# Data Mining Project: Physical Activity Recognition

Youssef Hosni
*Biomedical Engineering*
*University of Oulu*
youssef.hosni@oulu.fi
2601630

Aino-Kaisa Manninen
*Biomedical Engineering*
*University of Oulu*
ainokmanninen@gmail.com
1981814

Henna Mönttinen
*Computer Science*
*University of Oulu*
henna.monttinen@oulu.fi
2181260

***Abstract*—Activity recognition is widely used in several fields from healthcare to safety. Smartphones have nowadays several sensors that are useful on activity recognition and many people carry them almost constantly in their pockets or bags. Hence, they are very natural devices used for recognition task.**

**In our work we studied the effect of different classifiers, window sizes, axel combinations and positions to recognition accuracy. We also tested two different feature selection methods (PCA and correlation) and their effect on performance. Also, Hilbert-Huang transformation was tested for feature calculation. According to our findings Hilbert-Huang features did not bring significant improvement to activity recognition especially if taking account the increased complexity. The best window size was 4 seconds with KNN classifier and 8 seconds with LDA classifier. The best sensor position to be used is the left or right pocket sensor and the best axel combination was y and z axes of accelerometer, y and z axes of linear accelerometer and x, y and z axes of gyroscope.**

***Index Terms*—activity recognition, accelerometer, feature extraction, feature selection, gyroscope, Hilbert-Huang transformation, smartphone sensors**

## I. Introduction

Activity recognition is widely used for example in the fields of healthcare, safety, industry and well-being. In healthcare it can be used for diagnostic or disease monitoring where as in commercial products it is for example used to track people's daily activity in order to encourage maintaining healthy lifestyles.

There are three different approaches to recognize different activities: vision based, environment interactive sensor based and wearable sensor based approach. Activities are recognized with camera or video in the vision-based approach. This method is not often very convenient because of privacy issues particularly in indoor environments and the technical quality of photo of video varies a lot. In the environment interactive sensor-based approaches, activity recognition happens "by capturing interaction between the subjects and objects under the assumption that there exists the underlying relation between objects and activities". For example, if the sensor of object is placed on the bed and is staying there some time, we assume that the subject person is sleeping. In the wearable sensor based approach, the subject person is wearing or carrying a device with various sensing units. [4]

The recognition algorithms are mainly based on signals gathered from sensors or computer vision. This study belongs to the former category and exploits the sensors of Samsung galaxy S2 smartphones. According to Liu et al. [6] the benefits of sensor-based methods are their ease of computation, which makes real-time activity recognition easier to implement and the fact that it is more difficult to recognize person from sensor-data than from image data.

In this project an activity data collected and published by Shoaib et al. [1] is used for studying the effects of different feature selections, window sizes and sensor combinations on recognition of seven basic activities.

## II. Related work

The interest for detecting more complex human activities has inspired Liu et al. [6] to study temporal pattern-based method to detect normal daily activities like cooking. Idea of their method is based on discriminative pattern learning in computer vision and similar classification methods in other fields. The algorithm that they developed discovered frequently appearing higher level activities' patterns from the sets of lower lever activities. Their training set consisted labeled higher and lower level activities and their algorithm searched for discriminative temporal patterns in order to recognize the higher level activites like coffee time or cleaning. In total they had five higher level activity and several lower level activities. They used wearable sensor data from the open Opportunity challenge database that was presented by Chavarriga et al. [8]. According to Liu's et al. [6] results, patterns of three lower level activities were sufficient to classify higher level activities accurately.

Seto et al. [7] used dynamic time warping method for activity classification (walking, walking upstairs, walking downstairs, sitting, standing, lying). It is an algorithm for time series to calculate the distance and alignment of two different series. However, it is computationally heavy so they had to modify it by limiting the number of paths calculated during finding the optimal path between time series. From training data they built templates, to which the test data was compared by using the modified dynamic time warping method. The features were hence, the calculated distances to each template. Before classification dimensionality reduction was done by PCA and classification was done then by linear SVM. The data that they used in the study was from UCI HAR dataset and a synthetic dataset generated based on the same dataset. With the generated dataset they wanted to test algorithm robustness to noise. The classification results for both datasets were 0.890 and 0.67 respectively.

According Huile et al. [9] Hilbert-Huang transformation (HHT) has been used in many fields that study nonlinear systems but it is less used for activity recognition. In their study by using accelerometer data from three sensors from PAMAP2 dataset, they acchieved accuracies nearly 90% when using different combinations of HHT features. The importance of instantaneous energy features and features from marginal Hilbert spectrum were shown to be the most useful in classification. They noted that the usage of instantaneous frequency and amplitude features resulted poorer classification results as calculation of features from them faded their connection to time. However, when they were compared to features from marginal Hilbert spectrum, over 90% accuracy were reached, which was better than instantaneous energy and marginal spectrum features together. Hence, showing that they can complement each other.

Wanneburg and Malekian (2017) have researched physical activity recognition from smartphone accelometer data. They studied everyday activities like sitting, standing, laying, walking and jogging. The dataset was sampled 70% for training set and 30% were testing set. The data was segmented for 1 second windows with 50% overlap. They used both time domain and frequency domain features. Time domain features were: mean, maximum, minimum, median, standard deviation, signal magnitude area, mean deviation, principal component analysis (PCA) an interquartile Range. Frequency domain features were: mean, maximum, minimum, median, skewness, kurtosis and PCA. They tested ten different classifiers for they model: support vector machine, multilayer perceptron, Naive Bayes, Naive Bayes witn BayesNet parameter, k-nearest neighbour method with n = 1 and n = 5 parameter, Naive Bayes with Bagging method, Three method, multilayer perceptron and kStar. According the results, both KNN (n=1 and n=5) and kStar algorithms gave the best overall accuracy of 99%. [2]

Siirtola and Röning (2012) compared two different classifiers: KNN and quadratic discriminant analysis (QDA) in real time activity recognition on a mobile phone. They studied recognition of walking, running, cycling, driving a car and sitting/standing from mobile phone data. Phones were positioned in the pocket of each subject's trousers. They used sliding window technique with 7.5 seconds window size and 21 different features for their algorithm. As a result, the overall accuracies of real-time recognition gave as good rates as in the offline recognition rates. QDA classification gave a slightly better results than KNN classifier. The KNN took also slightly more time for computing. [3]

Wang et al. (2016) utilized accelometer and gyroscope data of smart phone for activity recognition. Their framework were tested for both offline and online prediction scheme. They studied six different human activities: walking, walking upstairs, walking downstairs, sitting, standing and lying They used sliding window method and various time domain and frequency domain features. Generally, feature selection can be categorized for filter-based methods and wrapper-based methods. Filter-based methods are not dependent on classifier, and they "have lower computational complexity and better generalization ability". Wrapper-based methods are specific for certain classifier and the computation are more time and cost consuming. These methods tend to obtain better classification performance. Wang et al. (2016) used in their study combination of filter-based and wrapper-based techniques in order to get better tradeoff betwween time, cost and accuracy.As a result, the combination of accelerometer and gyroscope data obtained better recognition performance than using only one data source. If comparing data from one single source, accelerometer provide more discriminant information than gyroscope. [4]

## III. Objectives

In our study we are going to answer to following research questions sets defined below.

Question sets 1:

What is the effect of changing of window size, as mentioned in [1] that the best window size is 3 seconds, so we decide to investigate and see how changing the window size will effect the performance and how this will change using different classifiers and features and sensor position and what is the best window length in each case.

Question sets 2:

What is the effect of different axel combinations? Each sensor has three different axis: x-axel, y-axel and z-axel and it will be tested different combinations of axis and how they affect on activity recognition.

Question sets 3:

What is the best sensor position to be used for recognizing the different activities and does this change with the different classifiers and the window size and features used. Finally what are the best activity recognized by each position.

Question set 4:

What features are the most important ones? Is there difference if principal component analysis (PCA) or correlation is used for feature selection? How features calculated from Hilbert-Huang transformed data compares to features in the original article?

The expectation is that we will find a few most important features that together will be capable to perform equally well than using all possible features. It will be interested if PCA and correlation would give different features but the most important features are expected to be the same nevertheless. On the contrary it will be interesting to see if the usage of Hilbert-Huang transformation brings any benefits compared to more commonly used features calculated from time space and frequency space. It is expected that some benefits might exist but as performing Hilbert-Huang transformation is computationally more heavy, it is doubtful if the effect is so remarkable that it pays to use more complex method.

## IV. Data

The data set given for the project was collected and introduced by Shoaib et al. in their arcticle "Fusion of Smartphone Motion Sensors for Physical Activity Recognition” [1]. Originally, it was collected for the purpose of studying the best positions and combinations of smartphonse sensors for different types of activity recognition tasks.

The dataset consisted data from 10 different volunteers who performed seven different activities while wearing five smartphones (Samsung Galaxy SII i9100) in different sites of their bodies. The activities were walking, running, standing, sitting, biking, walking upstairs and walking downstairs. Respectly, the smartphones were located in left and right jean’s pockets, at the right side on a belt, on the right upper arm and on the right wrist. In all other locations except on the belt the smartphones were in portrait orientation. All volunteers were 25 to 30 year old males and the length of each activity were between 3 to 4 minutes. [1]

The data itself was collected at 50 Hz sampling rate and smartphones calibrations were checked before data collection. Data from all three physical sensors (namely an accelerometer, gyroscope and magnetometer) were provided. In addition a fourth variable called linear acceleration was derived from accelerometer data by filtering gravitational acceleration off. [1]

The dataset was downloadable at https://www.utwente.nl/en/eemcs/ps/dataset-folder/sensors-activity-recognition-dataset-shoaib.rar and consisted separate .csv files for each participant. Each file contained data from all five smartphones and activities as well as a timestamp vector for each smartphone data and one column of activity labels for all smartphone locations. The smartphone locations were differentiated by using headers within a file. Timestamps were in UNIX time format and all measured data was in floating point numbers. The labels were in a string format.

## V. Methods

### A. Data pre-processing

Before any analysis or further pre-processing the data was imported to Matlab (The MathWorks, Massachuttes, USA) and the data label columns were changed from string format to integers into the range from 1 to 7. Also only accelerometer and gyroscope data were left for further analysis. Otherwise there was no pre-processing used before feature calculation as it was noticed that scaling were better to do in feature space instead of doing scaling on raw data. Before classification feature data was scaled between 0 and 1.

### B. Feature extraction

The data was segmented after pre-processing. We utilized variable sized sliding windows for segmentation with 50% overlap.

*1) Time and Frequency domain features:* We selected both time domain and Fourier transformed frequency domain features from presented in Fig 1. The eight of eleven features were from the study of Shoaib et al. [1] from 2014. We also added three frequency domain features. The all eleven features were combined together.

| Features | Time or Frequency Domain |
|---|---|
| Mean | Time domain |
| Standard Deviation | Time domain |
| Median | Time domain |
| Variance | Time domain |
| Zero crossings | Time domain |
| Root mean square value | Time domain |
| Sum of FTT coefficients | Frequency domain |
| Signal energy | Frequency domain |
| Mean | Frequency domain |
| Skewness | Frequency domain |
| Kurtosis | Frequency domain |

Fig. 1. Time and Frequency domain features.

Each segment values were stored in feature matrix which consisted of three axis of three sensors: accelerometer, linear accelerometer and gyroscope, all five positions which were left pocket, right pocket, belt, wrist and upper arm and all eleven features so there were 495 column in these features matrix which is 11(features)*3(sensors)*3(axis)*5(sensor position).

*2) Features from Hilbert-Huang transformation:* For new features, Hilbert-Huang transformation was used. HHT is an empirical method to separate a raw signal first to so called intrinsic mode functions by empirical mode decomposition and then do normal hilbert transformation to the calculated functions. According to articles that we found, the work by Huile et al. [9] is the only article that used it for activity recognition.

The overall principle of the transform is the following. In the empirical mode decomposition the upper and lower envelopes of raw signal are determined and the mean of them is calculated. After that, the mean is subtracted from the signal and if the result fulfills the conditions of intrinsic mode function, it is determined as one and decomposition continues with the rest of the data. If conditions for intrinsic mode function is not fulfilled, the loop is repeated again but the signal is replaced with the subtracted signal. Hence, the process is iterative and continues when a stopping criteria is fulfilled. After not finding any intrinsic mode functions anymore, the functions are then transformed to frequency-time-energy spectrum by Hilbert transformation. As noted by Zeiler et al. [10] there is problems in empirical mode decomposition and some improvements are suggested. Hence, in this work we used embedded empirical mode decomposition, which is noise assisted method. In this method each intrinsic mode function is calculated several times with random white noise applied to it. After some rounds, the final intrinsic mode function is got by averaging the functions from all rounds. This method should make the decomposition more general and reproducible.

Features were calculated from the first three intrinsic mode functions after hilbert transformation. The features extracted are presented in the Fig 2 on the next page.

| Marginal Hilbert Spectrum | 1st intrinsic mode function | 2nd intrinsic mode function | 3rd intrinsic mode function |
|---|---|---|---|
| Mean | Instantaneous energy : Mean | Instantaneous energy : Mean | Instantaneous energy : Mean |
| Standard deviation | Instantaneous energy: Standard deviation | Instantaneous energy: Standard deviation | Instantaneous energy: Standard deviation |
| Median | Instantaneous frequency : Mean | Instantaneous frequency : Mean | Instantaneous frequency : Mean |
| Variance | Instantaneous frequecy : Standard deviation | Instantaneous frequecy : Standard deviation | Instantaneous frequecy : Standard deviation |
| | | | |

Fig. 2. Features extracted from Hilbert-Huang transformed variables

### *C. Feature selection*

For feature selection, two different methods were used. Principal component analysis (PCA) was tested by taking all those features in that belong to first principal component and had coefficient value greater than 0.01. The second method was the usage of simple correlation. In that we used Spearman's rho value to indicate features that correlated the most with each other. It was decided to use absolute rho value of 0.8 to be the indicator for too high correlation between features. If the threshold was crossed, one of the features were left out.

### *D. Classification*

To recognize the activity from the feature matrix, classification model was used.Two classifier were used K nearest neighbors (KNN) and linear discriminant analysis (LDA) classifier.

*1) KNN classifier:* KNN classifier is one of the most straight forward classifier, it classify the given query depending on the class of the nearest k neighbors of this query as shown in fig 3 [11]. This classifier is used as it is one of the simplest classifiers and can be easily implemented and understood, although it has some drawback which is long computational time, but since we did not use in run time so it did not affect our work.

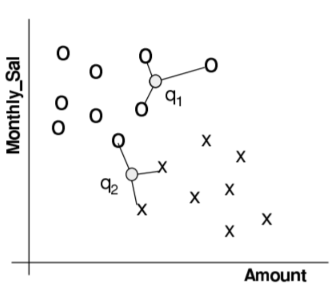


Fig. 3. simple example of 3 nearest neighbor classification

*2) LDA classifier:* Linear Discriminant classifier is a well know to learn discriminative features, and it can be used for dimensionality reduction and classification.In LDA two assumptions about the data set is made:The data of each class is normally distributed and the variance of each class is the same. Classification is done for new input point by calculating the probability that the new input belongs to a each class by assigning it with a linear discriminant function and the class that has the highest probability will be the class of the input data. [12]

To evaluate the performance of the classifier 10 fold cross validation is used in which the classifier is trained on 9 folds of the data and here it is corresponding to 9 participant data and then tested on the left out fold which is corresponding to one participant data, this is repeated ten times so each participant data is used once as test data and then the accuracy from all the ten iterations is being averaged and considered as the accuracy for the model.

## VI. Results

### *A. Effect of window length and sensor position*

Window lengths used are from 2s to 8 sec, and 2 seconds was chosen the minimum, because less than this some activity will not be able to recognized. These window sizes was tried with different cases as will be shown below.

*1) ALL sensor positions:* In this case all the senors data were used and the for each window size these cases was tried: All the features (time and frequency domain features) were used once with the KNN classifier and once with the LDA classifier then time domain features are used with both KNN and LDA classifiers and finally the frequency domain features used with each one of the two classifiers. The results of these are summarized in fig 4

*2) using single position:* In this section the results got from using each position on it own and using the data from the three sensor and all the axis of these sensors. Different window lengths was used and also different features domain were tried with both classifiers.

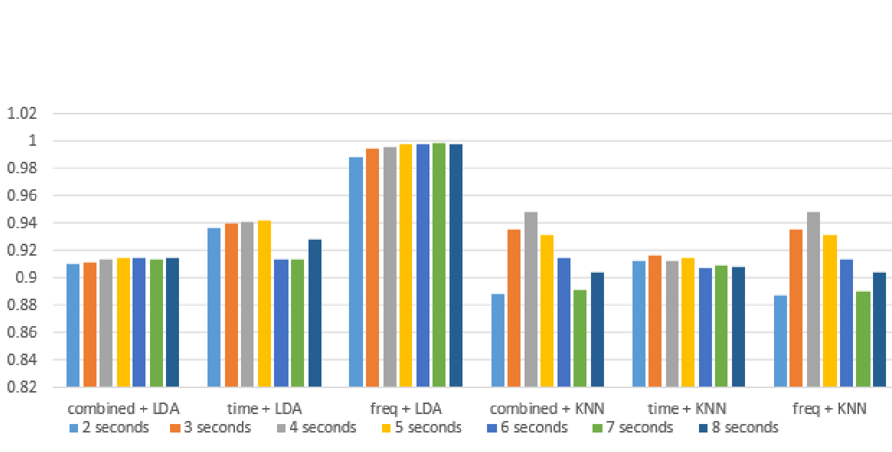


Fig. 4. effect of window size using data from all position with different combination of features and classifiers

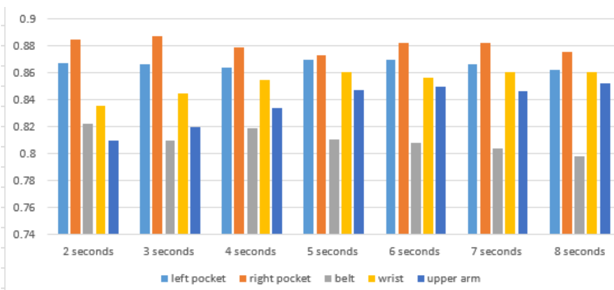


Fig. 5. effect of window size using data from single position with different time domain features and KNN(K=5) classifier

1).using different window lengths with the time domain features and the KNN classifier and the results for this is shown in fig 5.Based on this result based if all the sensor positions data are used the best window length to be used is 3 sec.

2).using different window lengths with the frequency domain features and the KNN classifier and the results for this is shown in fig 6.from the results it can be seen that the best position was the right pocket and the best window length was 4 seconds.

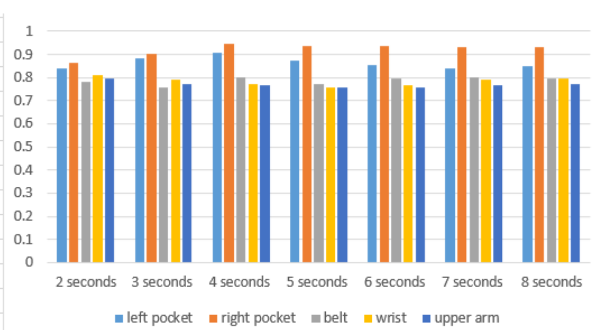


Fig. 6. effect of window size using data from single position with different time domain features and KNN(K=5) classifier

3). Using the combined features both the time features and frequency domain features with the KNN classifier with different window lengths.The result for this is shown in fig 7.from the results it can be seen that the best position was the right pocket and the best window length was 4 seconds.

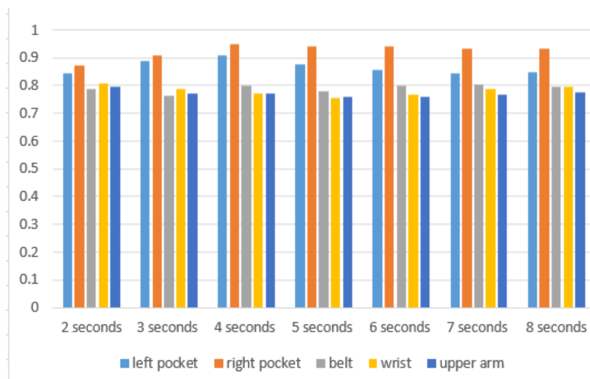


Fig. 7. effect of window size using data from single position with different all the features(time and frequency domain features) and KNN(K=5) classifier

4).using different window lengths with the time domain features and the LDA classifier and the results for this is shown in fig 8.from the results it can be seen that the best position was the left pocket and the best window length was 8 seconds.

5).using different window lengths with the frequency domain features and the LDA classifier and the results for this is shown in fig 9.from the results it can be seen that the best position was the left pocket and the best window length was 8 seconds.

6). Using the combined features both the time features and frequency domain features with the LDA classifier with different window lengths.The result for this is shown in fig10.from the results it can be seen that the best position was the left pocket and the best window length was 8 seconds.

7). For each position in different cases and different classifier and different features used, the activity classified with highest accuracy was recorded and in fig 11, in which for each

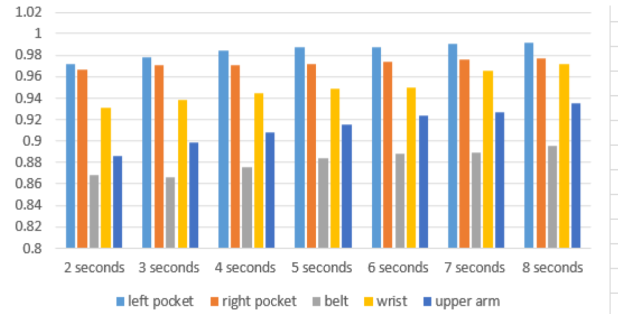


Fig. 8. effect of window size using data from single position with different time domain features and LDA classifier

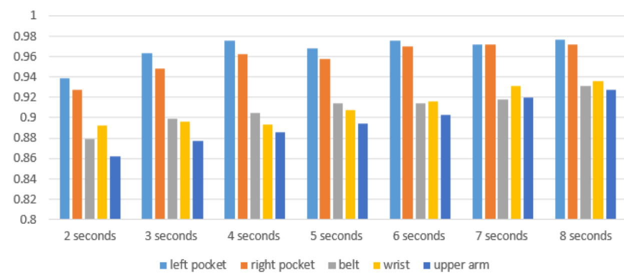


Fig. 9. effect of window size using data from single position with different time domain features and LDA Classifier

position it shows how frequency of each activity to be the best activity. From the results and as can be seen from the figure below that for the left and the right pocket the best activities are sitting , for the wrist and upper arm are biking and jogging and for the belt data it was jogging and siting.

The overall best accuracy can be achieved by using time domain features and the data from the left pocket position and window length of 8 seconds and LDA classifier.

### *B. Axel combination.*

In this part, the left pocket position was selected. Eleven features and LDA as a classifier were used. The window size was 8 seconds.

At the beginning, a single axel were used in computing (Fig 12). The y- and z-axel of accelerometer gave the best overall accuracy, an the x-axel of accelerometer gave the worst overall accuracy.

Some combinations were tested. If selecting only one axis of each sensor, the y-axes gave the best overall accuracy. With combination of only four axes: y and z-axis of accelerometer and linear accelerometer, the same overall accuracy was achieved than with all nine axes (Fig 13). The best combination of axis was combination of axes which leaved out the x-axis of accelerometer and x-axis of linear accelerometer (Fig 14).

### *C. Feature selection*

In this part all features were used. Positions that were selected were wrist and left pocket. Feature selection was then tested only with HHT features and then with all features

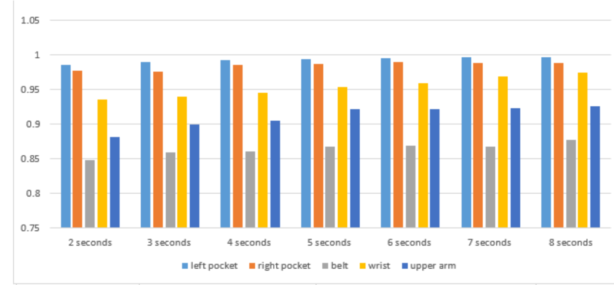


Fig. 10. effect of window size using data from single position with different all the features(time and frequency domain features) and LDA classifier

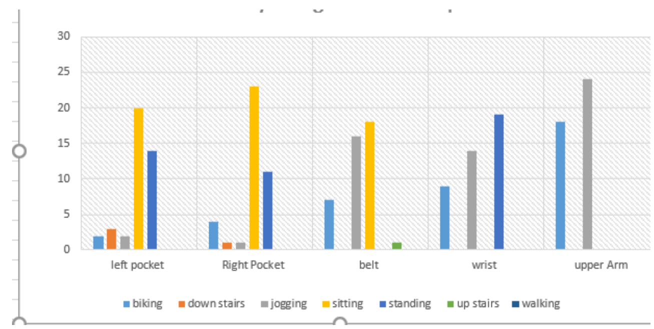


Fig. 11. frequency of the activity that best recognized by the data of each sensor position.

calculated for the two positions by using 3-second window. KNN with k-value of 5 was used as a classifier in this part. The results are presented in Fig15. As can be seen, PCA produced more stable number of variables included in every fold and changes between each fold were only amount of one feature. However, the features sets itself were still different in each fold and only a few features were repetitive in all folds. When it come to the usage of correlation in feature selection, the number of selected features varied more between folds. Anyway, in both methods it can be noticed that the number of features can be heavily decreased without that it affect to overall recognition performance.

### *D. Comparison to the original article*

In the original article of Shoaib et al. [1] not exact accuracies were given but they had to be estimated from the figures. The accuracies were presented for activities of walking upstairs and downstairs. As only the wrist location was the same as used in this part of the study and presented in the figure of the article, the following accuracies are then for accelerometer and gyroscope in wrist posistion. The estimated accuracies for these in the original arcticle were: walking upstairs: 0.94 and walking downstairs: 0.96-0.97 [1]. The results that we got by using KNN with k value of 5 and three second window were: walking upstairs: 0.76 and walking downstairs: 0.84. We used the correlation method for feature selection and leave-one-participant-out crossvalidation.

| Sensor | Axis | Accuracy |
|---|---|---|
| Accelometer | $x_1$ | 0.5169 |
| Accelometer | $y_1$ | 0.7605 |
| Accelometer | $z_1$ | 0.7296 |
| Linear Accelometer | $x_2$ | 0.6242 |
| Linear Accelometer | $y_2$ | 0.6417 |
| Linear Accelometer | $z_2$ | 0.6315 |
| Gyroscope | $x_3$ | 0.5803 |
| Gyroscope | $y_3$ | 0.5726 |
| Gyroscope | $z_3$ | 0.5809 |

Fig. 12. Overall accuracy with single axis.

| Axis Combination | Accuracy |
|---|---|
| all axis | 0.9506 |
| x1, x2, x3 | 0.6309 |
| y1,y2,y3 | 0.8971 |
| z1, z2, z3 | 0.8252 |
| y1, z1 | 0.9016 |
| y1, z1, y2 | 0.9175 |
| y1, z1, y2, y3 | 0.9392 |
| y1, z1, y2, z3 | 0.9411 |
| y1, z1, y2, z2 | 0.9580 |
| y1, z1, y2, z2, z3 | 0.9656 |
| y1, z1, y2, z2, x3 | 0.9678 |
| y1, z1, y2, z2, y3 | 0.9701 |

Fig. 13. Axel combination part 1.

## VII. Discussion

*1) Effect of changing window size different positions:* Based on these results based on fig 4 if all the sensor positions data are used the best window length to be used is 3 sec which agreed with what mentioned in [1]. if sensor from single position was used then depending on the classifier used the best window length will change, if the classifier used is KNN the best window size is 4 sec regardless the features used and if LDA classifier is used then the best window size is 8 seconds.

This could be as the LDA model the assume the data to be normally distributed then as the window length increase this assumption become more true, that's why LDA preforms better on longer window length than the KNN classifier.

The best position was always the right or the left pocket and this because they are the position that can catch all the activity and can separate between them as it is in the legs which is involved in all the activity the are being recognized from the data.

The best activity recognized from each activity are sitting and standing for the sensors of the left and right pocket and this because this are the most straight forward activity as there will be no signal so it will be very easy distinguishable from

| Axis Combination | Accuracy |
|---|---|
| all axis | 0.9506 |
| y1, z1, y2, z2, x3, y3, z3 | 0.9787 |
| y1, z1,y2, z2, y3, z3 | 0.9777 |
| y1, z1, y2, z2, x3, z3 | 0.9768 |
| y1, z1, y2, x3, y3, z3 | 0.9761 |
| y1, z1, x2, y2, z2, x3, y3, z3 | 0.9755 |
| y1, z1, x2, y2, z2, x3, z3 | 0.9732 |

Fig. 14. Axel combination part 2.

all other activities. For the wrist and upper arm data the best recognized activity from them are jogging and biking and this due to that they are involved in these two activities more than other activities.

Future work that can be done that is related to this objective is to see for each activity what are the best position and combination to be used.

*2) Axel combination:* Certain axel combinations were tested, and the best overall accuracy was achieved with combination of seven axes: y and z axes of accelerometer, y and z axes of linear accelerometer and x, y and z axes of gyroscope. The most meaningful axis was y-axis because the combination of y axes of each sensor gave the best results for activity recognition. All possible combinations were not tested and the process was manual. The code could have been more smart in this part.

The results might not be generalized and we did not tested axel combination with other positions or with different window size if the results would be same or different. All in all, the worst single axis might be good to eliminate from the computing in the future studies.

*3) Feature selection:* In feature selection it was a bit surprising that the selected features varied so much between each fold although only one participant data was different. Although the problem existed only in bigger scale with correlation method when HHT features were used. The possible reason for that could lie in HHT transformation and features that it gives. As mentioned in Zeiler et al. [10] empirical mode decomposition has it drawbacks, which make it a bit unstable when it comes to the uniqueness of extracted intrinsic mode functions. Especially for window type signal, it is not guaranteed that decomposed functions would be able to reconstruct the signal totally if window segmentation is used. This might cause some problems to feature extraction. Hence, as a future work it should be studied more detailed how different parameter setting affect extraction of features when HHT is used.

Also it can be noticed that when traditional features from time space and frequency space obtained by Fourier transformation, the number of features extracted dropped in both methods. It might suggest that the information got from HHT is at least partly similar to information got by traditional methods. Actually, according to Huile et al. [9] instantaneous frequency and energy partly miss their meaning when they are transformed to features like variance of them or mean value,

| PCA | | Correlation | |
|---|---|---|---|
| All features (324) | HHT features (192) | All features (324) | HHT features (192) |
| Mean: 6 features | Mean: 82 features | Mean: 63 features | Mean: 138 features |
| Std: 0.3 features | Std: 0.4 features | Std: 3 features | Std: 23 features |
| Overall acc: 0.7655 | Overall acc: 0.8226 | Overall acc: 0.9035 | Overall acc: 0.8238 |
| Most frequently selected features | | Most frequently selected features | |
| **Acceleration:** No features that would have been selected in every time | **Acceleration:** std of 1st IMF instantaneous frequency (always calc.) | **Acceleration:** mean of 3rd IMF instantaneous energy and mean of segment | **Acceleration:** variance of marginal Hilbert spectrum (always calc.) |
| **Gyroscope:** No features that would have been selected in every time | **Gyroscope:** std of 2nd IMF instantaneous frequency (always calc.) | **Gyroscope:** var and mean of marginal HS, std of 1st IMF inst frequency, var of segment, sum of FFT coeff. | **Gyroscope:** All marginal Hilbert spectrum feat. And all instantaneous energy feat. (always cal) |

Fig. 15. The effect of different feature selection methods to selected features. The features mentioned as the most frequently calculated are those, which appeared in every fold as selected ones. Moreover, if they are determined as always calculated, it means that their value for every axel was considered.

as their time dependence is lost. If their assumption is true, it might explain why similar information might be obtained by these more traditional methods. However, before making any assumptions about their validity, there should be performed more calculations and revise the calculations in this study also, as due to lack of time, calculations were not revised after they had performed. Especially, the revision of parameters for the mode decomposition should be checked again.

When it comes to feature selection methods, more careful inspection of criteria used for feature selection should be done. In case of PCA for all features, it can be seen that the criteria was too tight and too harsh feature selection was performed, which decreased the classification accuracy. As a future work, this should be tried again but now use the knowledge obtained from the work done in previous parts and use different classifier and different window length in feature calculations. Also different selection criteria should also be tested and in this way find the optimal number of features that provide the same accuracy as the usage of all possible features.

*4) Comparison to original article:* The comparison to the results of original article by Shoaib et al. [1], reveals that the method that we used to calculate the results performed worse than their method. However, the results are not totally comparable as they had used KNN classifier with k value of 1 instead of our 5. The window length used in the article was 2 seconds when the window used in calculation of our results was 3. Moreover, they were also calculated features from magnitude vector of both sensors. In this part the magnitude vectors of each sensor were not calculated. In addition, it was not told in detail, how cross-validation was done in their study. But it can be suspected if person-wise division to different folds were used. Anyway, if the results are compared to results in other parts in this study, it can be noticed that there are some difference found even in those. Partly it is explained by the position as wrist is not the most accurate place according to our findings and activity as walking upstairs and downstairs were not the best activities for wrist sensors. Anyway, there might still be some lack of consistency in calculations, which should be revised as a future work.

## VIII. Reflection on Group Work

*1) Aino-Kaisa's thoughts:* The project work was very interesting and I got familiar with the process of how activity recognition models can be made in practise. Some data mining concepts such as cross-validation become more concrete for me. I have not so much experience with Matlab coding, but with the help of Henna and Youssef and literature, I managed to code eleven features and test different axis combination. I think the most difficult part was to write this report. I am not familiar with writing scientific text in English and I feel uncertainty about terminology. I wrote some parts of introduction, related work, data, feature extraction, results and discussion.

*2) Henna's thoughts:* Working together was a nice experience although due to lack of common time to be used tasks had to divide between each other so that every one made their own things, which were then collected together. Due to that the study was a little bit hard to make absolutely consistent. For example, in different parts there is used different feature sets as all features were not available at the same time. Perhaps the most time consuming part was finding articles and reading them as you had to get to know new methods what was used and understanding of them took time. Although, I had theoretical knowledge about performing data mining, still the practical implementation of the methods was not that straight forward thing to do but I was able to find details that had to be thought more than once, in order to do things as correctly as possible.

Anyway, as a whole the process went quite smoothly and all of us made quite equal amount of work for the project. Tasks that I contributed for this article was parts of introduction, reading and writing about 3 articles for previous work part, studying HH transformation and deciding features to calculate from it, feature selection and comparison the results to original article.

*3) Youssef's thoughts:* It was really nice working in a team, everything went very smoothly between us and I think the main reason of that is the commitment from all the team members towards the tasks distributed among us and the meetings so no one was overloaded so the work went in a fluent way and we managed to finish the project with a good quality and satisfying results in the required time slot, although it was tight. So I am really happy with this experience and I hope it could be repeated again. I worked on creating the features maps from the raw data and the classifier and testing the effect of window lengths and the sensor positions and get the best recognized activity for each of this and plotting the results and writing these in the methods, results and discussion section in the report.

## REFERENCES


[1] M. Shoaib, S. Bosch, O. Durmaz Incel, H. Scholten, and P. J. M. Havinga, Fusion of Smartphone Motion Sensors for Physical Activity Recognition. Sensors, vol. 14, 2014.

[2] J. Wannenburg and R. Malekian, Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing. IEEE Transactions on Systems, Man and Cybernetics: Systems, vol. 47, no. 12, December 2017.

[3] P. Siirtola and J. Röning, Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data. Finland Special Issue on Distributed Computing and Artificial Intelligence, 2012.

[4] A. Wang, G. Chen, J.Yang, S. Zhao, and C-Y Chang, Member, A Comparative Study on Human Activity Recognition Using Inertial Sensors in a Smartphone, IEEE Sensors Journal, vol. 16, no. 11, June, 2016.

[5] J. Zhu, R. San-Segundo and J. M. Pardo, Feature extraction for robust physical activity recognition, Human-centric Computing and Information Sciences, vol. 7, no. 16, 2017.

[6] Y. Liu, L. Nie, L. Liu and D. S. Rosenblum, From action to activity: Sensor-based activity recognition. Neurocomputing, vol. 181, pp. 108—115, 2016.

[7] S. Seto, W. Zhang and Y. Zhou, Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition, IEEE Symposium Series on Computational Intelligence, 2015.

[8] R. Chavarriga, H.Sagha, A. Calatroni, S.T.Digumarti, G.Tröster, J.del R. Millán, D. Roggen, The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition, Pattern Recognition Letters, vol. 34, no. 15, pp. 2033–2042, 2013.

[9] X. Huile, L. Jinyi, H. Haibo, Y. Zhang, Wearable Sensor-Based Human Activity Recognition Method with Multi-Features Extracted from Hilbert-Huang Transform, Sensors, vol. 16, no.12, pp. 2048, 2016

[10] Zeiler, Angela Faltermeier, Rupert Keck, Ingo Tomé, Ana Puntonet, Carlos Lang, Elmar. Empirical Mode Decomposition - an introduction, Proceedings of the International Joint Conference on Neural Networks, 1-8, 2010 10.1109/IJCNN.2010.5596829.

[11] Cunningham, Padraig Delany, Sarah. (2007). k-Nearest neighbour classifiers. Mult Classif Syst.

[12] M. B. Abidine and B. Fergani, "Evaluating C-SVM, CRF and LDA classification for daily activity recognition," 2012 International Conference on Multimedia Computing and Systems, Tangier, 2012, pp. 272-277.