International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Detection of Heart Disease Using Machine Learning

[1]MOHAMMAD SADIQ B, [2]SAHANA K BALAGANUR

[1]Assistent professor,Department of Master of Computer Application,BIET,Davangere

[2]Student, Department of MCA,BIET,Davangere

Abstract: There are numerous aspects in daily life that have an impact on the human heart. According to the World Health Statistics report, one in three persons worldwide suffer from hypertension, or high blood pressure. This condition increases the risk of heart attacks and other heart-related diseases as well as the risk of dying from a stroke. Medical professionals face a challenging task when determining if a patient may have heart disease because it calls for extensive testing and years of experience. Machine learning is now widely used in a wide range of other industries as a result of its tremendous success in sectors like e-commerce, retail, and marketing. The healthcare sector is one of these. This sector is regarded for being "information rich" and offers considerable potential for making wise decisions and spotting hidden patterns. The leading cause of death worldwide is heart disease. In order to forecast cardiac disease, we have employed various parameters. They are Age, Gender, and Cerebral Palsy.Gender, cerebral palsy (CP), blood pressure (bp), fasting blood sugar test (fbs), and other information are included.

I.INTRODUCTION

One of the most important issues facing the globe today is heart disease. The prediction of cardiovascular illness presents a significant problem for clinical data analysis. Hybrid machine learning (ML) has demonstrated its ability to effectively support decision-making and prediction from the vast amounts of data generated by the healthcare sector and hospitals. Additionally, we have observed the employment of ML approaches in recent advancements across various IOT domains. Only a few research have used machine learning to predict cardiac disease. The narrative approach we suggest in this study tries to identify relevant features by utilizing machine learning techniques, which improves the accuracy of cardiovascular disease prediction

The forecasting system ,By combining a hybrid random forest with a linear model to create a prediction model for heart disease, we achieve an enhanced performance level with an accuracy level of 92%. The main difficulty with heart disease is detecting it. Although there are equipment that can forecast heart disease, they are either expensive or ineffective at calculating the likelihood of heart disease in humans. Analyzing the data for hidden patterns can help reduce mortality rates by early detection of cardiac problems. The hidden patterns can be used in medicine to diagnose illnesses.

The World Health Organization estimates that heart disease causes 12 million deaths worldwide each year. Since a few years ago, the prevalence of cardiovascular disease has been rising quickly throughout the world. Numerous studies have been carried out in an effort to identify the most important risk factors for heart disease and to precisely estimate the overall risk. Heart disease is also referred to as a silent killer because it causes a

person to pass away without any evident signs. In order to avoid complications in high-risk patients and make decisions about lifestyle changes, early detection of heart disease is crucial. By examining patient data that classifies whether or not they have heart disease, this initiative seeks to forecast the occurrence of heart disease in the future.

II. Related works

Heart disease is the world's largest cause of mortality, and preventing negative health effects depends greatly on early detection. The power of data analysis and pattern recognition has been increasingly used in machine learning approaches to detect cardiac disease. This note examines some significant related efforts in the area of machine learning-based cardiac disease prediction.By Dua and Du (2019), "Predicting Heart Disease Using Machine Learning Techniques": In this study, the effectiveness of different is compared, including k-nearest neighbors (KNN), support vector machines (SVM), decision trees, random forests, and artificial neural networks (ANNs).The Cleveland Clinic Foundation dataset was used by the researchers, and they saw encouraging findings, with ANN surpassing other algorithms with an accuracy of 86.19%.

Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organisation, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality [1].Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15,2017. Heart related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India have lost up to $237 billion, from 2005-2015, due to heart related or Cardiovascular diseases [2].

learning algorithms for heart disease prediction, including KNN, SVM, decision trees, and naive Bayes, was compared.The UCI Heart Disease dataset was used by the researchers, and they discovered that SVM beat other algorithms with an accuracy of 89.15%.In this, a new set of features is derived from the original feature set.Feature extraction involves a transformation of the features. This transformation is often not reversible asfew, or maybe many, useful information is lost in the process.In [3]and[4].

Sriwannawit and Ratanamahatana's "Feature Selection and Classification Model Construction on Heart Disease Prediction" (2018):The most crucial variables for heart disease prediction were identified by the authors by focusing on feature selection methods.They used the k-nearest neighbors algorithm and a hybrid feature selection method that combined correlation-based feature selection with genetic algorithm-based feature selection to increase prediction accuracy.Krittanawong et al. (2020), "Cardiovascular Disease Prediction Using Machine Learning Techniques: A Systematic Review and Meta-analysis":This comprehensive review and meta-analysis looked at how well machine learning methods predicted cardiovascular disease.The study covered a broad range of machine learning algorithms and datasets, showcasing how these methods may be used to enhance clinical risk classification and decision-making

III. Proposed work

After reviewing the results from the existing methods, we classified heart illness using Python and pandas operations for the data collected from the UCI repository. It provides an easily comprehendible visual depiction of the dataset, working conditions, and predictive analytics development. Data preprocessing is the first phase in the machine learning process, which is then followed by feature selection based on data cleaning, classification, and performance assessment. Using the random forest method improves the outcome.

precesion. Benefits of the proposed system increased cardiac illness diagnosis precision.

IV. Contribution

In general, many diseases can be accurately diagnosed using medically linked procedures.We'll use some of the algorithms in our suggested system to improve heart disease prediction.The mental weariness of the doctors will be lessened.

V. Methodology

This study attempts to estimate the likelihood of developing heart disease using computerized heart disease prediction, which may be useful for patients and medical professionals. We used a dataset and many machine learning methods to accomplish this goal, and the findings are presented in this study report. We intend to sanitize the data, get rid of extraneous details, and add new characteristics like MAP and BMI to improve the technique. The dataset will then be divided depending on gender, and k-modes clustering will be used. Finally, we will use the cleaned data to train the model. As seen in Figure 1, the enhanced methodology will result in more precise findings and greater model performance.

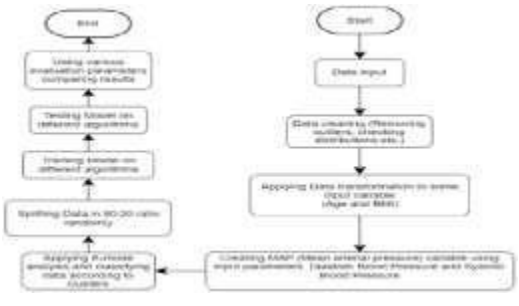

Fig 1: Architecture of Heart Disease Detection

According to [23], the dataset used in this investigation consists of 70,000 patient records with 12 distinctive traits, which are given in Table 2. Age, gender, systolic and diastolic blood pressure are some of these characteristics. The target class, "cardio," determines whether a patient has cardiovascular illness (expressed as 1) or is in good health (shown as 0).

A group of instances are grouped using a machine learning technique called clustering based on similarity indices. The k-means technique is a popular clustering algorithm, although it is useless when dealing with categorical data. The k-modes algorithm was created to get around this restriction. The k- modes approach was developed by Huang [29] in 1997 and is similar to the k-means technique in that it replaces the clusters' cluster means with their modes and uses dissimilarity measures for categorical data. This makes it possible for the algorithm to handle categorical data properly.

Given the major biological differences between men and women, which might affect how diseases manifest and advance, dividing the dataset based on gender can be useful for prediction. Men may have heart disease at a younger age than women do, and they may also experience different symptoms and risk factors. According to research, males are more likely than women to develop coronary artery disease (CAD), and there may be differences in the risk factors and symptoms of CAD between the sexes [30]. It is feasible to find particular risk variables and patterns of illness progression by studying the data individually for men and women, which may be possible when the data are aggregated.

VI. Experiment and Results

The "Heart Disease UCI" dataset, which includes several heart health parameters and a target variable indicating the presence or absence of heart disease, was used in the experiment. The objective was to create a machine learning model that, using the supplied features, can forecast cardiac

disease.The Random Forest algorithm, a well-liked ensemble learning technique that blends various decision trees to create predictions, was employed in the experiment. For classification jobs like predicting heart disease, Random Forest is effective. One-hot encoding was used to handle missing values in the dataset and transform categorical variables into numerical representations. Following that, training and testing sets were created from the dataset.

The training set, which consisted of 100 decision trees built and trained on various subsets of data, was used to train the Random Forest classifier. Throughout the training process, the model discovered patterns and connections between the characteristics and the target variable.

Using the trained Random Forest classifier, predictions were performed on the testing set after the model had been trained. By contrasting the predicted values with the actual values of the target variable, the accuracy of the model was determined. The percentage of accurate predictions the model made on the test set is reflected in the accuracy score.

Additionally, the confusion matrix was computed to provide more detailed insights into the model's performance. The confusion matrix shows the number of true positives, true negatives, false positives, and false negatives, which can be used to evaluate the model's predictive power and identify any potential biases or limitations.

It's vital to remember that the precise experiment outcomes could differ depending on the dataset, the algorithm of choice, the preprocessing methods used, and other elements. The supplied example serves as a broad blueprint for carrying out a heart disease prediction experiment using Python machine learning.

Visualization:

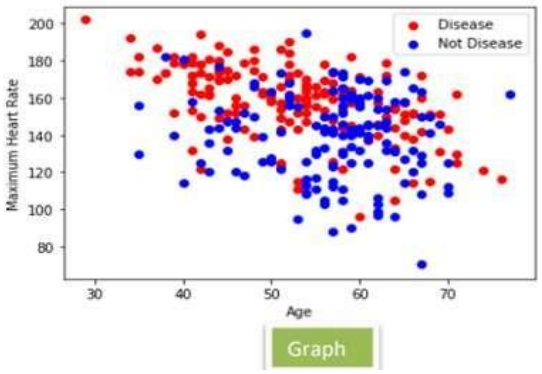

Fig 2:Graph of Heart Disease Detection

VI.Conclusion

The long-term saving of human lives and the early detection of irregularities in heart problems will be made possible by identifying the processing of raw healthcare data of heart information. In order to process the raw data and deliver a fresh and original insight towards heart disease, machine learning techniques were applied in this study. Prediction of heart disease is difficult and crucial in the medical industry. However, if the disease is discovered in its early stages and preventative measures are implemented as soon as feasible, the fatality rate can be significantly reduced. In order to focus the research on real-world datasets rather than only theoretical frameworks and simulations, further development of this study is extremely desirable. When using the proposed hybrid HRFLM technique, the traits of Random Forest (RF) .

REFERENCES

[1] Ramadoss and Shah B et al.“A. Responding to the threat of chronic diseases in India”. Lancet. 2005; 366:1744–1749. doi: 10.1016/S0140-6736(05)67343-6.

[2] Global Atlas on Cardiovascular Disease Prevention and Control. Geneva, Switzerland: World Health Organization, 2011

[3] Dhomse Kanchan B and Mahale Kishor M. et al. “Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis”, 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication.

[4] R.Kavitha and E.Kannan et al. “An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining “, 2016

[5] N. Al-milli, ‘‘Backpropogation neural network for prediction of heart disease,’’ J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[6] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and
R. Deepika, ‘‘Analysis of neural networks based heart disease prediction system,’’ in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233– 239.

[7] P. K. Anooj, ‘‘Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27– 40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.

[8] L. Baccour, ‘‘Amended fused TOPSIS- VIKOR for classification (ATOVIC) applied to some UCI data sets,’’ Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025