# HUMAN ACTIVITY RECOGNITION

**Saiteja Malyala**
Department of INFOTECH
University of Stuttgart
st164488@stud.uni-stuttgart.de

**Sampath Garuda**
Department of INFOTECH
University of Stuttgart
st169945@stud.uni-stuttgart.de

February 16, 2021

**ABSTRACT**

Human Activity Recognition has become more vibrant and important field of research and application in the recent times. Along with the growing number of sensors used in day to day consumer electronic devices, the flexibility of leveraging sensor data to recognize different human activities has also increased rapidly. Human activity recognition can be now widely seen in sports, health and human well being applications. In Health care, monitoring vital parameters is quite essential as it enables health care professionals to have a deeper insight about the patient's health condition. Also in Industrial applications, Human-Robot Interaction (HRI) systems can predict the next state of the workflow by observing the current state of workers (humans). In this paper, we propose a Long Short-Term Memory (LSTM) Neural Network which can classify twelve different human activities. Which, include three static activities, three dynamic activities along with six postural transitions. The publicly available Human Activities and Postural Transitions Dataset (HAPT) [1] was used for this classification. This Dataset has tri-axial data of accelerometer and gyroscope from a smartphone, captured at 50Hz frequency. Our proposed Neural Network architecture shows an overall accuracy of 90.4% on the test dataset.

## 1 Introduction (Section by Saiteja Malyala)

Human Activity Recognition (HAR) enables us to analyze and interpret the human activities properly. This can be achieved using Video or Sensor based systems. Body-worn sensor based HAR systems have a major advantage over other methods as the activity monitoring zone is not limited to a confined region.HAR systems can be employed to monitor human activities, detect a fall or abnormal change in parameters like blood pressure, heart rate and pulse rate.

This project is about how Deep Neural Networks can be employed to detect various human activities. The following sections describe the Data pre-processing methods (Section 2), Model architecture (Section 3), Evaluation metrics (Section 4) and Conclusion (Section 5).

## 2 Input Pipeline (Section by Saiteja Malyala and Sampath Garuda)

The HAPT dataset consists of accelerometer and gyroscope data separately for 30 users. Each user performed certain activities as a part of the experiment. The activities performed were Standing, Sitting, Walking, Walking Upstairs, Walking Downstairs, Laying, Stand to Sit, Sit to Stand, Lie to Sit, Sit to Lie, Lie to Stand, and Stand to Lie.

The entire DataSet had been split into Train, Validation, and Test sets. We have used the raw data files instead of the pre-processed data. Users 1 to 21 account for the training set, users 22 to 27 for the test set, and remaining users 28 to 30 are used for the validation set.

### 2.1 Data Preprocessing

**Noise Removal and One hot encoding:** In practice, the last and first five seconds of raw data from each experiment are removed to ensure high quality sensor data. The respective labels of denoised input data are one-hot encoded before feeding to the neural network. In Fig.1. activity samples of User 01 are plotted as a bar graph. From this figure, we can observe that there is quite a significant amount of data which is **Unlabelled**. The neural network should not update its weights and biases with Unlabelled data. So possible solutions would be removing such data or encoding it with all zero labels. We removed this unlabelled data.

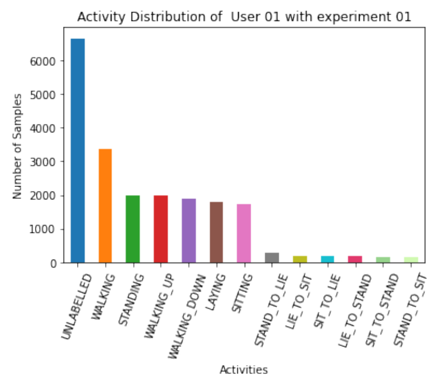


Figure 1:Distribution of Samples of User 1 with Experiment 1

**Input normalization:** Z-score normalization technique was employed to remove large offsets and strongly differing variances in-order to obtain better training performance. When dealing with multi-variate data, normalization gives a better performance. Normalization removes user biases and offsets in data, due to device placement.

**Sliding Window Technique:** As the model accepts input sequences of fixed length, data is windowed by a fixed length followed by a window shift. A window length of 250 samples with a shift/overlap by 125 samples was chosen. Window shifting is applied only on training data. Window length and window shift are hyper parameters, and can be considered for optimizing model performance.

## 3 Model Architecture (Section by Saiteja Malyala and Sampath Garuda)

The data points in HAPT data set are correlated in time hence, we require a model that can learn from time series data, for which a straight forwards approach is to use a Recurrent Neural Network(RNN) based model or its variants such as Gated Recurrent Unit(GRU) or Long Short-term Memory (LSTM) based networks. All RNNs have feedback loops in recurrent layer, this helps them maintain information in memory over time. But RNN has few shortcomings, Majorly, RNN donot remember input sequences of longer duration and it faces the issue of vanishing gradients.

These disadvantages of RNN can be overcome by using LSTM based networks. Unlike RNN, LSTM networks are a type of RNN that includes a memory cell that can maintain information over period of time. LSTMs can remember the long-term dependencies in the input signal effectively with help of gates. LSTM cell has three gates.

- **Forget Gate:** This gate determines whether the information is necessary or not.
- **Input Gate:** This gate determines what kind of new information should be updated to cell state.
- **Output Gate:**This gate decides what is the next hidden state.

A simple LSTM architecture (Fig.2.) comprising of two LSTM layers and two Dense layers are used to perform activity classification task. Batch Normalization and Dropout layers are employed as preventive over fitting measure to improve regularization. Six channel (tri axial accelerometer and tri axial gyrometer) windowed input along with labels were fed to the model with a batch size of 128.

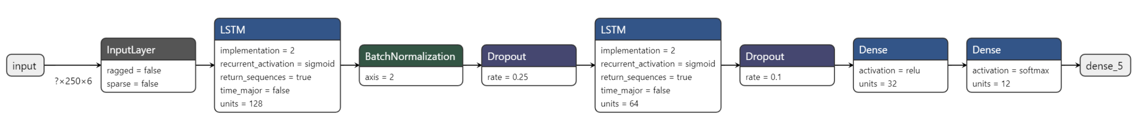


Figure 2 Model Architecture

The model was tuned to find the optimum Hyper-parameters. Keras Hyperband was used for tuning the hyper parameter set (LSTM Neurons in layers and the optimizer). The results obtained from various combination of hyper parameters could be found in Table 1.

Table 1: Hyper Parameters Tuning

| Hyperparameter Tuning | | | |
|---|---|---|---|
| LSTM_1 Neurons | LSTM_2 Neurons | Optimizer | Validation_Accuracy |
| **128** | **64** | **RMSProp** | **83.1%** |
| 112 | 16 | RMSProp | 81.4% |
| 96 | 48 | RMSProp | 80.8% |
| 80 | 64 | RMSProp | 80.1% |
| 128 | 80 | Adam | 81.1 |
| 112 | 32 | Adam | 80.3 |
| 128 | 64 | Adam | 79.1 |
| 80 | 80 | Adam | 77.1 |

Table 2: Performance on Test dataset

| Best Parameters Evaluation on Test DataSet | | | | |
|---|---|---|---|---|
| Optimizer | LSTM_1 Neurons | LSTM_2 Neurons | Epochs | Test_Accuracy |
| RMSProp | 128 | 64 | 50 | **90.38%** |

## 4 Evaluation Metrics(Section by Saiteja Malyala and Sampath Garuda)

Test accuracy of 90.3% is obtained after training the model on the best hyperparameters set (Table 2). Subplots in Fig.3. include visualization of raw accelerometer and gyroscope values overlay with respective ground truths, ground truths, predicted output, and the colormap for each activity. Performance of the model is evaluated using the following metrics: precision, recall, f1 score (Fig.4.). From confusion matrix (Fig.5.) we can infer that classification accuracy of postural transitions is relatively low compared to static and dynamic activities and the reason being lower volume of data samples from these activities during training.

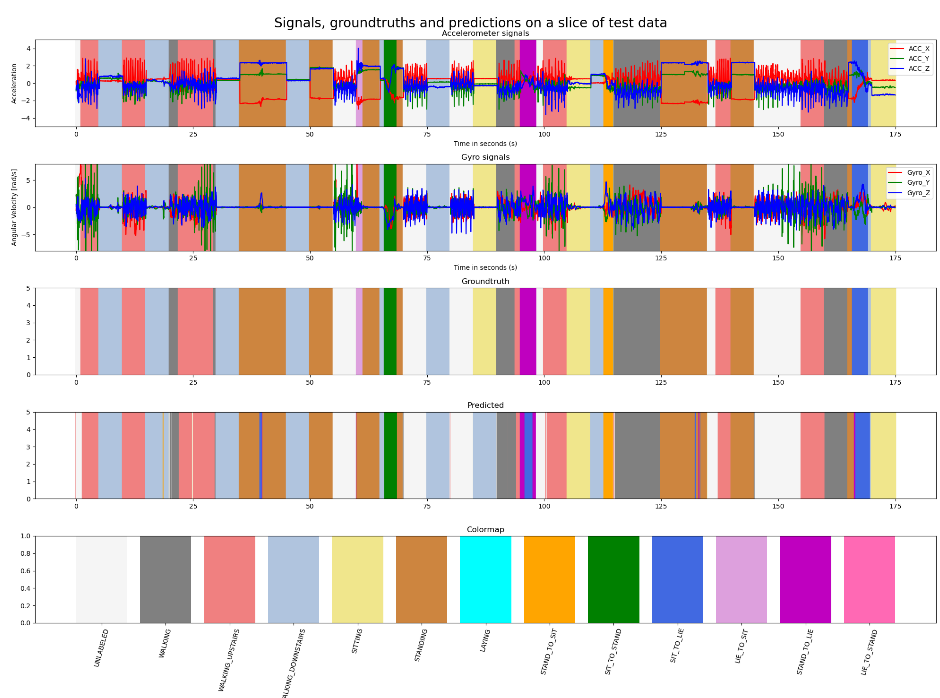


Figure 3: Visualization of a segment of Test DataSet with Predicted and True labels

| | precision | recall | f1-score |
|---|---|---|---|
| WALKING | 0.88 | 0.97 | 0.92 |
| WALKING_UPSTAIRS | 0.98 | 0.84 | 0.90 |
| WALKING_DOWNSTAIRS | 0.95 | 0.92 | 0.93 |
| SITTING | 0.87 | 0.99 | 0.93 |
| STANDING | 0.94 | 0.88 | 0.91 |
| LAYING | 0.97 | 0.98 | 0.97 |
| STAND_TO_SIT | 0.82 | 0.46 | 0.59 |
| SIT_TO_STAND | 0.85 | 0.66 | 0.74 |
| SIT_TO_LIE | 0.69 | 0.69 | 0.69 |
| LIE_TO_SIT | 0.64 | 0.56 | 0.60 |
| STAND_TO_LIE | 0.44 | 0.62 | 0.51 |
| LIE_TO_STAND | 0.54 | 0.53 | 0.53 |
| Accuracy | 0.90 | 0.90 | 0.90 |
| Macro avg | 0.80 | 0.76 | 0.77 |
| Weighted avg | 0.90 | 0.90 | 0.90 |

Figure 4: Classification Report

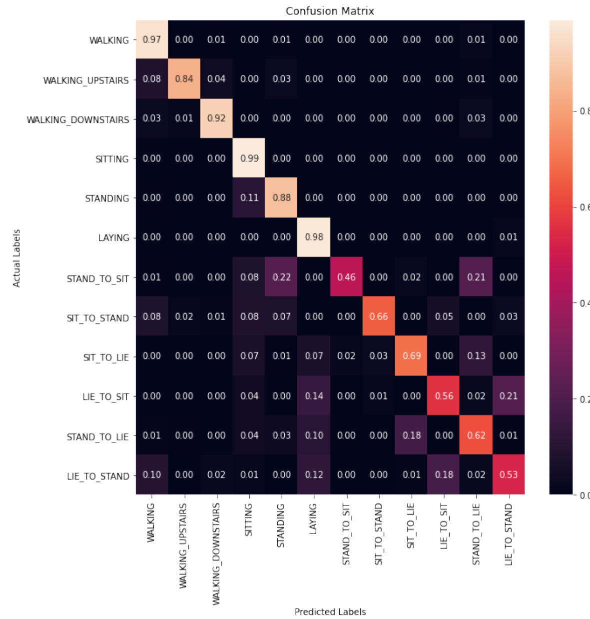


Figure 5: Confusion Matrix

## 5 Conclusion (Section by Sampath Garuda)

Human Activity Recognition owing to its numerous application is always an open research topic. Concerning our project, among all the 12 human activities of the HAPT data set dynamic and static activities were classified with an average accuracy of 93%. Improvements on test accuracy could be made in several stages by further employing different data preprocessing techniques like Spectrogram [2], improved model architecture with a combination of CNN (Convolutional Neural Network) and LSTM [3]. Thus the key to successful human activity recognition is multi-staged. This includes effective modeling, discriminating representation, accurate analysis of the features of the recorded data, with real-time detection results [4].

## References

[1] Dua, Dheeru and Graff, Casey archive.ics.uci.edu/ml/datasets/smartphone based recognition of human activities and postural transitions

[2] Xiaochen Zheng,Meiqing Wang,Meiqing Wang. publication/326164998 : Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0.

[3] Ronald Mutegeki and Dong Seog Han publication/340698806 A CNN-LSTM Approach to Human Activity Recognition.

[4] Beddiar Djamila Romaissa , Brahim Nini and Mohammad Sabokrou publication/343679855 Vision-based human activity recognition a survey..