Journal of Engineering Sciences
Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

MYOCARDIAL INFARCTION ANALYSIS SYSTEM USING MACHINE LEARNING

Pooja Sakhare,Dept of CSE, Tulshiramji Gaikwad Patil College of Engineering & Technology, Nagpur poojasakhare0406@gmail.com

Prof. Jayant Adhikari, Dept of CSE, Tulshiramji Gaikwad Patil College of Engineering & Technology, Nagpur jayant.cse@tgpcet.com

ABSTRACT: ML, a popular application of Artificial Intelligence (AI), is making a huge impact on the research sector. In this study, artificial intelligence is employed to determine whether or not a subject has cardiac disease. Many people around the world are afflicted with cardiovascular diseases (CVDs), which are sometimes fatal. In order to determine if a person is suffering from cardiovascular disease, machine learning can take into account various attributes, such as chest pain and cholesterol levels. supervised learning algorithms, a form of machine learning, can be used to diagnose cardiovascular problems. In this study, significant traits are discovered using machine learning methodologies, improving the accuracy of cardiovascular disease prediction. Various feature combinations and well-known classification methods are employed in the model's introduction.. EDA + Classification + Ensemble predicts heart disease at a 92% accuracy rate, which is an improvement over previous models.

Cardiovascular illnesses are the focus of this study. Algorithms for classification and machine learning

1. INTRODUCTION

Every organ in the body has a certain job to do in order to function properly. Failure of the heart to adequately pump blood throughout the body can be life-threatening. The main cause of death in the United States today is heart disease [1]. Therefore, it is imperative that we maintain good health in all systems of the human body, such as our cardiovascular system. Sadly, cardiovascular disease affects people all around the world. Any device that can detect and cure these diseases before they spread will save lives and money. Data mining tools can be used to forecast heart disease. [2] Predictive models can be built by discovering patterns and trends in databases that were previously unknown. Data mining is a technique for extracting information from large volumes of data. Early detection of heart issues using machine learning can help save patients' lives. According to recent developments in science and technology, a person's risk of heart disease can be predicted using machine learning.

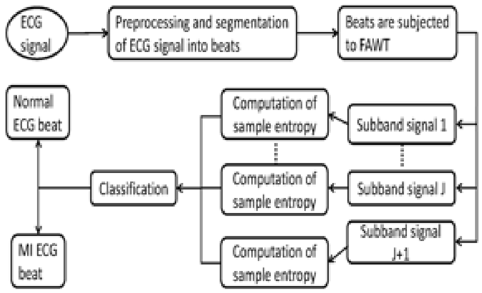

Fig.1: Example figure

Heart troponin testing is becoming more widely used in clinical settings. Due to improvements in test analytical precision, cardiac troponin concentrations are now finding new applications in early diagnosis and risk assessment for individuals being assessed for suspected acute coronary syndromes. Patients with a rise or decrease in cardiac troponin concentration have traditionally been classified as having myocardial infarction if their 99th percentile cardiac troponin concentration is greater than or equal to one of the median values. Using high-sensitivity cardiac troponin testing, some studies suggest that patients can be risk stratified to differing degrees of myocardial infarction risk using lower criteria. Accommodating earlier troponin measurements at presentation and 1 to 3 hours later as part of their diagnostic frameworks, accelerated diagnostic pathways advocate these strategies to help speed up diagnosis and treatment in patients with myocardial infarction or to hasten discharge in those who do not.

2. LITERATURE REVIEW

2.1 Responding to the threat of chronic diseases in India [1] :

At this point in India's health revolution, chronic diseases account for around 53% of deaths and 44% of lost productive years. Diabetes and cardiovascular disease are more common in urban areas. The number of people diagnosed with cancer as a result of cigarette usage is astronomical.. Cigarette smoking and non-smoking tobacco use is prevalent among rural and disadvantaged communities. In spite of the prevalence of hypertension and dyslipidemia, they remain underdiagnosed and undertreated. It is expected that by 2030, the number of people suffering from chronic illness would climb sharply due to demographic and socioeconomic factors. After being established in 1975, the national cancer control programme has built 13 registries and increased treatment capacity since then. Anti-tobacco legislation was passed in its entirety in 2003. A national strategy to prevent and treat heart disease and diabetes is currently being developed. In order to better allocate resources, coordinate multi-sectoral policy actions, and increase the involvement of the health system in activities related to chronic disease prevention and control, there is a need.

Machine learning techniques based on component analysis (PCA) have been studied in [3].

According to a global assessment, heart disease and its syndrome are the main causes of mortality. If current trends continue, 23.6 million people will die from heart disease by 2030. It's a pity that the healthcare industry collects so much data on heart disease yet doesn't do anything with it. According to a study in this paper, a minimal number of attributes is needed to improve accuracy in various supervised machine learning algorithms by using PCA (principal component analysis). The purpose of this research is to examine machine learning algorithms for predicting heart illness. Data mining relies heavily on categorization and pre-processing procedures.

Diabetes is a condition that can be fatal in many developed and developing countries, including India. 1865 datasets of diabetic patients were used to classify the data, all of which were retrieved from a hospital repository and have a wide range of attributes in common. The dataset includes two types of blood tests and one type of urine test. A variety of data mining methods used to forecast diabetes sickness are examined in this study. Data mining, a well-known technique for identifying diseases like diabetes and cancer, is commonly used in bioinformatics research.

2.3 An Efficient Framework for Extraction and Features [4]: For heart disease classification,

Pre-processing employs a high-dimensional data set to classify the heart disease dataset. There are many duplicates and irregularities in this raw dataset, making it difficult to search for and store. In order to achieve accurate categorization, it is required to remove extraneous and irrelevant data. In order to condense large amounts of data into smaller amounts, the dimensionality reduction method must be applied. Using a framework, heart disease can now be accurately predicted. The features are extracted and the most important ones are determined using PCA and a mathematical model. Increased efficiency, precision, and speed are achieved by the proposed effort. Image processing and matching patterns are just a few examples of how this technology could be put to use.

The Random Forest Classification Framework [5]: 2.4 Cardiovascular Risk Prediction Method.

Cardiovascular disease (CVD) is a major factor to the loss of both quality and quantity of life in many countries. Clinicians rely greatly on early diagnosis and prediction of danger in order to treat their patients and deliver correct diagnoses for them. An accurate and realistic risk prediction system for the auxiliary medical profession will be developed using data mining techniques. There are four main components to a healthcare information system: data interface, data preparation, feature selection and classification. Hospitals must respond to the data interface in order for raw data to be obtained; this includes pre-processing for data integration and data purification among other things. We utilised the CFS Subset Evaluation and the Best First Search method to identify the most essential attributes. Random forests have previously been used to predict cardiovascular disease (CVD) risk. PKU People's Hospital inpatient data and the Cleveland Heart-Disease Database were both examined to guarantee accuracy and usefulness.

Building a Cardiovascular Disease Prediction Model Using a Structural Equation Model and a Fuzzy Cognitive Map:

According to the Public Health Agency of Canada, cardiovascular disease (CVD) is the leading cause of mortality for both men and women (PHAC). Using machine learning/data mining techniques to predict CVD has a number of limitations, including a lack of transparency in the construction of the prediction model and an inability to include human wisdom. In this research, we provide a new approach to addressing these issues and construct a model that is exceedingly robust and quite accurate. Our method is based on SEM and FCM, both of which are based on structural equation modelling (FCM). Data from the 2012 Canadian Community Health Survey was used to conduct the tests. There is an accuracy level of

ISSN NO:0377-9254

74% in the model's predictions. It is our belief that adding additional traits and a panel of cardiac doctors would further enhance the accuracy of our method.

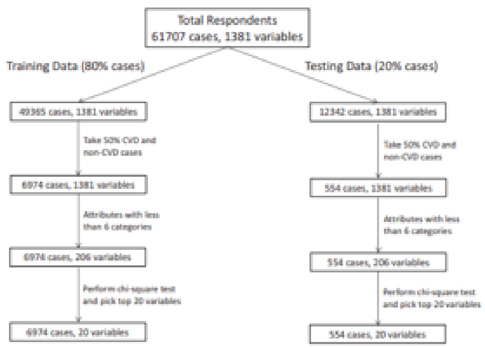

Fig.2: Data preprocessing model

2.6 Prediction of Heart Disease Using Hybrid Technique For Selecting Features [7]:

Despite the fact that the healthcare industry is often referred to as "information rich," not enough data is being mined to identify hidden patterns. Medical decisions are made more effectively thanks to advances in data mining. When predicting cardiac illness, we used naive bayes and random forest models. In order to improve the model's performance, it is also recommended that features be selected before categorization. In order to select features from a dataset and assign a weight to each one, methods like SVM-RFE and gain ratio are used. Using this approach, you can get better accuracy while simultaneously saving time on your computer. Experiments have shown that the proposed strategy of selecting characteristics enhances the accuracy of both models.

3. IMPLEMENTATION

To identify the Cleveland UCI repository for cardiac disease, we used a R studio rattle Visual representations are used to portray data, the working environment, and the process of creating predictive analytics. For each new set of inputs, a series of data pre-processing steps is employed, culminating in an iterative iteration that culminates in a model with improved precision. Features and model selection are ongoing for a wide range of quality.

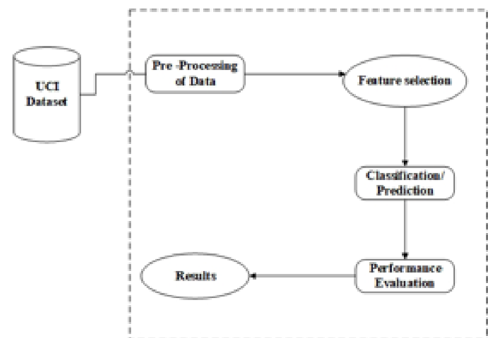

Fig.3: Model architecture

Data pre-processing:

Many records are obtained and pre-processed to gather information on cardiovascular disease. Six records in the dataset include incomplete data, bringing the total number of patient records in the dataset to 303. In pre-processing, only 297 patient records are used after excluding those six entries. Multi- and binary-category variables are used to classify dataset attributes. A person's risk of heart disease can be assessed using the multi-class variable. Having a heart condition results in a value of 1, while being heart-healthy results in a zero value. During the pre-processing stage, medical records are transformed into diagnosis values. One-third of the 297 patient records with data pre-processing results indicate the presence of heart disease, with one-third showing the absence of heart illness, with a value of one demonstrating the absence of heart disease.

Features can be pared down:

Two of the data set's 13 qualities, referring to the patient's age and gender, are used to determine his or her identity. It is important to note that the following 11 qualities contain critical medical records. Heart disease diagnosis and severity assessment rely heavily on medical records. As previously indicated, a variety of machine learning approaches, including NB, GLM, LR, DL, DT, RF, GBT, and SVM, are employed in this experiment. Experiments with all 13 attributes and all ML approaches were carried out in the experiment.

Description of the dataset:

We used the UCI Machine Learning repository's Cleveland Heart Disease Dataset [2]. There are 303 records in total, with 6 records being blank. In this dataset, the data are represented as both numerical and category. Preprocessing deleted the missing information, resulting in 137 individuals with heart disease and 160 patients without. Target class attribute is the final of the 14 attributes in the dataset.

4. ALGORITHMS

After removing the null values from the characteristics and samples, several Machine Learning methods are applied to the dataset in order to predict the outcomes. The algorithms employed in the model are outlined in the following section:

The Naive Bayes classification algorithm is a probabilistic classifier.. Strong independence assumptions are incorporated into the probability models that underlie this approach. The reality does not always support the notions of independence. Consequently, they are viewed as inexperienced. In order to estimate the likelihood of distinct classes based on various attributes, Naive Bayes employs a similar approach. Text categorization and problems involving many classes are common applications for this approach.

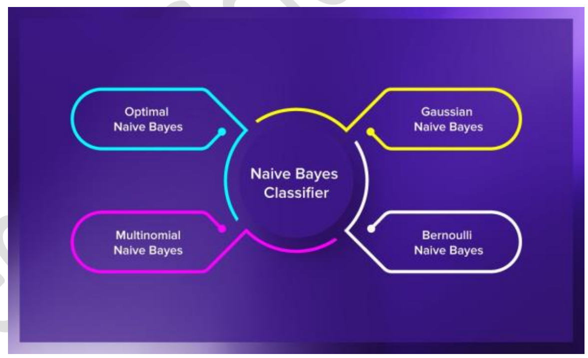

Fig.4: Naïve bayes model

Support Vector Machines (SVMs), a common supervised learning method, can be used to address classification and regression problems. This technique is typically used in Machine Learning to solve Classification problems. So that future data points may be easily classified, we need to discover the best line or decision boundary that divides n-dimensional space into classes for future use. The best suitable boundary is a hyperplane. When constructing the hyperplane, the SVM makes advantage of the dataset's most outlying points and vectors. The "Support Vector Machine" technique is also known as a "Support Vector Machine" because of the phrase "support vectors" (SVM). The

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

following two classifications in the diagram are denoted by a decision boundary or hyperplane:

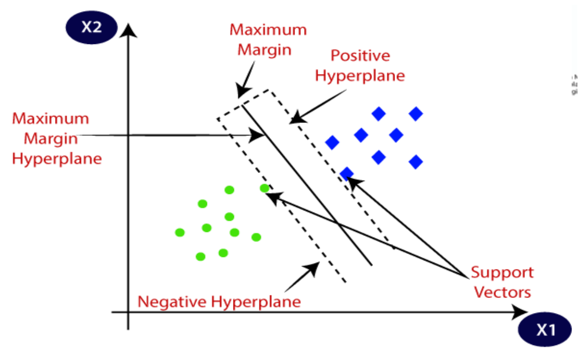

Fig.5: SVM model

K Nearest Neighbour is a straightforward algorithm for classifying incoming data or instances based on their degree of similarity to already stored data. Most commonly, it is used to classify data points based on the classification of their neighbours. If you want to use KNN for classification, it first calculates how far an example is from an input query, then it chooses K instances (the number of examples) that are closest to the query. Finally, it votes on the most frequent label, or it averages all of the labels (in the case of regression).

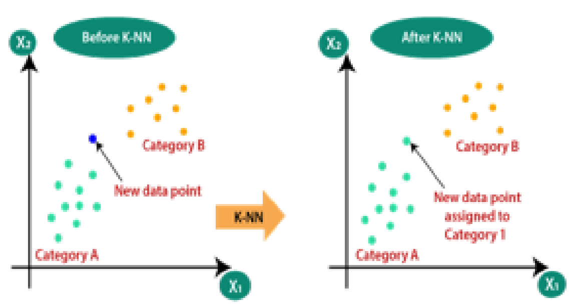

Fig.6: KNN model

It is up to decision trees, which use various strategies, to decide whether to divide up a node into multiple smaller nodes. Adding sub-nodes has the advantage of increasing the homogeneity of the generated sub-nodes. Purity improves as a node's value grows in relation to the target variable. Supervised Machine Learning (SML) is a technique in which the training data is continually segmented according to a particular parameter and the results are then analysed. Decision nodes and leaves make up the tree's structure.

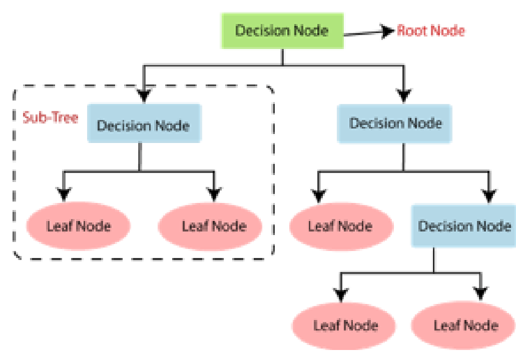

Fig.7: Decision tree model

As a Supervised Machine Learning Algorithm, random forest is utilised extensively in Classification and Regression. For classification and regression, it uses a combination of multiple samples to construct decision trees, which are voted on by a majority of the members of the sample. This approach is suitable for both classification and regression. Because of the cross validation, it has a better level of accuracy. It is possible to use a random forest classifier to handle missing values and keep a large amount of data accurate.

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

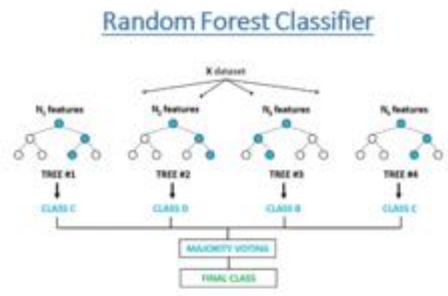

Fig.8: Random forest model

5. EXPERIMENTAL RESULTS

In this work, Jupyter Notebook was used to apply the algorithms for the Cleveland Heart Disease Dataset. The ROC curve and the f1-score are two measures used to gauge the test's diagnostic accuracy. Here are a few of the factors that play a role in this phenomenon:

Positivist Error: A false positive is when the model predicts a result as positive when it actually isn't. In other words, a true negative occurs when the model correctly forecasts a negative result. A false positive occurs when the model predicts the positive class incorrectly.

A true negative occurs when the model correctly predicts the negative class. A false positive occurs when the model predicts the wrong class. False negatives can also be produced when the model forecasts the negative class incorrectly.

When a test result wrongly suggests the presence of an illness (like a disease) when it is not there, false negatives occur when the test result incorrectly indicates the absence of an illness (like a disease) when it is present.

False negative mistakes or false negatives refer to test findings that suggest a condition doesn't exist. Pregnancy tests that suggest a woman is not pregnant while she is, for example, are examples of false negatives, as are not guilty judgments in criminal cases.

One must be able to distinguish between the real negative and true positive, or true negative plus false negative + false positive.

False Negatives + Negatives

True Positive + False Negative equals Sensitivity."

False Positive/(False Positive + True Negative) Equals Specificity. /

Journal of Engineering Sciences

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

Accuracy is the sum of false positive and true positive

Predicting the F1 score is as simple as multiplying the precision by the recall rate.

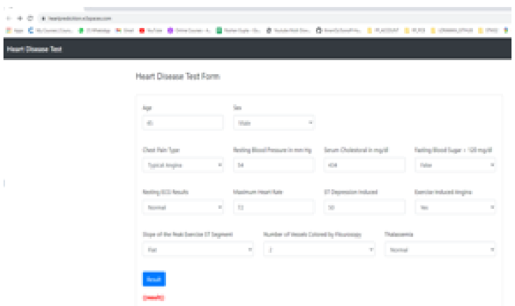

Fig.9: Predict HD page

Doctors only have access to the Predict HD page shown above. To demonstrate how the patient's data can be included in a prediction HD, look no further than this page. This form's input validation prevents spam and other unwanted data from being entered and resulting in an error message. The vast majority of input data must be entered in a drop-down list in order to limit the amount of junk that is entered. When a doctor correctly forecasts HD, the patient's progress is displayed on a special success page. In addition, the patient's report is entered into the database table as part of this process.

6. CONCLUSION

We'll be able to finish this project with all of its goals checked off our list. A web application for HDPS contains all of the features and functions that we had envisioned. This project's research requirements have been met in the system. The real-world problem is solved by this system. HD can be accurately predicted in patients. It can also be used to store and manage a doctor's predictions for a cardiac patient's prognosis. The admin user has the ability to establish new doctor accounts, manage existing ones, and examine patient reports. We can answer the project's issue statement and create a novel tool to forecast an HD in the hospital by using the whole system. As part of our domain and technical research, we were able to do a literature review and learn more about our subject matter. Every researcher had done their own research to forecast the HD at domain research, and we were able to identify 13 of these significant characteristics. In addition, we developed two algorithms, DT and NB, which are both highly accurate in heart disease datasets, and NB yields an overall accuracy of 88%. However, the HD dataset's dispersed results may jeopardise the heart dataset study. As a result, the Kaggle dataset has been modified. We were able to successfully complete the final project thanks to our technical and other research. This project requires a certain programming language, database, and technique. We had a hard time locating specific information regarding the HD's characteristics during our investigation. In the journal manuscript and dataset description, there was less information about the incorrect attributes name, their influence, and their sub-attributes. We were unable to gather additional information on HD because of the doctor's busy schedule. owing to a lack of time, we were not able to study other aspects and different heart disease types.

7. FUTURE SCOPE

The HDPS will soon be able to forecast a specific form of HD, such as heart attacks, CVD, CAD, etc. Potential uses for HDPS include hospitals, clinics,

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

smartphones and smart wearables as well as emergency systems and fitness apps. We plan to use this algorithm to forecast cardiac disease in the hospital and clinic system. The HDP Model will be included into smart clothing to detect and recommend HD precautions. Using this paradigm, we'll also be able to test ourselves on a mobile app. Smart clothing will be integrated into the hospital and police emergency systems in order to save the patient's life.

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] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan et al. “Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework”, 2017 IEEE 2nd International Conference on Big Data Analysis.

[6] Manpreet Singh, Levi Monteiro Martins, Patrick Joanis and Vijay K. Mago et al. “ Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map”, 978-1-5090-0626-7/16/$31.00 c 2016 IEEE.

[7] Kanika Pahwa and Ravinder Kumar et al. “Prediction of Heart Disease Using Hybrid Technique For Selecting Features”, 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON).

[8] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. “ A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017

[9] Hanen Bouali and Jalel Akaichi et al. “Comparative study of Different classification techniques, heart Diseases use Case.”, 2014 13th International Conference on Machine Learning and Applications

[10] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. “ A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017

[11] Houda Mezrigui, Foued Theljani and Kaouther Laabidi et al. “Decision Support System for Medical Diagnosis Using a KernelBased Approach”, ICCAD’17, Hammamet - Tunisia, January 19- 21, 2017.

Vol 13, Issue 05, May / 2022

ISSN NO:0377-9254

[12] Dr.(Mrs).D.Pugazhenthi, Quaid-E-Millath and Meenakshi et al. "Detection Of Ischemic Heart Diseases From Medical Images " 2016 International Conference on Micro-Electronics and Telecommunication Engineering.

[13] J. Hodges et al. "Discriminatory analysis, nonparametric discrimination: Consistency properties," 1981.

[14] S.Rajathi and Dr.G.Radhamani et al. "Prediction and Analysis of Rheumatic Heart Disease using kNN Classification with ACO ", 2016.

[15] Puneet Bansal and Ridhi Saini et al. "Classification of heart diseases from ECG signals using wavelet transform and kNN classifier", International Conference on Computing, Communication and Automation (ICCCA2015).