Project 2 Final Report

Josh Mandzak, Coby White, Tanner Thornton

Data Collection

Data collection for this process took place over the course of multiple weeks, collected by two different applications: AndroSensor and Sensor Logger. Multiple hours of walking were collected, equal samples of walking up and down stairs were collected, and fifty instances of standing up and squats were collected. The squats and instances of standing up were each collected in their own respective single session rather than having a single recording only contain one squat or one instance of standing.

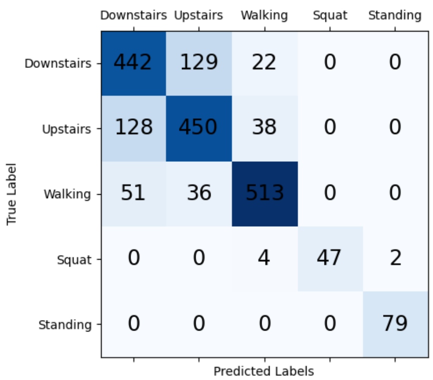
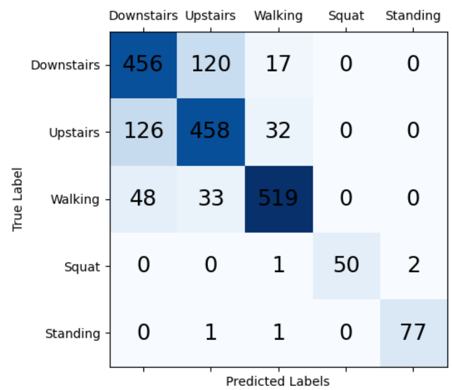
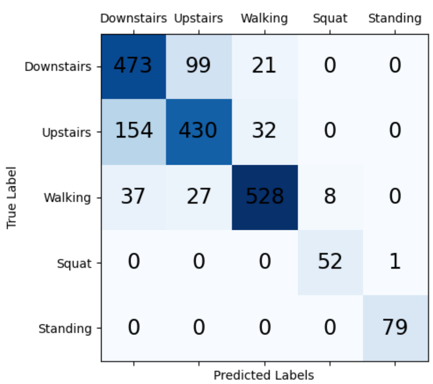
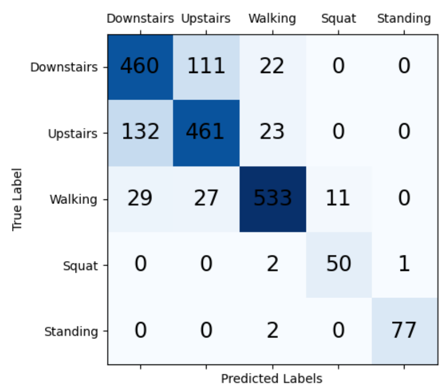
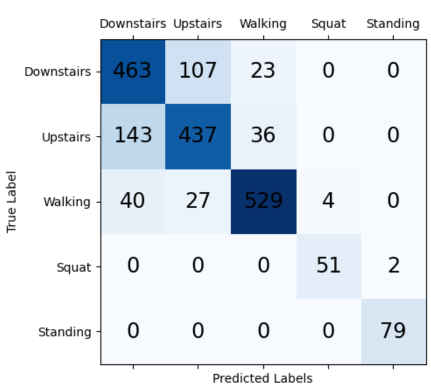
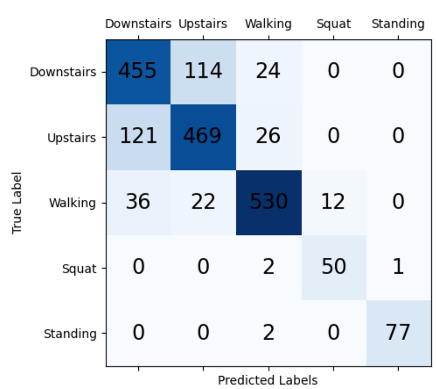
In order to split each activity into individual samples, it was decided to use windowing to split activities into samples. The specific windowing technique used eight time steps (half second change in time resulting in a total of four seconds per sample) with a fifty percent overlap of four time steps. This resulted in a significantly unbalanced dataset, with 593 downstairs samples, 616 upstairs samples, 7,138 walking samples, 53 squat samples, and 79 standing samples. Initially, nothing was done about this. However, after seeing poor model performance specifically because of walking, walking was then downsampled to only contain 600 samples. This improved model accuracy significantly and provided for a more even dataset. The decision was made to downsample because walking was the only class that contained significantly more samples than any other class, with a difference numbering in the thousands. It was not unnoticed that the standing and squatting samples were noticeably less than the other samples, however, the model performed extremely well identifying these samples and so it was decided that no further work needed to be done to balance the dataset for these classes.

Feature Engineering

Feature engineering was conducted to add features to the dataset, followed by a comparative analysis of how models did when using these features vs just using raw measurements. The specific features added to the dataset included median absolute deviation, mean, standard deviation, minimum, maximum, and sum of the squares divided by the number of samples. All of these features were calculated by taking the measurements of a specific feature (such as accelerometer in the x direction) for all time steps in the sample. This means that when mean is discussed, it is the mean of all eight time steps for a specific measurement, whether that be accelerometer x, y, or z data, or gyroscope or another unit of measurement altogether.

Results

In order to test the accuracy of a machine learning model on this dataset, two separate algorithms were tested. The first of these is a Random Forest model from Sci-Kit learn, while the second model is an XGBoost Classifier from the XGBoost library. Both of these models are well known and frequently used in machine learning competitions, hence why they were chosen for this classification task. Tests were conducted to measure the accuracy of the model under three different data circumstances. These circumstances are only using the actual measurements as inputs for a sample’s feature vector, only using the features described above for a sample’s feature vector, and using both for the sample’s feature vector. Ten-fold cross validation was used to evaluate each model. Since the dataset was still slightly unbalanced, a stratified k-fold approach was used to ensure all classes were present in training and validation for each fold. The results can be seen in the confusion matrices and accuracy matrices below. The accuracies shown in the chart are averages of all ten folds. Scores in each fold for all models ranged from the 70s to the 90s, with every model having at least one fold in the 90s and multiple folds in the 80s.

<table>
<tr><th></th><th>Random Forest</th><th>XGBoost</th></tr>
<tr><td>Raw Measurements Only</td><td>
</td><td>
</td></tr>
<tr><td>Features Only</td><td>
</td><td>
</td></tr>
<tr><td>Measurements + Features</td><td>
</td><td>
</td></tr>
</table>

	Random Forest	XGBoost
Raw Measurements Only	78.88%	80.37%
Features Only	80.48%	81.45%
Measurements + Features	80.32%	81.45%

As can be seen from the charts above, the best model is a two-way tie between the features only XGBoost and the measurements + features XGBoost models. More importantly, these charts show that the feature engineering conducted improved performance in all cases, albeit the increase was not a large amount. More work could be done to find more impactful features, but since the final project is overlapping this project it was decided to move on to work on it instead of fine-tuning these models further.