

## Myocardial Infarction Prediction Using Hybrid Machine Learning Techniques

Vaddi Niranjana Reddy<sup>a</sup>, Gopichand G<sup>b\*</sup>

<sup>a,b</sup>School of Computer Science & Engineering, Vellore Institute of Technology, Vellore, India

<sup>b\*</sup>gopichand.g@vit.ac.in

**Article History:** Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

**Abstract:** The myocardial infarction prediction is an important task in health care domain in the current days. So, Prediction of cardiovascular diseases is a critical challenge in the area of clinical data analysis. It is difficult to predict myocardial infarction prediction by physicians with huge health records. To overcome this complexity we need to implement the automatic heard disease prediction system to notify the patient and get to recovery from the disease. Here to gaining the automatic system we are using machine learning techniques to easily performing myocardial infarction prediction. The machine learning techniques can be split into multiple types like unsupervised and supervised learning classifier. The supervised learning techniques working with structured data which is recommended to implement this classifiers. So, in this system we are using supervised learning techniques namely KNN, RF, NN, DT, NB, and SVM classifiers. To predict myocardial infarction, this system is using training dataset which is accessing from UCI ML repository. As well as this system is comparing accuracy performance between various machine learning algorithms and accuracy results with graphical presentation. This makes the accessing of the risk of the disease in the early stages and can try to save the patient without having any loss.

**Keywords:** Dataset Collection, Classification, Prediction

### 1. Introduction

Day by data the huge of health records are raises in healthcare medical industry. So, there is a recommended to manage a data and make them as useful information for decision making. Due to this problem, healthcare industry wants to implement an automatic system technique which will deliver productive decision from a huge dataset. So, the ML techniques are effective of resolving these kinds of issues very well. Because it can provide effectible methods to retrieve meaningful information without analyze the huge database. In the medical industry, the important data can be gathered from various patient's manifestations and clinical reports for analysis by physicians. These days at stage of lifetime lot of people are getting heart failure symptoms. But comparing between old people and young people, the senior citizens are facing this type of problems. However, the machine learning techniques can find correlations between different features for predicting myocardial infarction status from training dataset. By using this kind of training models, it can detect the myocardial infarction patients without help of medical Practitioners. Then it can pretend as an automatic system to categorize between positive myocardial infarction patients and negative myocardial infarction patients with accurately, so then it reduces the diagnosis time and cost of treatment.

In the medical health care domain, providing qualities services and predict the diagnosis status accurately is a main challenge task. According survey a lot of people passed away with myocardial infarction even managed and controlled effectively by automatic system. Here any disease can be controlled by dependents of detection of that disease at right time. So in this system the proposed system can predict the heart disease status at advance stage to notify the patients and help them to recovery from that disease. The huge of medical records are generated by medical experts for analyze and retrieve the useful information form that database. The health care database contains mostly unattached information which is tedious task for prediction of myocardial infarction. So, in the health care domain if we implement the machine learning techniques which is understand structured information to prediction of various diseases. Therefore, this system proposed a automatic system to physicians for prediction of myocardial infarction at advance stage then they can provide treatment to patient and save them from rugged seriousness. So, the machine learning techniques have an important role in myocardial infarction prediction with supervised classifiers at advance stage to diagnoses the patients.

### 2. Related Work

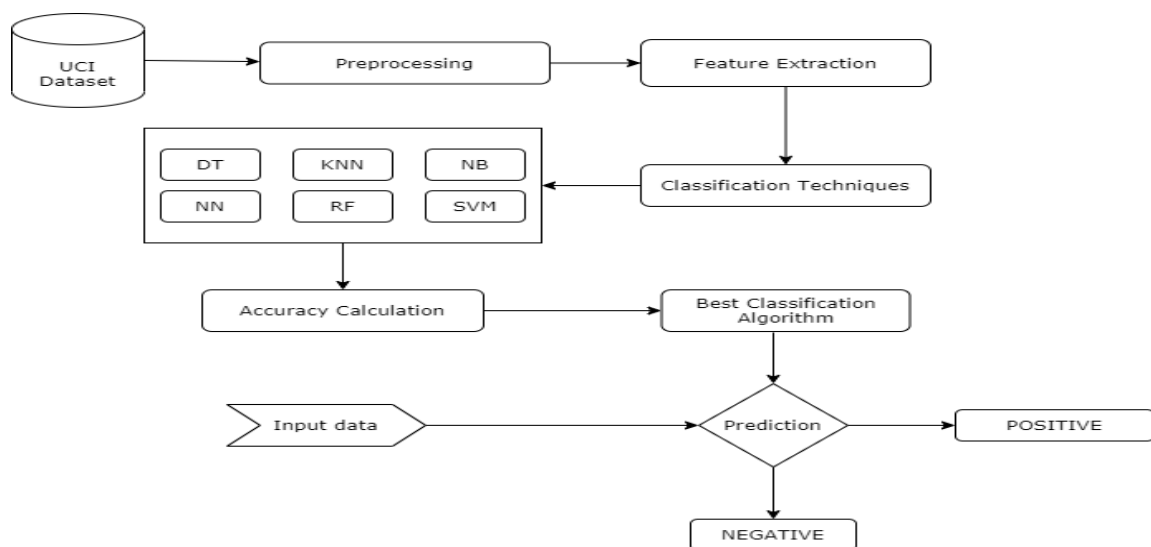
The detection or prediction of myocardial infarction is toughest task in the health care domain. The physicians can detect myocardial infarction with some symptoms such as smoking, high in taking of fat and consuming alcohol etc. But with this symptoms the physicians can detect disease may or may not be accurately. Due to this

reasons doctors cannot do treatment to patients at early stages then there is a chance to face the harmful outcomes by patients. So, to detect or prediction myocardial infarction we need to develop the tool for detection of any disease at early stages then the physicians will do the treatment to patients to preventing harmful consequences. Here the prediction tool can be implementing by the supervised machine learning techniques to myocardial infarction prediction. Many clinical or hospitals generate huge medical records which is unstructured format. So by using this prediction tool can easily fetch the useful information to make the training dataset for further references. The machine learning algorithms will take this training dataset as input and predict the myocardial infarction status with current patient details as testing dataset.

In this system we are proposing myocardial infarction prediction with six machine learning classification algorithms and accuracy results between these six algorithms. Here the main task of this system is predict the accurately when patient feels the pain with heart disease. Regarding implementation with help of heart disease training dataset and machine learning classifiers we build the train model file and then the physicians can enter the input values which is get from patients health records and giving to the training model as input for myocardial infarction prediction. The heart disease training dataset is downloading from UCI ML repository which is uploaded by the medical department experts. To build this system we had chosen python language which is has pre-trained libraries or packages to access the machine learning classifiers. Here all six algorithms providing best accuracies to prediction of myocardial infarction.

### 3. Implementation

#### System Model:



**Figure.1** System Architecture.

The figure.1 depicts about our proposed system model. In this system model they used heart disease training dataset which is downloaded from UCI repository. Later by preprocessing, it can read the training dataset and split the independent and dependent attributes by feature extraction and then build the training model with classification algorithm for myocardial infarction prediction by giving input data, finally calculate the accuracy between six machine learning classifiers.

#### Dataset Collection

In this system we are using UCI heart disease dataset shown in figure.2 which is accessing from Kaggle web repository (<https://www.kaggle.com/ronitf/heart-disease-uci?select=heart.csv>). This training dataset contain 14 attributes or features which are defined in Table.1 as well as it contains 303 records among them 164 records belong to NEGATIVE and 139 records belong to POSITIVE classes or targets.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Target							
2	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0							
3	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1							
4	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1							
5	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0							
6	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0							
7	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0							
8	62	0	4	140	268	0	2	160	0	3.6	3	2	3	1							
9	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0							
10	63	1	4	130	254	0	2	147	0	1.4	2	1	7	1							
11	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1							
12	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0							
13	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0							
14	56	1	3	130	256	1	2	142	1	0.6	2	1	6	1							
15	44	1	2	120	263	0	0	173	0	0	1	0	7	0							
16	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0							
17	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0							
18	48	1	2	110	229	0	0	168	0	1	3	0	7	1							
19	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0							
20	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0							
21	49	1	2	130	266	0	0	171	0	0.6	1	0	3	0							
22	64	1	1	110	211	0	2	144	1	1.8	2	0	3	0							
23	58	0	1	150	283	1	2	162	0	1	1	0	3	0							
24	58	1	2	120	284	0	2	160	0	1.8	2	0	3	1							
25	58	1	3	132	224	0	2	173	0	3.2	1	2	7	1							

Figure 2. Dataset

**Preprocessing:**

In the preprocessing, we primarily load or read the training dataset with help of pandas library and by importing the pandas library we can invoke read\_csv () method for read the entire dataset and store in a variable. The below snippet can show the training dataset loading processes.

```
from PyQt5 import QtCore, QtGui, QtWidgets
import pandas as pd
import sys
import numpy as np

df = pd.read_csv("heartdisease.csv")
```

**Feature Extraction:**

After completion of preprocessing such as loading the training dataset, we need to get features of given dataset. By using feature extraction method this system can separate the input and output attributes and both are storing in x\_train and y\_train variable respectively. The below snippet will shows that syntax.

```
from PyQt5 import QtCore, QtGui, QtWidgets
import pandas as pd
import sys
import numpy as np

df = pd.read_csv("heartdisease.csv")

x_train = np.array(df.drop(['class'], 1))
y_train = np.array(df['class'])
```

From the above snippet using drop () function we can remove the target column or class column to save the only input attribute values and convert it into array format by using numpy module as well as for only storing the output variable values or target column values will be retrieve from data frames.

Here we need to import numpy module from array conversion.

### Classification Techniques:

#### DT:

The DT classifier is supervised machine learning technique, in this classifier it has one root and multiple nodes and finished with leaf nodes. Here the DT classifier can prepare the dataset in the tree form. Later when the user enter the testing dataset for myocardial infarction prediction then it follows the IF and THEN rules which means it compares with every node, until it reaches to leaf node. The leaf nodes contain the target column values such as POSITIVE or NEGATIVE. Finally, the testing dataset value matches the final leaf node value that can be the system predictable values. In this system, we are using sklearn.tree package to import the DecisionTreeClassifier to build training model for heart disease prediction. The below snippet will shows the building of DT classifier model.

```
from sklearn.tree import DecisionTreeClassifier
rf = DecisionTreeClassifier()
rf.fit(x_train, y_train)
pre_cls = rf.predict(x_test)
```

#### RF:

The RF classifier is a collection of decision trees. It also belongs to supervised machine learning algorithm. Here the RF classifier will gathered number of decision trees randomly to taking the decision. While prediction of disease it takes the all output values of decision trees randomly and which class is voted more than other class then that become the system predictable output status. In this system RF classifier providing 98% best accuracy compare with remaining algorithms. This system will use sklearn.ensemble package to import the RandomForestClassifier to build training model for heart disease prediction. The below syntax will shows the preparation of build model.

```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
pre_cls = rf.predict(x_test)
```

#### NN:

The neural network classifier is advance of all classifiers, because it was follow the brain neurons working process. The NN classifier takes the three layers such as input layer, hidden layer and output layer. Here with the primary input layer, we provide the data from the dataset having various features and passing to hidden layers to features classification and finally the classification results can share to output layer. When the classifier return the predictable output where it matches highest percentage with output layers. This classifier also can import sklearn.neural\_network package MLPClassifier for myocardial infarction prediction. Follow the below snippet code:

```
from sklearn.neural_network import MLPClassifier
rf = MLPClassifier()
rf.fit(x_train, y_train)
pre_cls = rf.predict(x_test)
```

#### NB:

In this predictive model, the naive bayes algorithm can be used for prediction of heart disease. This algorithm follows the bayes rule to hear disease prediction. It is fastest and easily predictable classifier and it calculates posterior probability events with other events and this algorithms uses mostly for text classifications. This classifier MultinomialNB is importing from sklearn.naive\_bayes package. The classifier following the below snippet.

```
from sklearn.naive_bayes import MultinomialNB
nb = MultinomialNB()
nb.fit(x_train, y_train)
pre_cls = nb.predict(x_test)
```

#### SVM:

The support vector machine classifiers in an important classifier because of its classification advantages. The SVM classifier while classification of features, first it can draw the margins between different classes and the hyper plane line can separate with support vectors which means the nearest classes to that hyper plane line. Here this system can separate hyper plane with POSITIVE and NEGATIVE features and select the nearest support vectors and build the training model to predict status of the disease. The following is the syntax code of prediction.

```
from sklearn import svm
svm = svm.SVC()
svm.fit(x_train, y_train)
pre_cls = svm.predict(x_test)
```

#### KNN:

The K-nearest neighbor classifier is a different learning classifier compare with another machine learning classifier, because it follow the Euclidian distance formula to calculate the distance. This classifier while prediction it calculates the distance between with each records then it returns the distance and store it like this follow the last record and it can return the predictable output value which distance is less to compare with all distances and that one become our myocardial infarction predictable output like POSITIVE or NEGATIVE. It is also import the KNeighborsClassifier module from this sklearn.neighbors. Here we took K value is 1 for pick the nearest distance value as output.

```
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier()
knn.fit(x_train, y_train)
pre_cls = knn.predict(x_test)
```

#### Prediction

This module will be executing after build the training model with respective best classifier. For myocardial infarction prediction we need to invoke predict () method with testing dataset as input. This method will be available in every classifier. By calling this function it can start to comparing with training dataset with given testing dataset with respective classifier and it returns the target column as output which is matches to near with training dataset. The below syntax is used for prediction of myocardial infarction as POSITIVE or NEGATIVE.

```
from PyQt5 import QtCore, QtGui, QtWidgets
from sklearn.ensemble import RandomForestClassifier
import pandas as pd
import sys
import numpy as np

df = pd.read_csv("heartdisease.csv")

x_train = np.array(df.drop(['class'], 1))

y_train = np.array(df['class'])

tf = pd.read_csv("testing_dataset.csv")

testdata = np.array(tf)

testdata = testdata.reshape(len(testdata), -1)

rf = RandomForestClassifier()

rf.fit(x_train, y_train)

result = rf.predict(testdata)
```

### Calculation of Accuracy

Here the system can calculate accuracy between six supervised machine learning algorithms. For this we need to split the myocardial infarction dataset with 70% as training dataset and remaining 30% as testing dataset can be done with help of `train_test_split()` method which is importing from `sklearn.model_selection` package. Therefore it can return the `x_train`, `x_test`, `y_train`, `y_test` parameters then by taking the inputs as `x_train`, `y_train` and build the training model with respective classifiers and get the predicted classes `pre_cls` by invoke the prediction function by taking input as `x_test` then finally calling `metrics.accuracy_score()` with input `y_test` and `pre_cls` then it returns the accuracy of each respective classifier. The below snippet will show the process of accuracy calculation.

```
from PyQt5 import QtCore, QtGui, QtWidgets
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
import pandas as pd
import sys
import numpy as np

df = pd.read_csv("heartdisease.csv")

datainput = np.array(df.drop(['class'], 1))

y = np.array(df['class'])

x_train, x_test, y_train, y_test = train_test_split(datainput, y, test_size=0.3)

rf = RandomForestClassifier()

rf.fit(x_train, y_train)

pre_cls = rf.predict(x_test)

accuracy_rf = metrics.accuracy_score(y_test, pre_cls) * 100
```

#### 4. Experimental Results

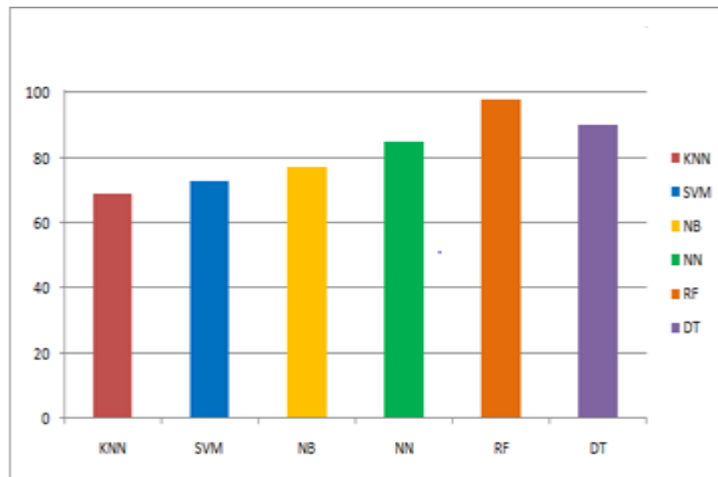


Figure.3 Accuracy Comparison between six Classifiers

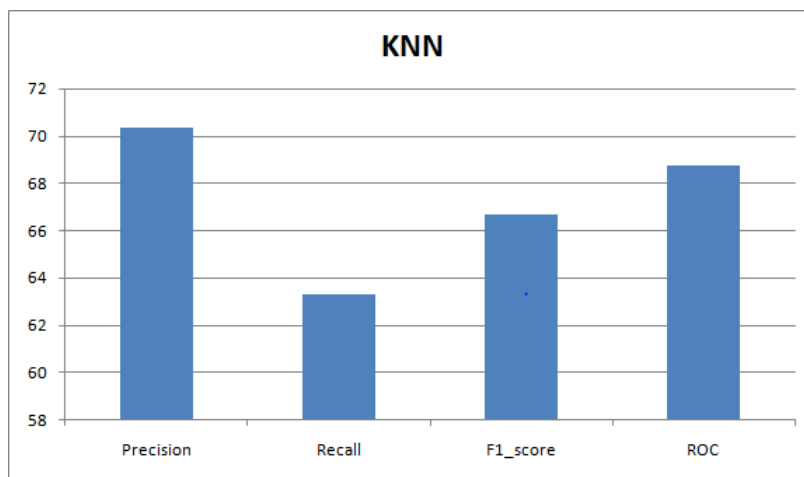


Figure.4 KNN Algorithm Metrics

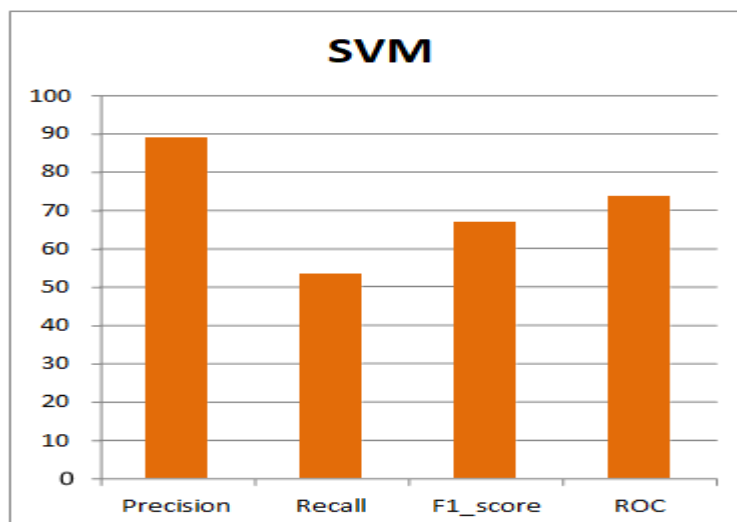
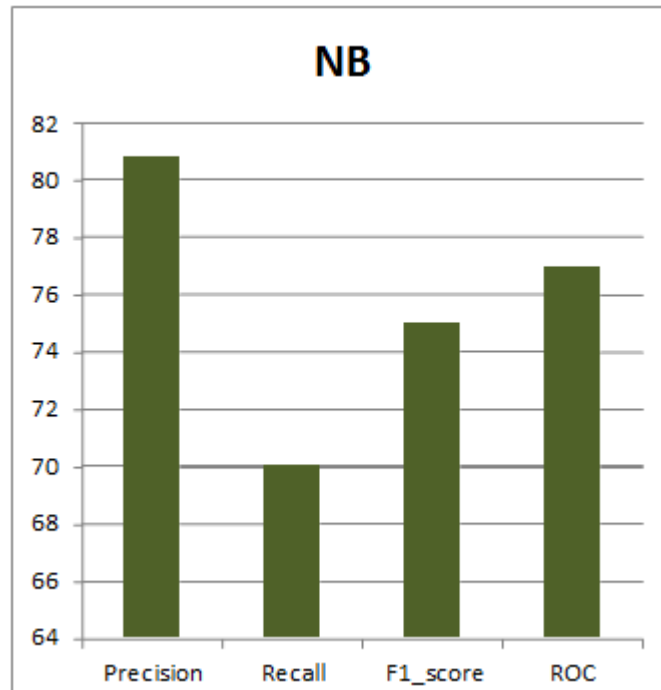
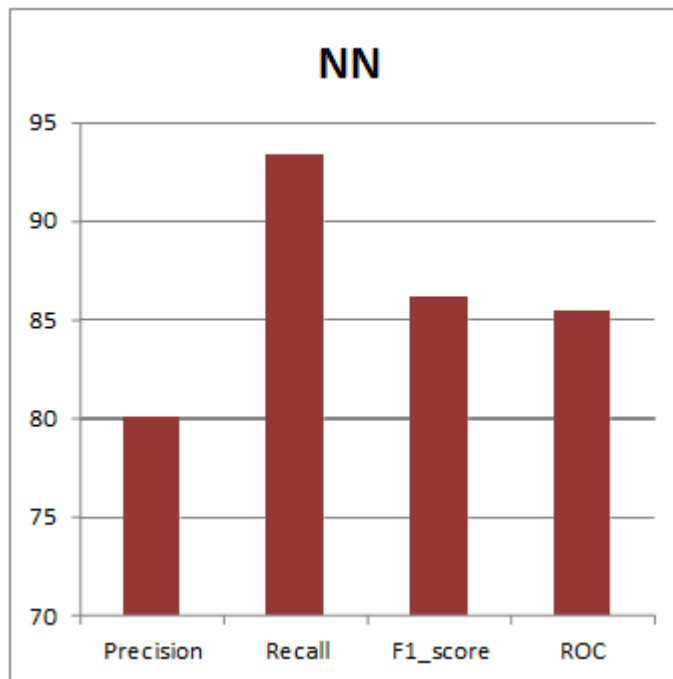


Figure.5 SVM Algorithm Metrics



**Figure.6** Naïve Bayes Algorithm Metrics



**Figure.7** Neural Networks Algorithm Metrics



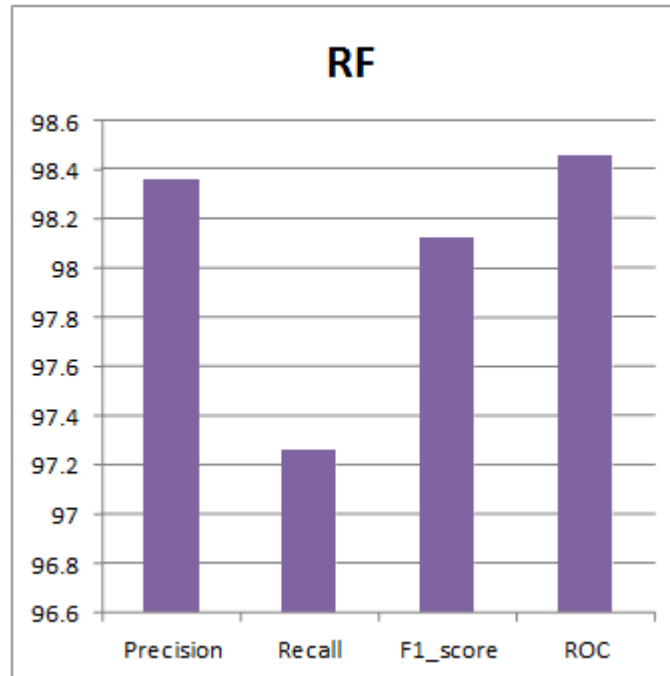


Figure.8 Random Forest Algorithm Metrics

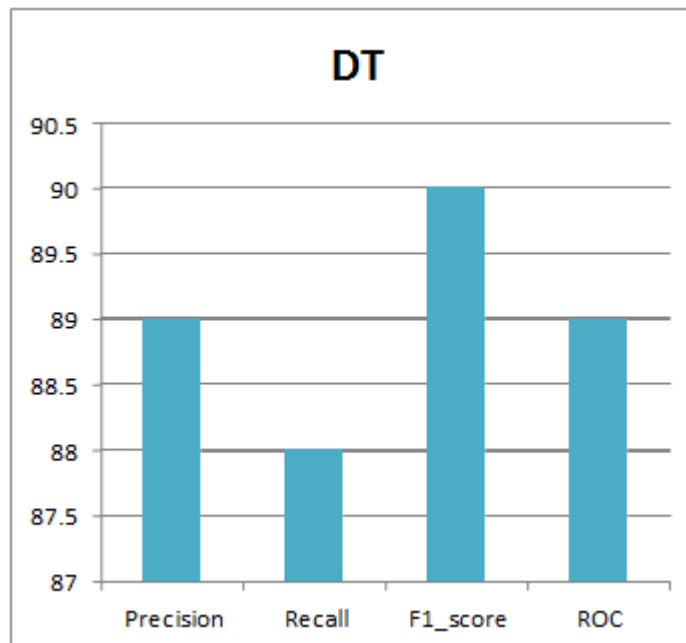


Figure.9 Decision Tree Algorithm Metrics

### 5. Conclusion

By taking the advantages of online world we have lot of medical history data is available. Extracting and analysis medical history data is become very necessary for the prediction of the diseases. Especially in myocardial infarctions, the rate of deaths due to heart attacks is increasing day by day.

This rate we can decrease by predicting disease by analyzing the heart patient's medical history data. In this project we have propose a comparative analysis of myocardial infarction prediction using popular classification algorithms. We classify and compare the results in terms of Accuracy calculation. Here we have used KNN, SVM, NB, NN, DT and Random Forest for classifying the heart attack medical data and calculate the accuracy score. In the algorithms used the accuracy of 98% has been achieved by random forest algorithm.

## References

- T.Nagamani, S.Logeswari, B.Gomathy, “Heart Disease Prediction using Data Mining with Mapreduce Algorithm”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278- 3075, Volume-8 Issue-3, January 2019.
- Fahd Saleh Alotaibi, “Implementation of Machine Learning Model to Predict Heart Failure Disease”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.
- Avinash Golande, Pavan Kumar T, “Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.
- Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, “Design And Implementation Heart Disease Prediction Using Naives Bayesian”,International Conference on Trends in Electronics and Information(ICOEI 2019).
- Radha B. K, Dr. Bharati Harsoor, Mr. Rupam Das “A Novel Framework for Prediction and Detection of Myocardial Ischemia with Single Lead ECG and PAT”
- Theresa Princy R,J. Thomas, “Human heart Disease Prediction System using Data Mining Techniques”, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.
- C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, —Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field,|| in Machine Learning Paradigms, 2019, pp. 71–99.
- Puneet Bansal and Ridhi Saini et al. “Classification of heart diseases from ECG signals using wavelet transform and kNN classifier”, (ICCCA2015).
- Hossein Yousefi-Banaema, Saeed kermania, Sasan Asiaeib, Hamid Saneic “Prediction of Myocardial Infarction by Assessing Regional Cardiac Wall in CMR Images Through Active Mesh Modeling”.
- S. Radhimeenakshi, “Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural network,” in Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom), New Delhi, India, Mar. 2016, pp. 3107–3111.