

Human Activity Recognition Using Smartphone Sensor Data: Analysis and Classification using Azure ML and Power BI

Séamus Knightly
IMECEI
University Of Galway
Galway, Ireland
s.knightly1@universityofgalway.ie

Seán Kelly
IMECEI
University Of Galway
Galway, Ireland
s.kelly178@universityofgalway.ie

Abstract—blah blah

Index Terms—IoT, Remote Patient Monitoring, Accelerometer, Gyroscope, Classification, Assisted Living

I. INTRODUCTION

Human Activity Recognition (HAR) is an important area in the domain of machine learning and medical computing. The focus is the identification and classification of human physical activities based on provided sensor data. The data can be collected by a wide range of devices, from stationary room sensing video or radar that can monitor multiple subjects, to wearable or mobile devices that can monitor individual subjects. HAR has become an important component in applications such as remote patient monitoring and assisted living.

There are a number of large, labelled datasets available for human activity recognition. For this project, the *Human Activity Recognition Using Smartphones* dataset [1] was chosen. This dataset contains sensor data collected from smartphone accelerometers and gyroscopes as participants performed activities such as walking, sitting, and standing. The observations are represented by multiple time and frequency features extracted from the motion signals. Sensor signals from the device's accelerometer and gyroscope were recorded at 50 Hz and processed into 561 time and frequency features.

This dataset was chosen over other datasets, such as HARTH dataset [2] [3] for its simplicity, requiring only 1 worn sensor. And for the simpler set of classifications, which correspond better with scenarios that would arise in a care facility or other healthcare setting.

This paper aims to analyse and classify human activities using the UCI HAR dataset by developing and evaluating multiple machine learning models in Azure Machine Learning Studio. Power BI is employed to visualize data characteristics, feature distributions, and model performance metrics. The results provide insights into the role of sensor based data and machine learning in Human Activity Recognition.

The introductory paper [4] demonstrated that a multiclass Support Vector Machine (SVM) model could achieve an overall classification accuracy of 96% on this dataset, comparable

to or exceeding the performance of systems using specialised wearable sensors. Their work highlighted the feasibility of using simpler accelerometer and gyroscope sensors, such as smart phones, as unobtrusive, affordable, and reliable sensing tools for HAR.

II. DATASET DESCRIPTION

A. Dataset Source

The dataset used in this study is the *Human Activity Recognition Using Smartphones* dataset [1], first introduced in 2013. It was created as part of a study on human centered computing at the University of Genova and the Universitat Politècnica de Catalunya [4]. The dataset consists of sensor data collected from 30 volunteers aged between 19 and 48 years, performing six basic activities of daily living.

Each participant carried a Samsung Galaxy S II smartphone on the waist while performing the activities. The smartphone's embedded triaxial accelerometer and gyroscope were used to capture linear acceleration and angular velocity at a constant sampling rate of 50 Hz. The experiment was carried out in a controlled laboratory environment, with participants following a standardised sequence of activities to ensure consistency across samples. The signals were normalised and bounded within [1, -1], and each feature vector is also standardised.

The dataset is publicly available from the UCI Machine Learning Repository under an open license.

B. Feature Overview

The dataset provides both raw sensor readings and pre-processed feature vectors based on that raw data. Each observation represents a 2.56 second window of sensor data with 50% overlap between adjacent windows, there are 128 readings per window. From each window, 561 features were extracted from the time and frequency domains.

The features include statistical and signal-based measures such as mean, standard deviation, median absolute deviation, signal magnitude area, energy, entropy, interquartile range, and correlation between sensor axes. Frequency-domain features were obtained using Fast Fourier Transform (FFT) on the

windowed signals. These features capture important motion characteristics for each activity and are normalized to ensure comparability across subjects and sessions.

Each record in the dataset is labeled with one of six activity classes:

- Walking
- Walking Upstairs
- Walking Downstairs
- Sitting
- Standing
- Laying

Additionally, subject identifiers (1–30) are included, allowing for subject specific or subject independent analysis.

C. Data Preparation

Prior to model training, a subset of the original dataset features was selected to simplify the analysis and reduce computational overhead. The full UCI HAR dataset contains 561 features derived from both time and frequency transformations. This feature set provides detailed motion characterisation, but it is computationally expensive to process, particularly on the edge in the Raspberry Pi.

To address this, a representative subset of features was selected, focusing on descriptive statistics (mean, maximum, and minimum values) from the accelerometer and gyroscope signals in both body and gravity components. The retained features include the following categories:

The retained features include the following categories:

- **Time domain features:** body acceleration, body acceleration jerk, gyroscope, and gyroscope jerk (mean, max, and min for each axis)
- **Magnitude features:** signal magnitude for acceleration, jerk, and gyroscope signals
- **Frequency domain features:** selected mean, max, and min values for acceleration, jerk, and gyroscope signals

This reduced dataset comprises 101 variables in total, including 100 descriptive features and one target label representing the activity class. The selected features preserve the essential statistical and dynamic characteristics of human motion while reducing data dimensionality. This reduction improves computational efficiency, especially on the edge where limited processing power and energy consumption are key considerations.

Furthermore, the original study [4] employed a Support Vector Machine (SVM) classifier, which required extensive pre-processed features to achieve high accuracy. In contrast, this work explores more modern machine learning techniques capable of capturing non-linear relationships and implicit feature interactions. It is therefore expected that the reliance on an extensive set of features is reduced without sacrificing accuracy.

On the Azure Machine Learning side, the subject feature was excluded from model training to prevent the classifier from learning participant specific patterns. This ensures that the model generalises to unseen individuals rather than over-fitting to sensor characteristics unique to particular subjects.

The dataset was then partitioned into training (80%) and testing (20%) subsets for model evaluation.

III. DATA ANALYSIS

A. Characteristics

B. Trends and Patterns

IV. FEATURE COMPARISON

A. Visual Comparison

B. Discussion

V. BOX PLOT ANALYSIS

A. Box Plots (or Violin Plots idk change this)

B. Findings

VI. MACHINE LEARNING MODEL DEVELOPMENT AND EVALUATION

A. Single Feature Models

B. Combination of Features

C. All Features

D. Feature Selection Methods

VII. RESULTS COMPARISON AND DISCUSSION

A. Performance Comparison

B. Discussion of Findings

VIII. CONCLUSION

REFERENCES

- [1] J. Reyes-Ortiz, D. Anguita, A. Ghio, L. Oneto, and X. Parra. "Human Activity Recognition Using Smartphones", UCI Machine Learning Repository, 2013 [Online]. <https://doi.org/10.24432/C54S4K>.
- [2] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, "HARTH: A Human Activity Recognition Dataset for Machine Learning," *Sensors*, vol. 21, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/s21237853.
- [3] K. Bach, A. Logacjov, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, "A Machine Learning Classifier for Detection of Physical Activity Types and Postures During Free-Living," *Journal for the Measurement of Physical Behaviour*, vol. 1, no. aop, pp. 1–8, Dec. 2021, doi: 10.1123/jmpb.2021-0015.
- [4] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition Using Smartphones," *Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013)*, Bruges, Belgium, pp. 437–442, Apr. 2013.