CV Project 2 Report

Below are the steps that I followed for the classification of images into face and non-face.

Dataset

I used the dataset that was provided along with the project description. I used 1000 face images and 1000 non face images for the training sample and 100 face images and 100 non face images for testing. Each image is of 16*16 dimensions. Training samples and testing samples are mutually exclusive.

HAAR Feature Extraction

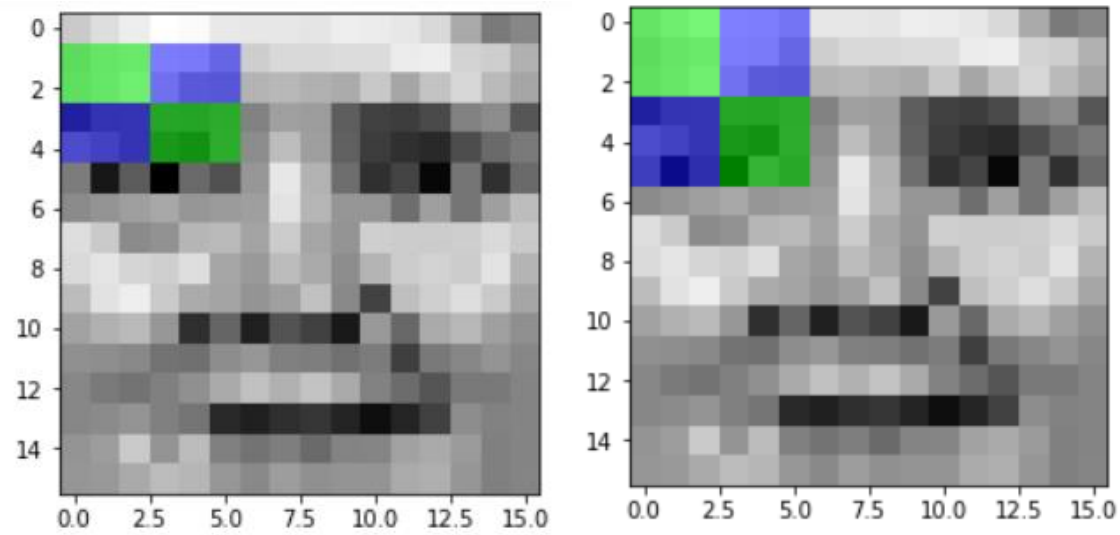
I used integral_image method of the skimage library to get the Integral image. I then used the haar_like_feature and haar_like_feature_coord of the skimage library to extract the haar features and haar feature coordinates respectively. Each image has a haar feature which is a vector of length 32176. Hence the training set will have 2000*32176 has the input matrix to the algorithm.

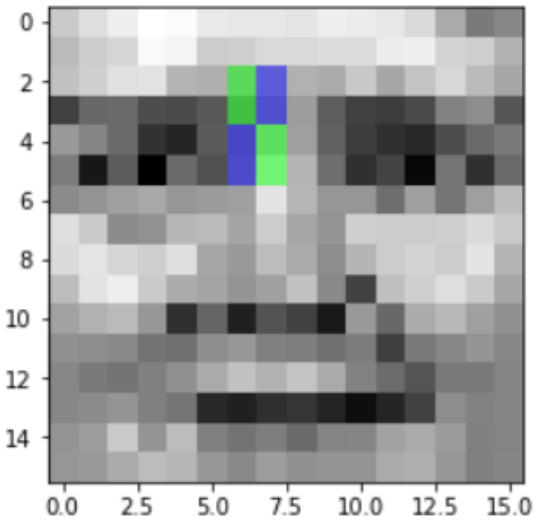
Weak Classifier

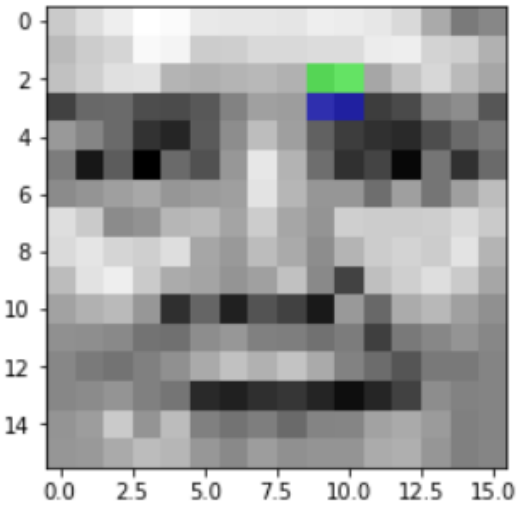
I built a decision stump as the weak classifier which has threshold, polarity, alpha as classifier parameters.

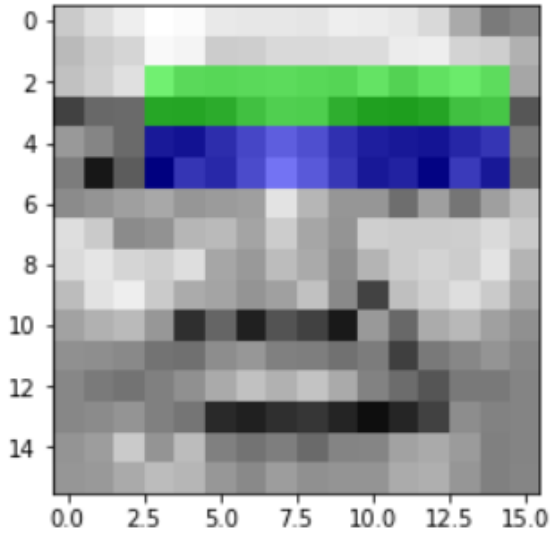
Before Boosting

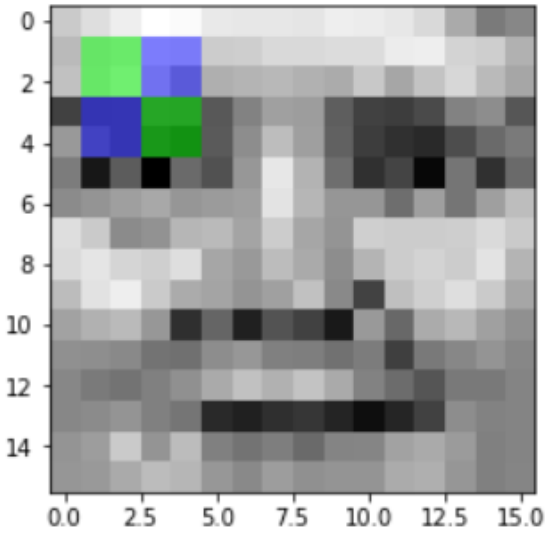
Each haar feature acts as a weak classifier and I have computed the threshold, polarity and minimum error for each weak classifier. Since there are 32176 haar features, I have 32176 weak classifiers. Based on the minimum error criteria I have picked best 10 haar features. Below are the plots of the best 10 features before boosting:

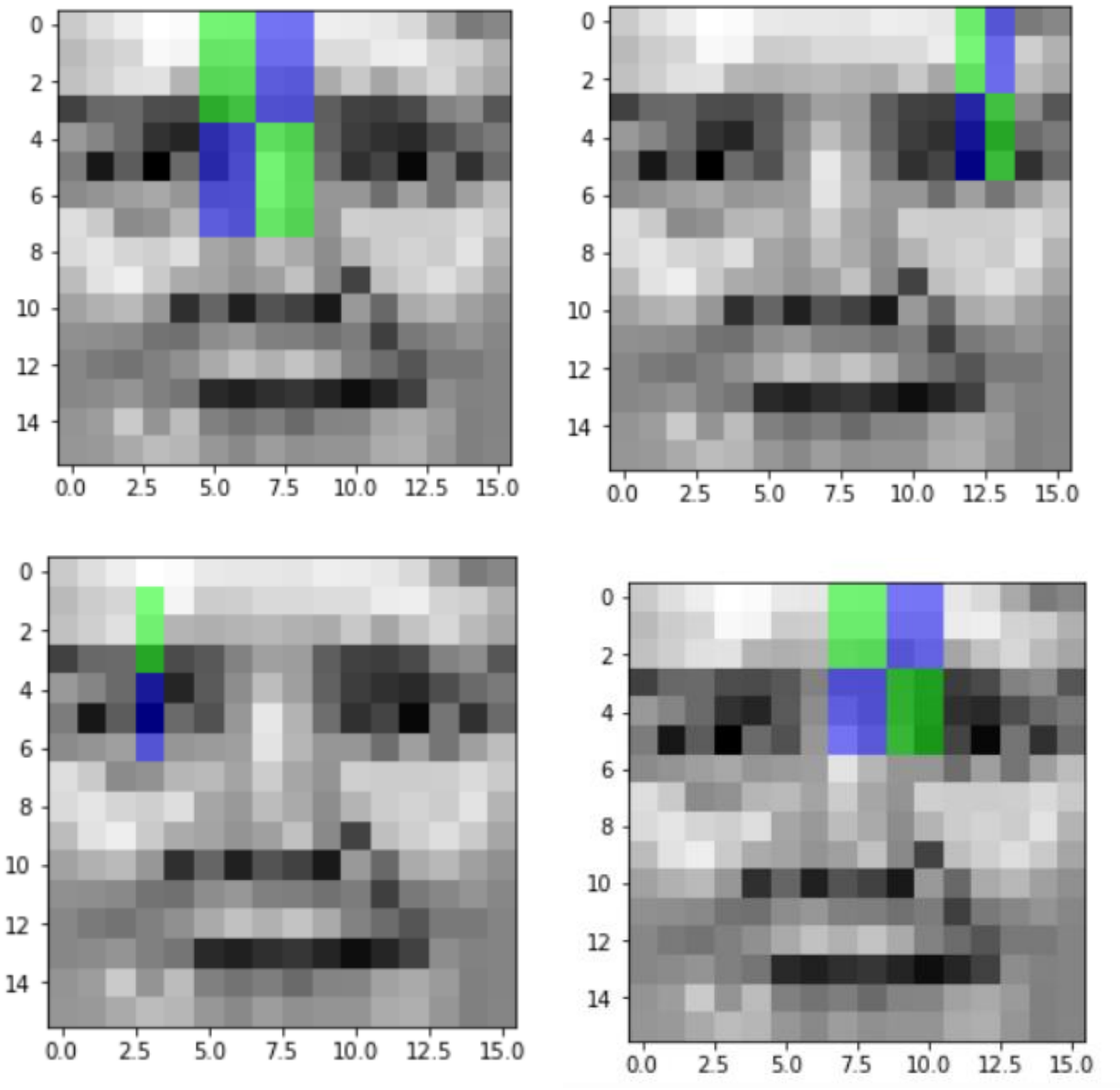



0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0


0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

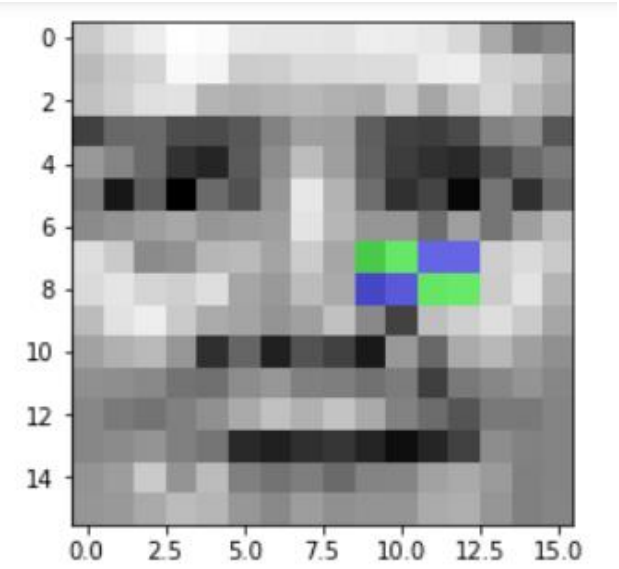

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

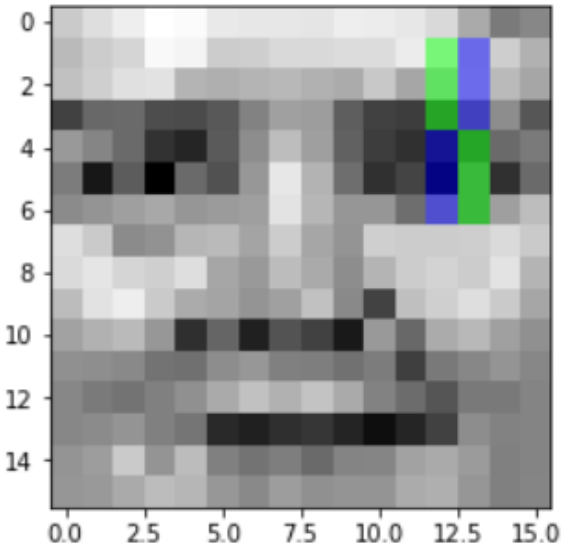

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

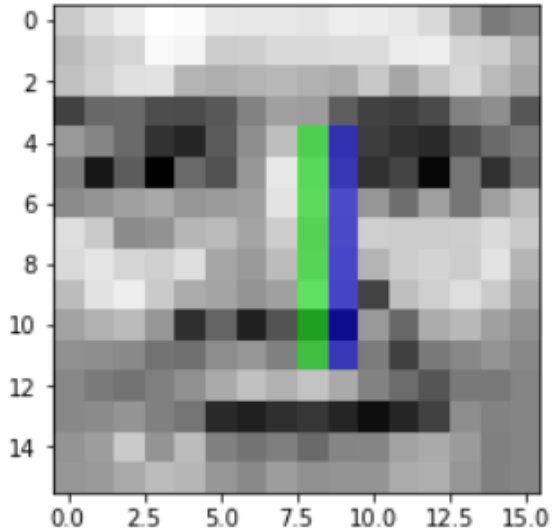


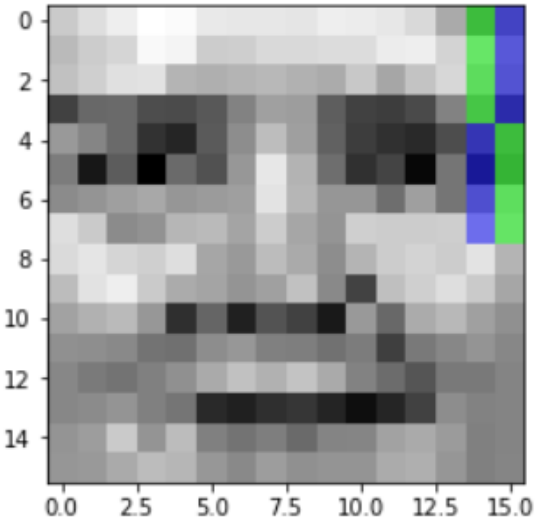
Ada Boost

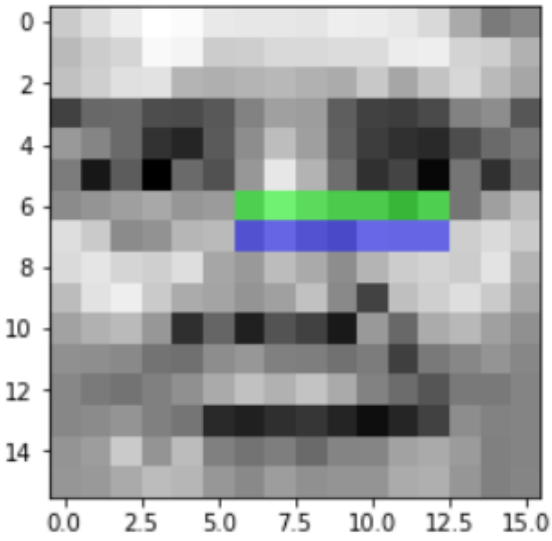
Based on the algorithm given in the project description I have built the ada boost model which is based on the weak classifier mentioned earlier. Each haar feature acts as the weak classifier. The ada boost algorithm computes alpha, weights, threshold, polarity and minimum error for each classifier. I then picked the best 10 haar features based on the minimum error criteria. Below are the plot of the best 10 features obtained from ada boost algorithm:

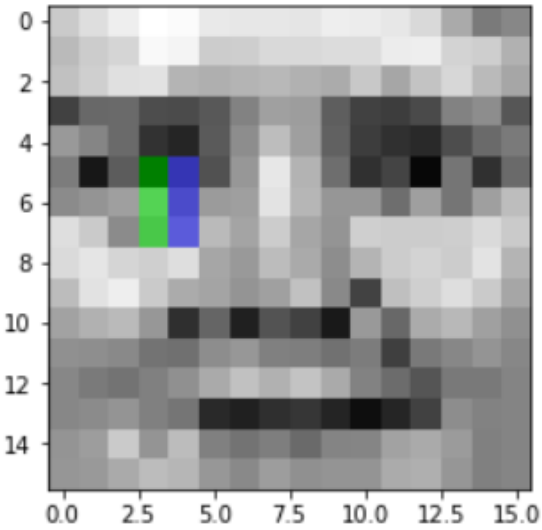

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

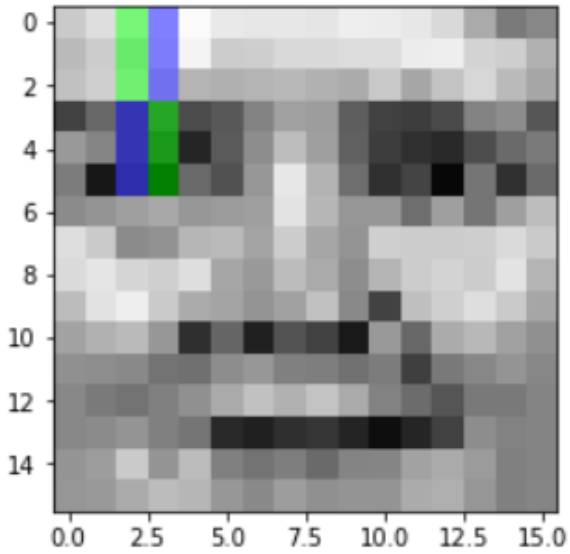

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

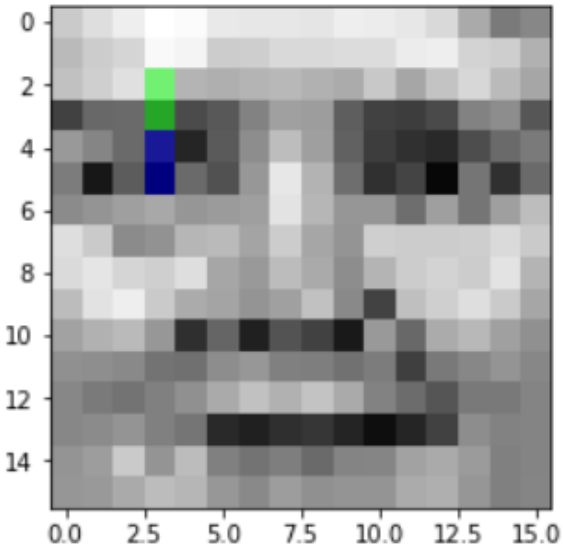

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

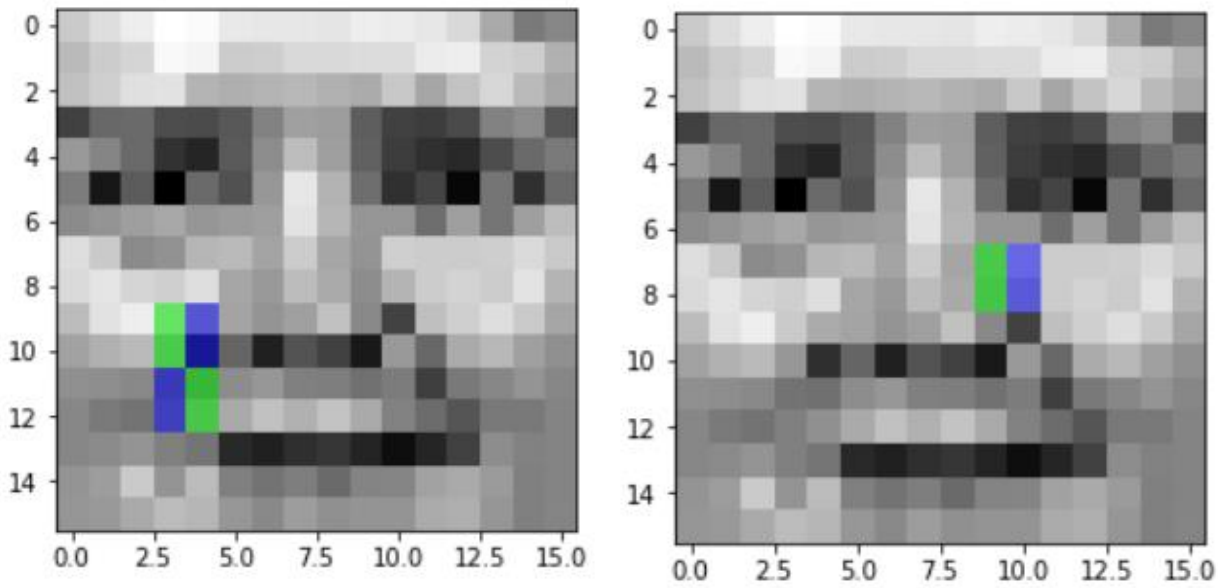

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0


0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0


0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0

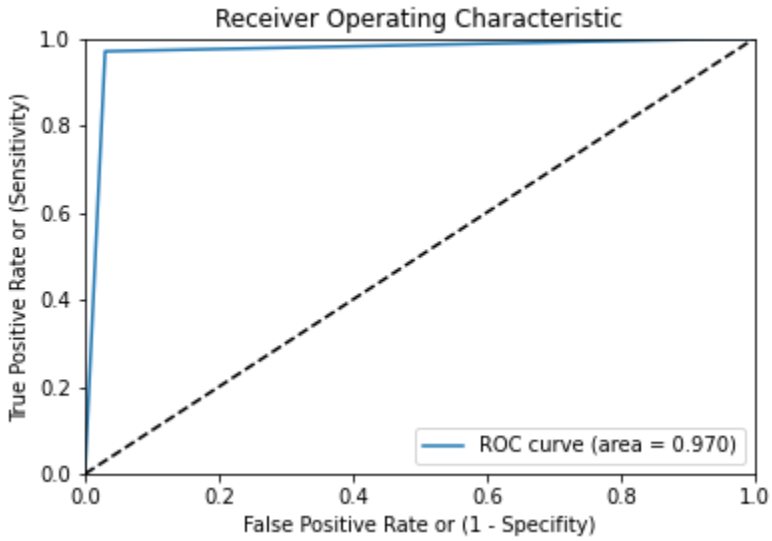

0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0


0
2
4
6
8
10
12
14
0.0
2.5
5.0
7.5
10.0
12.5
15.0



ROC Curve

Below is the ROC curve of the ada boost model that I built.



Accuracy

I obtained the accuracy of 97% on the testing sample by predicting the testing samples using the ada boost model.

Observation of Features

After observing the features before boosting and features obtained after boosting, I observed that the features obtained from ada boost are fine-tuned and hence helps in obtaining a better accuracy.