IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE PREDICTION SYSTEM

[1]Priyanka Gaikwad, [2]Anil Gujar,

Department of Computer Engineering
TSSM Bhivarabai Sawant College of Engineering Research
Savitribai Phule Pune University , Pune, India

Abstract: The healthcare is one of the important research field in the current situation with the rapid improvement of technology and data. The primary concern of this research work is the usage of machine learning techniques for predicting heart diseases using clinical and symptom-related information. Considering the fast-growing volume of medical records, it becomes difficult to rely on conventional approaches to the analysis of data. Machine learning is used to examine huge amounts of data quickly and detect patterns linked to diseases. To build predictive models, the present study employs the dataset related to Cleveland Heart Disease, provided by the UCI Machine Learning Repository available at Kaggle. Several approaches to classifying information and developing predictive models will be examined, such as the Support Vector Machine, Decision Tree, and Random Forest. In particular, the Support Vector Machine (SVM) technique is important to discuss because it is highly effective in solving classification problems by defining an optimal hyperplane to differentiate points in multidimensional space. This System will be implemented using SVM Algorithm. It is classification based machine learning algorithm.

Index Terms: **support vector machine, heart disease, machine learning, supervised, unsupervised.**

I.INTRODUCTION

The human heart is considered the key organ of the human body. In essence, it controls the flow of blood in our body. The malfunctioning of the human heart might lead to troubles with other parts of the body. The abnormal functioning of the heart can be categorized under Heart Diseases. As per the reports of the World Health Organization, over ten million deaths occur because of heart diseases annually worldwide. A healthy living pattern and early detection can only prevent heart-related diseases. Heart disease refers to a number of medical disorders concerning the heart. There are various reasons for increasing the risks of heart disease, including high cholesterol, obesity, high triglyceride level, hypertension, and others.

Heart Disease Prediction system is a desktop application for predicting heart disease based on the symptoms entered by the user. This prediction system has **d**atasets gathered from various websites dedicated to health issues. Using Heart Disease Prediction system, the user can determine his chances of having this disease based on his entered symptoms.

The existing system which has been developed for the prediction of heart disease operates on a limited dataset. Our proposed system will operate on large datasets to improve the efficiency of the entire system. It is quite easier to heart disease predictions with the help of our system, which takes input from natural language processing of the user's problems related to health issues.

The Past researches use 13-attribute data sets for their analysis. Classification is the method used in all cases to determine whether the person is having heart disease or not. The results achieved have high accuracy of about 88 percent using SVM.

The rest of the paper is segmented into six parts. section 1 contains the introduction, section 2 contains the literature review, section 3 contains the methodology applied, section 4 contains the discussion, section 5 contains the analysis of results, and section 6 contains conclusions and future scope.

II.LITERATURE REVIEW

The prediction of heart disease using machine learning techniques has been widely explored in recent years, with various approaches proposed to improve diagnostic accuracy and efficiency. Researchers have applied multiple classification algorithms and hybrid models to analyze clinical datasets and identify patterns associated with cardiovascular diseases.

In heart disease prediction using machine learning method research paper by Mafia Rasheed, Muhammad Adnan Khan, Ghasan F Issa, Taher M. Ghazal. The purpose of this paper is to propose a light and easy methodology for predicting heart disease using machine learning. Prediction of heart disease can be done using machine learning. In this paper, various machine learning methods have been employed and the result compared based on different performance measures. The objective of this study is to make comparative analysis of prediction of heart disease by using publicly available datasets taken from UCI machine learning repository. Various publically available datasets are available like Swiss dataset, Hungarian dataset, Cleveland dataset, etc. Here we use Cleveland dataset having 303 entries and total 14 variables for our study and test. These datasets are cleaned by removing the noisy and missing values from it. The pre-processing dataset is then used for the analysis of the research. Six different machine learning models have been compared based on the various performance criteria and the one that shows better accuracy has been selected as the model for prediction of the heart disease. Finally, a graphical user interface is designed for prediction of heart disease.

In Efficient Heart Disease Prediction System using Decision Tree by Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma Cardiovascular disease (CVD) is a big reason of morbidity and mortality in the current living style. Identification of Cardiovascular disease is an important but a complex task that needs to be performed very minutely, efficiently and the correct automation would be very desirable. Every human being cannot be equally skillful and so as doctors. All doctors cannot be equally skilled in every sub specialty and at many places we don't have skilled and specialist doctors available easily. An automated system in medical diagnosis would enhance medical care and it can also reduce costs. In this study, we have designed a system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the user's requirement. The performance of the system is evaluated in terms of classification accuracy and the results shows that the system has great potential in predicting the heart disease risk level more accurately.

In Analyzing the impact of feature selection on the accuracy of heart disease prediction research paper by Muhammad Salman Pathan, Abhishek Nag, Muhhammad Pathan. The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to pro- vide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. In this paper, we apply different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modelling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of disease.

III ALGORITHM

Support Vector Machine (SVM) is a supervised machine learning algorithm primarily used for classification and regression tasks. It works by finding an optimal hyperplane that separates data into different classes with the maximum possible distance between them.

The main goal of SVM is to maximizes the margin between the two classes. The larger the margin the better the model performs on new and unseen data. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Hyperplane :

A decision boundary separating different classes in feature space and is represented by the equation $wx + b = 0$ in linear classification. SVM works by finding the hyperplane that maximizes the margin—the distance between the boundary and the closest data points, known as "support vectors". There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM. The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2

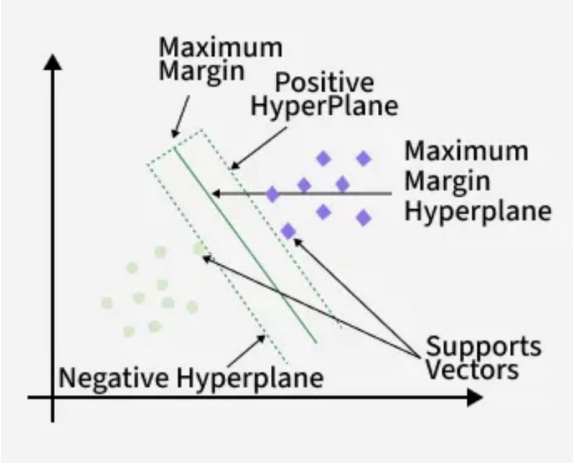
features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

Support Vectors:

The hyperplane is determined solely by the points closest to it, known as support vectors. These points define the boundary's location and orientation, and moving them affects the hyperplane.

Margin:

The distance between the hyperplane and the nearest support vectors from each class. SVM's primary goal is to maximize this margin to ensure the model generalizes well to new, unseen data.



Support vectors are the closet data point to the hyperplane that define the class boundary.

A hyperplane is a plane that separate different classes.

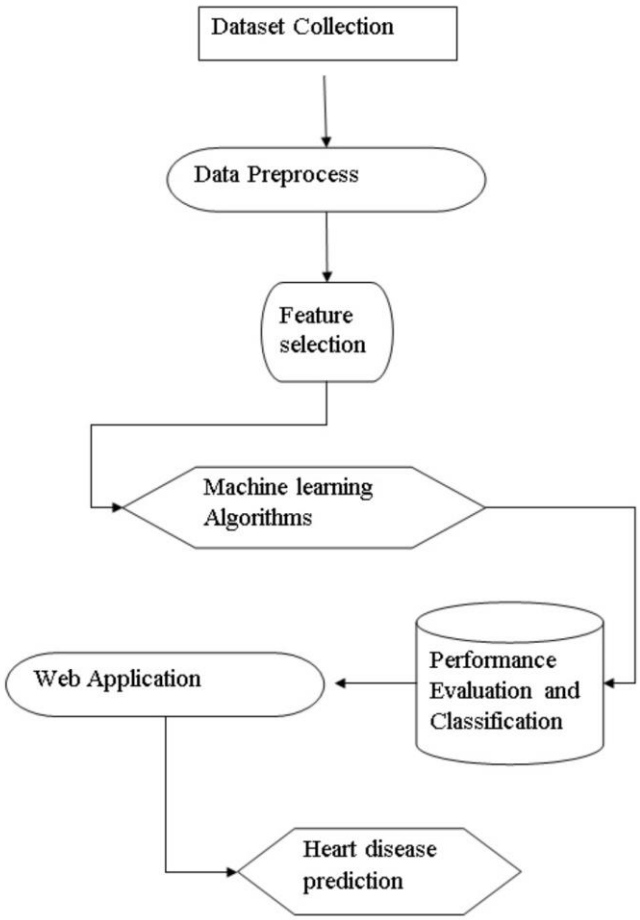
Steps:

1. Import the dataset
2. Explore the data to figure out what they look like
3. Pre process the data
4. Split that data into attributes and labels
5. Divide the data into training and testing sets.
6. Train the SVM algorithm
7. Make predictions
8. Evoluate the result of the algorithm.
9. Support vectors are the data points that are close to the decision boundary, they are the data points most difficult to classify, they hold the key for SVM to be optimal decision surface. The optimal hyperplane comes from the function class with the lowest capacity i.e minimum number of independent features/parameters.

IV. PROPOSED SYSTEM

Following Figure depicts the project plan where the first step involves problem statement identification. Here, the problem statement to be identified is whether the person has heart disease or not. Next is publicly available datasets. Various types of datasets are available in UCI machine learning repository which includes Hungarian dataset, Switzerland dataset and Cleveland dataset among others. The next step involves selection of the dataset. For example, in this thesis the Cleveland dataset was used. Next is formulation of problem solving strategies. In formulating problem solving strategy, various classifiers will be considered, they include SVM i.e. support vector machine, Random forest, Decision tree, Naïve Bayes, KNN and logistic regression. The next step is data preprocessing. Data preprocessing involves removal of noise and missing values in the dataset. Next is model building for prediction. Model is built using publicly available datasets like sex and age. Then training is done using different machine learning algorithms and performance metric measures are used in obtaining results. The next step is application of the model to the user's

data and prediction of heart disease. Last step is comparison of various machine learning models based on their classification result. In the last step it will compare all the classifiers and find out whether a person has heart disease or not.



1. Problem Definition

Predict likelihood of heart disease based on patient health parameter.Predict person has heart disease or not. If person has heart disease it represent by 1 and if person has no heart disease it represent by 0.

2. Data Collection

Gather dataset(heart disease dataset, hospital record)

3. Dataset Selectioon

Select relevant feature like name, age ,gender, chest pain type, blood pressure, cholesterol, heart rate etc. and perform exploratory data analysis.

4. Data Preprocessing

Handle missing values, Normalise Data, Encode Categorial Variable.

5.Feature Selection

Identify important feature using correlation, statistical test and create derived features if needed.

6. Split Dataset

Train /Test (60/40)

7. Input Details

User input details like name, age, gender chest pain, Thestbps, Chol, FBS, exang, Thalach, oldpeak, slope, CA, Thal, Restecg.

8. Model Trainning

I n this step the comparison is done between the classifier. Such as SVM, KNN,Naive Bayes, Logistic Regression. Compare all model based on the accuracy, precision, recall and f1 score. and select the best performing one.

9. Prediction Module

Apply train module to user input. GUI is created in Python using tkinter that generates a dialogue box asking for inputs regarding all the parameters to perform evaluation. The input data is taken from the user and there comes a prompt that determines if the person is suffering from heart disease or not.

V. RESULT

For evolution process confusion matrix, accuracy score, precision, recall and f1 score are used. Confusion matrix used to evaluate how accurately a machine learning model identifies patient with or without heart disease. The matrix breaks down prediction into four categories. The first one is True Positive(TP). The model correctly identifies a patient with heart disease. The second one is false positive(FP) in which the values identified are false but are identified as true. The Third one is false negative (FN). In which the value was true but was identified as negative. The forth one is true negative . In which the value was negative and was truly identified as negative.

TP	FN
FP	TN

Accuracy score is used to measure how often model makes correct prediction. Accuracy calculated as

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

After accuracy there is specificity which is the proportion of true negative cases that were classified as negative thus , it is measure of how well a classifier identifies negative cases.

Specificity=TN/(TN+FN)

Sensitivity also known as Recall or True Positive Rate (TPR) measures a machine learning models ability to correctly identified actual positive instance from the total positive cases.(TP/TP+FN).

VI. CONCLUSION

Heart is one of the essential and vital organ of human body and prediction about heart diseases is also important concern for the human beings so that the accuracy for algorithm is one of parameter of analysis of performance of algorithm.Accuracy of the algorithm depend on dataset that isa used for training and testing purpose.

The main goal of this project is to predict heart disease based on the symptoms. The system takes symptoms from user as input and generate output I.e. predict heart disease. This system will reduce medical error and saves time. This application will help in faster diagnosis, reduce medical error and easy to apply. This system successfully predict heart disease with 98% accuracy.

FUTURE WORK

The similar prediction system web application can be built with integrated AI chatbot. It will provide detail analysis.This application will be accessible to the doctor as well as patient. That web application can be accessible from any device.

REFERENCES

1. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 04 | Apr 2024
2. Zohaib Ahmed , "Heart Disease Prediction using Machine Learning Algorithm(2025)
3. Mafia Rasheed, Muhammad Adnan Khan, Ghassan F. Issa, Taher M. Ghazal "Heart Disease Prediction Using Machine Learning Method". ICCR,2022
4. Rahul Katarya, Polipireddy Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey", Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE.
5. Alive Kor, [Online]Available: https://www.alivecor.com/how-it-works.
6. Rahul Katarya, Polipireddy Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey", Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE.
7. Aditi Gavhane, Gauthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD), "Prediction of heart disease using machine learning", International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record
8. Archana Singh, Rakesh Kumar,"Heart Disease Prediction using Machine Learning Algorithm:,International Conference on Electrical and Electronics Engineering,2020.
9. Akash Chauhan, Aditya Jain,Purushottam Sharma, Vikas Deep "Heart Disease Prediction using Evolutionary Rule Learning ",International Conference on Computational Intelligence and communication technology(CICT 2018).