping.
12	Ca	No. of major vessels (0-3) colored by fluoroscopy
13	Thal	Inherited blood disorder that causes your body to have lesser HB than normal. Values:3=normal, 6=fixed defect, 7=reversible defect.

Software Used

The processing of the completely experimental work has been done using open-source Anaconda 2020. Anaconda is a free and conditional open-source distribution of the python, R programming languages which is used for scientific computing (machine learning and data science applications, pre-processing of huge amount of data, predictive analysis and so on), that aims at simplification of deployment and package management. In addition, for programming tasks and measurements Python (3.7.6) along with Spyder as an integrated development environment is used. After the dataset is split into 70:30 ratio that is 70% for testing and 30% for training.

Results Analysis

Here table 5.5 shows the results obtained for the accuracy parameter after the simulation of different classifier such as SVM, DT and RF classifier and it is illustrated through bar graph in fig. 5.8. After analysis it is found the accuracy value of random forest classifier (99.39%) is more than the SVM classifier (84.93%) and equal to DT classifier (97.59%), it means our proposed methodology (random forest) is better in prediction if heart disease with respect to accuracy parameter.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001009 |

Table 2 Heart disease classification using accuracy measuring parameter

Classifier	Accuracy
SVM	84.93 %
DT	97.59 %
RF	99.39 %

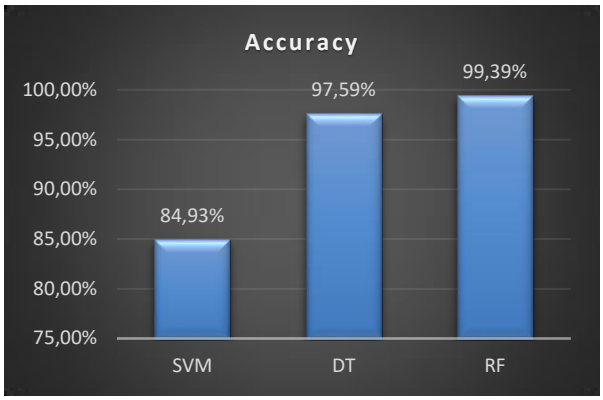

Fig. 3 Graph for Heart disease classification using accuracy measuring parameter

V. CONCLUSION

With the growing interest in heart disease prediction systems, it has become necessary to provide a summary of the still-incomplete studies. Cardiovascular disease (CVD) is a term used to describe disorders involving restricted or obstructed blood arteries, which can result in heart attacks (myocardial infarctions), chest pain (angina), and stroke.In dissimilar experiments, different datasets of heart disease patients are used. The majority of the studies used data from the UCI repository's online Cleveland database. The main contribution of this study was a comparison of various ML algorithms for prediction of CVD at the early stages. The quality of dataset was improved by using preprocessing techniques, where removing outliers and to handle corrupted and missing values was the main concern. In addition, we execute the three different machine learning algorithms to predict the disease and their results were compared with various statistical measures. The experimental results show that the accuracy of for our dataset is (100% for training set) and (97.29% for testing set) which is the highest among all classifiers. The robustness validation of SVM, DT, and RF was verified by using a 10fold cross-validation technique.

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), 43-8

[2] Dangare C S &Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-8.

[3] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.

[4] Senthilkumar Mohan et al. "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", In IEEE access 2019.

[5] Akram Ahmed Mohammed et al. "Random Forest Machine Learning technique to predict Heart disease", European Journal of Molecular & Clinical Medicine, Volume 7, Issue 4, 2018.

[6] Harshit Jindal et al. "Heart disease prediction using machine learning algorithms", IOP Conf. Series: Materials Science and Engineering 1022 (2018) 012072

[7] RiddhiKasabe and RiddhiKasabe, "Heart Disease Prediction using Machine Learning", International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 08, August-2018.

[8] InduYekkala et al. "Prediction of Heart Disease Using Random Forest and Rough Set Based Feature Selection", International Journal of Big Data and Analytics in Healthcare Volume 3, Issue 1, January-June 2018.

[9] Amin UlHaq et al. "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Hindawi Mobile Information Systems Volume 2018, Article ID 3860146, 21 pages.

[10] VaseemNaiyer, JitendraSheetlani, Harsh Pratap Singh, "Software Quality Prediction Using Machine Learning Application", Smart Intelligent Computing and Applications, 2020, Pages 319-327, Springer.

[11] https://archive.ics.uci.edu/ml/datasets/heart+disease

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details