# HUMAN ACTIVITY RECOGNITION

**Gauri Dilip Rasane**
Institute of Signal Processing and Systems Theory
University of Stuttgart
Stuttgart, 70569
st164260@stud.uni-stuttgart.de

**Kaiwalya Kalyan Belsare**
Institute of Signal Processing and Systems Theory
University of Stuttgart
Stuttgart, 70569
st164262@stud.uni-stuttgart.de

February 11,2020

**ABSTRACT**

As the number of elderly people will grow rapidly in the upcoming years, 'aging in place' (which means living at home irrespective of age) is gaining importance [1]. In this paper, we put forth a Long Short-Term Memory (LSTM) neural network based human activity classifier. It is a 12-class classifier, which classifies 3 static activities (sitting, standing, laying), 3 dynamic activities (walking, walking downstairs, walking upstairs) and 6 postural activities (stand-to-sit, sit-to-stand, lie-to-stand, stand-to-lie, stand-to-sit, sit-to-stand). We have used publicly available Human Activity and Postural Transitions (HAPT) raw data set for this purpose which consists of data collected at 50Hz frequency from 3-axial accelerometer and gyroscope embedded in a smartphone [4]. Experimental results show overall test accuracy of 77.23% with especially good results obtained for laying, walking and sitting.

## 1 Introduction

For people to age at home, it is necessary to monitor their health status and detect emergency situation like a fall. To manage such emergency situations Human Activity Recognition (HAR) systems are employed [1]. Human-Robot Interaction (HRI) systems also use HAR that recognize the current human activity to be able to predict future events in the workflow.

In this paper, we focus on the significant part of 'aging in place' and HRI, which is activity recognition. It is the ability of the HAR classifier to recognize actions performed by user based on a set of observations recorded through sensors. HAPT dataset features data collected from sensors namely 3-axial gyroscope and accelerometer, which are embedded in a smartphone. This is effective as it relieves the elderly people from carrying any additional sensor devices.

In section 2 of this paper, we describe the HAPT dataset and discuss the data preprocessing techniques used to build the input data pipeline. In section 3 we focus on RNN and explain the need to use LSTM units in RNN layers. We also put forth the LSTM model architecture implemented for this classification task. The results obtained through evaluation metrics such as test accuracy and confusion matrix are discussed in Section 4, after which we conclude the paper in Section 5.

## 2 Input Pipeline

The HAPT dataset contains separate accelerometer and gyroscope readings for each user, with total number of users being 30. Two experiments have been conducted per user to record linear acceleration and angular velocity.

We split the dataset into train, test and validation sets, as per the ratio 70:20:10. Train dataset contains data for user 1 to user 21, test set contains data for user 22 till user 27. The remaining users 28 till 30 account for the validation dataset.

### 2.1 Data Visualization

Figure 1 and Figure 2 show sensor data for walking activity.

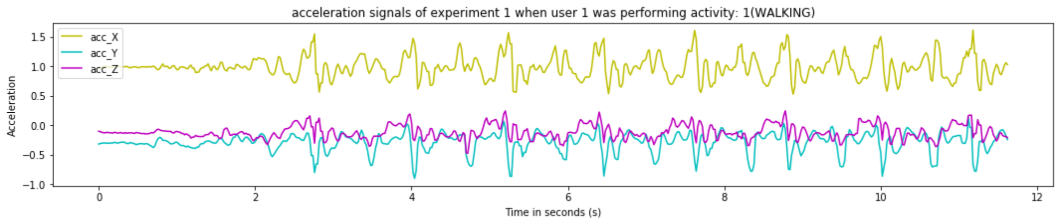


Figure 1: 3-Axial linear acceleration for walking activity.


Figure 2: 3-Axial angular velocity for walking activity.

### 2.2 Data Preprocessing

The time series sensor data is preprocessed by removing noisy rows from the data set and using Z-Score normalization technique. Also, there is a lot of unlabeled data in the data set. Instead of discarding this unlabeled data, we label it as 0 and use it during training. This is done to in order to avoid random initialization of weights in the hidden layers. Additionally, we one hot encode the activity labels. The train data distribution as per the activity is shown in Figure 3. It can be seen that train data set is imbalanced and has a lot of unlabelled data.

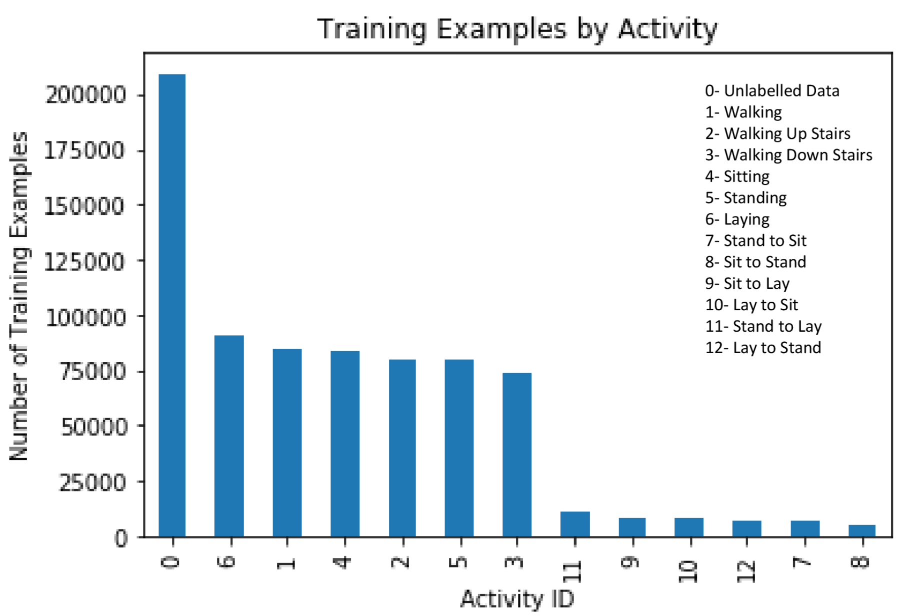


Figure 3: Training data distribution as per activity type.

We also perform window operation on the data set and group the time series data into windows of size 250 samples. A 50% overlap between the consecutive windows has been implemented as part of data pipeline.

## 3 Model Architecture

RNN is used to model sequence of data in which each sample is dependent on the previous samples. It has short term internal memory. It produces output, copies it and loops it back to the network. Thus, every neuron in RNN has 2 inputs: the present and the recent past. Simply put, RNN adds immediate past to the present. However, due to backpropagation of gradients through time in addition to backpropagation through layers, the problem of vanishing gradient occurs more profoundly in RNN.

Vanishing gradient problem of RNN, is resolved through the use of LSTM units in RNN layers. LSTM units extend memory of RNN and are thus able to remember inputs for longer duration. LSTM unit can read, write and delete from its memory just like memory of a computer. In LSTM unit, there are 3 gates: input gate, forget gate and output gate. Input gate determines whether to read new input, forget gate deletes information because it is unimportant and output gate decides if the new input should be allowed to impact the current output. Hence, an LSTM network has been used on the time series raw HAPT dataset.

A very simple LSTM network as shown in Figure 4 is used. Dropout layer has been added as a regularization technique. Input is presented in batches of size 16, window size of 250 and 6 channels (3-axial accelerometer and gyroscope). The output is obtained as batches of size 16 with window size of 250 and class probabilities for the 12 different activity types.

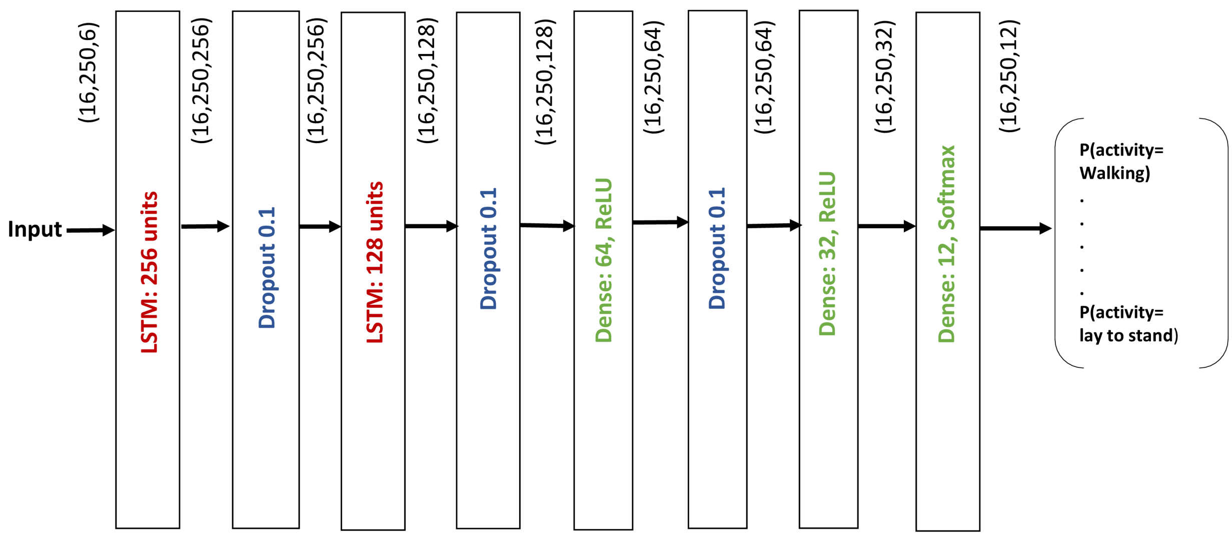


Figure 4: Model architecture

The model has been tuned for hyperparameters such as optimizer, number of neurons in the first LSTM layer and dropout rate using Tensorboard. Table 1 represents the accuracy obtained for different combinations of hyperparameters used. As is evident from Table 1, RMSProp optimizer, 256 LSTM units in the first recurrent layer and dropout rate of 0.1 were chosen as hyperparameter values.

| Optimizer | Number of LSTM units | Dropout | Accuracy |
|---|---|---|---|
| RMSProp | 128 | 0.1 | 74.27 |
| RMSProp | 128 | 0.2 | 74.34 |
| RMSProp | 256 | 0.1 | 78.09 |
| RMSProp | 256 | 0.2 | 75.83 |
| Adam | 128 | 0.1 | 73.78 |
| Adam | 128 | 0.2 | 74.44 |
| Adam | 256 | 0.1 | 75.12 |
| Adam | 256 | 0.2 | 58.01 |

Table 1: Tuning for hyperparameters: optimizer, number of LSTM units in first recurrent layer, dropout rate

# 4 Evaluation

Test accuracy of 77.23% has been achieved with the above model after training it for 20 epochs. The curves for training and validation accuracy and loss are shown in Figure 5.

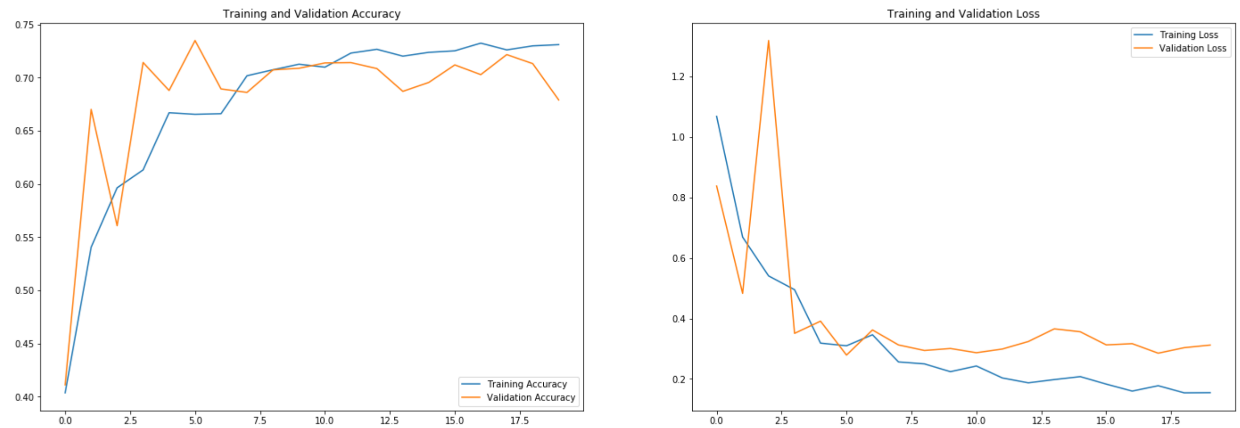


Figure 5: Accuracy and loss curves for train and validation datasets.

The model shows good accuracy as high as 95% for laying, 94% for walking and 91% for sitting. Activities such as walking downstairs/upstairs and some postural transitions also exhibit decent results. This can be clearly seen from the confusion matrix shown in Figure 6. Precision, Recall and F1-Score obtained for each class is tabulated in Table 2.

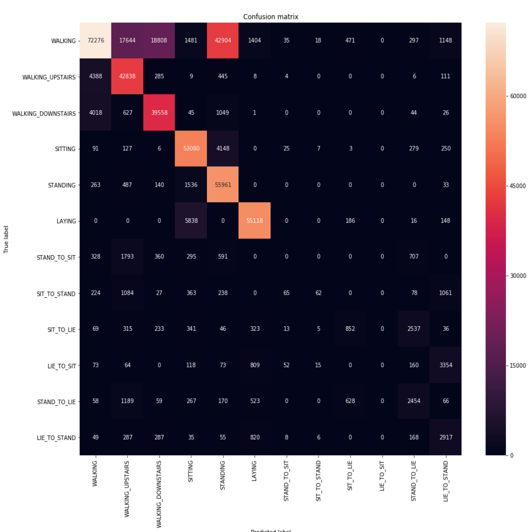


Figure 6: Confusion matrix obtained on test dataset.

| Activity | Precision | Recall | F1-Score |
|---|---|---|---|
| WALKING | 0.94 | 0.51 | 0.66 |
| WALKING UPSTAIRS | 0.68 | 0.97 | 0.80 |
| WALKING DOWNSTAIRS | 0.77 | 0.91 | 0.83 |
| SITTING | 0.91 | 0.92 | 0.92 |
| STANDING | 0.54 | 0.95 | 0.69 |
| LAYING | 0.95 | 0.97 | 0.96 |
| STAND TO SIT | 0.49 | 0.20 | 0.29 |
| SIT TO STAND | 0.78 | 0.63 | 0.69 |
| SIT TO LAY | 0.76 | 0.23 | 0.36 |
| LAY TO SIT | 0.38 | 0.31 | 0.34 |
| STAND TO LAY | 0.39 | 0.61 | 0.47 |
| LAY TO STAND | 0.44 | 0.50 | 0.46 |

Table 2: Precision, Recall and F1-Score for each activity

# 5 Conclusion

Human activity recognition has broad applications in medical research, health care monitoring and HRI [3]. In our project, we have obtained satisfactory results for identifying 12 types of human activities on time series HAPT dataset.

Test accuracy could be further improved through the implementation of Convolutional Recurrent Neural Network (CRNN) which takes advantages of both Convulational Neural Netowrk (CNN) and RNN [2]. Future work may also consider more activities like riding bike, driving and falling. Today, 'aging in place' is no longer a distant dream.

## References


[1] R.-A. Voicu, C. Dobre, L. Bajenaru, and R.-I. Ciobanu. Human Physical Activity Recognition Using Smartphone Sensors. In *Sensors*, vol. 19, no. 3, p. 458, Jan. 2019.

[2] Y. Zhang, N. Suda, L. Lai, and V. Chandra. Hello Edge: Keyword Spotting on Microcontrollers. In *arXiv*:1711.07128 [cs, eess], Feb. 2018.

[3] A. Rasekh, C.-A. Chen, and Y. Lu. Human Activity Recognition using Smartphone, 2014.

[4] Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set. https://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions


# Appendix

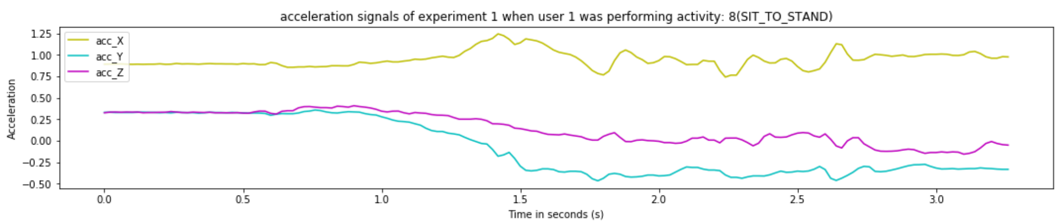


Figure 6: 3-Axial linear acceleration for sit-to-stand activity.

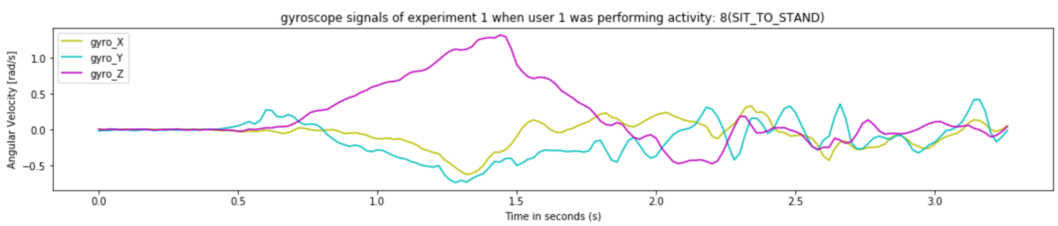


Figure 7: 3-Axial angular velocity for sit-to-stand activity.

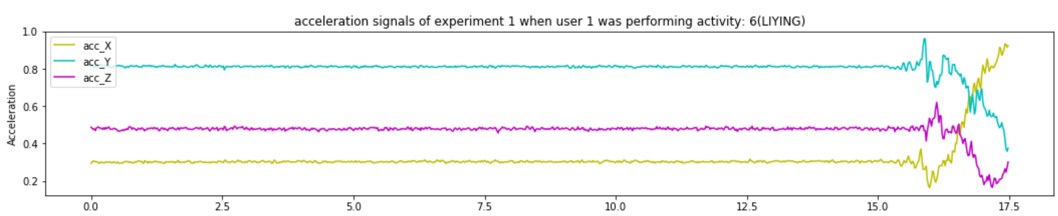


Figure 8: 3-Axial linear acceleration for laying activity.