International Journal of Innovative Research in Engineering
Volume 3, Issue 4 (July- August 2022), PP: 47-50.
www.theijire.com
ISSN No: 2582-8746

Heart Disease Prediction Using Machine Learning

Joseph George[1], K Arjun Sasi[2], K A Althaf Kareem[3], Joe Aju[4]
[1,2,3,4]*Department of computer science, Adi Shankara Institute of Engineering and Technology, Kerala, India.*

How to cite this paper: Joseph George[1], K Arjun Sasi[2], K A Althaf Kareem[3], Joe Aju[4], "HEART DISEASE PREDICTION USING MACHINE LEARNING", IJIRE-V3I04-47-50.

Abstract: *Heart attack disease is one of the leading causes of the death worldwide. In today's common modern life, deaths due to the heart disease had become one of major issues, that roughly one person lost his or her life per minute due to heart illness. Predicting the occurrence of disease at early stages is a major challenge nowadays. Machine learning when implemented in health care is capable of early and accurate detection of disease. In this work, the arising situations of Heart Disease illness are calculated. Datasets used have attributes of medical parameters. The datasets are been processed in python using ML Algorithm i.e., Random Forest Algorithm. This technique uses the past old patient records for getting prediction of new one at early stages preventing the loss of lives. In this work, reliable heart disease prediction system is implemented using strong Machine Learning algorithm which is the Random Forest algorithm. Which read patient record data set in the form of CSV file. After accessing dataset, the operation is performed and effective heart attack level is produced. Advantages of proposed system are High performance and accuracy rate and it is very flexible and high rates of success are achieved.*

Key Word:*— Decision Tree, Naive Bayes, Logistic Regression, Random Forest, Heart Disease Prediction*

I.INTRODUCTION

The work heart disease effects the functioning of the heart. World Health Organization had made a survey and made a conclusion that 10 million people are affected with heart disease and lost their lives. The problem that the Healthcare industry faces in today's life is early prediction of disease after a person is affected. Records or data of medical history is very large and the data in real world might be incomplete and inconsistent. In past predicting the disease effectively and treatment to patients might not be possible for every patient at early stages under these circumstances.Many scientists tried to build a model which is capable of predicting the heart disease in the early stage, but they are not able to build a perfect model. Every proposed system has disadvantages in its own way. In the existing system, Shen et al. had initially, proposed a system which is based on self-appliedquestionnaire. In this system the user needs to enter all the symptoms which he is suffering from, based on that the result is predicted. This study is based on the analysis data collected in SAQ.Chen et al. came up with an idea to predict heart disease. He used the technique of Vector Quantization which is one of the artificial intelligence techniques for classification and prediction purpose. Training of neural networks is performed using back propagation to evaluate the prediction system. In the testing phase approximately 80% accuracy is achieved on testing set. Practical use of data collected from previous records is time consuming. Low accuracy rate.So to overcome this we are implementing Random forest algorithm in order to achieve accurate results in less time. Machine learning is given a major priority in modern life in many applications and in healthcare sector. Prediction is one of area where machine learning plays a vital role, our topic is to predict heart disease by processing patient's dataset and a data of patients i.e., user of whom we need to predict the chances of occurrence of a heart disease.

II.RELATED WORK

[1]'Prediction and Analysis of Heart Disease using SVM Algorithm' heart disease is a fatal disease by its nature. This disease makes a life-threatening complexity such as heart attack and death. The importance of Data Mining in the Medical Domain is realized and steps are taken to apply relevant techniques in the Disease Prediction. We are implementing a system which will help to predict heart disease depending on the patient's clinical data related to the factor associated with heart disease. By using medical dataset of the patients such as age, sex, blood pressure, overweight and blood sugar and by appling SVM classifier we can predict that the patients getting a heart disease or not. In addition, classification accuracy, sensitivity, and specificity of the SVM have been found to be high thus making it a superior alternative for the diagnosis. We are alsodoing analysis on the data from which we are getting at which age it mostly occur or which region gets influenced by that disease. So precaution can be taken to avoid the death due to the heart disease.

[2] 'Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms By Python Jupiter' The aim of

this study is to evaluate the risk of 10-year CHD using 14 IVs. The attributes are selected after the backward elimination process considering the P values which are lower than 5%. Therefore, the logistic regression model is derived through P values of the variables <0.05 (sex, age,cigsPerDay, totChol, sysBP, glucose). According to the logistic regression outcome, men are more suspectable to heart disease than women. Age, number of cigarettes per day and systolic blood pressure are the odds of CHD.However, There is no significance change in the total cholesterol level and the glucose level. But, the level of glucose has a negligible change in odds. The model is more specific than sensitive. Further, the accuracy of the moel is 0.87. The value under the ROC curve is 73.5which is somewhat satisfactory. Moreover, the model could be improved by using more data.

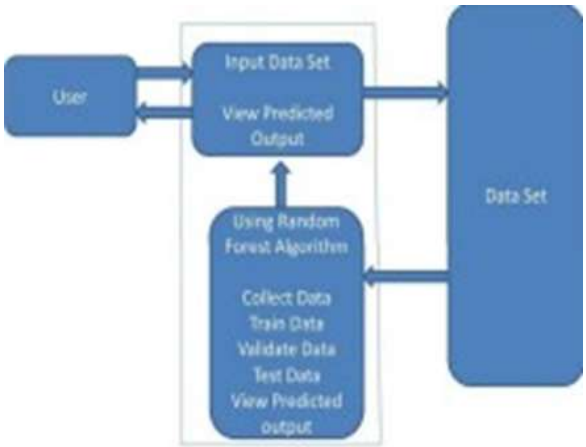
[3] 'prediction system for heart disease using nave baise' Decision Support in Heart Disease Prediction System is developed using Naive Bayesian Classification technique. The system extracts hidden knowledge from a historical heart disease database. This is the most effective model to predict patients with heart disease. This model could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy. HDPS can be further enhanced and expanded. For, example it can incorporate other medical attributes besides the above list. It can also incorporate other data mining techniques. Continuous data can be used instead of just categorical data. HDPS can be further enhanced and expanded. Another area is to use Text Mining to mine the vast amount of unstructured data available in healthcare databases. Another challenge would be to integrate data mining and text mining.

[4] 'Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters' Data mining technics have been used in many fields, one of them is healthcare. This paper's objective is to check whether heart attack prediction can be based on fewer parameters than what recommended on previous studies. We use 8 parameters (out of 13 recommended), which are: (1) Age, (2) Sex, (3) Chest pain, (4) Resting blood pressure systolic, (5) Resting blood pressure diastolic, (6) Resting ECG, (7) Resting heart rate, and (8) Exercise induced angina. The reasons to choose those parameters for this study are: they are simple measurements and consistently recorded in Harapan Kita Hospital, the biggest cardiovascular hospital in Indonesia. Experiments using 8 parameters with KNN shows good accuracy if we compared with 13 parameters, even with other data mining algorithms like Naive Bayes and Decision Tree we use Simple CART). The benefit as the result from this study is: we can proof those 8 simple parameters are good enough to be used in heart attack prediction. In our future researchit can be used as parameters in remote patient monitoring using machine-to-machine (M2M) technology, especially for patients treated at home or remote clinics. The end-to-end M2M will be built and a prediction system will be embedded as the novel feature.

[5] 'Prediction of Heart Disease Using Decision Tree Approach'The decision-tree algorithm is one of the most effective and efficient classification methods available. It has been shown that, by using a decision tree, it is possible to predict heart disease vulnerability in diabetic patients with reasonable accuracy. Classifiers of this kind can help in early detection of the vulnerability of a diabetic patient to heart disease. Preprocessing of a data set for the removal of duplicate records, normalizing the values used to represent information in the database. Clustering technique, simple k-means algorithm is used. Thus, the patients can be forewarned to change their lifestyles. This will result in preventing diabetic patients from being affected by heart diseases, thereby resulting in low mortality rates as well as reduced cost on health care for the state. This can be extended in future to predict other types of ailments which arise from diabetes, such as visual impairment. The proposed work can be further enhanced and expanded, to use stacking techniques to increase the accuracy of decision trees and reduce the number of leaf nodes.

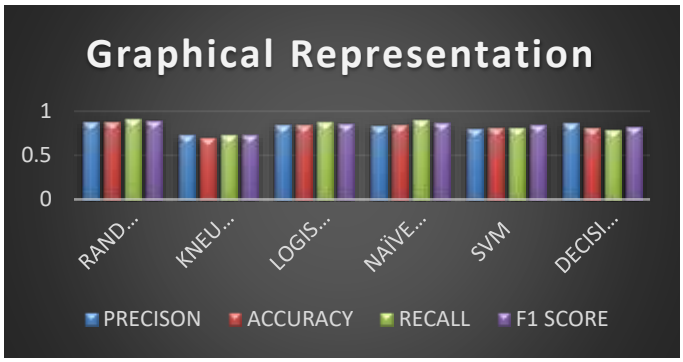
III.ARCHITECTURE OF PROJECT

The working principle of the system is shown in fig. The user enters the input which is compared with the data present in the existing data set by using the Random Forest Algorithm. Random Forest algorithm is an efficient ML algorithm that comes under supervised learning technique. It is be used for both Regression and Classification problems. To solve a complex problem, it uses a process of combining multiple classifiers, to increase the accuracy and performance of the model.



IV.COMPARISON TABLE

Algorithm	Precision	Accuracy	Recall	F1 Score
Random forest	0.886	0.885	0.915	0.889
KNeighbor classifier	0.735	0.705	0.735	0.735
Logistic Regression	0.857	0.852	0.882	0.870
Naïve Bayes	0.838	0.852	0.912	0.873
SVM	0.811	0.820	0.882	0.845
Decision tree	0.871	0.820	0.794	0.831



The project involved analysis of the heart disease patient dataset with proper data processing. Then, different models were trained and predictions are made with different algorithms KNN, Decision Tree, Random Forest,SVM,Logistic Regression From This Comparison its Cleared That the accuracy rate of Random Forest algorithm is much greater than all other algorithms. Random Forest has an accuracy rate above 87%. In predicting heart disease.

V.RESULT AND ANALYSIS

The results obtained by applying Random Forest, Decision Tree, Naive Bayes and Logistic Regression are shown in this section. The metrics used to carry out performance analysis of the algorithm are Accuracy score, Precision (P), Recall (R) and F-measure. Precision (mentioned in equation (2)) metric provides the measure of positive analysis that is correct. Recall [mentioned in equation (3)] defines the measure of actual positives that are correct. F-measure [mentioned in equation (4)] tests accuracy.

Precision = (TP) / (TP +FP) (2)
Recall = (TP) / (TP+FN) (3)
F– Measure =(2 * Precision * Recall) / (Precision +Recall) (4)

- TP True positive: the patient has the disease and the test is positive.
- FP False positive: the patient does not have the disease but the test is positive.
- TN True negative: the patient does not have the disease and the test is negative.
- FN False negative: the patient has the disease but the test is negative.

In the experiment the pre-processed dataset is used to carry out the experiments and the above-mentioned algorithms are explored and applied. The above-mentioned performance metrics are obtained using the confusion matrix. Confusion Matrix describes the performance of the model.
Feature selected from dataset is given in table:

Sl. No.	Attribute Description	Distinct Values of Attribute
1.	*Age*- represent the age of a person	Multiple values between 29 & 71
2.	*Sex*- describe the gender of person (0-Female, 1-Male)	0,1
3.	*CP*- represents the severity of chest pain patient is suffering.	0,1,2,3
4.	*RestBP*- It represents the patient's BP.	Multiple values between 94 & 200
5.	*Chol*- It shows the cholesterol level of the patient.	Multiple values between 126 & 564
6.	*FBS*- It represent the fasting blood sugar in the patient.	0,1
7.	*RestECG*- It shows the result of ECG	0,1,2
8.	*MaxHeart*- shows the max heart beat of patient	Multiple values from 71 to 202
9	*Exang*- used to identify if there is an exercise induced angina. If yes-1 or else no=0	0,1

10.	*OldPeak*- describes patient's depression level.	Multiple values between 0 to 6.2.
11.	*Slope*- describes patient condition during peak exercise. It is divided into three segments(Unsloping, Flat, Down sloping)	1,2,3.
12.	*CA*- Result of fluoroscopy.	0,1,2,3
13.	*Thal*- test required for patient suffering from pain in chest or difficulty in breathing. There are 4 kinds of values which represent Thallium test.	0,1,2,3
14.	*Target*-It is the final column of the dataset. It is class or label Colum. It represents the number of classes in dataset. This dataset has binary classification i.e. two classes (0,1)In class "0" represent there is less possibility of heart disease whereas "1" represent high chances of heart disease. The value "0" Or "1" depends on other 13 attribute.	0,1

VI.CONCLUSION

Random Forest algorithm is an efficient algorithm which is an ensemble learning method for regression and classification techniques. The algorithm constructs N of Decision trees and outputs the class that is the average of all decision trees output. So accuracy of prediction at early stages is achieved effectively. Processing of healthcare data i.e., data related to heart will help in early detection of heart disease or abnormal condition of heart which results in saving of long-term deaths. Heart disease prediction is a major challenge in the present modern life. With this application if the patient/user is away from reach of doctor, he/she can make use of the application in prediction of disease just by entering the report values. And can proceed further whether to consult a doctor or not.

VII.FUTURE SELECTION

In future this application can extended by updating some features like, if the user is effected with heart disease all his family members will be notified with a message in early. And also the information should be passed to the nearest hospital. Another feature is there should be online doctor consultation with the nearest doctor available.

In this regard, it is important to note that, ML applications using various efficient algorithms are utilized not only in disease prediction and diagnosis but also in the field of radiology, bioinformatics and medical imaging diagnosis etc.
In future if we use neural network algorithm for predicting heart disease it will give more accurate result.

References

[1]. Prediction and Analysis of Heart Disease using SVM Algorithm Madhura Patil1, Rima Jadhav2, vishakha Patil3, Aditi Bhawar4, Mrs. Geeta Chillarge5 1,2,3,4Students, Dept. of Computer Engineering, Marathwada Mitra Mandal's College Of Engineering, Pune, Maharashtra, India

[2]. Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms By Python Jupyterlab A. S. ThanujaNishadi University of Colombo, Faculty of Graduate Studies, Sri Lanka

[3].] prediction system for heart disease using nave baise Shadab Adam Pattekari and Asma Parveen Department of Computer Science and Enginering Khaja Banda Nawaz College of Engineering, RouzaBuzurg, Gulbarga-585 104, Karnataka, India.

[4]. Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters I Ketut Agung Enriko, Muhammad Suryanegara, DadangGunawan Dept. of Electrical Engineering, Universitas Indonesia, Indonesia.

[5]. Prediction of Heart Disease Using Decision Tree Approach R. Vijaya Kumar Reddy1 , K. Prudvi Raju2 , M. Jogendra Kumar3 , CH. Sujatha4 , P. Ravi Prakash5 1, 2, 5 Asst. Professor, Department of IT, PVPSIT, Kanuru, Vijayawada, Andhra Pradesh, India 3, 4 Asst. Professor, Department of ECM, PVPSIT, Kanuru, Vijayawada, Andhra Pradesh, India