Pattern Recognition Semester Project

Report 1

SHUDDHANK PUSHOLA 20170020239
ADITYA MEHRA 20170020189
PRATYUSH RAJ 20170020232

Activity Recognition from Single Chest-Mounted Accelerometer Data Set

December 11, 2019

Problem

Uncalibrated Accelerometer Data are collected from 15 participantes performing 7 activities. The dataset provides challenges for identification and authentication of daily activities performed by people.

Data Set Information

- **The dataset collects data from a wearable accelerometer mounted on the chest**
- **Sampling frequency of the accelerometer: 52 Hz**
- **Accelerometer Data are Uncalibrated**
- **Number of Participants: 15**
- **Number of Activities: 5 (The original Dataset has 7 activities but for our project we're using 5 activities)**
- **Data Format: CSV**

Attribute Information

- Data are separated by participant
- Each file contains the following information
- sequential number, x acceleration, y acceleration, z acceleration, label
- Labels are codified by numbers
    - 1: Working at Computer
    - 2: Standing
    - 3: Walking
    - 4: Going UpDown Stairs
    - 5: Talking with Someone

Feature Selection For Motion Data

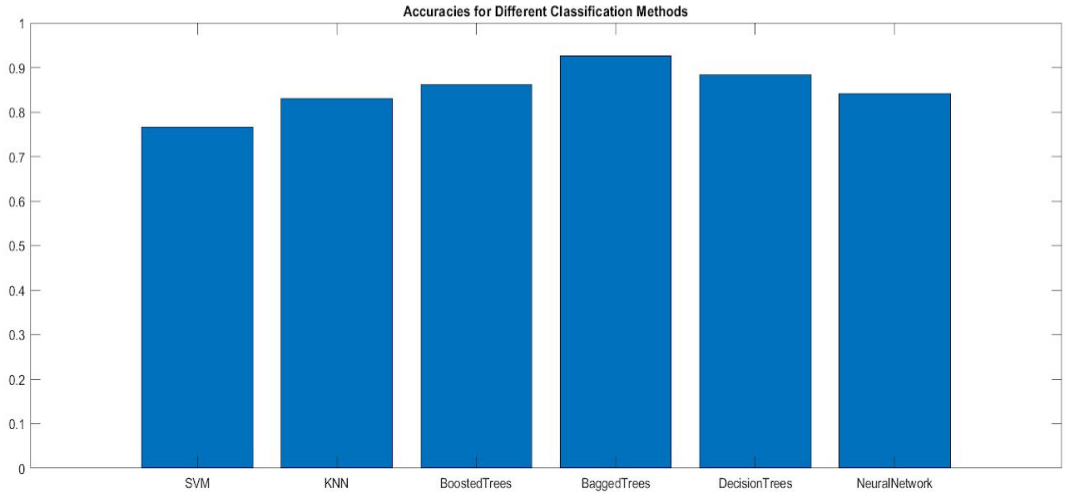
Recognizing human activities depends directly on the features extracted for motion analysis. Accelerometers provide three separate accelerometer data time series, one time series for acceleration on each axis Ax,Ay,Az. We will calculate an additional time series Am which is the magnitude of acceleration. From research it has been found that detection of movement is much more dependent on low frequency components of accelerometer data. We filter the time series data into low pass and high pass signal with the cutoff frequency set to 1Hz. We use these three time series data for feature extraction.

Features Extracted

- **Time Series Data : We extract time series measures like mean, variance, kurtosis, skewness, min and max value of the data**
- **Cross Correlation : We take the max values of the cross correlation between the x,y and z axis**
- **Velocity : We calculate the velocity by taking integrals of window size 10 and averaging the obtained values**
- **Energy Values of Wavelet Decomposition : We calculate the energy of coefficients of 7 level wavelet decomposition of the data**
- **After performing Feature Extraction we get a feature vector of size 125**

Classification

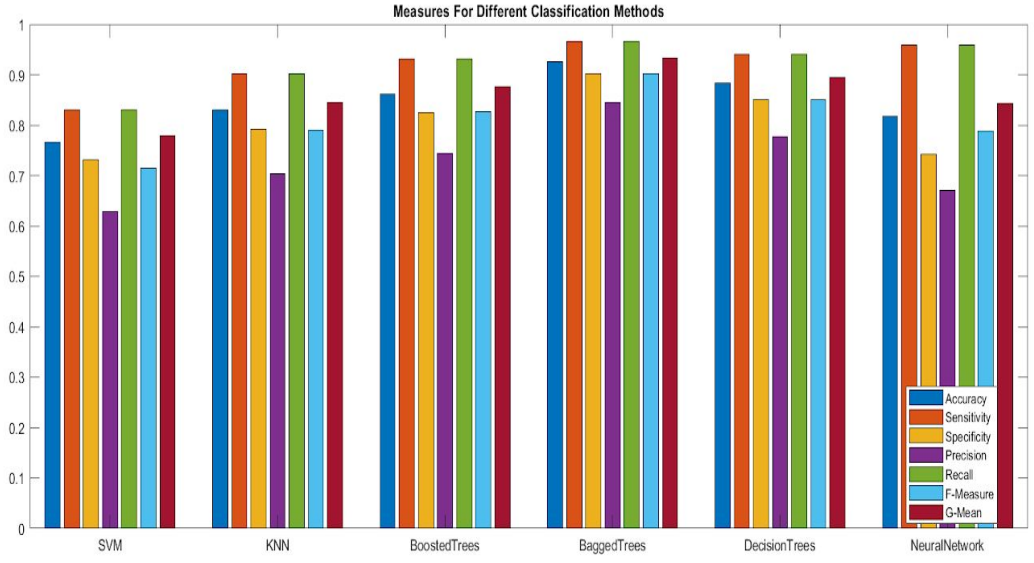
We carried out classification using different classifiers like KNN, SVM, Trees and Ensemble classifiers like Bagged Trees and Boosted Trees and also used Neural Networks. We got the Following Results



We also calculated different Measures for each Classification Method.

Classification Method	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure	G-Mean
SVM(Gaussian)	0.7659	0.8309	0.7303	0.6281	0.8309	0.7154	0.7789
KNN	0.8304	0.9022	0.7910	0.7029	0.9022	0.7902	0.8447
Boosted Trees	0.8621	0.9316	0.8240	0.7437	0.9316	0.8271	0.8761
Bagged Trees	0.9252	0.9662	0.9027	0.8447	0.9661	0.9014	0.9338
Fine Tree	0.8830	0.9408	0.8513	0.7763	0.9408	0.8506	0.8949
Neural Network	0.8520	0.9587	0.7414	0.6702	0.9586	0.7889	0.8431

For Better Visualization We have plotted the Data in the Form of a Bar Graph

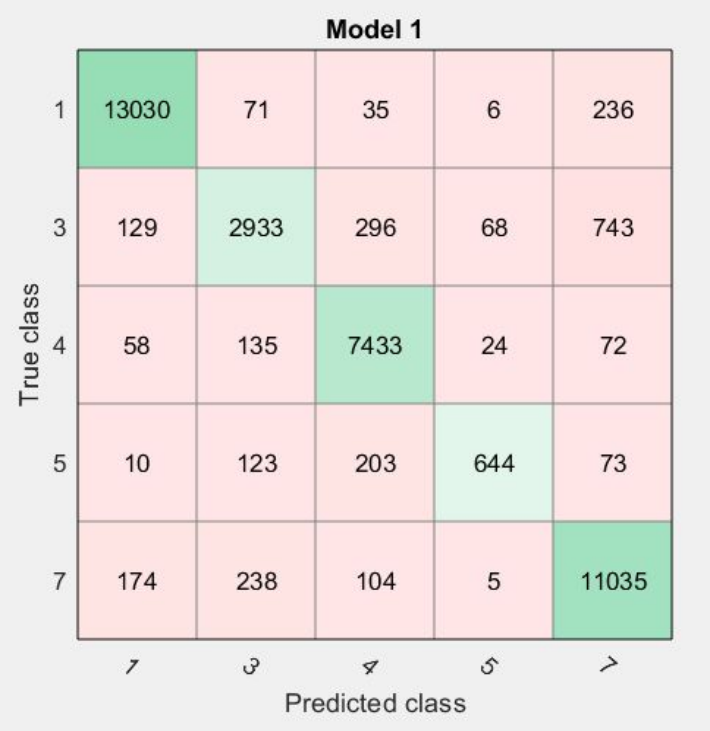


Best Classifier (Bagged Trees/Random Forest):

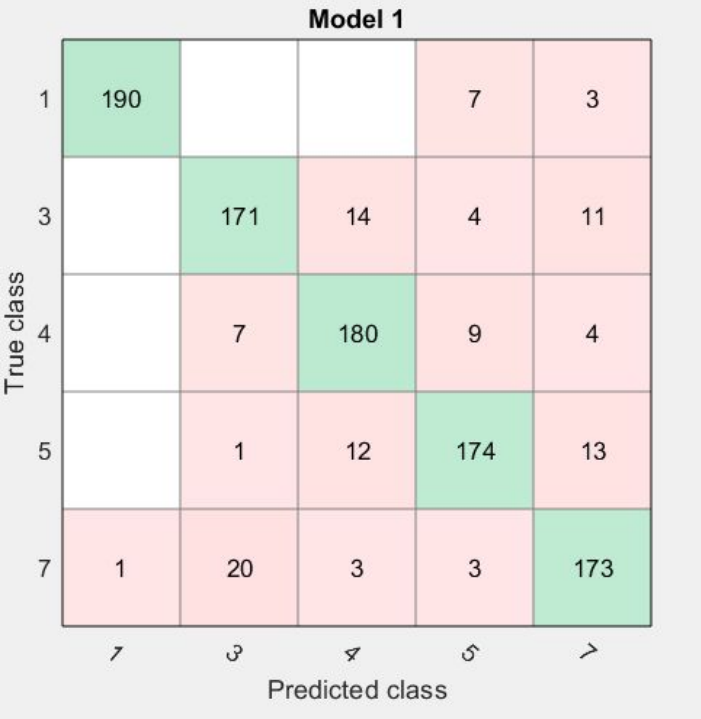
Accuracy = 93%

After Carrying out classification using different classifiers it was evident that Bagged Trees Classifier gave the best accuracy at about 93%. In Bagged Trees/ Random Forest we create many decision trees. We send the feature data to each tree and get the predicted class. We chose the overall predicted class by vote.

Confusion Matrix:



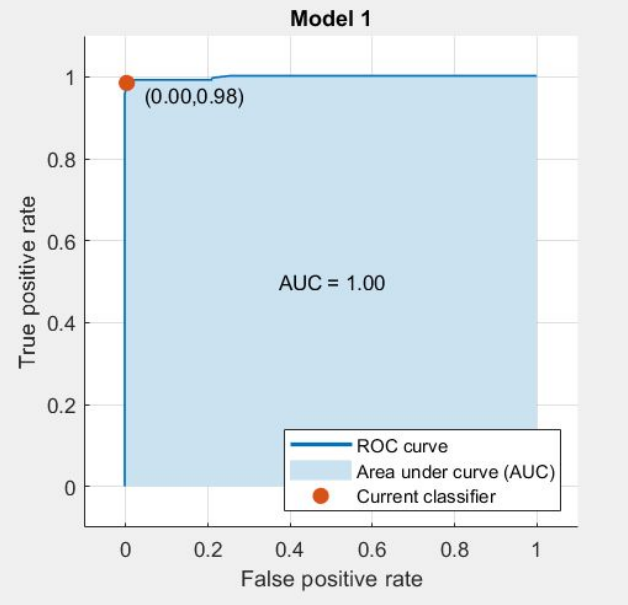
We observe that the no. of samples for each class vary in sizes which can affect the classification. So we take another set of data where the number of occurrences of each event is equal.



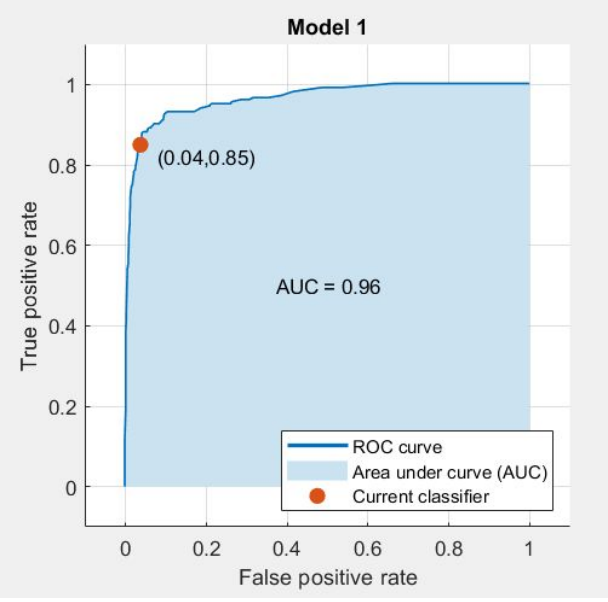
Accuracy for Each Class:

ROC Curves:

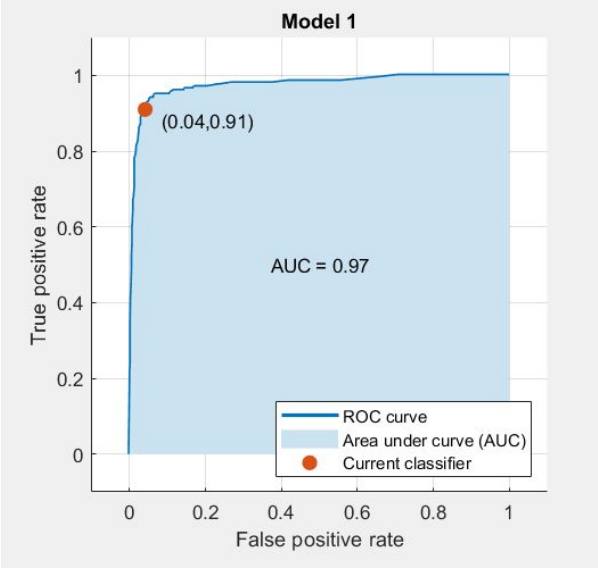
Class 1:



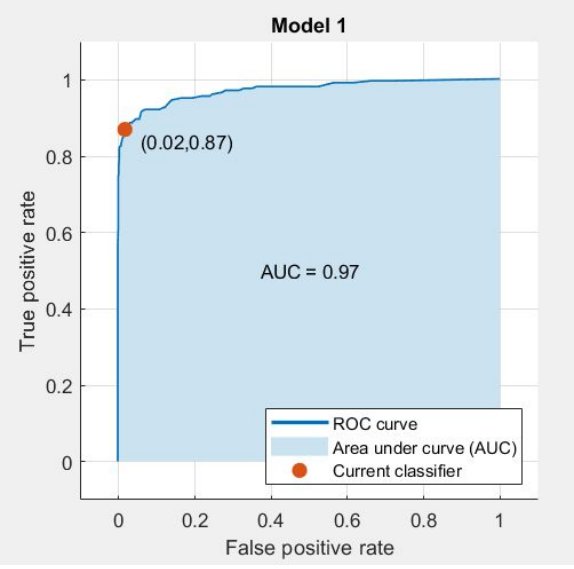
Class 2:



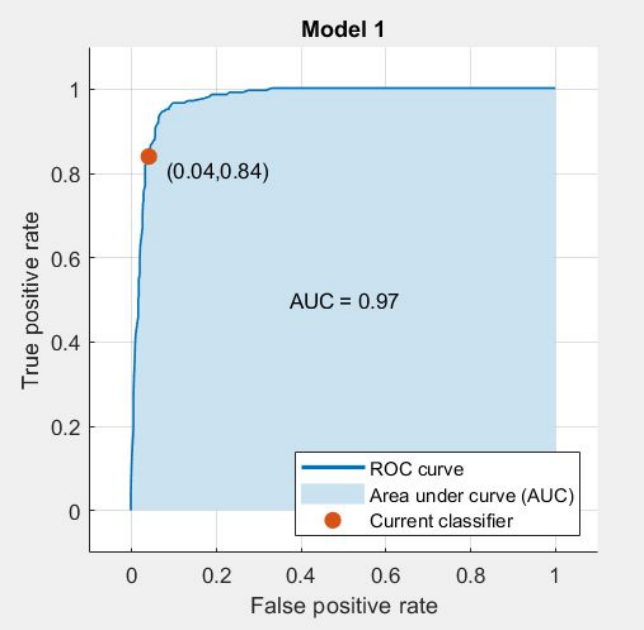
Class 3:



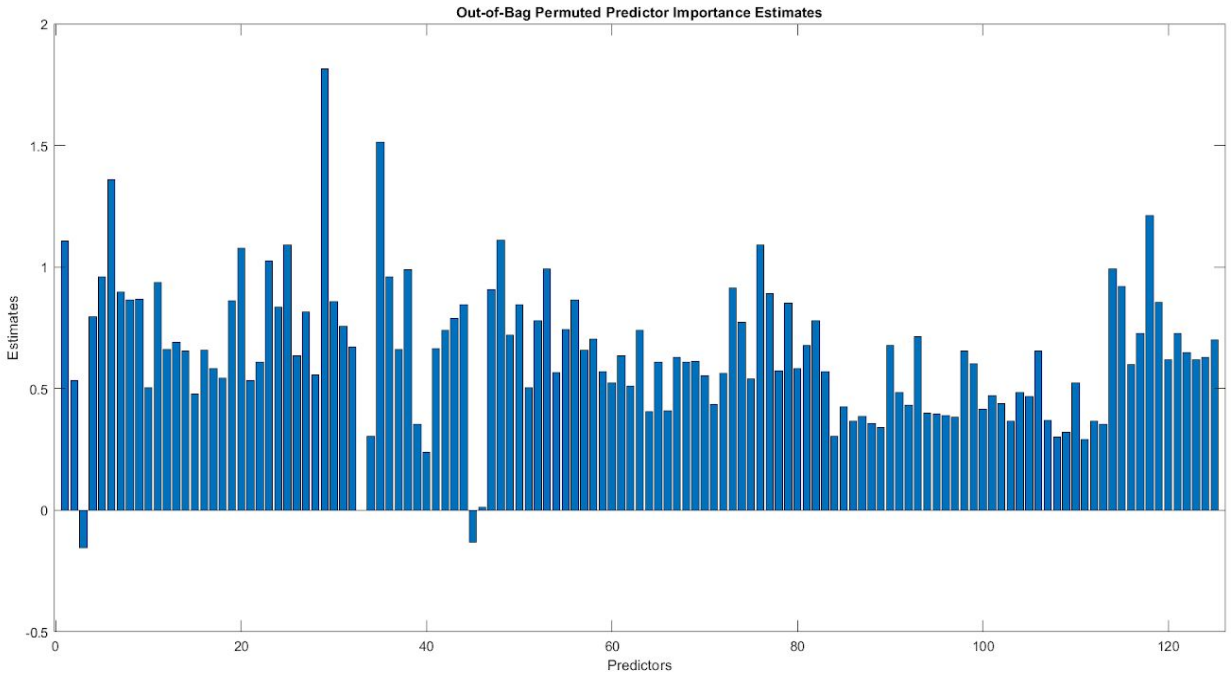
Class 4:



Class 5:



In Ensemble Classifier we can also calculate the importance of particular features. For this we use Out-of-Bag Permuted Predictor Importance Estimates to measure the importance of the different features that we used for the classification.



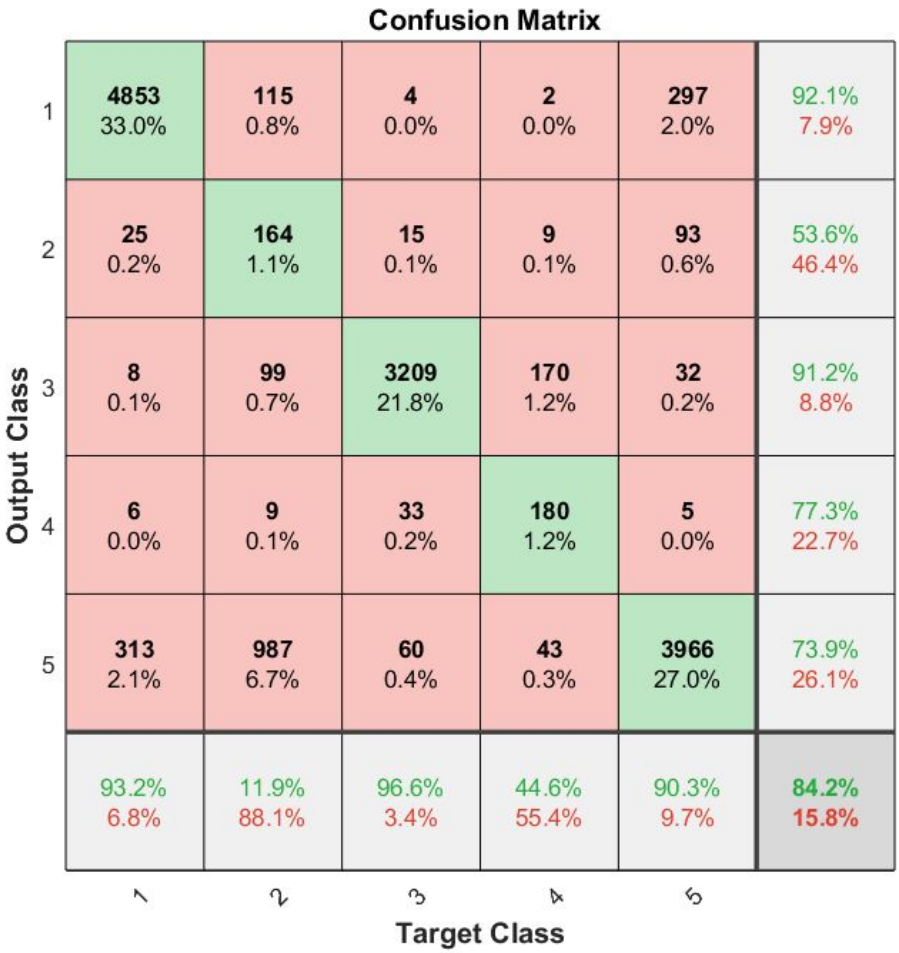
From The data we were able to conclude that Velocity and Features in z axis were some of the more prominent features that were used for classification.

On classifying with the smaller feature vector using only important features like time series and velocity we still get an accuracy of 92.3%.

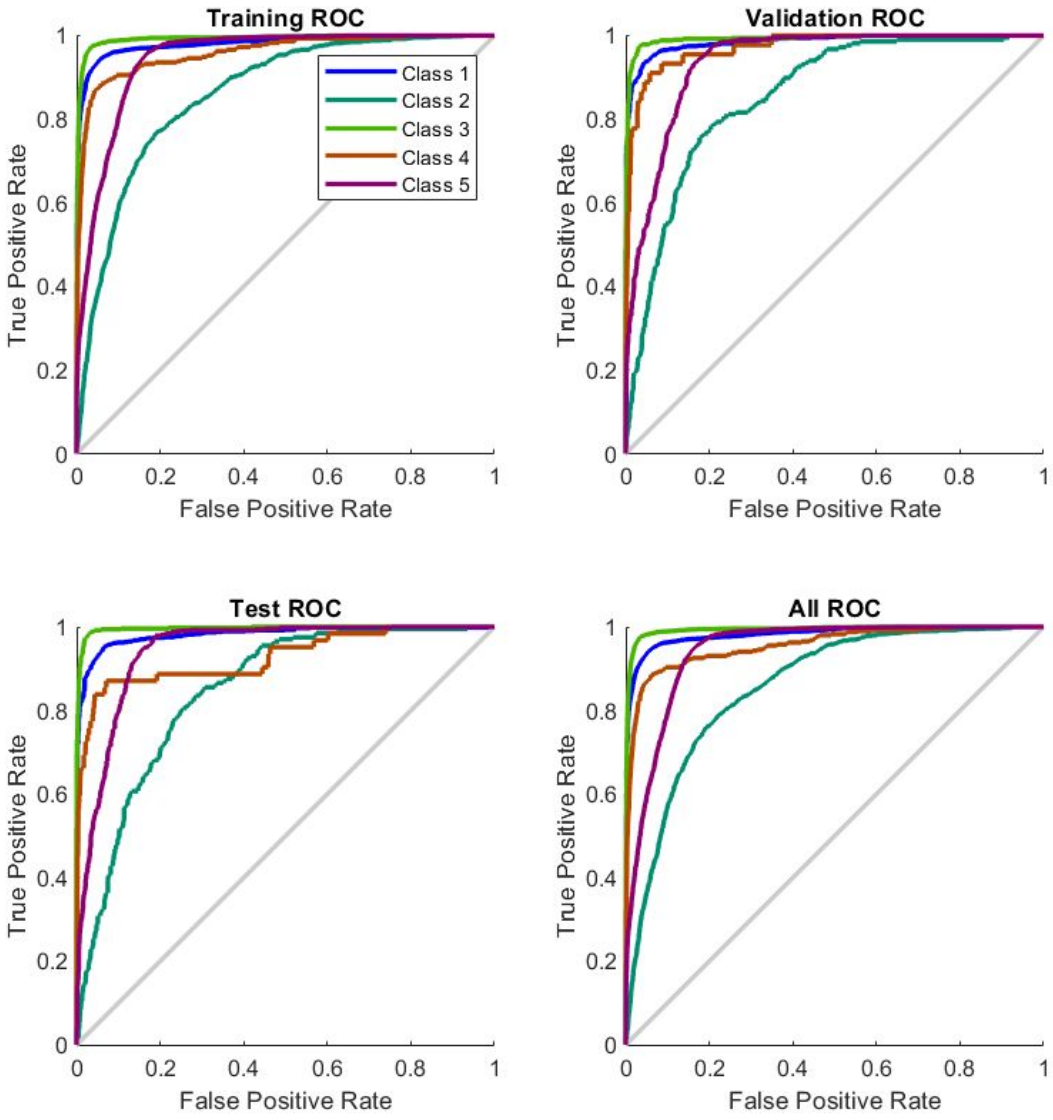
Neural Networks

For Neural Networks we used 30 hiddens layers and checked the performance

Confusion Matrix:



ROC Curves:


Training ROC
Validation ROC
Test ROC
All ROC
True Positive Rate
False Positive Rate
Class 1
Class 2
Class 3
Class 4
Class 5
0
0.2
0.4
0.6
0.8
1

Conclusions:

After testing out various Classification Techniques we came to the conclusion that Random Forest/Bagged Trees is the best classifier for Movement based detection. Also most of the important features related to motion detection are related to the Z axis. Velocity is another one of the more important features which is useful for classification.