



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

“HEART DISEASES PREDICTION”

Partha sai K
3rd year

Department of AI & ML
New Horizon College of Engineering
Bengaluru, India

Dr. Sonia Maria D'Souza
Assistant Professor
Department of ML
BMS College of Engineering
Bengaluru, India

Bhuvaneshwari k
3rd year

Department of AI& ML
New Horizon College of Engineering
Bengaluru, India

yogitha Reddy D
3rd year

Department of AI & ML
New Horizon College of Engineering
Bengaluru, India

Abstract— Heart disease is a commonly occurring disease and is the major cause of sudden death now a days. This disease attacks the person instantly. Most people are not aware of the symptoms of heart disease. Timely attention and proper diagnosis of heart disease will reduce the mortality rate. Medical data mining is to explore hidden patterns from the data sets. In this project, we have applied big data tools namely spark, and also using ML algorithms like LogisticRegression, Decision Tree, Random Forest, and Navie Bayes predict the heart disease. Result shows that KNN gives high accuracy score with 85% compare to other algorithms LR is 76%, RF is 78%, DT is 73%and NB is 79%.In order to improve the accuracy, we have done Feature Selection. Feature selection measure discards redundant features to improve the accuracy of the classifier. Our proposed method effectively identifies the redundant features compared to other existing features to effectively predict heart disease. This project helps to predict the future possibility of heart diseasein patients

Keywords—Gestures; KNN; ML; Algoritms;

I. INTRODUCTION

The contents of this report primarily concentrate on predicting heart disease using machine learning technique. If the heart is not functioning properly, this may affect other parts of the body such as the brain, kidney, etc. Heart disease is a condition which affects the hearts functioning. Different individuals will show different symptoms of heart disease which can vary consequently. They often have back pain, jaw pain, neck pain, abdominal disorders, and breath weakness, chest pain, pain to the arms, and pain to the shoulders. There are a number of common heart diseases including heart failure and stroke and coronary artery disease. In today's era heart problems are the primary reason for deaths. Some heart diseases are heart failure and coronary cardiomyopathy. Among varied serious diseases, cardiomyopathy acts as an excellent deal of attention in medical analysis. Its tough for doctors to predict the heart attack because it could be an advanced task that needs a lot of expertise and high information. The diagnosis of heart can give machine-driven prediction regarding the heart condition of patients at earliest. It's necessary to appear at the signs, symptoms and physical examination of the patient. There are

several factors that increase the possibility of heart condition, like smoking habits, body cholesterol level and case history, obesity, high force per unit area and lack of physical exertion. Due to lack of resources in the medical field, the prediction of heart disease occasionally may be a problem. This issue can be resolved by adopting machine learning techniques. The Dataset contains the major symptoms of heart disease which will be used for prediction by applying machine Learning models. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine learning.

II. LITERATURE SURVEY

M Mari Muthu, M Abhinaya, KS Harish, K Madhan Kumar, V Pavithra used Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K- Nearest Neighbour (KNN), Naïve Bayes and Support Vector Machine (SVM) provides an insight of the existing algorithm and it gives an overall summary of the existing work.it works on already existed algorithm, the accuracy as same as the other research.

V Krishnaiah, G Narsimha, N Subhash Chandra uses Fuzzy Intelligent Techniques In these it is observed that Fuzzy Intelligent Techniques increase the accuracy of the heart disease prediction system. The generally used techniques for Heart Disease Prediction and their complexities are summarized in this paper The accuracy of the model may be not more, because of the algorithm used.

M Mari Muthu, M Abhinaya, KS Harish, K Madhan Kumar, V Pavithra uses Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K- Nearest Neighbour (KNN), Naïve Bayes and Support Vector Machine (SVM) in these paper provides an insight of the existing algorithm and it gives an overall summary of the existing work. As this paper works on already existed algorithm, the accuracy as same as the other research.

M Albania, Mahmoud H Qutub, Mohammad Hijazi uses Machine learning classification technique start by overiewing the machine learning and describing brief

definitions of the most commonly used classification techniques to diagnose heart disease. Here we use unsupervised algorithm techniques, so the accuracy is not much.

Adith Gavhane Gouth a mi Kokkula uses Neural network. In this it can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc. Neural network algorithm doesn't provide more accuracy..

III. EXISTING SYSTEM

Machine learning techniques have been around us and have been compared and used for analysis for many kinds of data science applications. The major motivation behind this research-based project was to explore the feature selection methods, data preparation and processing behind the training models in machine learning. With first hand models and libraries, the challenge we face today is data where beside their abundance, and our cooked models, the accuracy we see during training, testing and actual validation has a higher variance. Hence this project is carried out with the motivation to explore behind the models

IV. Proposed System

Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data. And the main challenging thing is doing this in big data tool as this should have a very large data to perform

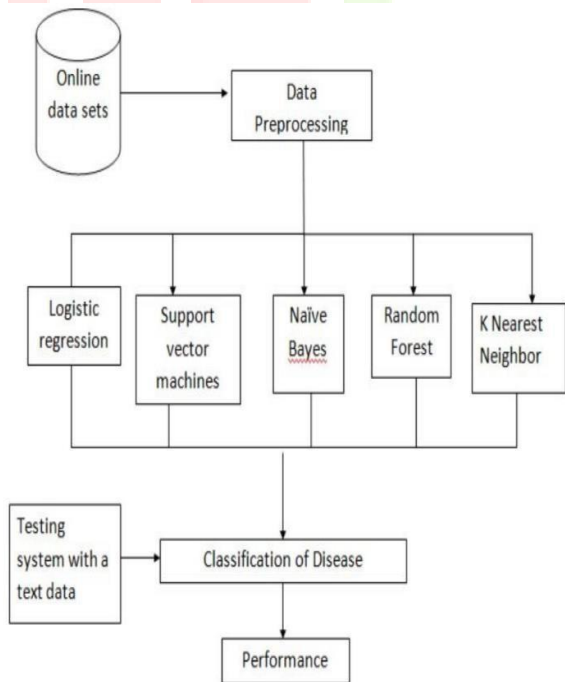


Figure 1 Flow chart

V. ALGORITHM

- Step 1: Start
- Step 2: open Google collab in chrome
- Step 3: creating new file
- Step 4: uploading patient database form the files
- Step 5: installing big data tool
- Step 6: install spark
- Step 7: perform the data preprocessing steps on the selected database
- Step 8: applied few ML algorithms on the data
- Step 9: based accuracy of all the algorithms one is selected perfect algorithm
- Step 10: the prediction will be displayed.

VI. Results

```

input=(78,1,3,45,33,7,5,78,8,1,65,7,8)
input_array=np.asarray(input)
input_reshape = input_array.reshape(1,-1)
pred_LR=LR.predict(input_reshape)
print(pred_LR)
if (pred_LR[0]==0):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')
  
```

[1]
The Person has Heart Disease

FIG. 6.1 PREDICTIVE SYSTEM FOR LR

```

input=(63,1,3,145,233,4,5,778,8,1,5,8,7,8)
input_array=np.asarray(input)
input_reshape = input_array.reshape(1,-1)
pred_NB=NB.predict(input_reshape)
print(pred_NB)
if (pred_NB[0]==0):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')
  
```

[0]
The Person does not have a Heart Disease

FIG.6.2 Predictive System for NB

```

input=(63,1,3,145,233,1,0,150,0,2,3,0,0,1)
input_array=np.asarray(input)
input_reshape = input_array.reshape(1,-1)
pred_dt=dt.predict(input_reshape)
print(pred_dt)
if (pred_dt[0]==0):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')
  
```

[0]
The Person does not have a Heart Disease

FIG.6.3 Prediction System for DT

```

input=(63,1,3,145,233,1,0,150,0,2,3,0,0,1)
input_array=np.asarray(input)
input_reshape = input_array.reshape(1,-1)
pred_knn=knn.predict(input_reshape)
print(pred_knn)
if (pred_knn[0]==0):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')

[0]
The Person does not have a Heart Disease
    
```

FIG.6.4 Prediction System for KNN

```

Predictive System for RFC

input=(63,1,3,145,233,1,0,150,0,2,3,0,0,1)
input_array=np.asarray(input)
input_reshape = input_array.reshape(1,-1)
pred_rfc=rfc.predict(input_reshape)
print(pred_rfc)
if (pred_rfc[0]==0):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')

[0]
The Person does not have a Heart Disease
    
```

FIG.6.5 Prediction System for RFC

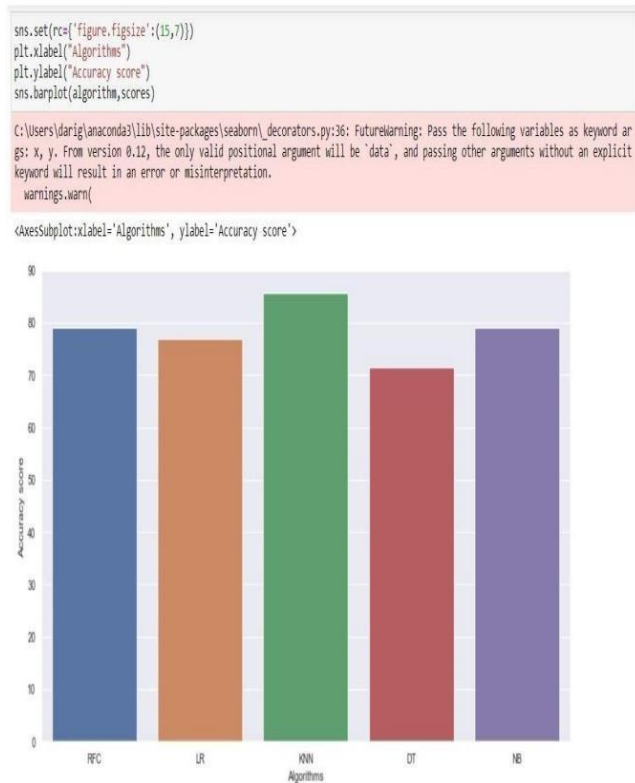


FIG.6.6 COMPARISON

```

cust_data=np.array([[56,0,1,140,294,0,0,153,0,1,3,1,0,2]])
predict_knn=knn.predict(cust_data)
predict_knn[0]

1

cust_data=np.array([[56,0,1,140,294,0,0,153,0,1,3,1,0,2]])
predict_NB=NB.predict(cust_data)
predict_NB[0]

1

cust_data=np.array([[56,0,1,140,294,0,0,153,0,1,3,1,0,2]])
predict_dt=dt.predict(cust_data)
predict_dt[0]

0

cust_data=np.array([[56,0,1,140,294,0,0,153,0,1,3,1,0,2]])
predictRFC = RFC.predict(cust_data)
predictRFC[0]

0

cust_data=np.array([[56,0,1,140,294,0,0,153,0,1,3,1,0,2]])
predict_LR=LR.predict(cust_data)
predict_LR[0]

0
    
```

FIG.6.7 Verification

VII. CONCLUSION

This project addressed the prediction of heart disease based on ML algorithms. Our approach uses KNN, LR, RFC, DT and NB as a classifier to reduce the misclassification rate. This paper also investigates feature selection measures to select a small number of features and to improve the classification performance. The results suggest that the proposed approach can significantly improve the learning accuracy. From simulation results, it is concluded that Particle Swarm Optimization (PSO) based feature selection is important for classification of heart disease. This model helps the physicians in an efficient prediction of diseases with predominant features.

VIII. FUTURE SCOPE

According to the survey, there are still chances that the result might get improved. During the research which we have carried out, it is observed that the most of the algorithm and methods which has been used by most of the authors does not use a collective approach where all the features that cause the heart disease can be simultaneously be used for predicting the heart disease and hence deciding the possibility of disease. There can be further improvements because in some analyses very few amounts of datasets have been used. Therefore by using a large dataset. During prediction of disease the patients dataset could be extended with more symptoms which may cause disease. The future work would be as follows:

- Integrate ensemble classifiers with PSO to develop a decision support system for early diagnosis of heart disease.
- We can use the Deep learning to get the more accuracy for the prediction

References

- 1) M. Preethi, Dr. J. Selva Kumar .heart disease prediction by various data mining techniques(2002).
- 2) Gagandeep Kaur, Anshu Sharma, Anurag Sharma, Heart Disease Prediction using KNN classification approach(1998).

- 3) M Albania, Mahmoud H Qutub, Mohammad Hijazi, V Krishnaiah, G Narsimha, N Subhash Chandra. fuzzy intelligent techniques(2016).
- 4) Lamido Yahaya, N David Oya, Etamin Joshua Garba, Adithi Gavhane,Goutha mi Kokkula, Dynamic KNN is a simple algorithm used for disease prediction(2017).
- 5) Kummita Saravan Kumar Reddy Department of Computer Science and Engineering , Saveetha School of Engineering used Dynamic KNN is a simple algorithm for disease prediction(2020).
- 6) M Mari Muthu, M Abhinaya, KS Harish, K Madhan Kumar, V Pavithra used Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K-Nearest Neighbor (KNN), Naïve Bayes and Support Vector Machine (SVM)(2022).
- 7) V Krishnaiah, G Narsimha, N Subhash Chandra used Fuzzy Intelligent Techniques(2022).

