EFFECTIVE HEART DISEASE PREDICTION USING HYBRID MACHINE LEARNING TECHNIQUES

M Megha Ganesh Chandra, A. Harshitha, Uday Kiran, N. Harsha vardhan

ABSTRACT

Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM)

INTRODUCTION

It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Various techniques in data mining and neural networks have been employed to find out the severity of heart disease among humans. The severity of the disease is classified based on various methods like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), and Naive Bayes (NB).

The nature of heart disease is complex and hence, the disease must be handled carefully. Not doing so may affect the heart or cause premature death. The perspective of medical science and data mining are used for discovering various sorts of metabolic syndromes. Data mining with classification plays a significant role in the prediction of heart disease and data investigation.

Various methods have been used for knowledge abstraction by using known methods of data mining for prediction of heart disease. In this work, numerous readings have been carried out to produce a prediction model using not only distinct techniques but also by relating two or more techniques. These Amalgamated new techniques are commonly known as hybrid methods.

We introduce neural networks using heart rate time series. This method uses various clinical records for prediction such as Left bundle branch block (LBBB), Right bundle branch block (RBBB), Atrial fibrillation (AFIB), Normal Sinus Rhythm (NSR), Sinus bradycardia (SBR), Atrial flutter (AFL), Premature Ventricular Contraction (PVC)), and Second degree block (BII) to find out the exact condition of the patient in relation to heart disease. The dataset with a radial basis function network (RBFN) is used for classification, where 70% of the data is used for training and the remaining 30% is used for classification.

Neural networks are generally regarded as the best tool for prediction of diseases like heart disease and brain disease. The proposed method which we use has 13 attributes for heart disease prediction. The results show an enhanced level of performance compared to the existing methods in works like. The Carotid Artery Stenting (CAS) has also become a prevalent treatment mode in the medical field during these recent years. The CAS prompts the occurrence of major adverse cardiovascular events (MACE) of heart disease patients that are elderly. Their evaluation becomes very important.

LITERATURE REVIEW

3.1 TITLE: HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES: A SURVEY

AUTHORS: V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja

ABSTRACT: Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Manyresearchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques and analyze their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers.

3.2 TITLE: A STUDY ON ROLE OF MACHINE LEARNING IN DETECTING HEART DISEASE.

AUTHORS: Preet Chandan Kaur

ABSTRACT: A fist size muscle occupies an important in the human body by supplying oxygen to all the body organs. According to study of demography from WHO (World Health organization), the main cause of increasing death rate is due to the cardiac failure of human being. The main challenge for data analysis is to predict and prevent the heart disease. Machine learning has been developed to perform impressive predictions and make appropriate decision from abundant data originated by healthcare centers. In this paper numerous machine learning techniques are surveyed by using the knowledge collected from preprocessing data (clinical knowledge), which comprises many medical features to perform heart disease detection. The comparative study states that the prediction of heart disease has been improved by combining various machine learning algorithms to perform early disease investigation in a cost-effective manner. The proposed research work primarily focuses on preparing a review of the research done by different professionals and compiling it into one paper and creating a direction for future research in this domain. In this paper many techniques are surveyed where best predictions are performed.

3.3 TITLE: A MACHINE LEARNING APPROACH TO LOW-VALUE HEALTH CARE: WASTED

TESTS, MISSED HEART ATTACKS AND MIS-PREDICTIONS

AUTHORS: Sendhil Mullainathan Ziad Obermeyer

ABSTRACT: We use machine learning to better characterize low-value health care and the decisions that produce it. We focus on costly tests, specifically for heart attack (acute coronary syndromes). A test is only useful if it yields new information, so efficient testing is grounded in accurate prediction of test outcomes. Physician testing decisions can therefore be benchmarked against tailored algorithmic predictions, which provide a more precise way to study low-value care than the usual approach—looking at average test yield. Implemented in a large national sample, this procedure reveals significant over- testing: 52.6% of high-cost tests for heart attack are wasted. At the same time, it also reveals significant under-testing: many patients with predictably high risk go untested, then experience frequent adverse cardiac events including death in the next 30 days. At standard clinical thresholds, these event rates suggest that testing these patients would indeed have been highly cost-effective. Of the potential welfare gains from more efficient testing, 42.8% would come from addressing under-use. Existing policy levers, however, appear too blunt a tool to address both over- and under-use inefficiencies. We find that they cut testing across the board, for low-risk (reducing overuse) and high-risk patients (exaggerating under- use). Finally, we uncover two behavioral mechanisms for physician testing errors: (i) bounded rationality, in which physicians use an overly narrow set of variables, but make effective use of that set; and (ii) representativeness, in which they overweight how \representative" heart attack is for a patient, above and beyond the conditional probability. Together, these results suggest the need for models of low-value care that incorporate mis-prediction so as to account for both over- and under testing.

3.4 TITLE: HEART DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

AUTHORS: Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti

ABSTRACT: Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression and KNN to predict and classify the patient with heart disease. A quite helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using KNN and Logistic Regression which showed a good accuracy in comparison to the previously used classifier such as naive bayes etc. So a quiet significant amount of pressure has been lift off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease.

PROBLEM IDENTIFICATION & OBJECTIVES

Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis.

It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors.

For this project, the dataset has been taken from the UCI repository. Classification techniques like Artificial Neural Network (ANN),, logistic regression, support vector machine (SVM), K- nearest neighbors (KNN) are used in the project..

The available information points to the deduction that females have less of a chance for heart disease compared to males. In heart diseases, accurate diagnosis is primary.But, the traditional approaches are inadequate for accurate prediction and diagnosis.

HRFLM makes use of a combination of Random Forest as well as Linear method along with 13 clinical features as the input.

The obtained results are comparatively analyzed against traditional method

OVERVIEW OF TECHNOLOGIES

The technologies or the classification techniques used in this project are Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, Decision tree' and
Deep Learning ANN Algorithm.

SVM ALGORITHM:

Machine learning involves predicting and classifying data and to do so we employ various machine learning algorithms according to the dataset. SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyper plane which separates the data into classes. In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

RANDOM FOREST ALGORITHM:

The random tree is a type of supervised classifiers. It produces lots of distinct learners. The stochastic process is used to form the tree. It is a type of ensemble learning technique for classification. It works the same as decision tree, but a random subset of attributes uses for each split. This algorithm uses for both classification problems and regression problems. A group of random trees is known as a forest. The random trees classifier takes the input feature set and classifies input for every tree in the forest. The output of the random tree selects from the majority of votes. In the tree, every leaf node holds a linear model. The bagging training algorithm is used to train the model.

DECISION TREE ALGORITHM:

This algorithm will build training model by arranging all similar records in the same branch of tree and continue till all records arrange in entire tree. The complete tree will be referred as classification train model.

LOGISTIC REGRESSION:

Logistic regression is also a type of supervised learning algorithm. It is a statistical model. The probability of target value is predicted from logistic regression. It is divided the target attribute into two-classes: success or not success. For success, it returns 1 whereas it returns 0 for not succeeding. Logistic regression is represented by the equation:

P **= 1/ (1 + e^ (- (b0 + b1x + b2x^2))**

where P is the predicted value, b0, b1, b2 are biases and x is an attribute. It is used in various field of machine learning application in social sciences and medical arena, for example, for spam detection, diabetes detection, cancer detection, etc. Logistic regression is the advanced version of linear regression. Through this technique, we only concern about the probability of the outcome variable.

.

DEEP LEARNING ANN ALGORITHM:

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output.

IMPLEMENTATION

DATA ANALYSIS:

Data analysis is the process of collecting, modeling, and analyzing data to extract insights that support decision-making. There are several methods and techniques to perform analysis depending on the industry and the aim of the analysis.

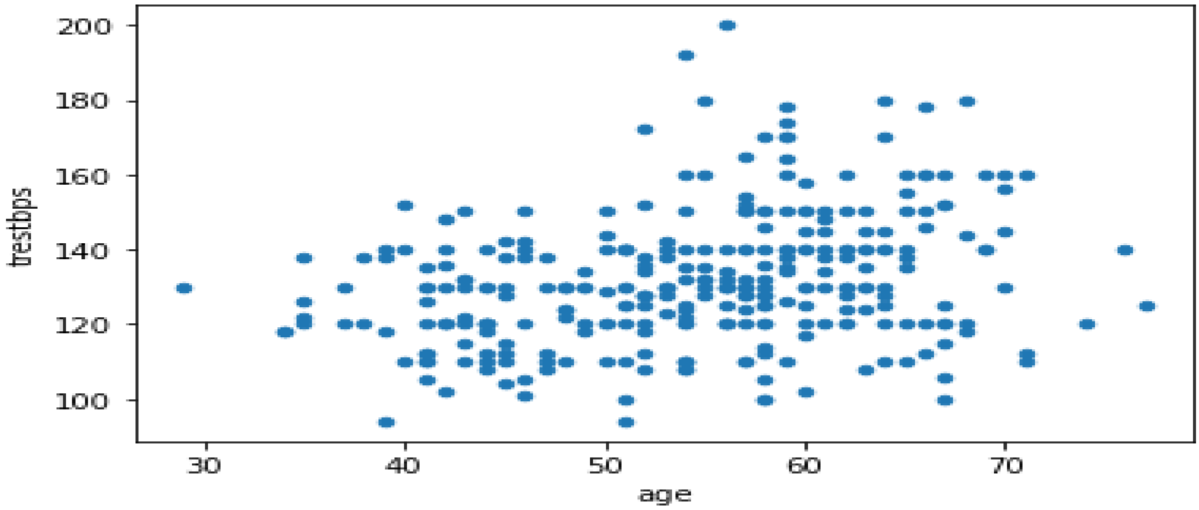
We use data analysis to portray the correlation between various features of the dataset and show the proportionality ,if any, to deduce how one feature might be affecting the other .

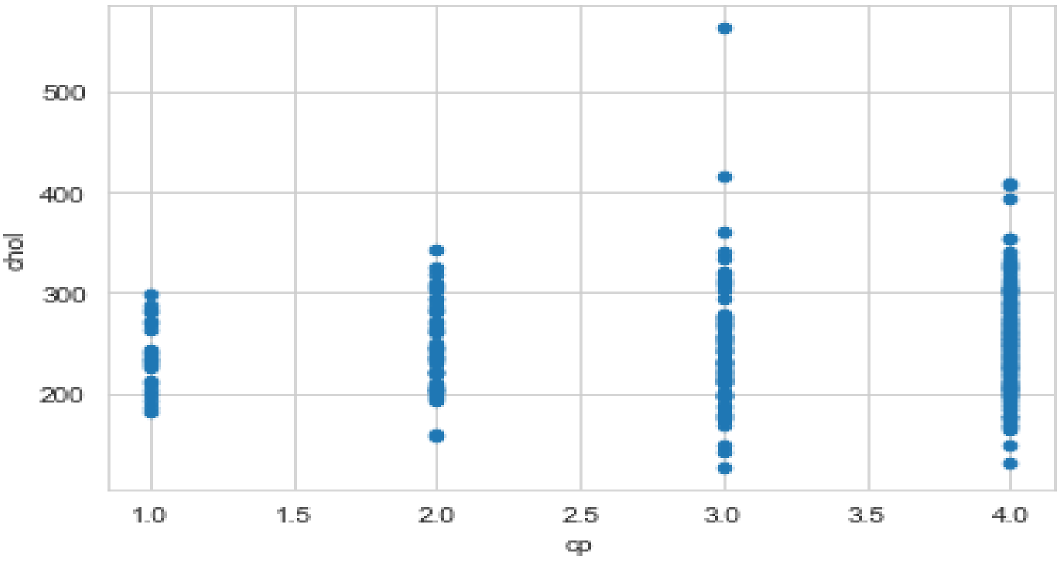
This also signifies which features are acting as indicators to the model’s performance .
A pair plot plot is a pairwise relationship in a dataset. The pair plot function creates a grid of Axes such that each variable in data will be shared in the y-axis across a single row and in the x-axis across a single column.

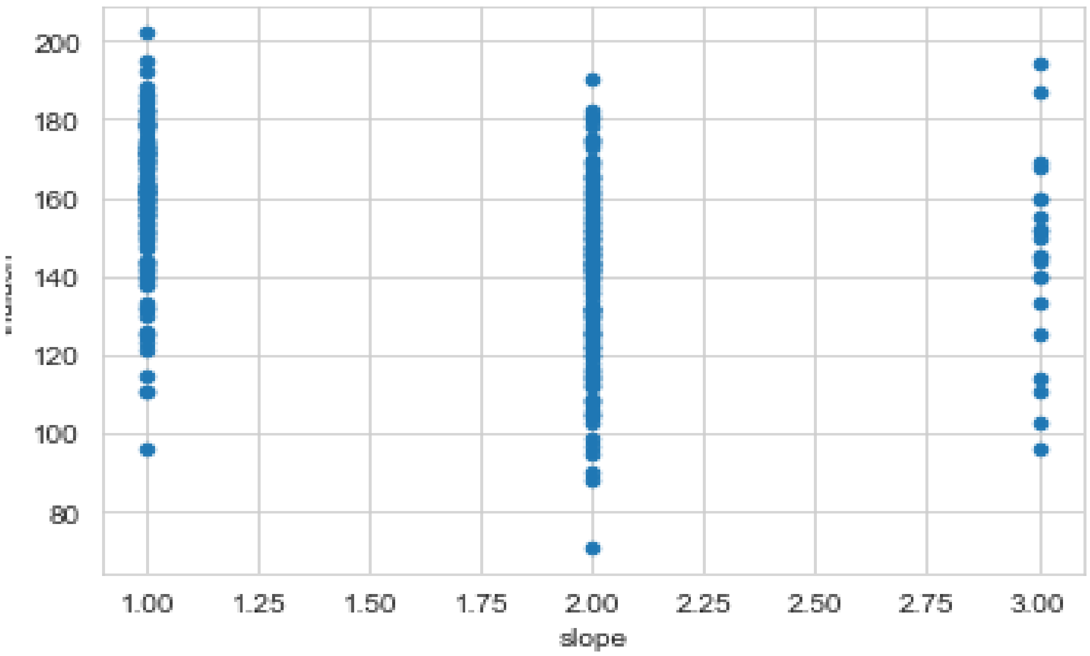
The **Seaborn Pairplot** function allows the users to create an axis grid via which each numerical variable stored in data is shared across the X- and Y-axis in the structure of columns and rows. We can create the Scatter plots in order to display the pairwise relationships in addition to the distribution plot displaying the data distribution in the column diagonally

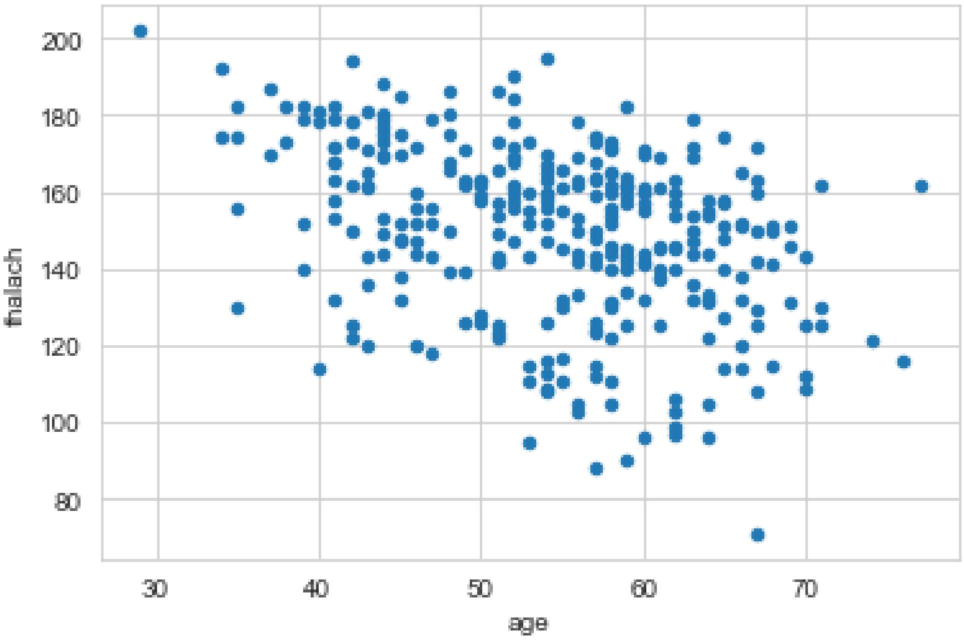
A Probability Density Function is a tool used by machine learning algorithms and neural networks that are trained to calculate probabilities from continuous random variables.

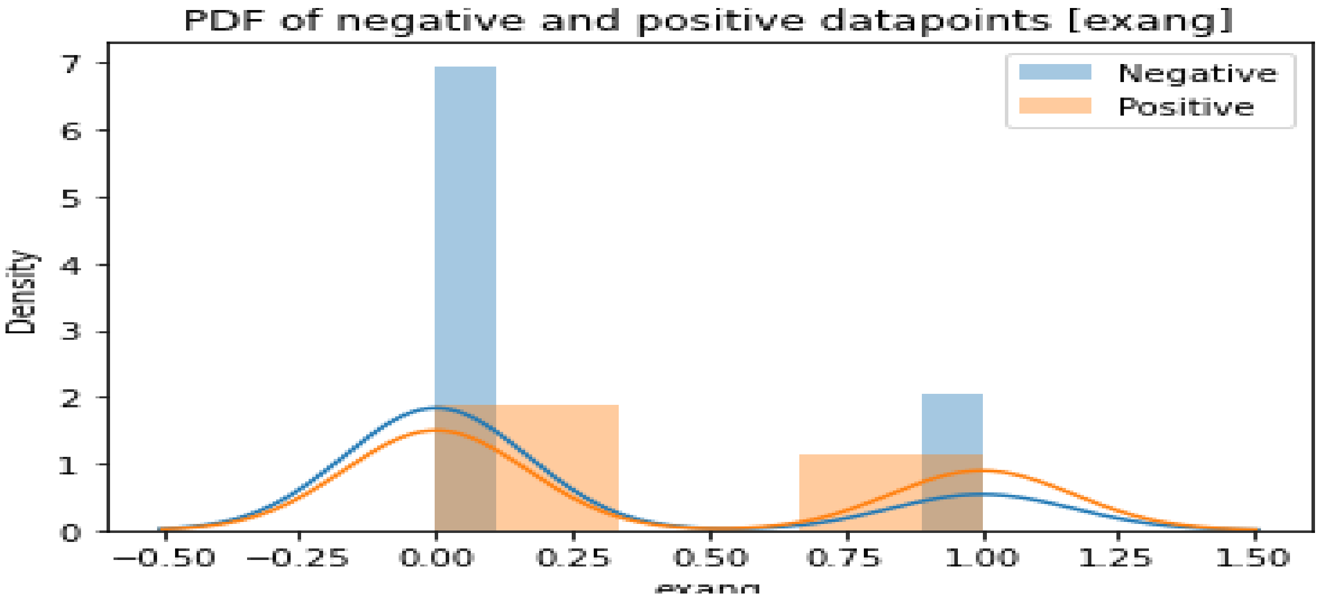
```
<AxesSubplot:xlabel='age', ylabel='trestbps'>
```

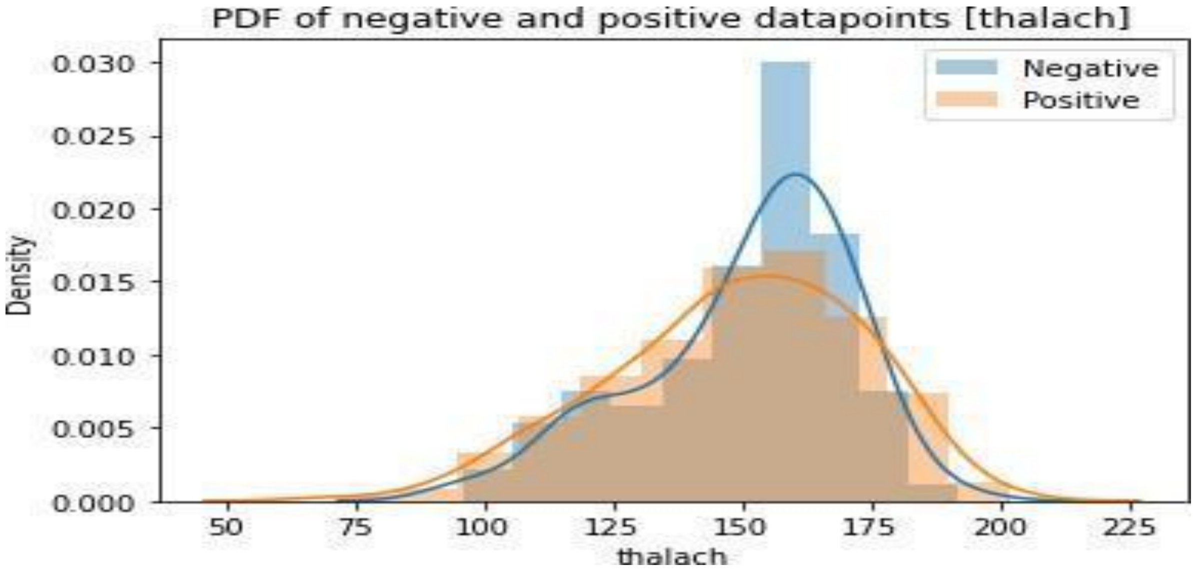

trestbps
200
180
160
140
120
100
30
40
50
60
70
age

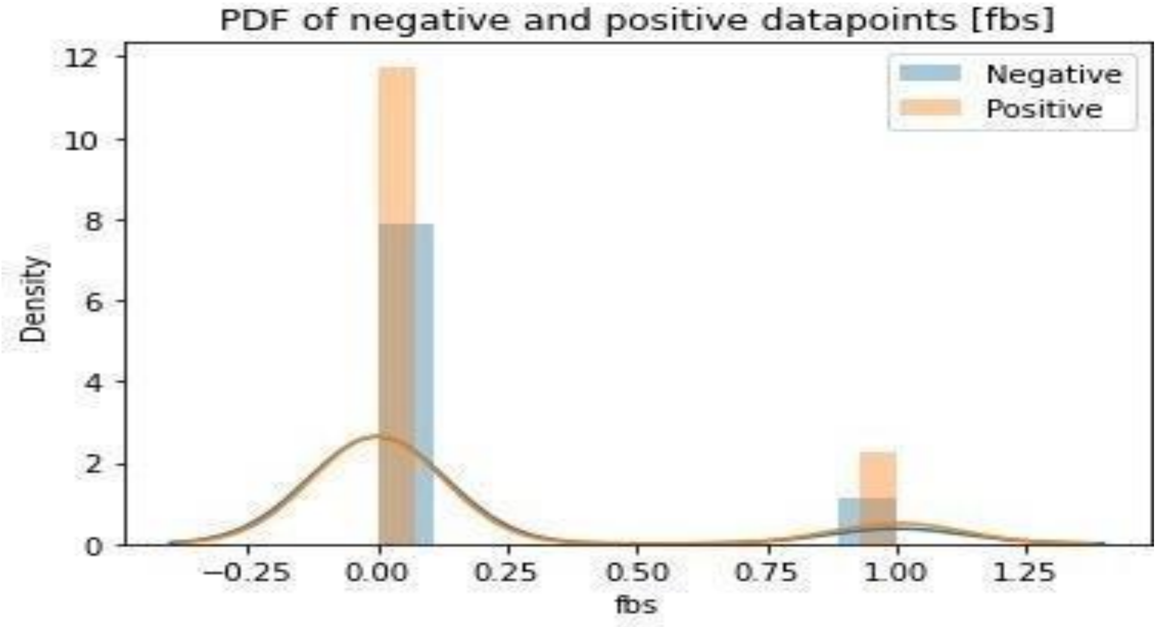

chol
500
400
300
200
1.0
1.5
2.0
2.5
3.0
3.5
4.0
cp

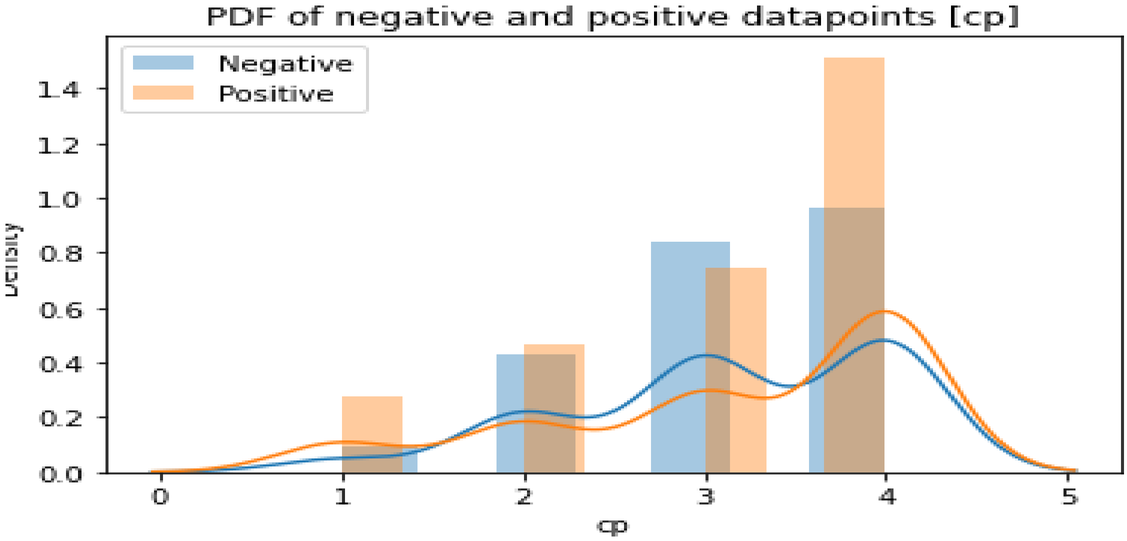

200
180
160
140
120
100
80
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
slope

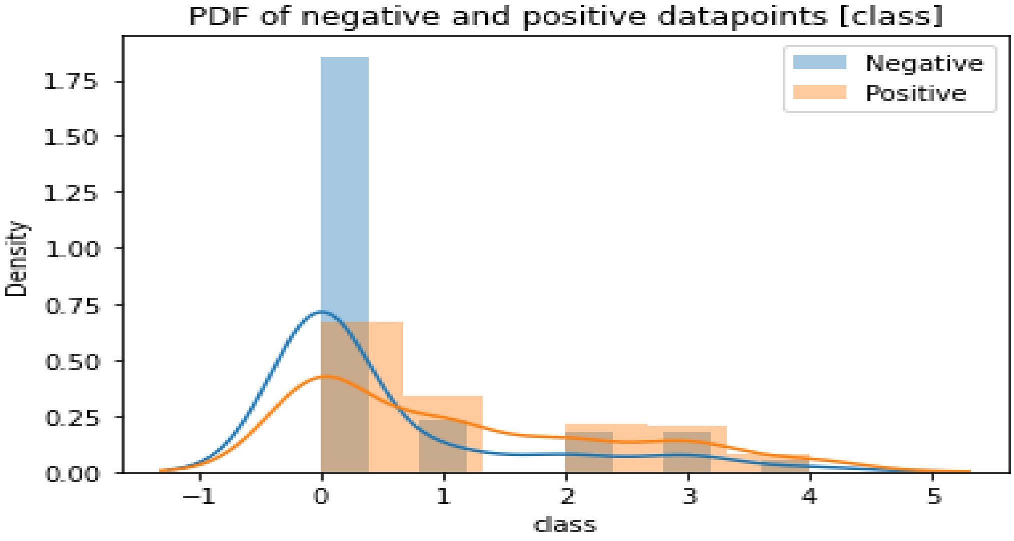

thalach
200
180
160
140
120
100
80
30
40
50
60
70
age

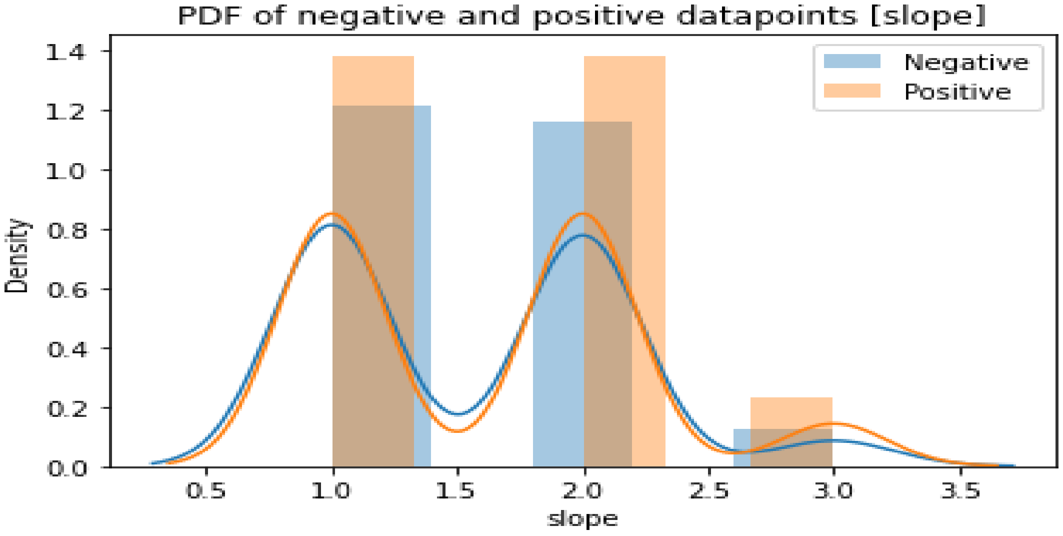

PDF of negative and positive datapoints [exang]
Negative
Positive
Density
0
1
2
3
4
5
6
7
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
exang

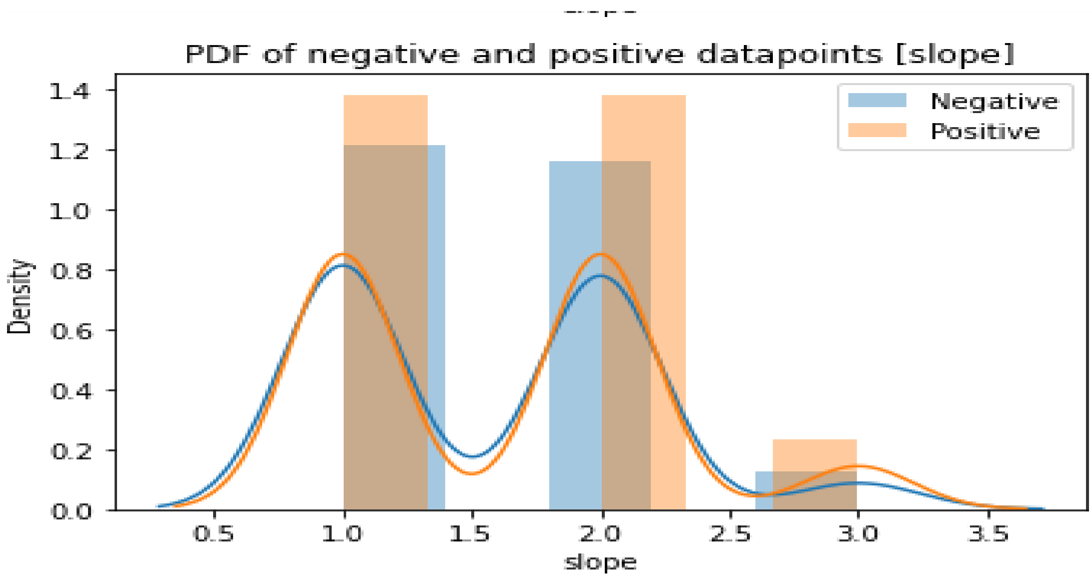

PDF of negative and positive datapoints [thalach]
Negative
Positive
Density
0.000
0.005
0.010
0.015
0.020
0.025
0.030
50
75
100
125
150
175
200
225
thalach

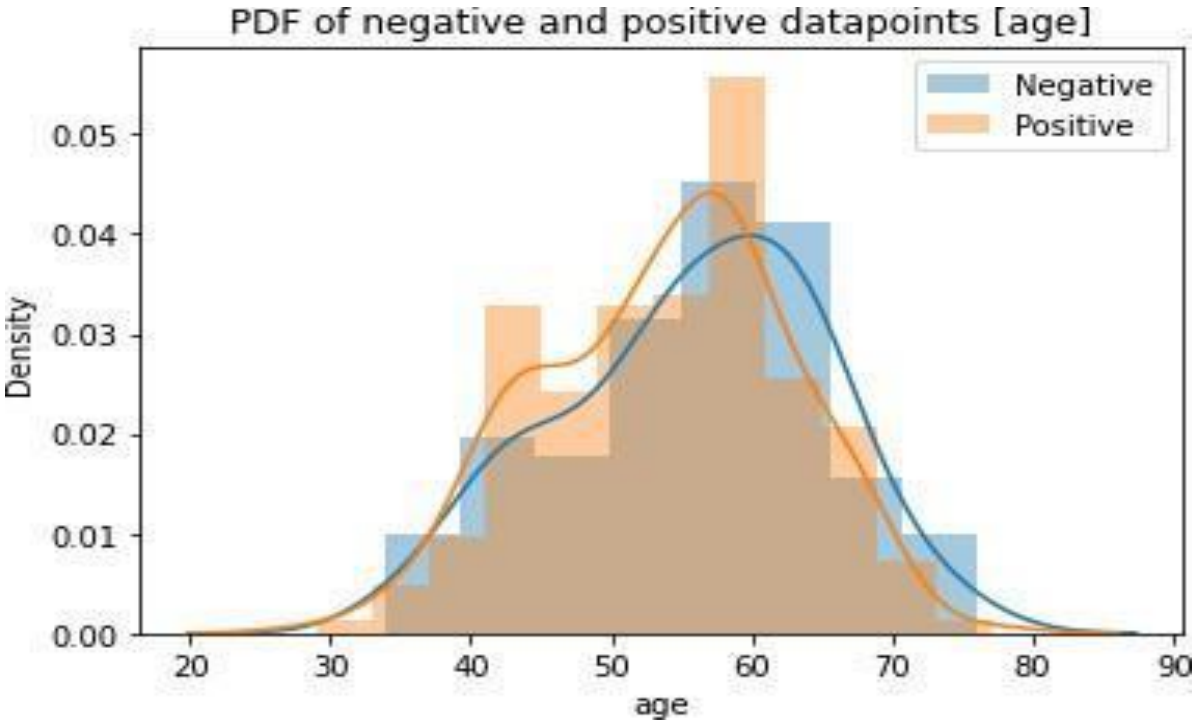

PDF of negative and positive datapoints [fbs]
Negative
Positive
Density
0
2
4
6
8
10
12
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
fbs


PDF of negative and positive datapoints [cp]
Negative
Positive
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
1
2
3
4
5
cp

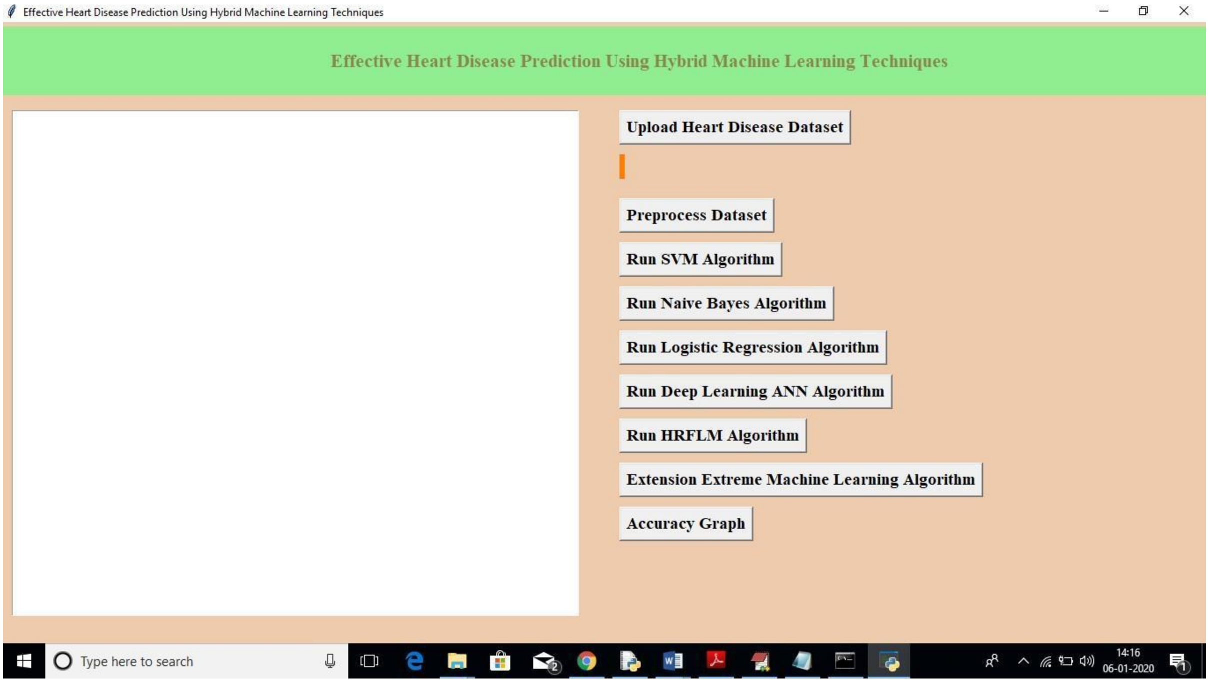

PDF of negative and positive datapoints [class]
Negative
Positive
Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
−1
0
1
2
3
4
5
class


PDF of negative and positive datapoints [slope]
Negative
Positive
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.5
1.0
1.5
2.0
2.5
3.0
3.5
slope

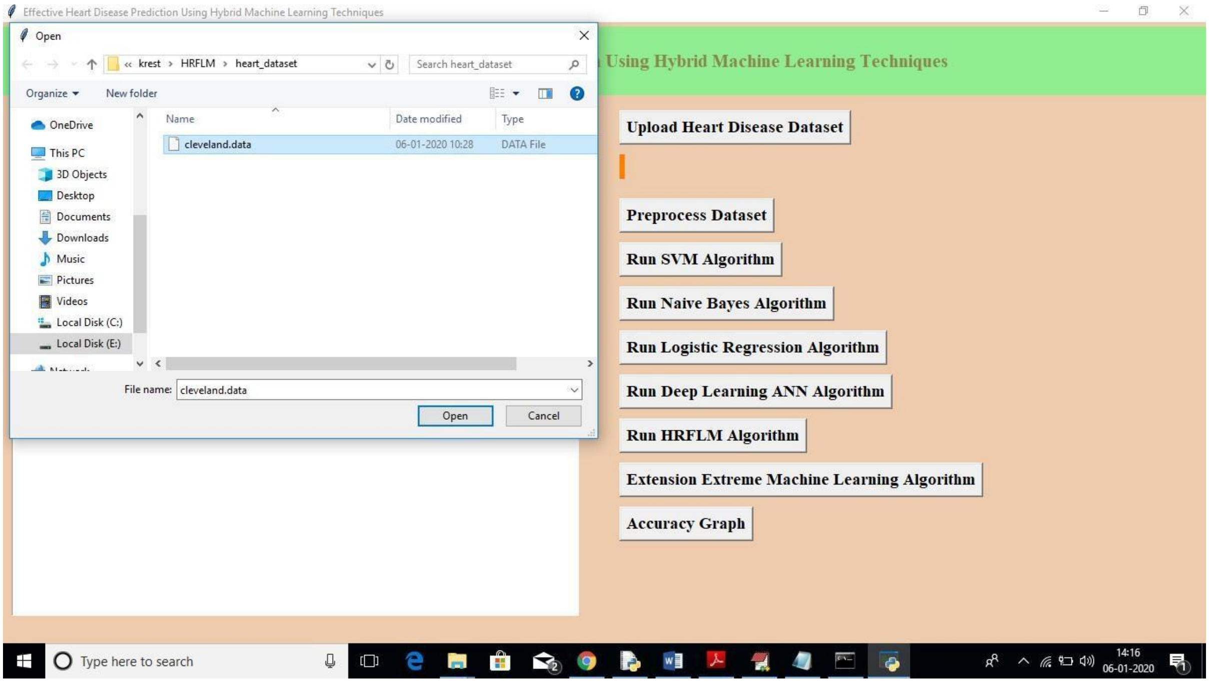

PDF of negative and positive datapoints [slope]
Negative
Positive
Density
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
slope


PDF of negative and positive datapoints [age]
Negative
Positive
Density
0.05
0.04
0.03
0.02
0.01
0.00
20
30
40
50
60
70
80
90
age

To run this project double click on ‘run.bat’ file to get below screen


Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Upload Heart Disease Dataset
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:16
06-01-2020

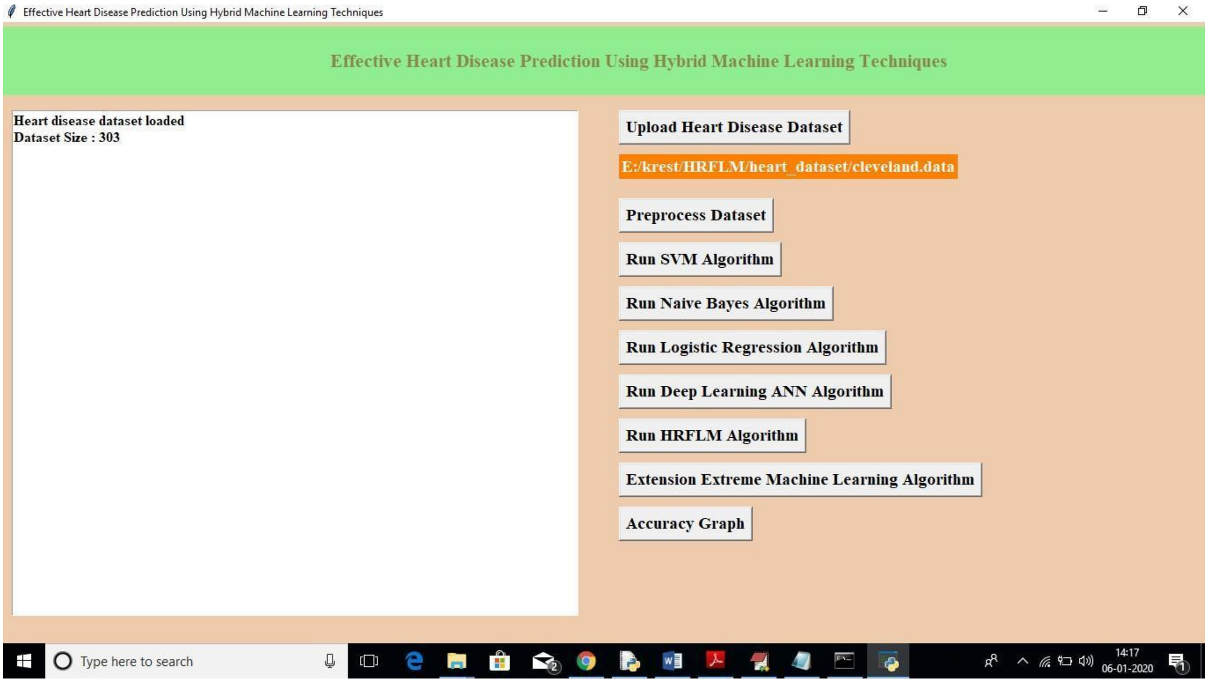
In above screen click on ‘Upload Heart Disease Dataset’ button to upload heart dataset


Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Open
« krest › HRFLM › heart_dataset
Search heart_dataset
Organize ▾
New folder
OneDrive
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Local Disk (E:)
Name
Date modified
Type
Open
Cancel
Using Hybrid Machine Learning Techniques
Upload Heart Disease Dataset
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:16
06-01-2020

Name	Date modified	Type
cleveland.data	06-01-2020 10:28	DATA File

File name: cleveland.data

In above screen I am uploading 'cleveland.data' dataset, after uploading dataset will get in below screen



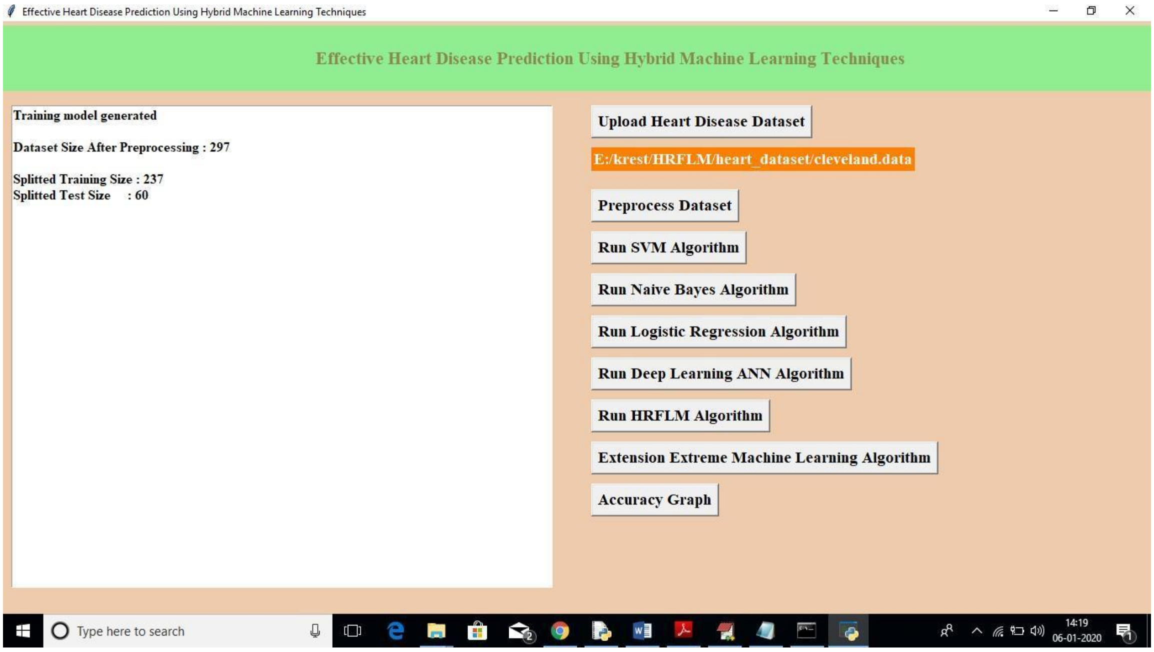
Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

```
Heart disease dataset loaded
Dataset Size : 303
```

E:/krest/HRFLM/heart_dataset/cleveland.data

Upload Heart Disease Dataset
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:17
06-01-2020

In above screen we can see dataset contains total 303 records, now click on 'Pre-process Dataset' button to apply pre-processing technique to remove out all non-numeric data.



Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

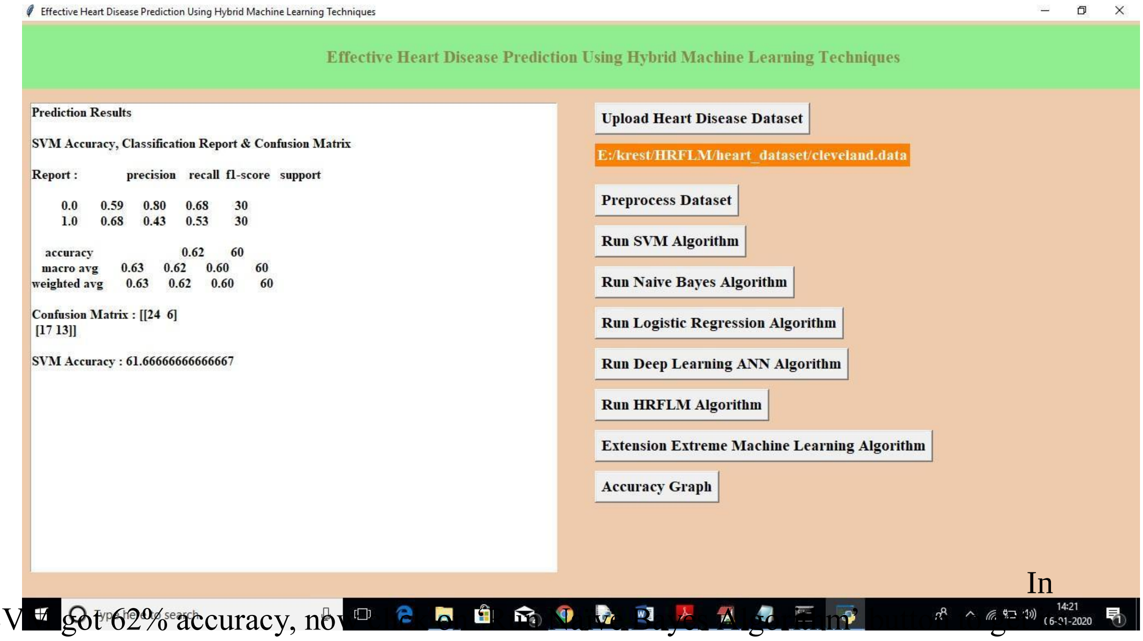
```
Training model generated

Dataset Size After Preprocessing : 297

Splitted Training Size : 237
Splitted Test Size     : 60
```

E:/krest/HRFLM/heart_dataset/cleveland.data

Upload Heart Disease Dataset
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:19
06-01-2020

In above screen after applying pre-processing dataset size reduced to 297 records and we can see application randomly split complete dataset in to tow parts called train and test. For training application using 237 records and for testing application using 60 records. Application will choose random 60 records so always accuracy of same algorithm will be different as records for testing are randomly chooses.

Now click on ‘Run SVM Algorithm’ button to generate SVM model on train dataset and to apply test data to get SVM classification accuracy.

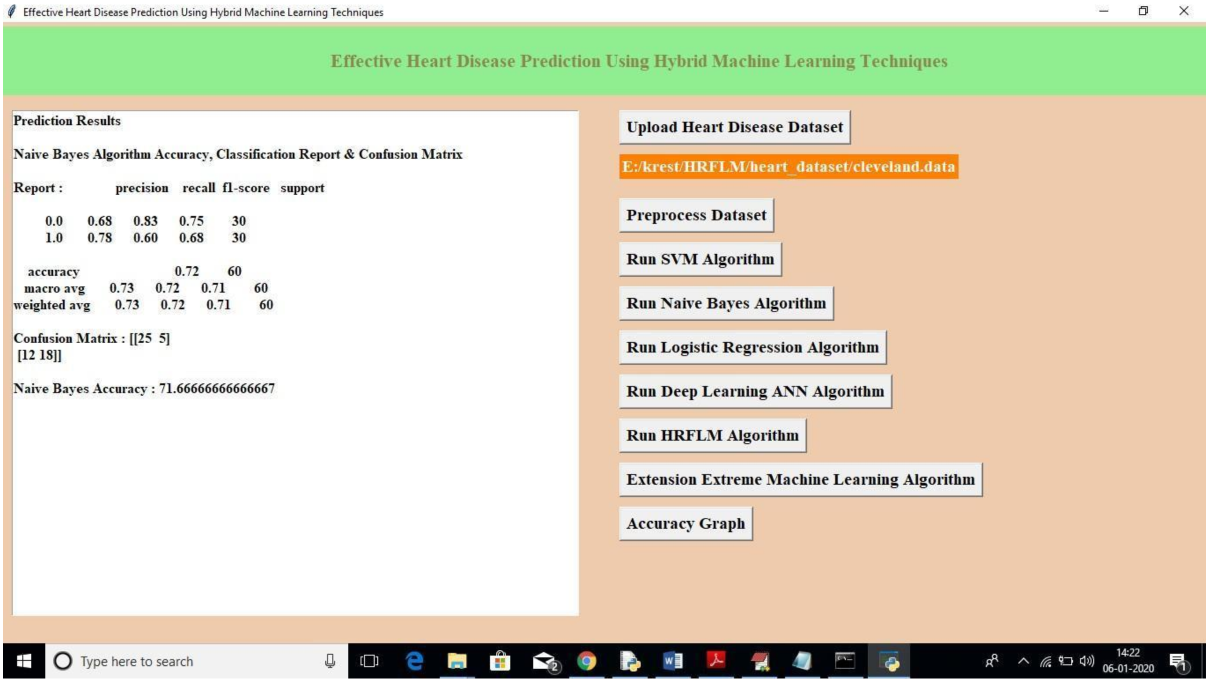

Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

```
Prediction Results

SVM Accuracy, Classification Report & Confusion Matrix

Report :           precision   recall  f1-score   support

       0.0      0.59      0.80      0.68      30
       1.0      0.68      0.43      0.53      30

    accuracy                        0.62      60
   macro avg      0.63      0.62      0.60      60
weighted avg      0.63      0.62      0.60      60

Confusion Matrix : [[24  6]
 [17 13]]

SVM Accuracy : 61.66666666666667
```

Upload Heart Disease Dataset
E:/krest/HRFLM/heart_dataset/cleveland.data
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:21
06-01-2020

In above screen SV[illegible] got 62% accuracy, no[illegible] Bayes Algorithm’ [illegible] accuracy

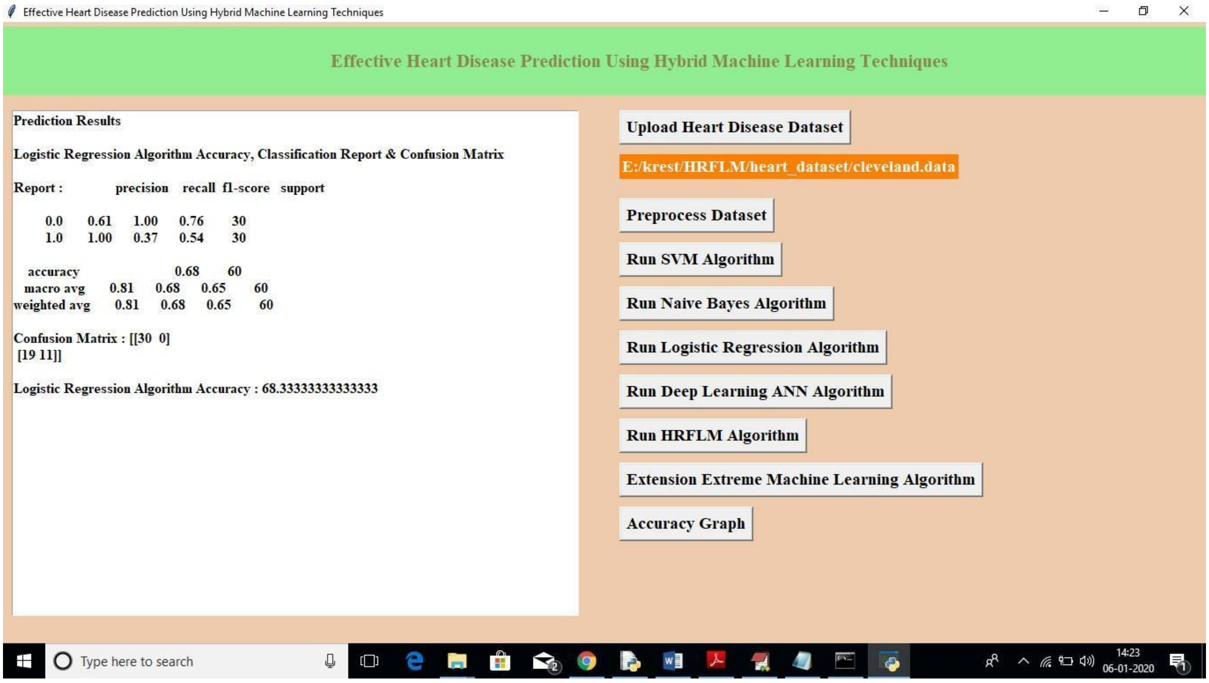

Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

```
Prediction Results

Naive Bayes Algorithm Accuracy, Classification Report & Confusion Matrix

Report :           precision   recall  f1-score   support

       0.0      0.68      0.83      0.75      30
       1.0      0.78      0.60      0.68      30

    accuracy                        0.72      60
   macro avg      0.73      0.72      0.71      60
weighted avg      0.73      0.72      0.71      60

Confusion Matrix : [[25  5]
 [12 18]]

Naive Bayes Accuracy : 71.66666666666667
```

Upload Heart Disease Dataset
E:/krest/HRFLM/heart_dataset/cleveland.data
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:22
06-01-2020

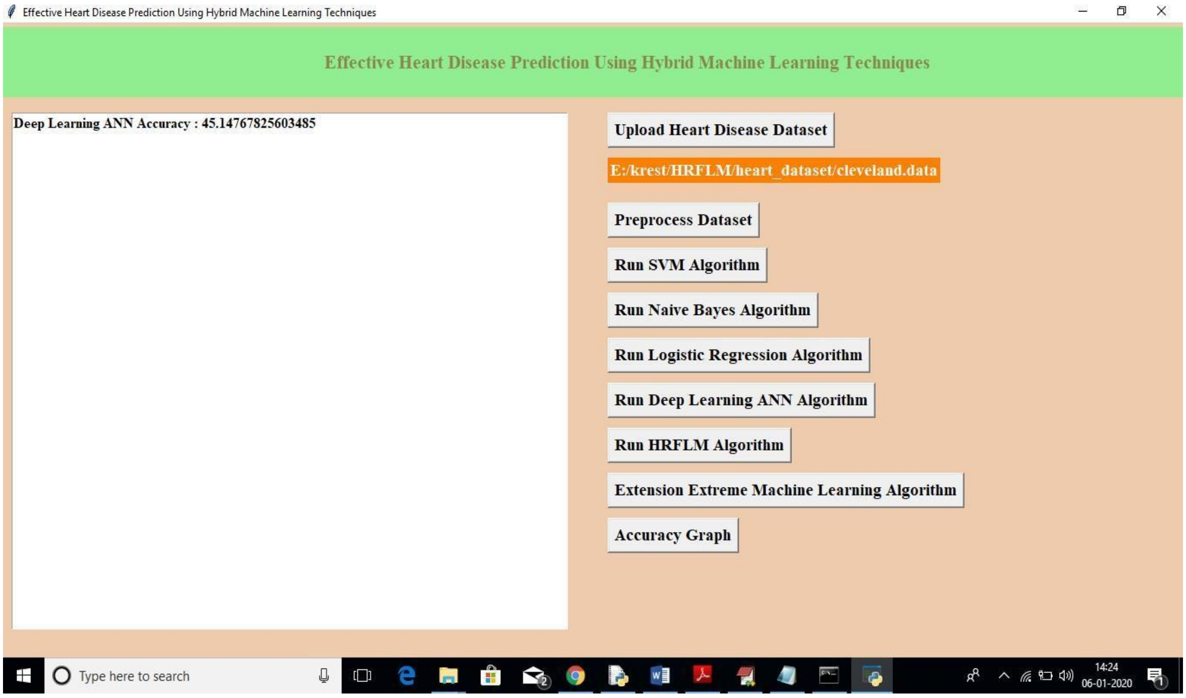
In above screen we can see Naïve Bayes got 72% accuracy, now click on 'Run Logistic Regression Algorithm' to get its accuracy.


Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Upload Heart Disease Dataset
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:23
06-01-2020

E:/krest/HRFLM/heart_dataset/cleveland.data

```
Prediction Results

Logistic Regression Algorithm Accuracy, Classification Report & Confusion Matrix

Report :           precision   recall  f1-score   support

      0.0      0.61     1.00     0.76       30
      1.0      1.00     0.37     0.54       30

   accuracy                           0.68       60
  macro avg      0.81     0.68     0.65       60
weighted avg      0.81     0.68     0.65       60

Confusion Matrix : [[30  0]
 [19 11]]

Logistic Regression Algorithm Accuracy : 68.33333333333333
```

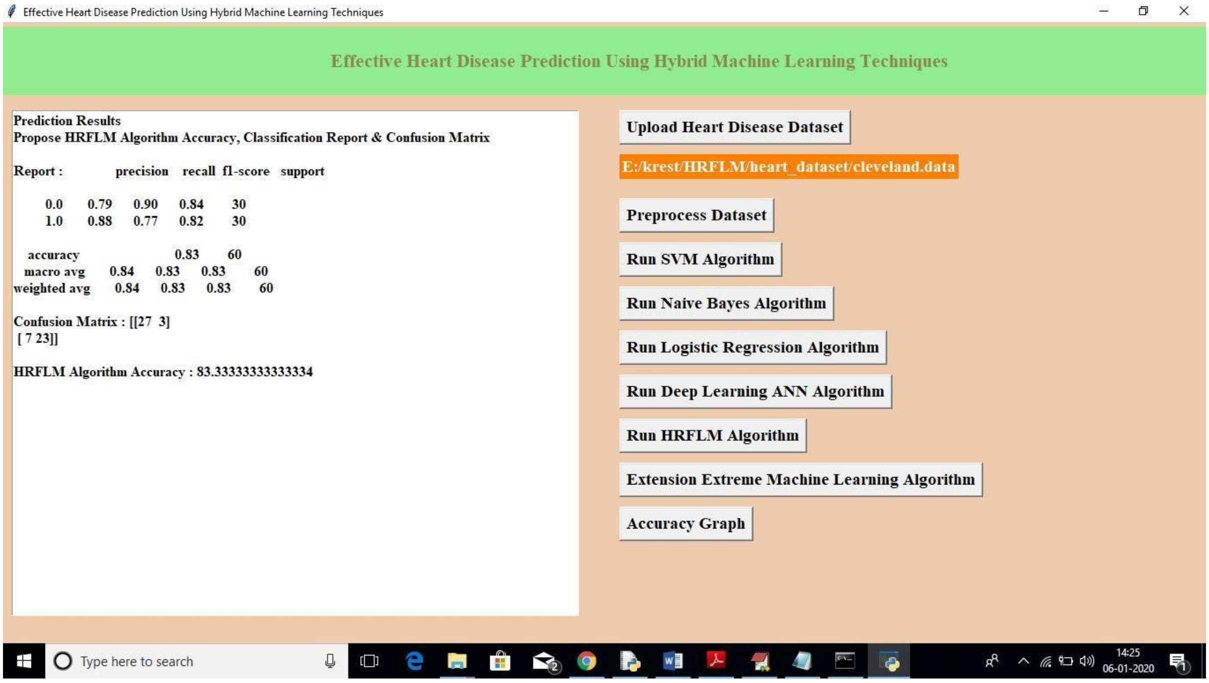
In above screen logistic regression got 69% accuracy, now click on 'Run Deep Learning ANN Algorithm' button to get its accuracy.


Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Upload Heart Disease Dataset
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:24
06-01-2020

E:/krest/HRFLM/heart_dataset/cleveland.data

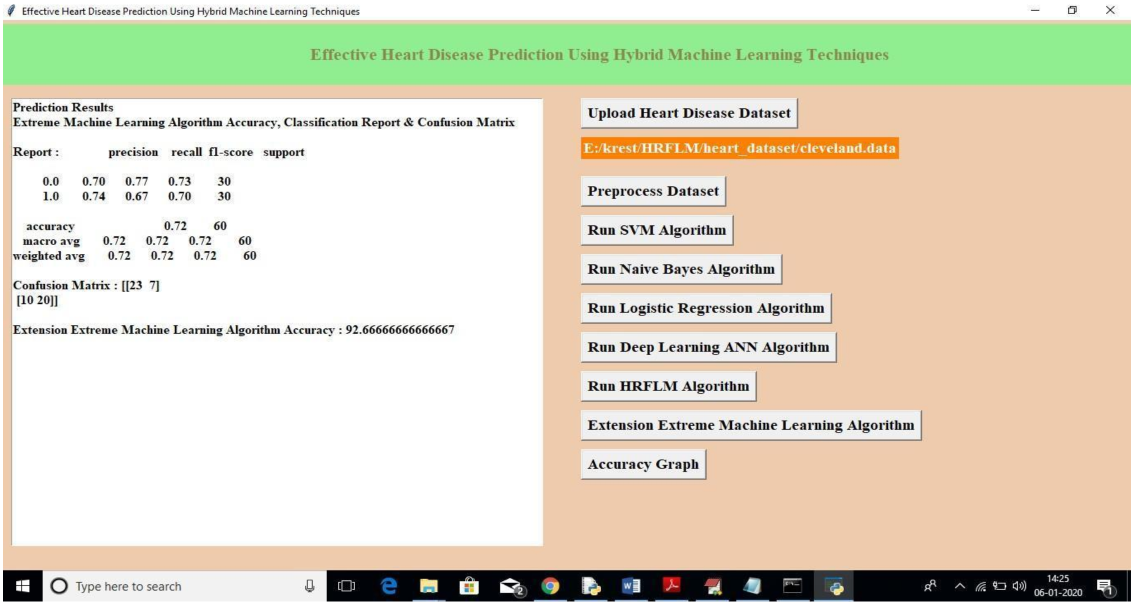
```
Deep Learning ANN Accuracy : 45.14767825603485
```

In above screen we can see ANN got 46% accuracy, now click on ‘Run HRFLM Algorithm’ button to get propose work accuracy.


Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Upload Heart Disease Dataset
E:/krest/HRFLM/heart_dataset/cleveland.data
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:25
06-01-2020

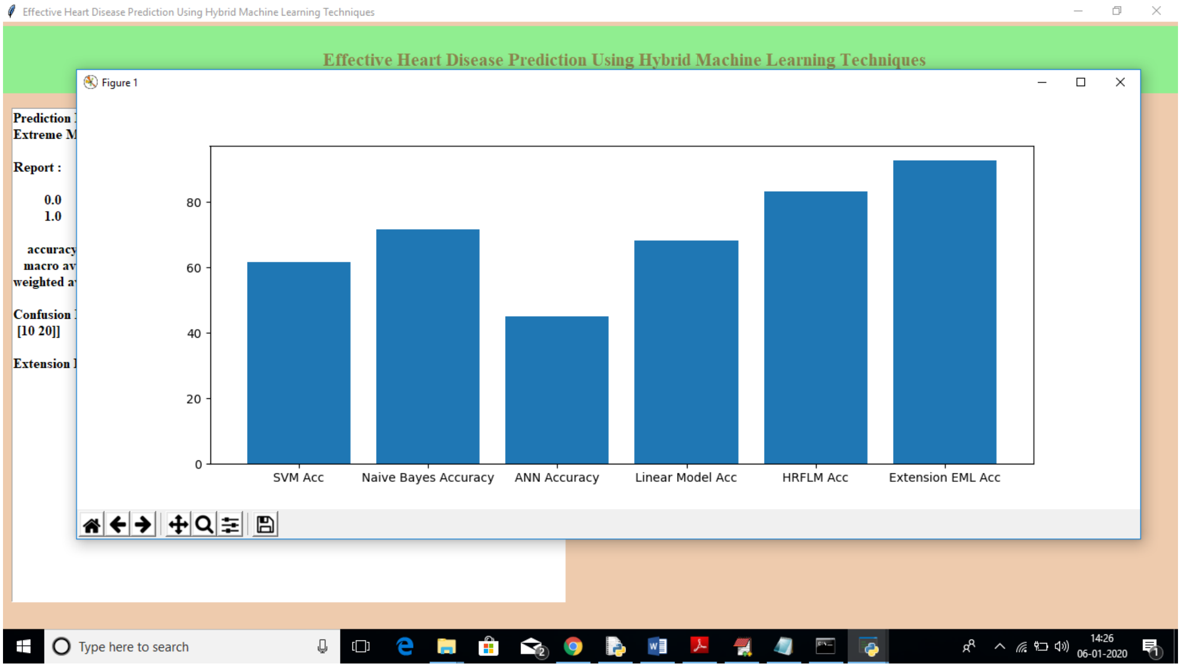
```
Prediction Results
Propose HRFLM Algorithm Accuracy, Classification Report & Confusion Matrix

Report :          precision   recall  f1-score   support

     0.0      0.79      0.90      0.84       30
     1.0      0.88      0.77      0.82       30

   accuracy                       0.83       60
  macro avg      0.84      0.83      0.83       60
weighted avg      0.84      0.83      0.83       60

Confusion Matrix : [[27  3]
 [ 7 23]]

HRFLM Algorithm Accuracy : 83.33333333333334
```

In above algorithm we can see HRFLM got 84% accuracy, now click on ‘Extension Extreme Machine Learning Algorithm’ button to check EML extension accuracy.


Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques
Upload Heart Disease Dataset
E:/krest/HRFLM/heart_dataset/cleveland.data
Preprocess Dataset
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Logistic Regression Algorithm
Run Deep Learning ANN Algorithm
Run HRFLM Algorithm
Extension Extreme Machine Learning Algorithm
Accuracy Graph
Type here to search
14:25
06-01-2020

```
Prediction Results
Extreme Machine Learning Algorithm Accuracy, Classification Report & Confusion Matrix

Report :          precision   recall  f1-score   support

     0.0      0.70      0.77      0.73       30
     1.0      0.74      0.67      0.70       30

   accuracy                       0.72       60
  macro avg      0.72      0.72      0.72       60
weighted avg      0.72      0.72      0.72       60

Confusion Matrix : [[23  7]
 [10 20]]

Extension Extreme Machine Learning Algorithm Accuracy : 92.66666666666667
```

In above screen we can see extension EML algorithm got 93% accuracy which is better than all algorithms. Now click on 'Accuracy Graph' button to get below graph.



```
Prediction
Extreme M

Report :

    0.0
    1.0

  accuracy
 macro av
weighted a

Confusion
 [10 20]]

Extension
```

In above graph x-axis represents algorithm names and y-axis represents accuracy of that algorithm. In all algorithms propose HRFLM and extension algorithm got better accuracy.

FUTURE ENHANCEMENT

The future course of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction.

CONCLUSION

Identifying the processing of raw healthcare data of heart information will help in the long term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM). HRFLM proved to be quite accurate in the prediction of heart disease.

REFERENCES

[1] A. S. Abdullah and R. R. Rajalaxmi, ''A data mining model for predicting the coronary heart disease using random forest classifier,'' in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heart disease,'' in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.

[3] N. Al-milli, ''Backpropagation neural network for prediction of heart disease,'' J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ''Analysis of neural networks based heart disease prediction system,'' in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

[5] P. K. Anooj, ''Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,'' J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.