# A Deep Learning Approach for Human Activity Recognition

Susana Benavidez, Derek McCreight {sbenavid, dmccreig}@stanford.edu; *Mentor: Advay Pal*

## Introduction

- Wearable devices and smartphones are ubiquitous, and many of these devices contain Inertial Measurement Units (IMUs) such as gyroscopes and accelerometers
- This data can be used to perform human activity recognition (HAR) to recognize the motion characteristics of a smartphone/watch user
- We use both Convolutional Neural Networks and LSTMs to classify 18 unique activities ranging from eating chips to dribbling a ball from labelled, time-series data

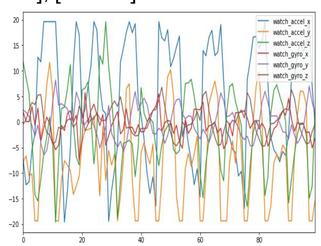
## Dataset and Baseline

- Our dataset consists of approximately **47,000** labelled, tri-axial sensor samples from the phone and watch IMUs from the **WISDM** [1] dataset
- The gyroscope and accelerometer tri-axial data was sampled at a rate of **20 Hz**
- Weiss used five distinct algorithms: Random Forests, J48 decision trees, B3 instance-based learning, Naive Bayes, and multi-layer perceptron. Using these algorithms, Weiss was able to obtain an overall accuracy rate of 25.3% with the phone accelerometer data, and 64% accuracy with the watch accelerometer data.

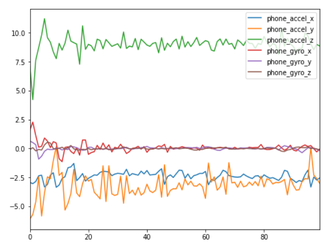
## Example Accelerometer & Gyroscope Data

```
1600,A,252207666810782,-0.36476135,8.793503,1.0550842;
```

Participant ID (beginning at 1600)], [timestamp(unix-based)], [x value], [y value], [z value] as shown above




**Accel/Gyroscope data for jogging**




**Accel/Gyroscope data for eating a sandwich**

## LSTM & CNN

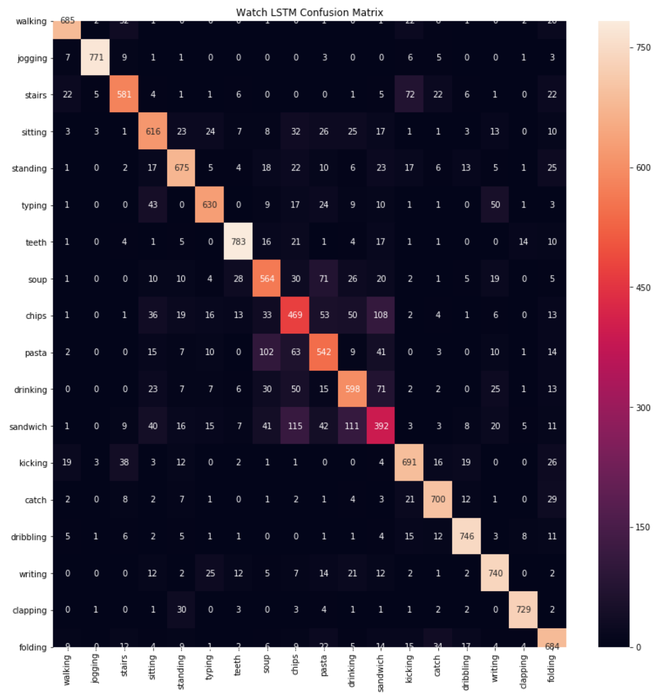
- Our Convolutional Neural Network consists of 4 Conv1D layers, accompanied with Batch Normalization and Max Pooling layers and lastly a flatten and two dense layers
- For the LSTM structure, we decided to use a stacked-LSTM structure. Specifically, we have two LSTM layers with 64 memory cells each, followed by a Dense Layer
- We chose to stack multiple recurrent states with multiple memory cells because it allows the model to determine more complex abstractions from the input data

[1] G. M. Weiss. Wisdm smartphone and smartwatch activity and biometrics dataset. UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset, Sep 2019

## Network Architecture & Analysis

| Layer (type) | Output Shape | Param # |
|---|---|---|
| lstm_5 (LSTM) | (None, 50, 64) | 18176 |
| lstm_6 (LSTM) | (None, 64) | 33024 |
| dense_3 (Dense) | (None, 18) | 1170 |

| Layer (type) | Output Shape | Param # |
|---|---|---|
| conv1d_1 (Conv1D) | (None, 98, 8) | 152 |
| batch_normalization_1 (Batch | (None, 98, 8) | 32 |
| max_pooling1d_1 (MaxPooling1 | (None, 49, 8) | 0 |
| conv1d_2 (Conv1D) | (None, 47, 16) | 400 |
| batch_normalization_2 (Batch | (None, 47, 16) | 64 |
| max_pooling1d_2 (MaxPooling1 | (None, 23, 16) | 0 |
| conv1d_3 (Conv1D) | (None, 21, 32) | 1568 |
| batch_normalization_3 (Batch | (None, 21, 32) | 128 |
| max_pooling1d_3 (MaxPooling1 | (None, 10, 32) | 0 |
| conv1d_4 (Conv1D) | (None, 8, 64) | 6208 |
| batch_normalization_4 (Batch | (None, 8, 64) | 256 |
| max_pooling1d_4 (MaxPooling1 | (None, 4, 64) | 0 |
| flatten_1 (Flatten) | (None, 256) | 0 |
| dense_1 (Dense) | (None, 128) | 32896 |
| dense_2 (Dense) | (None, 18) | 2322 |




## Evaluation

- The resulting confusion matrix shows that the models on both devices struggle to differentiate between activities that require similar hand movements such as eating chips versus eating soup.
- Comparing LSTM and CNN on both watch and phone data sets, we see that the LSTM performs better than the CNN with 79% vs. 72% accuracy and 74% vs. 50% for each respective dataset.

## Results

Watch Results with LSTM for Test Set

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| catch | 0.88 | 0.85 | 0.87 | 820 |
| chips | 0.57 | 0.56 | 0.56 | 841 |
| clapping | 0.93 | 0.95 | 0.94 | 767 |
| dribbling | 0.91 | 0.89 | 0.90 | 835 |
| drinking | 0.70 | 0.69 | 0.69 | 871 |
| folding | 0.80 | 0.76 | 0.78 | 903 |
| jogging | 0.96 | 0.98 | 0.97 | 786 |
| kicking | 0.83 | 0.79 | 0.81 | 874 |
| pasta | 0.66 | 0.65 | 0.66 | 830 |
| sandwich | 0.47 | 0.53 | 0.50 | 743 |
| sitting | 0.76 | 0.74 | 0.75 | 831 |
| soup | 0.71 | 0.68 | 0.69 | 835 |
| stairs | 0.78 | 0.83 | 0.80 | 703 |
| standing | 0.79 | 0.81 | 0.80 | 829 |
| teeth | 0.89 | 0.90 | 0.89 | 874 |
| typing | 0.79 | 0.85 | 0.82 | 740 |
| walking | 0.89 | 0.90 | 0.89 | 760 |
| writing | 0.86 | 0.82 | 0.84 | 897 |
| accuracy | | | 0.79 | 14739 |
| macro avg | 0.79 | 0.79 | 0.79 | 14739 |
| weighted avg | 0.79 | 0.79 | 0.79 | 14739 |

Watch Results with CNN for Test Set

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| catch | 0.80 | 0.88 | 0.84 | 319 |
| chips | 0.41 | 0.55 | 0.47 | 310 |
| clapping | 0.90 | 0.88 | 0.89 | 429 |
| dribbling | 0.85 | 0.88 | 0.87 | 393 |
| drinking | 0.50 | 0.63 | 0.56 | 363 |
| folding | 0.84 | 0.69 | 0.76 | 485 |
| jogging | 0.97 | 0.98 | 0.97 | 416 |
| kicking | 0.76 | 0.73 | 0.75 | 429 |
| pasta | 0.63 | 0.56 | 0.59 | 434 |
| sandwich | 0.36 | 0.36 | 0.36 | 400 |
| sitting | 0.65 | 0.57 | 0.61 | 495 |
| soup | 0.60 | 0.70 | 0.65 | 361 |
| stairs | 0.82 | 0.74 | 0.78 | 375 |
| standing | 0.73 | 0.77 | 0.75 | 361 |
| teeth | 0.82 | 0.88 | 0.85 | 369 |
| typing | 0.82 | 0.67 | 0.74 | 520 |
| walking | 0.84 | 0.92 | 0.88 | 371 |
| writing | 0.75 | 0.69 | 0.71 | 449 |
| accuracy | | | 0.72 | 7279 |
| macro avg | 0.73 | 0.73 | 0.72 | 7279 |
| weighted avg | 0.73 | 0.72 | 0.72 | 7279 |

Phone Results with LSTM for Test Set

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| catch | 0.66 | 0.68 | 0.67 | 534 |
| chips | 0.61 | 0.74 | 0.67 | 427 |
| clapping | 0.66 | 0.71 | 0.69 | 501 |
| dribbling | 0.77 | 0.68 | 0.72 | 610 |
| drinking | 0.68 | 0.58 | 0.62 | 628 |
| folding | 0.64 | 0.69 | 0.67 | 493 |
| jogging | 0.98 | 0.97 | 0.97 | 529 |
| kicking | 0.72 | 0.72 | 0.72 | 557 |
| pasta | 0.60 | 0.72 | 0.65 | 386 |
| sandwich | 0.68 | 0.62 | 0.65 | 552 |
| sitting | 0.76 | 0.77 | 0.77 | 503 |
| soup | 0.66 | 0.65 | 0.66 | 532 |
| stairs | 0.86 | 0.92 | 0.89 | 501 |
| standing | 0.81 | 0.73 | 0.77 | 566 |
| teeth | 0.71 | 0.70 | 0.71 | 503 |
| typing | 0.81 | 0.75 | 0.78 | 531 |
| walking | 0.95 | 0.94 | 0.95 | 575 |
| writing | 0.76 | 0.81 | 0.78 | 454 |
| accuracy | | | 0.74 | 9382 |
| macro avg | 0.74 | 0.74 | 0.74 | 9382 |
| weighted avg | 0.74 | 0.74 | 0.74 | 9382 |

Phone Results with CNN for Test Set

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| catch | 0.57 | 0.48 | 0.52 | 644 |
| chips | 0.26 | 0.37 | 0.31 | 372 |
| clapping | 0.47 | 0.47 | 0.47 | 545 |
| dribbling | 0.54 | 0.62 | 0.58 | 462 |
| drinking | 0.16 | 0.42 | 0.23 | 208 |
| folding | 0.43 | 0.53 | 0.47 | 427 |
| jogging | 0.92 | 0.95 | 0.93 | 509 |
| kicking | 0.50 | 0.53 | 0.52 | 532 |
| pasta | 0.32 | 0.37 | 0.34 | 392 |
| sandwich | 0.30 | 0.32 | 0.31 | 473 |
| sitting | 0.47 | 0.40 | 0.43 | 606 |
| soup | 0.44 | 0.28 | 0.34 | 829 |
| stairs | 0.73 | 0.64 | 0.68 | 611 |
| standing | 0.70 | 0.54 | 0.61 | 674 |
| teeth | 0.47 | 0.46 | 0.47 | 511 |
| typing | 0.41 | 0.46 | 0.43 | 438 |
| walking | 0.84 | 0.80 | 0.82 | 594 |
| writing | 0.48 | 0.42 | 0.45 | 555 |
| accuracy | | | 0.50 | 9382 |
| macro avg | 0.50 | 0.50 | 0.50 | 9382 |
| weighted avg | 0.53 | 0.50 | 0.51 | 9382 |

## Conclusion & Future Work

- We will use the expanded WISDM dataset that will be released later this year which includes more activities and a larger set of participants.
- We believe that with further fine-tuning of hyper parameters we can continue to improve the prediction accuracy of our approach.
- We will investigate creating unique models for each study participant as previous research has shown this approach to be more accurate