INTERNATIONAL JOURNAL OF COMPUTATIONAL LEARNING AND INTELLIGENCE
Vol. 01, Issue. 01, August 2022

Heart disease prediction using remote consultation and machine learning techniques

Hemadri A D[1] . Kumar Raja D R[2] . G. Hemanth Kumar[3]

[1]School of Computing and Information Technology, REVA University, Bengaluru, India.
[2]School of Computer Science and Engineering, REVA University, Bengaluru, India.
[3]School of Electronics Communication and Engineering, REVA University, Bengaluru, India.

Abstract – Heart disease is one of the major problems facing today lifestyle. There have been too many instances when you or someone else needed immediate medical attention, but for some reason you could not. Heart Disease Prediction Application is a support and online consultation application for end users. Here, we proposed heart disease prediction application that allows users to receive guidance from the heart through the process of machine learning algorithms. This application can also be used for free online heart disease counselling and the system acts as a decision support system and proved to be an aid for the physician for the diagnosis with the help of the data set.

Index Terms – Accuracy, Algorithm, Dataset, Decision support, Diagnosis, End user, Heart disease prediction, Machine learning, Risk factor, User.

I. INTRODUCTION

The correct prediction of heart disease can prevent life threats and incorrect prediction can also prove to be fatal at the same time. we have proposed a model using machine learning in this model we have used attributes causing heart disease the data is collected from the source consist of reasons for causing heart attacks these includes Age,sex,cp,trestbps,chol.restecg,exang,oldpeak,slope,target,ca,tcal these data is used to find patient has a possibility of getting heart disease.

In this model for the trained data we use machine learning algorithms to predict the heart disease these predictions are represented in the form of graphs by using this technique can differentiate the patient of getting heart disease and we can overcome the heart attacks.

II. RELATED WORK

There is a lot of work in areas directly related to this article. The author proposed a new method to improve the accuracy of cardiovascular disease prediction by applying machine learning technology to find important features. Identifying raw health data processing of heart information can save lives in the long run and help detect heart disease abnormalities early [1]. An overview of the numerous models are implemented on these algorithms and analyse their concert. Concludes that machine learning algorithms have great scope for envisaging cardiovascular disease. To each of the algorithms mentioned above worked very well in some cases, but not so well in others [2]. The models based on various properties associated to heart disease an supervised learning algorithms such as, decision trees, Naive Bayes, random forest and K-nearest neighbour's algorithms. Uses

existing data sets from the Cleveland Heart Disease Patient Database in the UCI Repository [3]. The superiority of cardiac disease classifying by filtering redundant features using the Fast Correlation Based Feature Selection (FCBF) method. It performs classifying based on various classification algorithms. The outcomes validate the effectiveness and robustness in dispensation for different types of information for heart disease classification [4]. Although several data mining classification algorithms for envisaging heart disease exist, data for envisaging heart disease in diabetic patients are insufficient, the existing models are not performed consistently, so they optimized it to perform best in predicting the probability of heart disease in diabetic patients[5]. When comparing K-Means neighbour classifiers, enhanced k-means neighbours are represented with fewer information sets and produce output with higher accuracy compared to real k-means neighbour classifiers [6]. To predict cardiovascular disease based on traits they proposed a machine learning technique. BMI is used for predicting cardiovascular disease. This prototypical proposed various features like classification and regression methods [7]. Use ANN and logistic regression to predict signs of heart disease in specific individuals. This indicates superior accuracy compared to earlier used classifiers as naive Bayes. Significant pressure is relieved to find the probability that the classifier will adequately and precisely recognize heart disease [8]. Although there are insufficient studies of each algorithm on meta analysis methodologies for mutually heart failure and arrhythmias, no differences are revealed due to overlapping confidence intervals among the dissimilar methods. Though, there is heterogeneity between ML algorithms about several parameters. This data can help doctors interpret the statistics and implement optimal algorithms for the data set [9]. A novel computational model for envisaging premature heart disease has been proposed. The predictive model is included in a new normalization based on decomposing the weights according to the standard deviation of the weight matrix and comparing the result to the parent [10].

III. METHODLOGY

A. DATA SET

The Data set in the table 1,2 are the selected data attributes from the patient record have taken as a data Samples of 304 patients samples of different age group, in Fig 1,2 it gives more information of the data set the data set contains 13 selected medical attribute .This gives us the much needed information explained in the Fig 1,2 these helps the doctor to detect weather the patient have a risk of getting heart attack and classify the patient who have a high risk of getting heart attack these record include two stages training and testing is explained in steps involved.

	Age	Sex	Cp	Trestbps	chol	fbs	Thal
0	63	1	3	145	233	1	1
1	37	1	2	130	250	0	2
2	41	0	1	130	204	0	2
3	56	1	1	120	236	0	2
4	57	0	0	120	354	0	2

TABLE 1:DATA SET

	Restecg	Exang	Oldpeak	slope	Target	Ca
0	1	0	2.3	0	1	0
1	0	0	3.5	0	1	0
2	0	0	1.4	2	1	0
3	1	0	0.8	2	1	0
4	0	1	0.6	2	1	0

TABLE 2:DATA SET

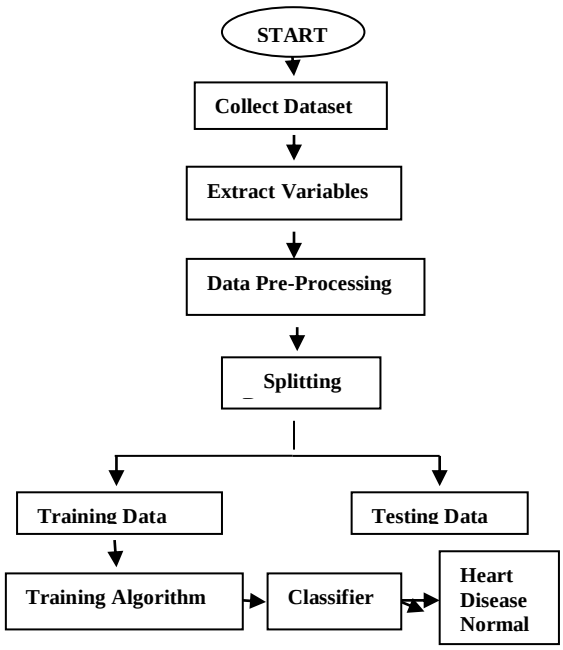

Fig. 1. Flow Chart of the proposed model

B. STEPS INVOLVED

Start The process involves in selection of algorithm data attributes:
The description in the Fig.1. is explained in these line of text Collection of Data set The data is loaded and the raw text data is converted into a separated File Input/file Heart .CSV. Extract significant variable the data is pre-processed and normalized the data in to a tabular column. Data Pre-Processing is normalized and the system can understand the data attributes. Splitting Data and the 80% data is taken as a test data set and validation data set is taken to check how accurate the model can classify the unknown data and 20% of data used as a test data. Training Data we use algorithms to train the data cleaning and normalization of data is done. Testing Data Finally the Proposed model is undertaken where evaluation of model by their performance and accuracy using the performance matrix and Check which model gives a proper accuracy among all the proposed algorithms.

C. EQUATION

Over all count of all the algorithm

The code is executed j times each time the inner loop is run. So for each time the middle loop is run, we execute the statement 1 + 2 + 3 + ... + i - 1 + i times. Machine recognize that equation is equal to i * (i + 1) / 2. Or (i^2 + i) / 2 For each time we run the outer loop, we execute the statement ((1^2+1) + (2^2+2) + ... (n^2+n))/2 times. By the equation we came across how the machine run the inner loop of all the algorithm to get accuracy of all the algorithm.

IV. RESULTS

From the results the accuracy of the previous research papers and in present proposed system algorithms are more accurate to detection of patient diagnosed with the Heart disease. The algorithm used is cost efficient and user- friendly algorithms than previous algorithm. The proposed Algorithms K-Nearest Neighbor,Decision tree Algorithm the accuracy among these is very low all the selected algorithm in the proposed system mentioned in Table 2 give the same results.

```
target      1.000000
exang       0.436757
cp          0.433798
oldpeak     0.430696
thalach     0.421741
ca          0.391724
slope       0.345877
thal        0.344029
sex         0.280937
age         0.225439
trestbps    0.144931
restecg     0.137230
chol        0.085239
fbs         0.028046
Name: target, dtype: float64
```

Fig. 2. Loading of dataset

In Fig.2. Loading of data set data is executed by machine the values present in the Fig.2. it differentiate all the values and provide accurate values present in data set and represent data values belong to the category of Int,Float,VarChar.

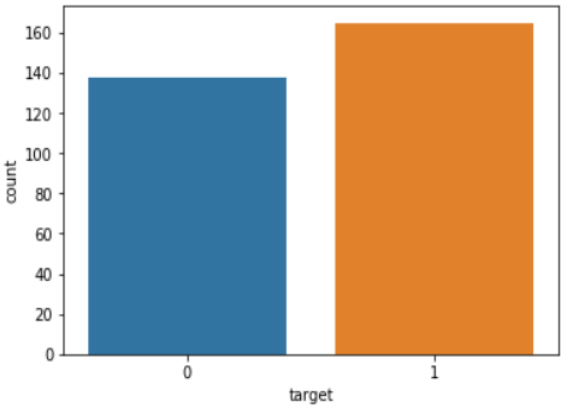

Fig. 3. Heart disease count

In Fig.3. Heart disease count the Fig.3. the loaded data in the machine learning algorithm the attribute heart disease count the algorithm reads the data of heart disease count only the individual attribute is executed the output of the attribute vary with index value 1 and target value 165 and they represent 0 to 160 in count and 1 and 0 heart disease of patient it consist of heartdiseasecountandavergeofpatientaffecte dbythe disease or not the patient belongs to 1 has a possibility of getting the disease of 160 and patient belongs to 0 has a possibility of getting the disease is 140 percentage The Fig.3. helps to know the average of the patient getting the heart disease.

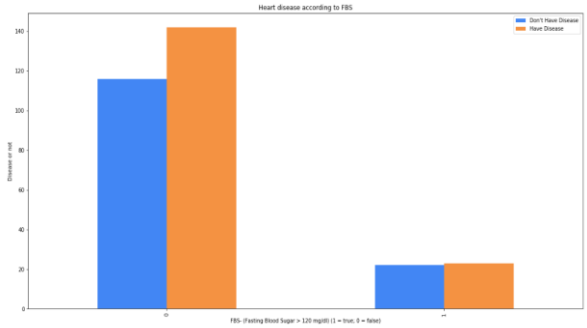

Fig. 4. Heart Disease by chest Pain Type

In Fig.4. heart disease by chest pain type the machine learning algorithm executes the individual attribute the the data is visualized with respect to count and chest pain type the heart beat rate of a patient is counted from 0 to 100 and the age of the patient is taken as a average of 0,1,2,3 the colour code of the orange represent the patient is having the possibility of getting the heart disease in case 1 the patient has a possibility of 40 percentage and no possibility of getting the heart.
disease in is 60 percentage in case 2 the possibility of getting the chest pain is 20 and the possibility of not getting the heart disease is 80percentage in case 3 the possibility of the getting the heart disease is 10 and the possibility of not getting the heart disease is very less in case 3 the possibility of getting the heart disease is very less than previous and the possibility of not getting the heart disease is also less than the previous output in visualization these visualization technique

is very easy to find the the average of the individual attribute.

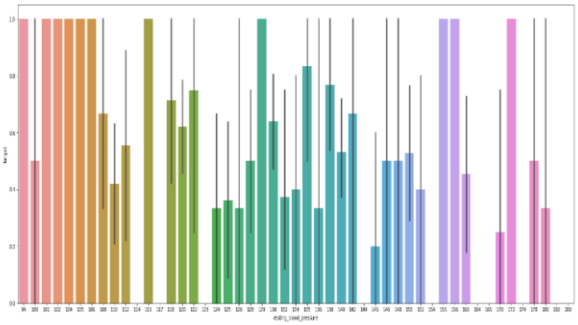

Fig.5. Heart disease by resting electrocardiographic results.

In Fig.5. ECG results the electrocardio graphic results the count of the ECG is taken as a average of0to100and0to20to100isthecountofEC G and the points are based on individual patients the colour represent the possibility of getting the heart disease and lines represent the possibility of not getting heart disease in the case of electro cardio graphs results.

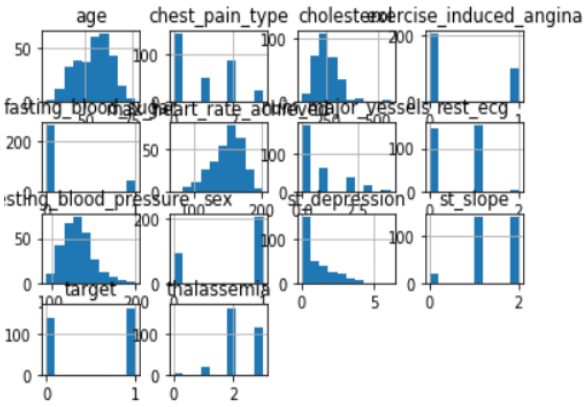

Fig.6.Visulazation of all the attribute

In Fig.6. visualization of all the attribute they are counted in the average of 0.0 and 100 all the attribute are visualized to 100 and predict the patients have a possibility of getting the heart disease the colour code for identify the the average number of possibility to get a heart attack sex attribute provide the average of 60 percentage of agehasapossibilityofgettingtheheartdiseaseand 10 percentage of patient does not have a possibility of getting the heart disease in cp average count 60 percentage of patients has a possibility of heart disease and average of 40 to 100 percentage does not have a possibility of getting the heart disease in the case 0,1,2,3,4 in fbs the possibility of getting the heart disease is higher than the other attribute the possibility is upto 110 percentage there is average of for the patient not getting the heart disease in fbs in restecg the average of getting the heart is counts as 70-80 percentage in average and patient does have a possibility of getting the heart disease is 20-30 percentage in average in exang the possibility of getting the heart disease is up to 140 percentage in average in case 1 and 10-20 percentage has a possibility in case 4 and 20-70 percentage does not have a possibility getting the heart disease in slope attribute the possibility of getting the heart disease in case 1,3 is 0-50 and 40- 80 does not have a possibility of getting the heart disease case 5the possibility is high from 0-100 the visualization technique of the attribute is used to find the maximum number of patient has a possibility of getting the heart disease and does not have a possibility of getting the heart disease.

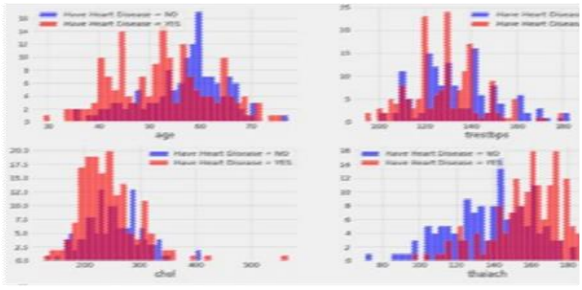

Fig. 7. Attributes had affected by Disease or not

In Fig.7. Attribute had affected by disease or not in thisvisualizationbasedontheinputinputparameter of a particular attribute. In age the possibility is taken as 0-16 and the age parameter is taken as 70 aveargeagefromtheaverageof0-70among60age group has a highest possibility of getting the heart related problem in trestbps the average of the attribute is from 0-25 and trestbps 120-180 in this the possibility of getting the heart disease in 120- 130 and in chol the average is taken has 0.0-20.0 and1.0-500theaverageofgettingtheheartaffected diseaseinthisattributeupto0.00to200percentage of getting the heart related disease in thalach the count 0-16 and 00-180 in this attribute there is equal number of possibility in getting the heart disease. The visualization technique is used to calculate the individual attribute parameter and possibility of getting the heartdisease.

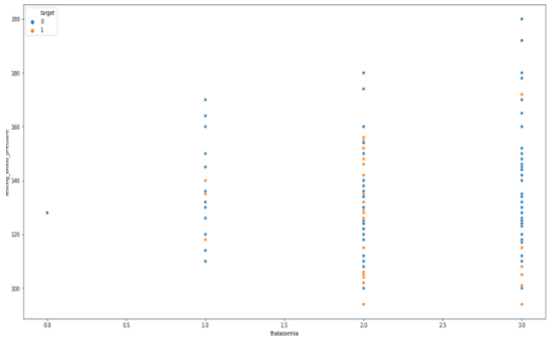

Fig. 8. Heart disease in function of Age and Max Heart Rate

IntheFig.8.Heartdiseasefunctionofageandheartrate in the visualization technique the highest possibility of getting the heart disease is taken in age and heart rate the average count is taken as 80-200 max heart rate and 30-70agethecolourredrepresentthepatientisaffected by the heart disease and blue colour does not have possibility of getting the heart disease this visualization technique is used to find the average using the scatter plot with maximum number of disease affected and maximum number of not disease affected in heart rate and age.

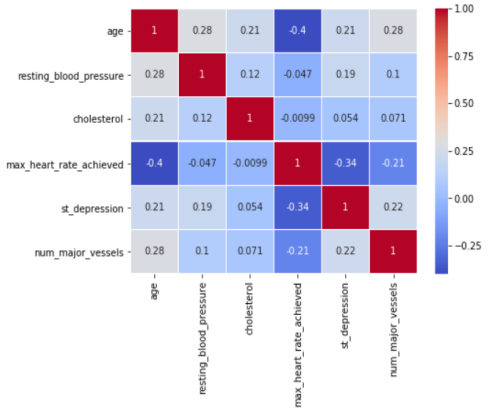

Fig. 9. Co- Relation Matrix

Here, we have plotted a co-relation matrix using the trained data set Fig.9. This provides this co- relation matrix provides the true labels and predicted values in heart attack prediction. Other than the activities in down stairs and upstairs having the same Sing master rest of all the activity can be identified as 100% accuracy. The matrix is used to check the model if it is confused among two classes .

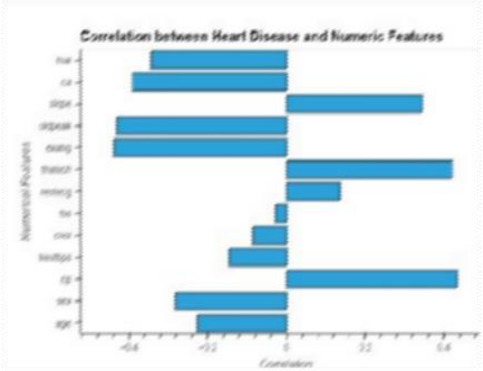

Fig.10. Co-Relation Matrix Representation of Heart Disease Numeric Features

In Fig.10. co relation matrix representation the visualization graph is used to find the possibility of getting the heart disease in attribute overall results in the numeric technique the count of -0.4 to 0.4 – indicates the high amount of amount of possibility of getting the heart related problems in +0.2 to 0.4 does not have a high risk of getting the heart disease this technique is used to find the possibility of risk of getting the heart related disease and low level of getting the heart disease in co-relation matix.

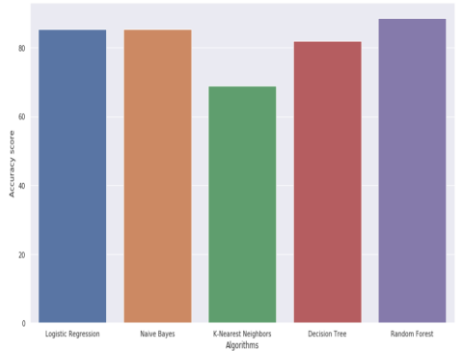

Fig.11.Accuracy of all the algorithm

In Fig.11 the prediction of all the algorithm accuracy is executed in the medium of 0-80 the bar graph represent the minimum accuarcy given by the algorithm in this k-nearest gives the low prediction of 80 and Decision tree gives a accuracy of 95 and other logistic

regression naive bayes,decision tree,random forest gives 100 in the slope count this Fig.11.represent the high accuracy algorithm for prediction.

TABLE 3. ACCURACY OF ALL THE ATTRIBUTE

	Model	**Training Accuracy %**	**Testing Accuracy %**
0	Logistic Regression	100	100
1	Naive bayes	100	100
2	K-Neighbour Algorithm	80	80
3	Decision Tree Classifier	95	95
4	Random Forest Classifier	100	100

In table 3. The accuracy of all the attribute
The machine has used the algorithms' to predict the heart disease using the data repository the following results shown in the table.3. is used to provide the accuracy of heart disease and to predict the most efficient and high accuracy algorithm in the table in table.3 training of data set K-Nearest Algorithm gives low accuracy of 80,

Decision tree gives accuracy of 95, Logistic regression, Naive Bayes, Random Forest gives 100 percent of accuracy in prediction these algorithm are more efficient than other algorithm

V. CONCLUSION

The analysis of the data set and accrued the results by the Machine learning algorithms. , we propose heart disease prediction that allows the user to get the guidance of the heart through the machine learning algorithm process. The system can also be used for free heart disease consulting online and the system acts as a decision support system and proved to be an aid for the physician for the diagnosis with the help of the dataset.

REFERENCE

1. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.
2. Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*(6), 1-6.
3. LK, S. S., Ahmed, S. T., Anitha, K., & Pushpa, M. K. (2021, November). COVID-19 Outbreak Based Coronary Heart Diseases (CHD) Prediction Using SVM and Risk Factor Validation. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-5). IEEE.
4. Khourdifi, Y., & Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *International Journal of Intelligent Engineering and Systems*, *12*(1), 242-252.
5. Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2021). Multiple disease prediction using Machine learning algorithms. *Materials Today: Proceedings*.
6. Waris, S. F., & Koteeswaran, S. (2021). Heart disease early prediction using a novel machine learning method called improved K-means neighbor classifier in python. *Materials Today: Proceedings*.
7. Barhoom, A. M., Almasri, A., Abu-Nasser, B. S., & Abu-Naser, S. S. (2022). Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms. *International Journal of Engineering and Information Systems (IJEAIS)*, *6*(4).
8. Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012072). IOP Publishing.
9. Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., ... & Tang, W. H. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. *Scientific Reports*, *10*(1), 1-11.
10. Albahr, A., Albahar, M., Thanoon, M., & Binsawad, M. (2021). Computational learning model for prediction of heart disease using machine learning based on a new regularizer. *Computational Intelligence and Neuroscience*, *2021*.
11. Al-Shammari, N. K., Alzamil, A. A., Albadarn, M., Ahmed, S. A., Syed, M. B., Alshammari, A. S., & Gabr, A. M. (2021). Cardiac Stroke Prediction Framework using Hybrid Optimization Algorithm under DNN. *Engineering, Technology & Applied Science Research*, *11*(4), 7436-7441.