

   Department of Robotics and Mechatronics Engineering

       University of Dhaka

**Project report on**

**Human Activity Recognition Using Classical Machine Learning Approach**

**Course name:** RME 3211 (Intelligent Systems and Robots Lab)

**Group number:** 9

**Group Members:** Humayra Rashid (SK-092-011)  
                               Safayat Hasan (SH-092-021)

**Submitted to:** Dr. Sejuti Rahman  
                         Chairperson  
                         Assistant Professor  
                         Department of Robotics and Mechatronics Engineering  
                         University of Dhaka

**Date of submission:** 15th November 2022

**ABSTRACT**

Human activity recognition adds a new dimension to different kinds of automation systems related to the movement on daily basis. But this recognition is challenging for researchers because of the lack of proper dataset. In this project we have used the dataset of MMAct named ‘trimmed’ which has four dataset of four sensors, ‘acc\_phone\_clip’, ‘acc\_watch\_clip’, ‘gyro\_clip’ and ‘orientation\_clip’. We have used the accelerometer sensor of phone named ‘acc\_phone\_clip’ to build our model. This dataset contains 35 activities which we tried to classify through our model. In our method, we have resampled our data to deal with a variable sample frequency of dataset. We have taken some samples of definite window-length and used the feature extraction to extract different more features from the given features. We have used random forest classifier which is a classical machine learning approach. Applying this approach, we got 27% validation accuracy on the dataset. We have trained our model on the train dataset and validated them on the test dataset.

**1. INTRODUCTION**

Human activity recognition tries to interpret human motions and the movement of human along a frame or environment. Due to the availability of the wearable technologies, the task of activity recognition has become very easier nowadays. Because the data related to the movement of human has become easier through cell-phones or sensors in the body. Activities like walking, swimming, running, cycling etc. can easily be tracked using these. The activity recognition of human has become very important in the healthcare sector. The accelerometer of phone is used in our project as its usability makes it easier to record the data. The accelerometer data makes the co-ordinate of the 3 axes through taking data. Through research it is found that, the ideal positions for attachment of the accelerometer are ankle, knee, biceps, wrist and hip.

Recording data and estimating activities have proven to be challenging as there is no automated system in the hospital which can monitor activities if the personnel. Sometimes the work becomes tougher as the activity of the person becomes similar. For example, giving injection and collecting blood samples contains almost similar features. We have tried to avoid these types of activities in our project. When some activity matched the features of another, we took only one of the activities to train our data.

Most of the work in human activity recognition assumes a figure-centric scene of uncluttered background, where the actor is free to perform an activity. The development of a fully automated human activity recognition system, capable of classifying a person’s activities with low error, is a challenging task due to many problems.

The goal of our human activity recognition project is to examine different activities from the accelerometer of the phone such as- looking around, kicking, closing, drinking, carrying, throwing, exiting, jumping, standing etc. and classify them correctly.

The working flow diagram of our project is illustrated below:

**Raw Data**

**Data Pre-Processing**

**Feature extraction**

**Classification**

The rest of the report is organized as follows:

In section 3 we will cover the literature review or previous surveys of the related project. Section 4 presents the dataset description of our project. Section 5 covers the work methodology. In section 6 we have discussed our proposed process in the project. In the section 7 we have discussed about the classification model and in section 8 tried to focus on the result. In section 9 we have drawn the conclusion finally.

**2. LITERATURE REVIEW**

There have been several surveys in the human activity recognition literature. **Kadir et al.** [19] proposed a feature extraction based method on motion capture data using K-Nearest Neighbors (KNN) classifier to predict nursing activities. In another study, **Haque et al.** [13] provided a Gated Recurrent Unit (GRU) model with an attention mechanism to recognize the nursing activity, where it used location-based as well as motion-based features. Besides, a classical approach using only acceleration data and random forest classifier performs better on recognizing nursing activity [24]. **Mantyjarvi et al.** [22] recognized human posture and ambulation using acceleration data that was collected using a band attached to the hip. **Antar et al.** [6] provided a comprehensive study on some comparative approaches to classify smartphone accelerometer data.

In the “**Pragmatic signal processing approach for Nurse care activity recognition**”, they used the Nurse care activity recognition challenge 2020 dataset to recognize nursing activities. For this task, the data from the accelerometer attached to the body of the personnel had been used.

In **Aggarwal and Cai** (1999), a new taxonomy was presented focusing on the human motion analysis, tracking from single view and multi-view cameras and recognition of human activities.

The survey of **Moeslund et al.** (2006) mainly focused on pose-based action recognition methods and proposed a four-fold taxonomy, initialization of human motion, tracking, pose estimation and recognition methods.

**Turaga at el.** (2008) proposed the separation between the meanings of ‘action’ and ‘activity’, where the activity recognition methods were categorized according to their activity complexity.

**James and Sebe**(2007) proposed a survey for multimodal human computer interaction focusing on affective interaction methods form poses, facial expressions and speech.

**Pantic and Rothkrantz**(2003) performed a complete study in human affective state recognition methods that incorporate non-verbal multimodal cues, such as facial and vocal expressions.

Finally, a thorough analysis of the ontologies for human behavior recognition from the viewpoint of data and knowledge representation was presented by **Rodriguez et al.** (2014).

**3. DATASET DESCRIPTION**

The dataset contains daily, abnormal and desk work activity data collected from the 3-axis accelerometer sensor of a mobile phone. The data was collected using semi naturalistic collection protocol. It is a new large scale multi modal dataset for action understanding based on diverse modalities. There are 35 different activities and those are: ‘Carrying’, ‘Checking time’, ‘Closing’, ‘Crouching’, ‘Entering’, ‘Exiting’, ‘Fall’, ‘Jumping’, ‘Kicking’, ‘Loitering’, ‘Looking around’, ‘Opening’, ‘Picking up’, ‘Pointing’, ‘Pulling’, ‘Pushing’, ‘Running’, ‘Setting down’, ‘Standing’, ‘Talking’, ‘Talking on phone’, ‘Throwing’, ‘Transferring object’, ‘Using phone’, ‘Walking’, ‘Waving hand’, ‘Using PC’, ‘Carrying light’, ‘Carrying heavy’, ‘Pocket in’, ‘Pocket out’, ‘Drinking’, ‘Sitting’, ‘Sitting down’, ‘Standing up’. The data is given in time series and the sensor data of x, y and z axis are given. Sensor data is taken under total 20 subjects where each subject has 4 scenes and 5 sessions under each scene.

**4. METHODOLOGY**

The goal of our approach was to create a straightforward learning methodology with minimal computational loss. So, we tried to employ a traditional machine learning algorithm on carefully selected features. We have worked more on data pre-processing and feature extraction so that the learning system can categorize activities using the optimum feature combination.

**4.1 Data cleaning & Pre-processing**

The dataset of accelerometer-phone-clip consist of the sensor data in such a way that there are total 20 subjects, 4 scenes under each subject and 5 sessions under each scene where each session contains the activity files named according to the activities. Each of the activity file contains the accelerometer sensor data in 3 axes with the timestamps. There was no column headings in the data file. So,-

* We rename the columns of each file.
* We converted the given type of time in first column in the date-time data type.
* We added 4 new columns named- ‘Subject\_Number’, ‘Scene\_Number’, ‘Session\_Number’ and ‘Activity’.
* We also determined the files that contained nothing and deleted those files.

**4.2 Feature Extraction**

We know that the sensor data was given in 3 channels x axis, y axis and z axis. Before extracting features of the whole data we tried split the data in several parts named number of windows. The window length we took is 40. That means we took first 40 rows of the merged raw data and extracted new features from those for each axis. The features we got out of each window are-

**Minimum:**

**Maximum:**

**Standard deviation:**

**Average:**

**Variance:**

**Peak to peak range:**

**Maximum rate of change:**

**Average rate of change:**

**Standard deviation of rate of change:**

**Interquartile range:**

**Correlation:**

**Root mean square:**

**VEL:**

With these features being extracted we created a new file of data consisting of these features. After this we got a file with 153901 rows and 40 columns. The last column is kept as the activity column. Activity is what we are supposed to predict here.

**5. PROPOSED PROCESS**

Our main idea was to get all the activity data together at first and then split each file of the all activity data according to the window length. After that from each sliding window we would process out the features of each axis and add them in a new file one by one. Thus we would get a new file which is totally different from our former data. Now rather than having a list of files we would have a single file which we could pass in our model and predict the activity.

**Data Pre-processing & cleaning**

**Bad files**

**List of files**

**Raw**

**Data**

**All Files**

**Removing bad files**

**ӯ**

**Classifier**

**Windowing and Feature Extraction**

**New**

**File**

**Windowing**

**Feature Extraction**

**Fig: Block diagram of our proposed method**

**6. CLASSIFICATION**

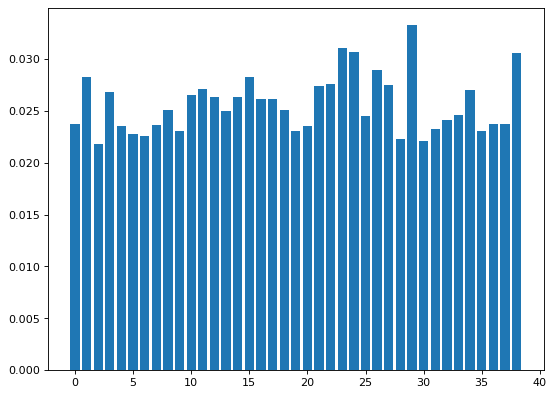
We have used a 70% of the reprocessed data-frame as training set and rest 30% as test set. Our objective is to assign a label ӯ which is any integer number between 0-34 representing any activity and observe that the predicted activity or label matches the actual activity or label to which extent. We have tested the processed data using different classifiers such as ‘Logistic Regression Classifier’, ‘Extra Trees Classifier’, ‘Decision Tree Classifier’ etc. and compared their results. ‘Random Forest Classifier’ worked best among all the classifiers mentioned. So, we finally used the ‘Random Forest Classifier’ for categorization in this project. Random forest classifier is a well-known and significant classifier. It is renowned for offering strong overall predictive performance, little over-fitting, and simple interpretability, which guarantees that it is simple to ascertain the importance of each variable on the tree choice. It makes the calculation of the contribution of each feature to the final conclusion easier.

**7. RESULTS & DISCUSSION**

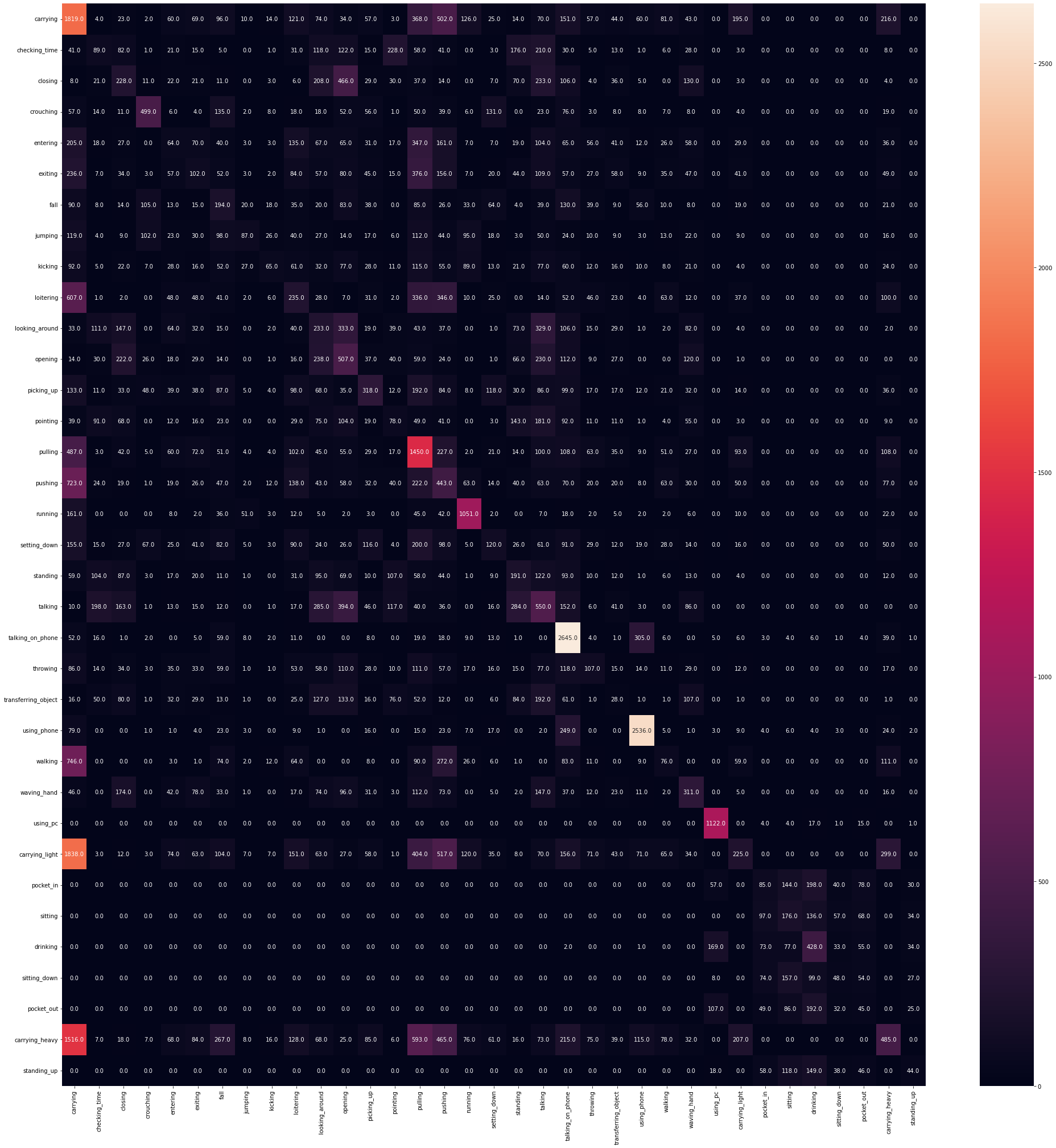
The project was very much challenging. It is mostly because of the dataset. The dataset we dealt with is given in an unstructured pattern. Also it is not a single csv file. There are way more csv files for each type of activity under many subjects. That’s what made our first job tougher. We gathered all the data together and windowed all files taking 30 as window length before extracting features. Using Random Forest Classifier we got 100% train accuracy and 27% test accuracy whereas 100% weighted average of F1 score in train and 29% in test set. The result is not very much satisfactory. The reason could be that while windowing any file there remained rows which didn’t fit in the window length. We didn’t handle these rows. As a result, these remained unprocessed and we missed some data that might be proved to be important in prediction.

The following figure shows the confusion matrix of our classification model on our test set. We can see that there are activities which are predicted more than some other data. Some of those are ‘Carrying’- TP=1819, ‘Using phone’- TP=2536. At the same time there are some activities in which case TP is 0. The reason behind this could be that some of the activities have very less amount of data so our data has not been trained uniformly. As a result, there is bias towards some activities. In the result, ‘Using PC’ has high precision of 0.96, whereas ‘Walking’ has very bad precision of 0.05. ‘Using Phone’ has a good recall of 0.77 but ‘Transferring Object’ has a poor recall of 0.05. So, we can see the variation in the different scores of the different activities.

The following figure shows the feature importance graph.



We also tried to plot the confusion matrix of the



**8. CONCLUSION**

We have tried to propose a very simple and classical machine learning approach to predict for any human activity mentioned in the report. We have worked with only accelerometer data from a mobile phone’s sensor. Our experimental result is not very promising. However, this method can be improved further to get a better result. The potential changes that can be made are changing the window size, try to process each and every data, using n-fold cross validation in place of 80-20 split and the overall performance can be more improved using deep learning techniques.

**9. REFERENCES**

* Aggarwal, J.K. and Chai, Q. (1999). Human motion analysis: a review.
* Md. Ahasan Atick Faisal, Md. Sadman Siraj, Md. Tahmeed Abdullah, Omar Shahid, Farhan Fuad Abir, M.A.R. Ahad. Pragmatic signal processing approach for Nurse care activity recognition using Classical Machine Learning
* Michalis Vrigkas, Christophoros Niklou and Ioannis A. Kakadiaris. A Review of Human Activity Recognition Methods
* Moeslund, T.B., Hilton, A. and Kruger, V. (2006). A survey of advances in vision-based human motion capture and analysis.
* Turaga P.K., Chellapa R., Subrahmanian V.S. and Udrea O. (2008). Machine recognition of human activities: a survey.
* Jaimes A. and Sebe, N. (2007). “Multimodal human-computer interaction: a survey”.
* Pantic M. and Rothkrantz L. (2003). “Towards an affect-sensitive multimodal human-computer interaction”.
* Rodriguez N. D., Cuellar M. P., Lilius J. and Calvo-Flores M.D. (2014). A survey on ontologies for human behavior recognition.