CSE 572 DATA MINING

Activity Recognition using Myo Sensor

Arizona State University

GROUP NUMBER: 14

Group Members

Simran Singh	1214917639
Pratik Sahoo	1215063460
Krishna Babuji	1215039332
Krithika Mahalingam	1215193486
Shubham Vipul Majmudar	1215200298

1. INTRODUCTION

Armbands, wrist watches and other body sensors are now ubiquitous. This opens up a world of new data that is unstructured, and highly complex data that if utilized properly, could solve a lot of medical problems for many. Capturing a lot of data brings with itself a lot of problems, which data do we use, how do we process it and what models do we use such that we get meaningful insights. In this project we deal with data captured from Myo armband sensors and try to answer all these problems**.**

1.1 Keywords

Feature processing, PCA, Decision Tree, SVM, Neural Network

1.2 Terminology

- *EMG* : Electromyography
- *IMU* : Inertial Measurement Unit
- *PCA* : Principal Component Analysis
- *SVM* : Support Vector Machines
- *ANN* : Artificial Neural Networks

2. MODELS USED

Decision Trees:

Decision trees are currently one of the most popular methods used for data modelling. They have the advantage of being conceptually simple, and have been shown to perform well on a variety of problems. Decision trees are constructed recursively from training data using a top-down greedy approach in which features are sequentially selected. During this feature selection process, the features that partition the set of instances into a more pure subset are preferred. The feature that has maximum information gain is said to be purest.

Support Vector Machines:

Logistic Regression and Decision Trees fail in one aspect; they cannot construct complex decision boundaries. Support Vector Machines intend to overcome this very shortcoming. SVMs aim to find a hyper-plane $W^T X + b = 0$ that best separates the data. Here W is the vector of weights that is to be learned, X is the feature matrix and b is a bias. Solving for W and b results in a constrained Lagrangian optimization problem.

Neural Networks:

The artificial neural networks (ANN) are like the neurons in the biological brain, it consists of a collection of nodes that are representing layers in the network. The nodes in a particular layer are all connected to the nodes in the previous as well as the consecutive layer. They are all connected together with ‘edges’ and each edge has a weight associated with it. The weights get adjusted as the learning process proceeds and until it reaches the optimal value. The signals are transmitted from the first layer to the last layer through these edges and nodes so as to solve problems like a human brain.

3. TASKS PERFORMED

3.1. Feature Creation

The groundtruth folder contained the data that was sampled at 50 frames/second. But, the data collected from MYO armband was sampled at 30Hz. Since the sampling rates were different, we had to multiply the frame number in groundtruth folder by a factor 50/30 to get the corresponding row number in the Myodata folder. After extraction of the corresponding rows, we took minimum, maximum, average, standard deviation and root mean square of each of the features of the IMU and EMG, and created another matrix with these features. For assignment 2, we took the 60% data of each user for training and 40% of data of each user for testing for 15 users. For assignment

3, we took total of 15 users and used the data of 9 users for training and the data of another 6 users for testing.

3.2. Feature Extraction using PCA

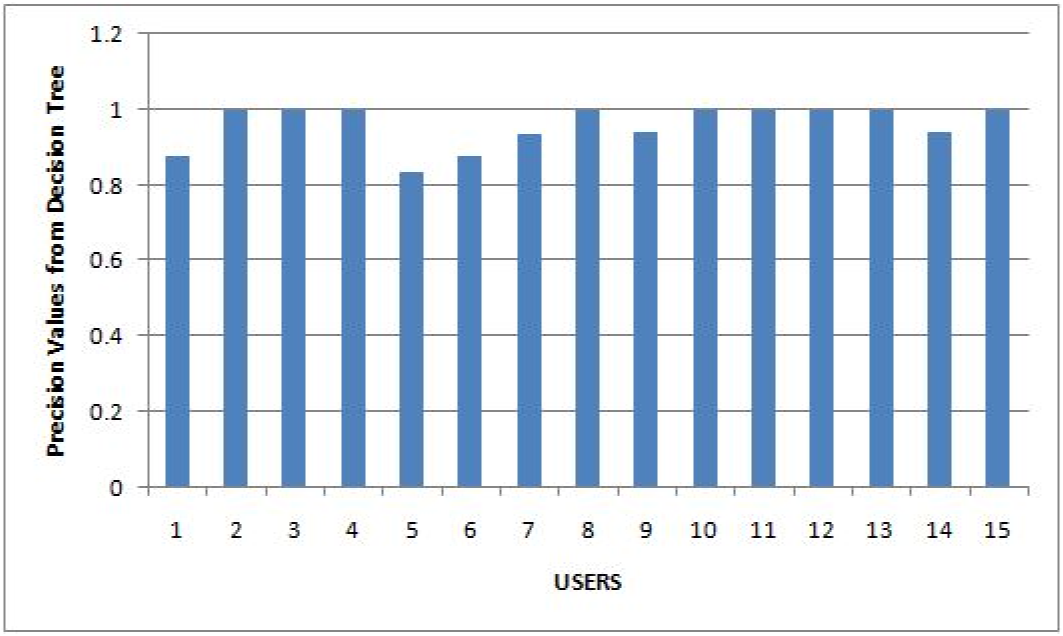
PCA is an acronym for Principal Component Analysis. PCA is a form of eigen decomposition that returns the direction where data has the maximum variance. It takes in the set of input features and returns an equal number of features that explains the data arranged in descending order of their eigenvalues. Hence, we can use these eigenvalues to identify the more important eigenvectors that best represents the data. For the purpose of our project we are using a reduced number of features from the projected matrix of our data (PC scores) for each user -- thus reducing the computation required for our task.

3.3. Prediction using Decision Trees, SVM, and Neural Nets

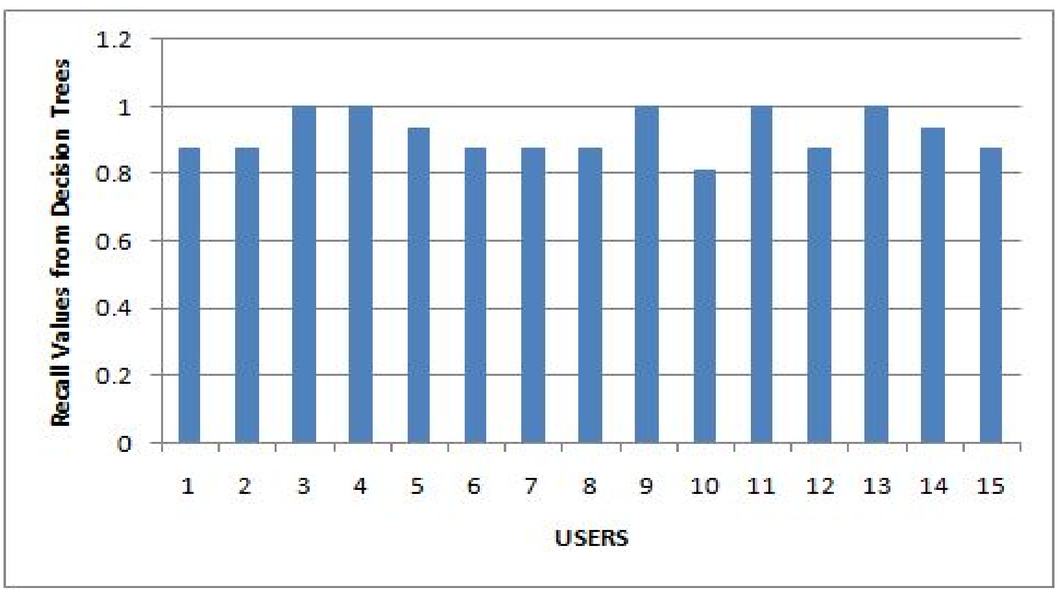
Once we get the new extracted features using PCA, we will now use Decision Trees, SVM and Neural nets to train the 60% of the records from each user and test the remaining 40% records of each user as a part of phase 2. In phase 3, we will do the same, train and test for 15 users, where eating and non eating activity of 9 users are taken as training dataset and its classifier is used to test the activities of 6 remaining users. The accuracy of the algorithms for each user and the combined dataset of 15 users are reported as values of Recall, Precision and F-Score. We have performed Decision tree using fitctree function, SVM using fitcsvm function and Neural Networks using Toolbox in MATLAB. The accuracy metrics and plots are documented in the following Result Section.

4. RESULTS

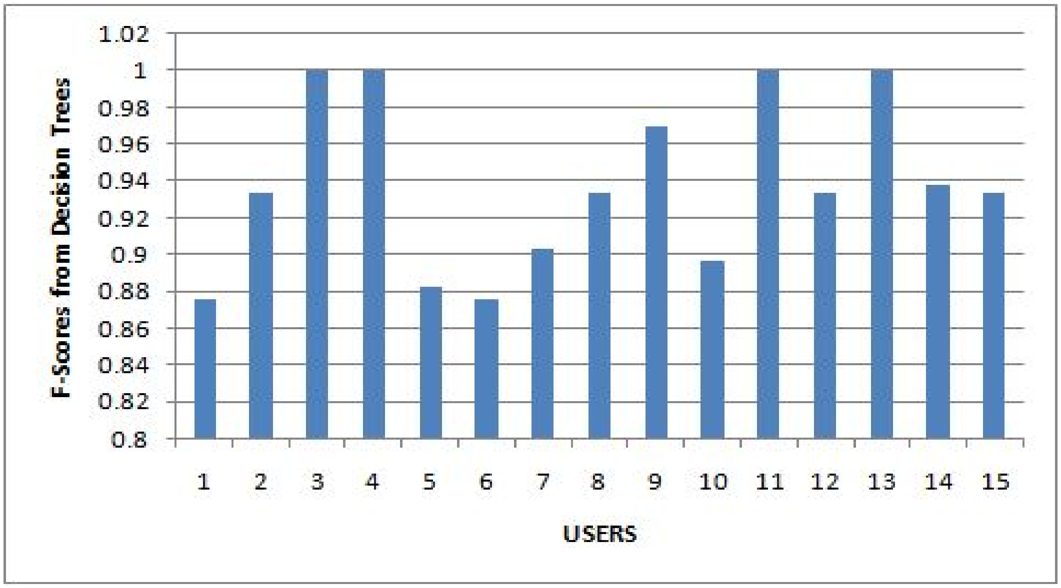
Assignment 2: For each user--



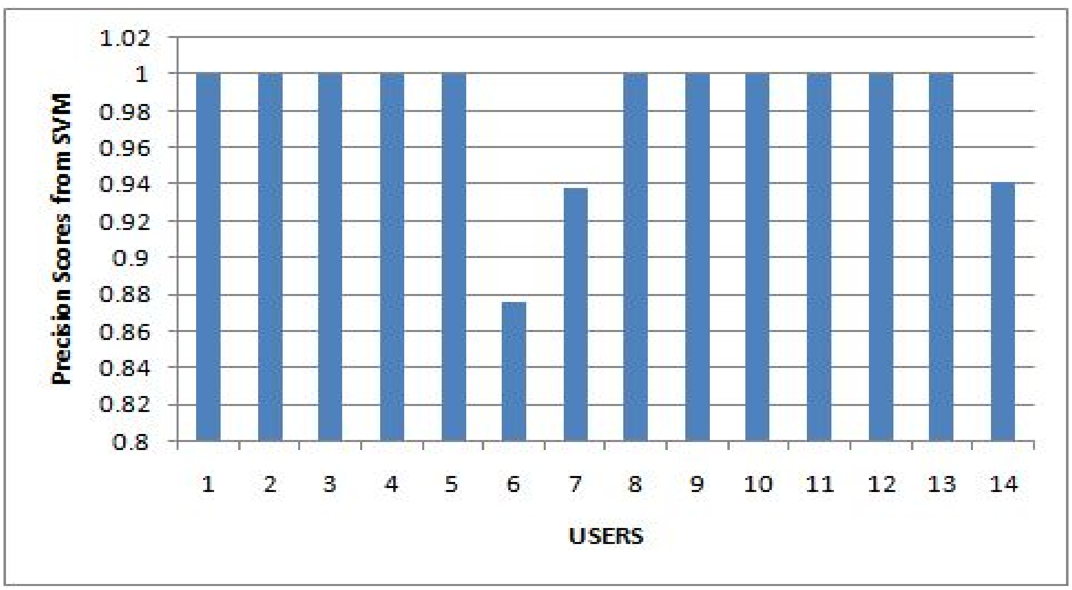
Precision Values of 15 users trained on Decision Tree



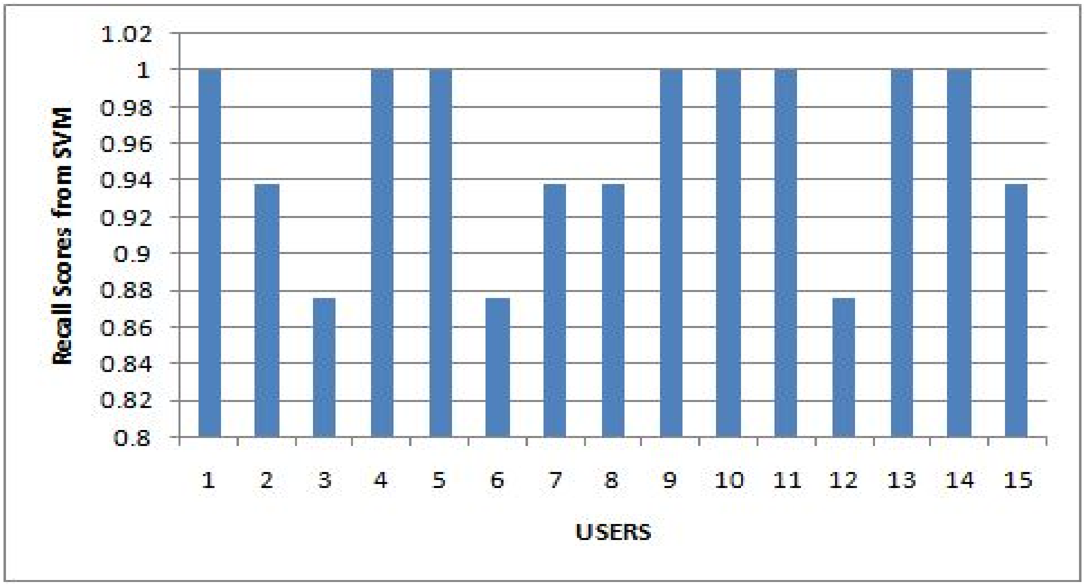
Recall Values of 15 users trained on Decision Tree



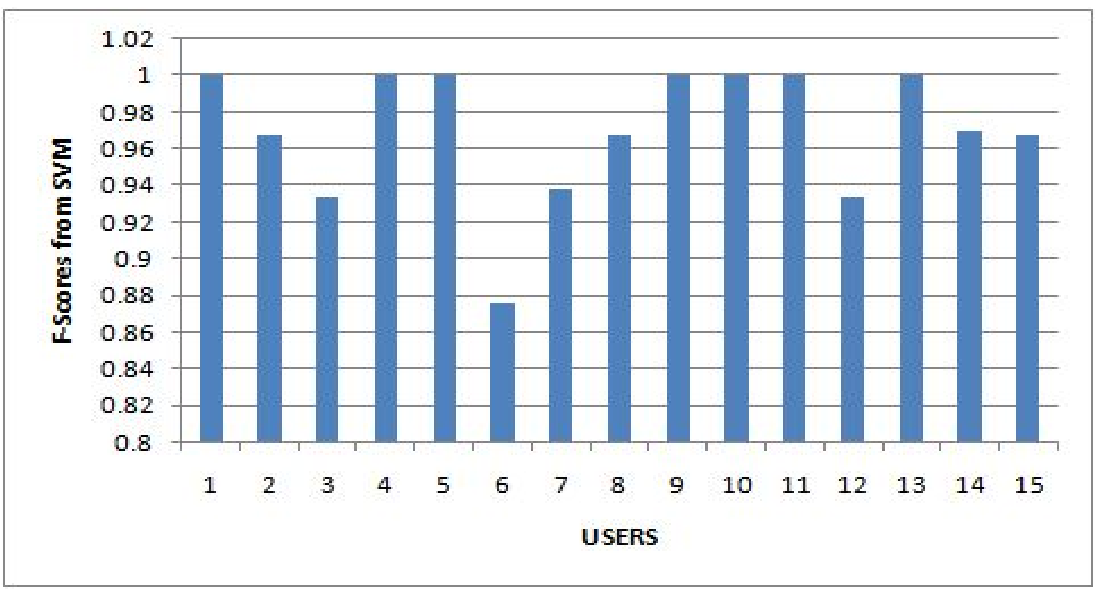
F-Scores of 15 users trained on Decision Tree



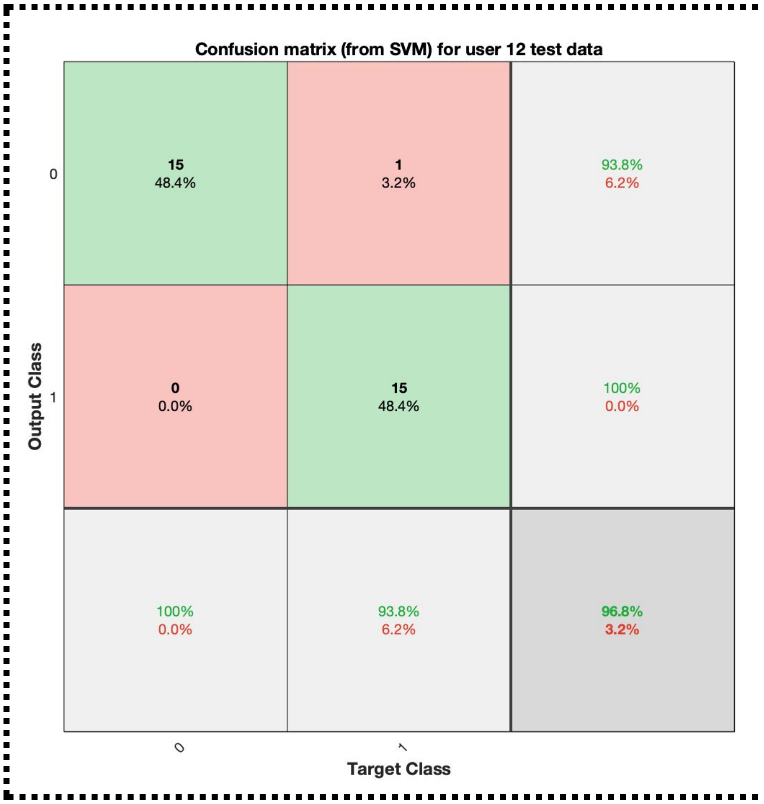
Precision Values of 15 users trained on SVM

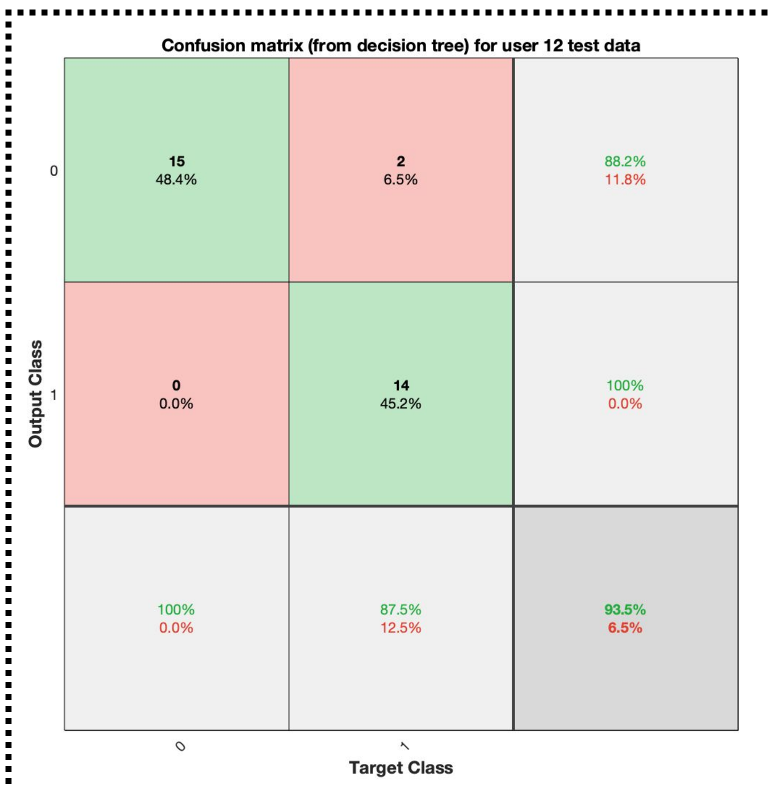


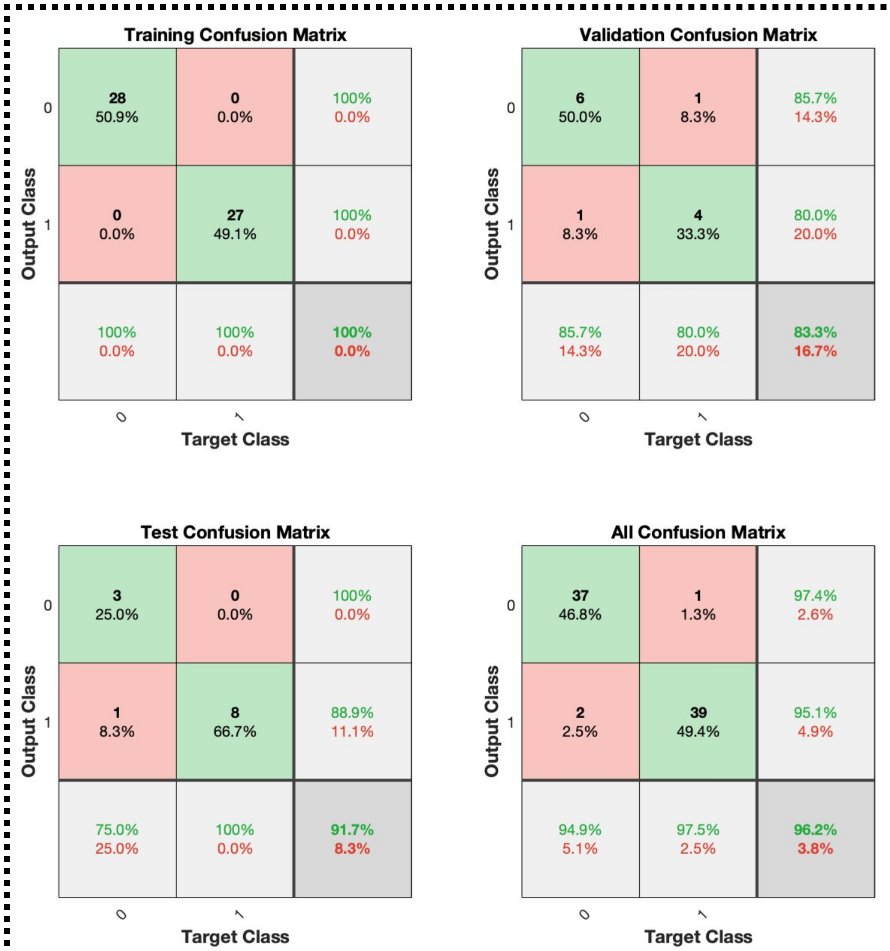
Recall Values of 15 users trained on SVM

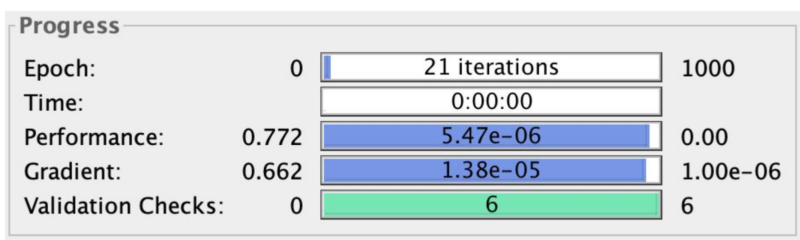


F-Scores of 15 users trained on SVM


Confusion matrix (from SVM) for user 12 test data
15
48.4%
1
3.2%
93.8%
6.2%
0
0.0%
15
48.4%
100%
0.0%
100%
0.0%
93.8%
6.2%
96.8%
3.2%
Output Class
0
1
Target Class
0
1


Confusion matrix (from decision tree) for user 12 test data
15
48.4%
2
6.5%
88.2%
11.8%
0
0.0%
14
45.2%
100%
0.0%
100%
0.0%
87.5%
12.5%
93.5%
6.5%
Output Class
0
1
Target Class
0
1

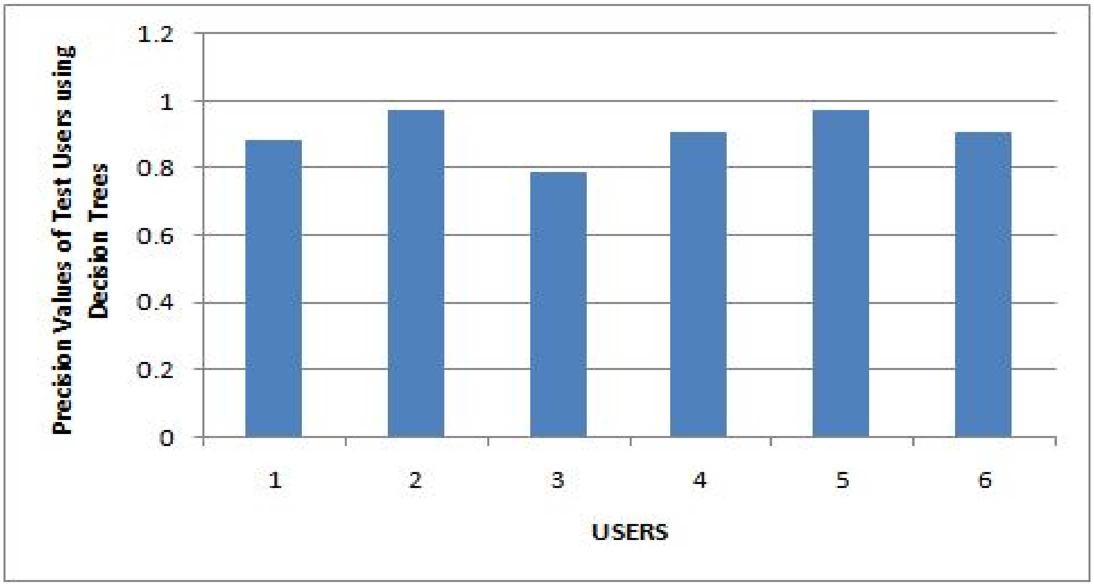




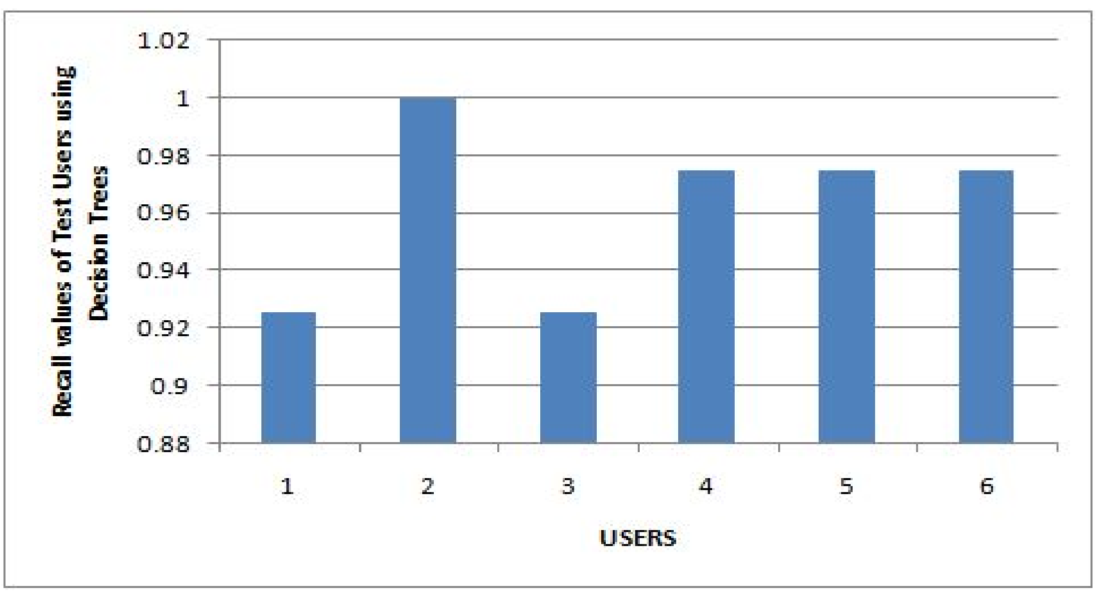
Confidence matrices for user-1 using SVM, decision tree and neural network

Assignment 3:

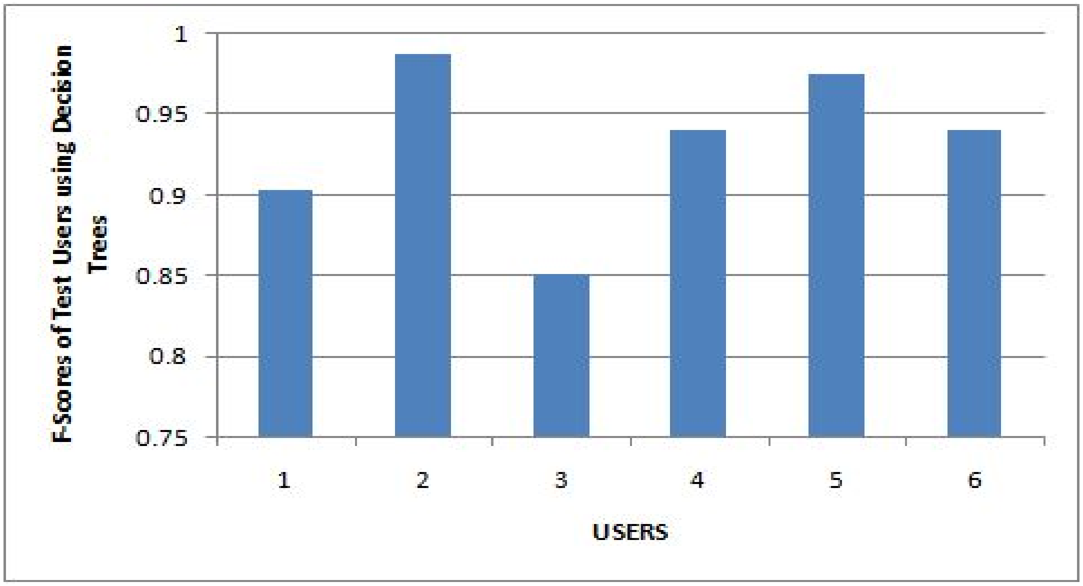
Combined matrix for all test users (in our case, 40% i.e. 6 test users)



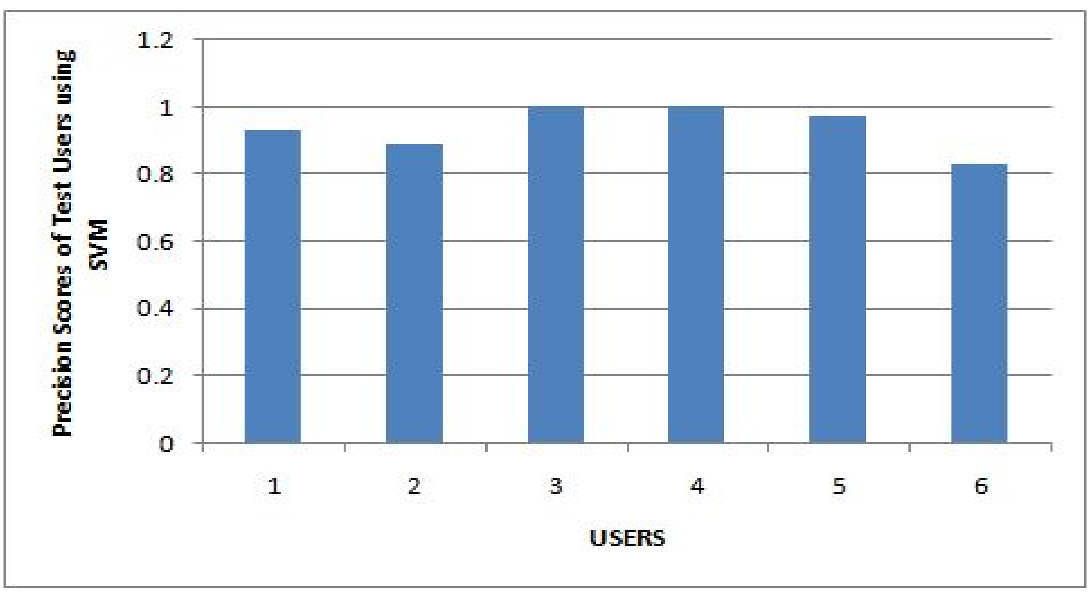
Precision Values of Test Users trained on Decision Trees



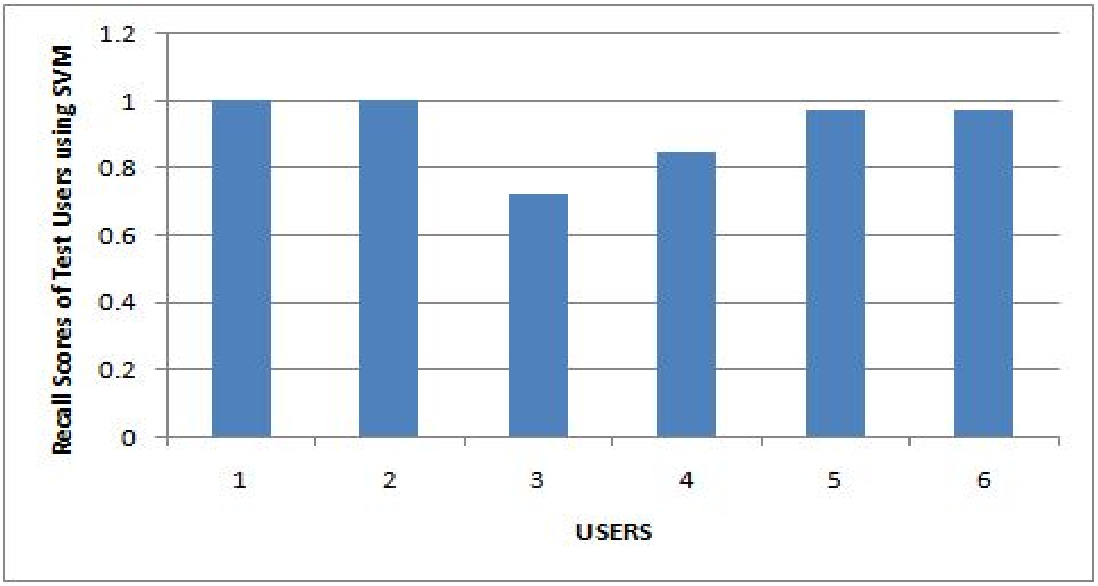
Recall Values of Test Users trained on Decision Trees



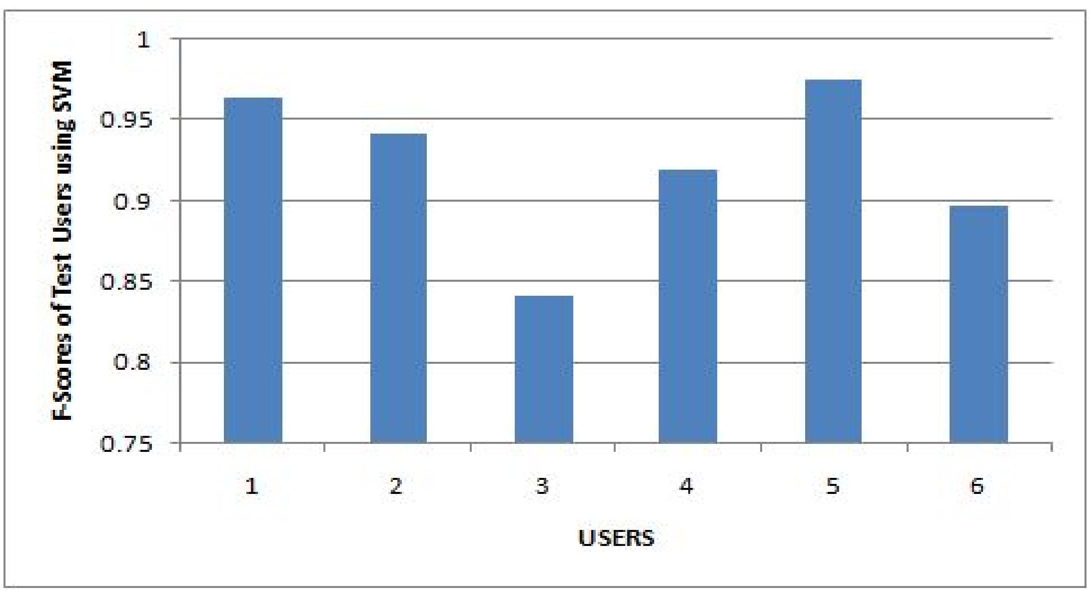
F-Scores of Test Users trained on Decision Trees



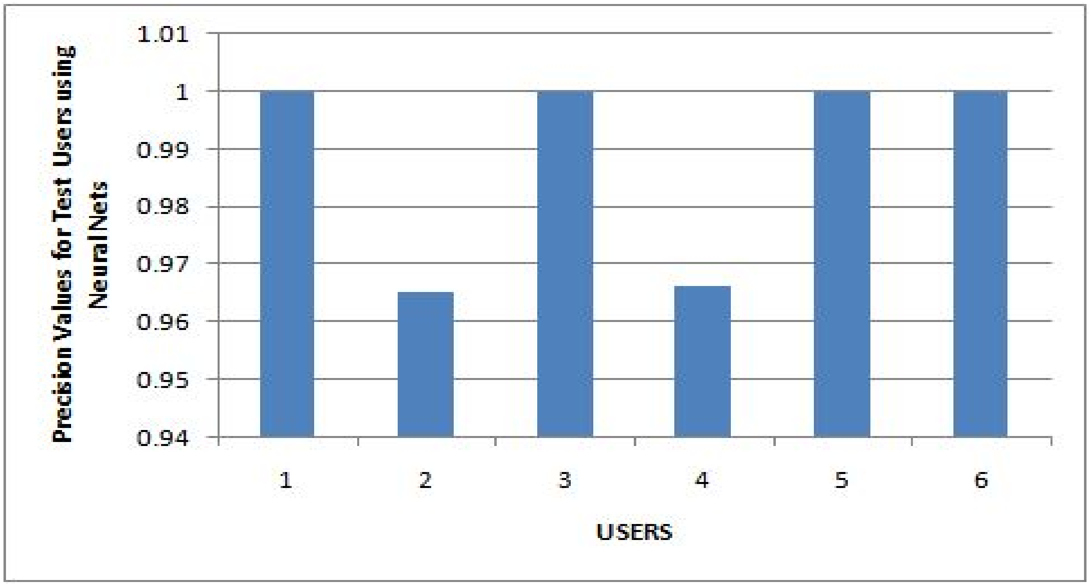
Precision Values of Test Users trained on SVM



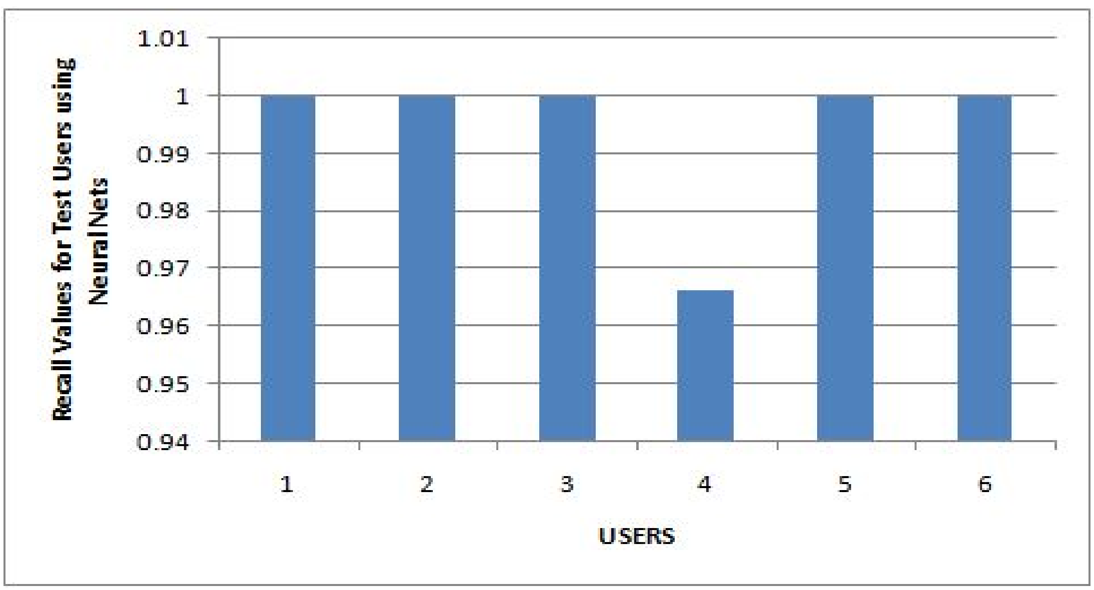
Recall Values of Test Users trained on SVM



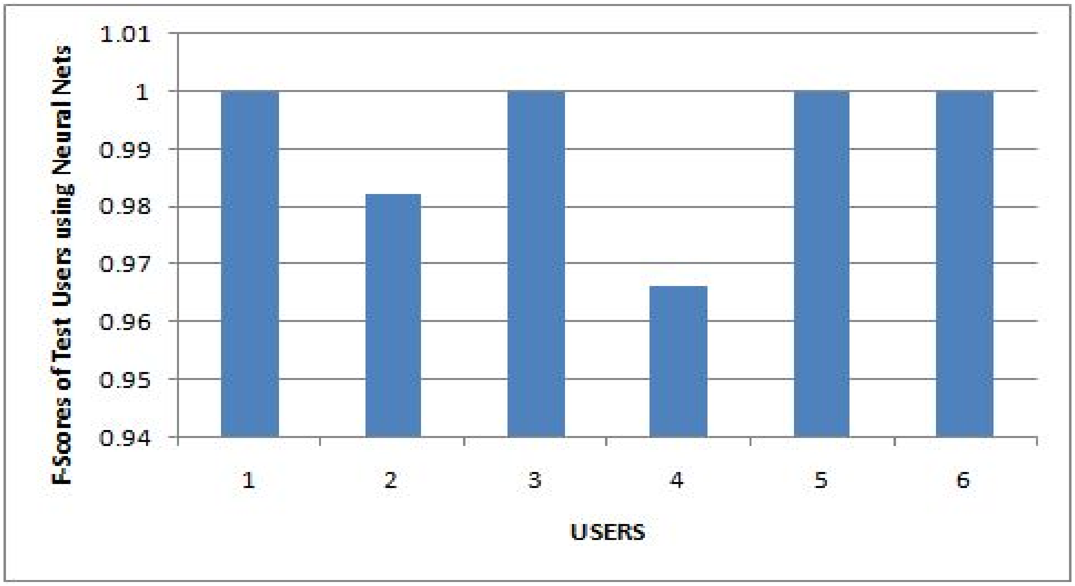
F-Scores of Test Users trained on SVM



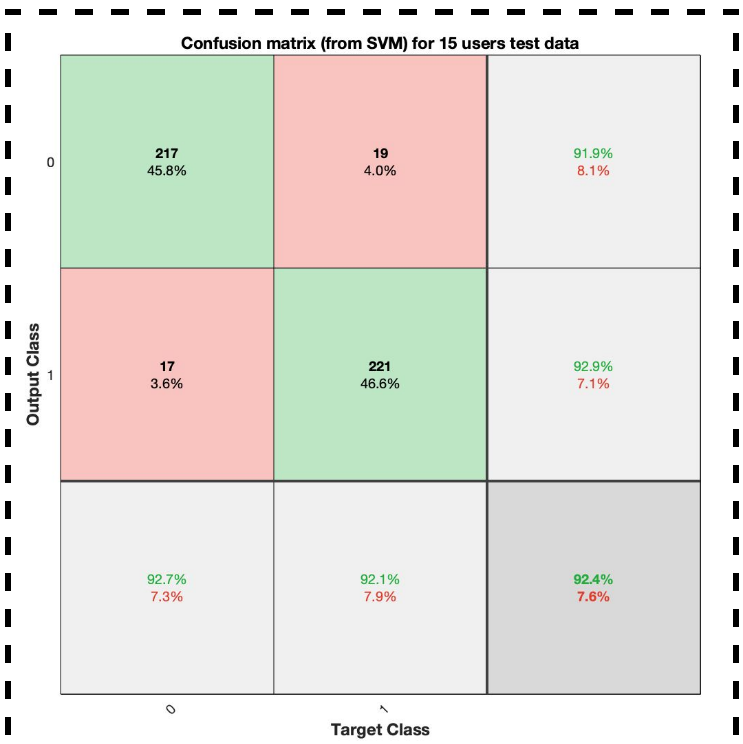
Precision Values of Test Users trained on Neural Nets

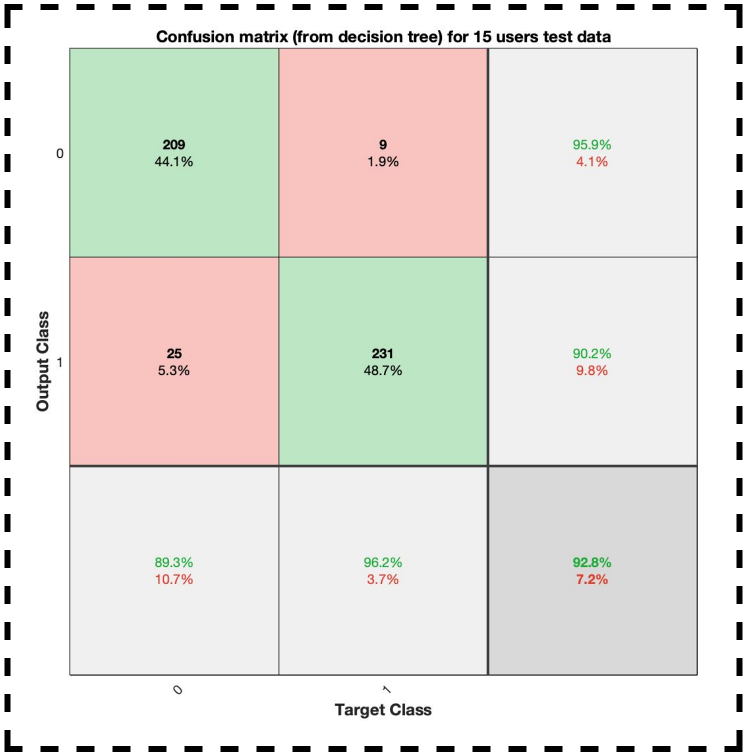


Recall Values of Test Users trained on Neural Nets

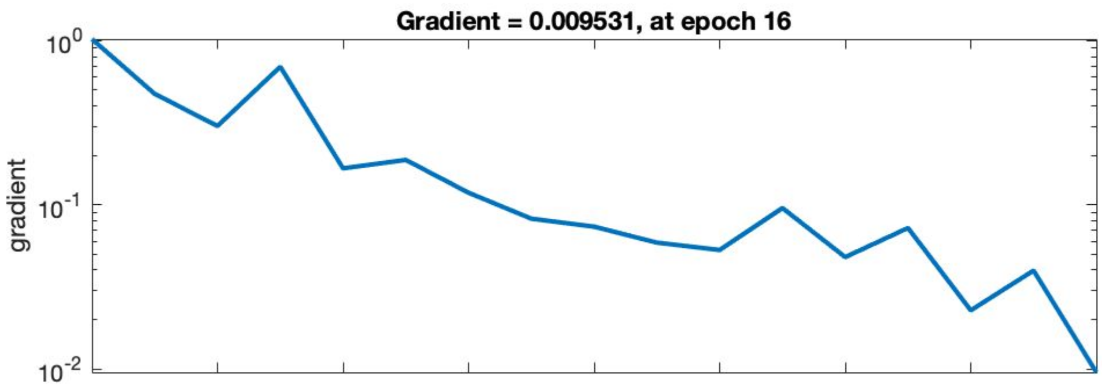


F-Scores of Test Users trained on Neural Nets

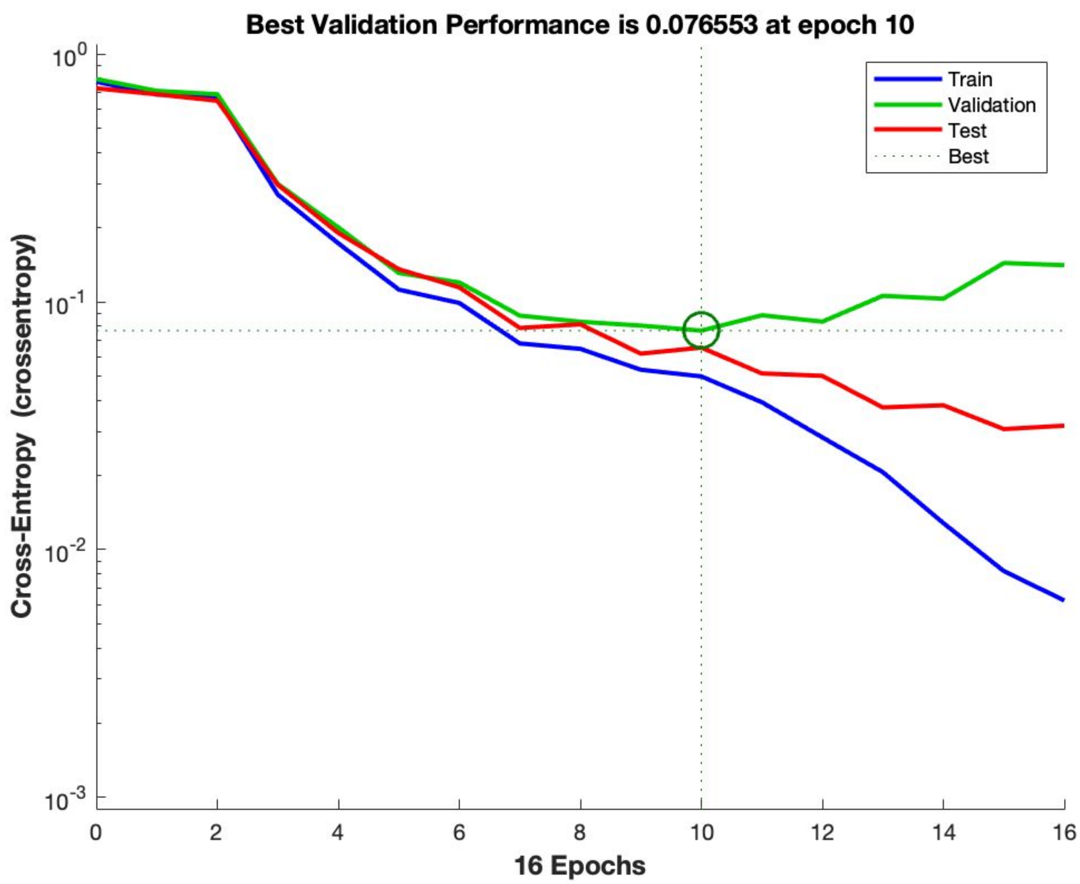



Confusion matrix (from decision tree) for 15 users test data
Output Class
0
1
209
44.1%
9
1.9%
95.9%
4.1%
25
5.3%
231
48.7%
90.2%
9.8%
89.3%
10.7%
96.2%
3.7%
92.8%
7.2%
0
1
Target Class

Confidence matrices for all 40% users using SVM, decision tree and neural network



Gradient descent for master dataset -- each unit is 10 iterations



Cross-entropy loss for master dataset

Confidence matrices for 6 test users on the master (all user) data

5. CONCLUSION

For single user data, all models seem to perform well on the training data, with hardly one or two instances of action being misclassified. However, since the number of instances are very small, this will result in bad generalization and the model will tend to perform badly for action instances of other users, or sometimes even an action instance of the same user.

Neural networks tend to perform better than decision trees and Support Vector Machines in test data. This means that they have a good capacity to generalize the classifier which helps in performing well for unseen data. SVM performs well in the training data, but not as well in the test data. This suggests that this model tends to overfit to an extent. On the other hand, the decision tree model generalizes too well for both, seen (training) and unseen (test) data, which is an indication of underfitting.

6. REFERENCES

[1] Toutountzi, Theodora & Collander, Chris & Phan, Scott & Makedon, Fillia. (2016). EyeOn: An Activity Recognition System using MYO Armband.
1-3.10.1145/2910674. 2910687

[2] Mathworks pca() documentation

[3] Social Media Mining : An Introduction *- Reza Zafarani, Mohammad Ali Abbasi, Huan Liu*

[4] PCA: http://www.lauradhamilton.com/introduction-to-principal-component-analysis-pca

[5] Neural Networks: http://neuralnetworksanddeeplearning.com

7. APPENDIX

Confidence matrices for user-2 using SVM, decision tree and neural net

Fig 2: Confidence matrices for user-3 using SVM, decision tree and neural net