Human Activity Recognition with Smartphone Data

***Abstract*—In this paper, we present a Human Activity Recognition (HAR) project using a public dataset collected from smartphone sensors. The goal of this project is to recognize 12 activities in our daily life, including 6 basic activities and 6 postural transition activities. We proposed our approaches for signal pre-processing, feature extraction, dimensionality reduction and classification, and achieved above 90% overall accuracy. Furthermore, we compared the results and analyzed the impact of parameters in different stages of the project. Our work can provide useful suggestions for other HAR systems.**

Keywords—Human Activity Recognition, Smartphone sensors, Machine Learning, Multiclass classification

I. INTRODUCTION

Human Activity Recognition (HAR) technology is used to recognize people's physical activities by analyzing the data collected from videos, sensors or other sources. It plays an important role in pervasive and wearable computing. Among different approaches, smartphone-sensor-based HAR is well-suited for daily life scenarios due to the ubiquity of smartphones. Nowadays, a variety of sensors are embedded in smartphones, such as accelerometer, gyroscope, GPS, thermometer. For HAR tasks, accelerometers and gyroscopes are most frequently used, since they are designed to monitor physical movement.

The quantity of data collected form smartphone sensors is generally very huge, so machine learning algorithms play another key role in HAR for the ability of analyzing big data. Many researchers have studied smartphone based HAR with different combinations of sensor setups and machine learning algorithms. Kwapisz et al. [1] collected data from a smartphone kept in pocket and used Logistic Regression, Neural Network, and Decision Tree as the classifiers. In [2], Anguita et al. attached the smartphone to the subject's wrist and applied Support Vector Machine to classify the activities. Shoaib et al. [3] used the fusion method of accelerometer, gyroscope and magnetometer to improve the performance of HAR.

In this project, we perform a HAR task using a public dataset from UC Irvine Machine Learning Repository [4]. It is a supervised machine learning problem. Our goal is to classify 12 classes of daily activities using raw signal data collected from the build-in accelerometer and gyroscope of a smartphone. The motivation of the project is to get an acceptable classification performance and study the impact of different parameters and algorithms in HAR process. As shown in Fig. 1, the main process of this project contains 6 steps.

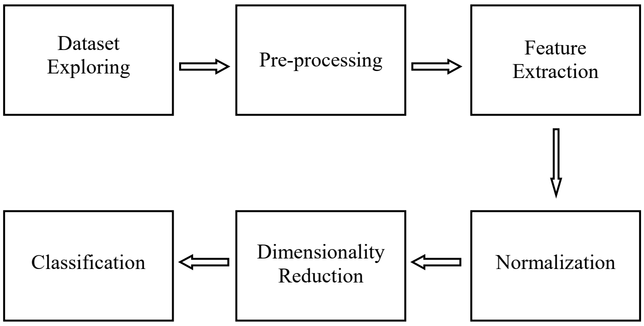

Fig. 1. Overview of the process

We design this process based on other researches [5] [6], it is a typical process for HAR problem. In section II, we will describe the dataset and research approaches in detail. Then the result of this project will be compared and discussed in section III. We conclude the project in section IV.

II. APPROACHES

A. Dataset exploring

In this project, we work on the Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set (HAPT) [7]. The dataset was collected from 30 subjects aged between 19-48, doing 6 basic activities and 6 postural transition activities. All the 12 activities are listed in Table I.

TABLE I. ACTIVITIES IN THE HAPT DATASET

Basic Activities	Postural Transition Activities
1. WALKING	7. STAND_TO_SIT
2. WALKING_UPSTAIRS	8. SIT_TO_STAND
3. WALKING_DOWNSTAIRS	9. SIT_TO_LIE
4. SITTING	10. LIE_TO_SIT
5. STANDING	11. STAND_TO_LIE
6. LAYING	12. LIE_TO_STAND

All the data were obtained by a Samsung Galaxy S II smartphone mounted on the subject's wrist. The dataset is provided with ready-to-use 561-dimensional features, as well as the raw sensor signals. We only use the raw sensor data in this project, which comprises the output of a 3-axial accelerometer and a 3-axial gyroscope, sampled at 50Hz.

B. Pre-processing

A low-pass filter is applied to smooth the signal and remove high frequency noise. We choose a moving average filter whose window size is set to 3 [8] and all data points are evenly weighted. Fig. 2 demonstrates the signal shapes before and after filtering.

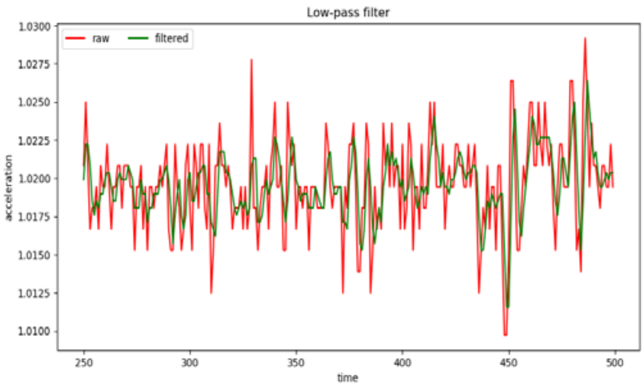

Fig. 2. Effect of low-pass filter

According to other researchers' work [5] [9] [10], a high-pass filter with very low cut-off frequency is usually used in this stage to remove or separate the DC component that represents the gravity. In this case, since the smartphone is fixed on subject's wrist uniformly, the gravity can provide very useful information, so it is retained in the signal. The impact of the gravity will be discussed in section III.

C. Feature extraction

After filtering the raw data, we apply a sliding window with 3-second length (50Hz * 3s = 150 data points) and 50% overlapping to all the signals for 6 axes (3 axes for accelerometer and 3 axes for gyroscope). The data fragment in the tail whose size is less than 3 seconds is also calculated. In each window, 9 features are extracted for every axis, and the correlation between the 3 axes within the accelerometer and gyroscope are calculate respectively. All the features are listed in Table II.

In total, there are 9 * 6 + 3 + 3 = 60 features extracted for each window, both in time domain and frequency domain. By doing this step, we have 10251 samples, each sample is a 60-dimensional vector.

D. Normalization

Machine learning algorithms are sensitive to actual values, so we need to transform all features to a common scale. In this step, all features are normalized to 0-1 by min-max normalization method.

TABLE II. FEATURES AND THEIR DEFINITIONS

Feature	Definition
MEAN	mean of the values for the requested axis in the window
MAX	maximum of the values
MIN	minimum of the values
MEDIAN	median of the values
SKEW	unbiased skew over requested axis, normalized by N-1
KURT	unbiased kurtosis over requested axis using Fisher's definition of kurtosis, normalized by N-1
STD	standard deviation normalized by N-1
ENERGY	normalized sum of squared amplitude for each data point
MEAN_FREQUENCY	weighted mean frequency calculated by PSD of the signal
CORRELATION_ACC	Pearson correlation between 3 acc axes, 3 features in total
CORRELATION_GYRO	Pearson correlation between 3 gyro axes, 3 features in total

E. Dimensionality reduction

In this step, we use Principle Component Analysis (PCA) to reduce the dimensionality of features.

According to Fig. 3, we can see 30 dimensions can provide more than 95% of the variance, so we reduce the feature dimensionality from 60 to 30.

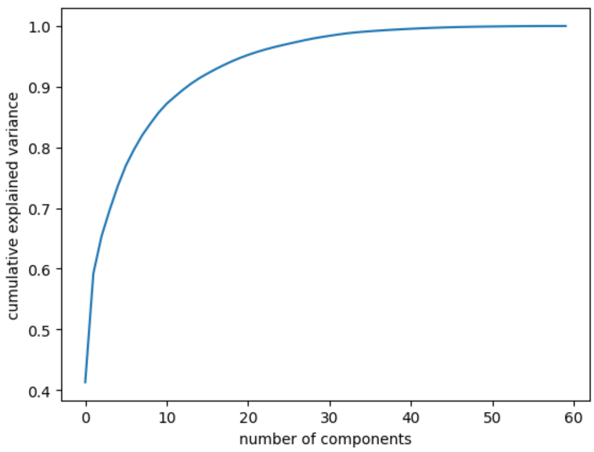

Fig. 3. Cumulative explained variance

F. Classification

We choose 4 different machine learning models to perform this 12-class classification HAR task. The data are divided to training and testing subsets by 80/20. We employ 10-fold-cross-validation strategy on the training subset to evaluate the models and tune the parameters. After that, the testing subset is used to get the final results in section III.

1) Naïve Bayes

The number of hyperparameters of Naïve Bayes is very limited. Here we simply assume the distribution of features is Gaussian. The Gaussian Naïve Bayes model is suitable for continuous features and multiclass classification problem.

2) K-Nearest Neighbors

The influence of parameter K is shown in Fig. 4. If K is small, the model will be more susceptible to noise. Based on the analysis, we set K to 9.

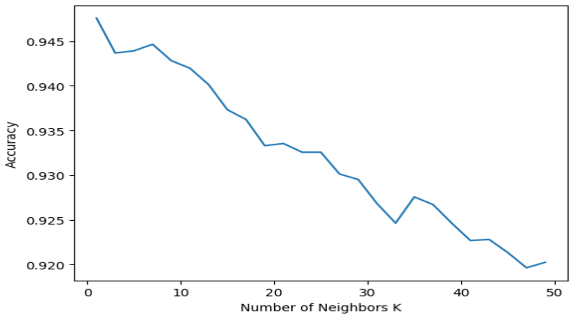

Fig. 4. Influence of parameter K

3) Support Vector Machine

The SVM model is a binary classifier itself, so we use one-vs-one strategy for multiclass classification. Other important hyperparameters are the kernel type and coefficient, penalty parameter of the error term. By performing grid search [11] on a list of hyperparameters, we choose linear kernel and set penalty parameter to 1000.

4) Random Forest

Similarly, we use grid search to tune the hyperparameters of Random Forest model, the optimal number of decision trees is 120, and the number of features when looking for the best split is $\log_2 N$ (N is the number of features).

III. Results and Discussion

A. Results

To evaluate the performance of 4 classifiers, we use the overall accuracy (correctly classified samples / all samples) as the main metric. The result is shown in Table III.

TABLE III. OVERALL ACCURACY OF THE 4 CLASSIFIERS

	Naïve Bayes	KNN	Random Forest	SVM
Accuracy	87.0%	93.9%	94.7%	**94.8%**

The SVM and Random Forest outperform others in accuracy (about 95%). On the other side, Naïve Bayes model performs the worst among the 4 classifiers, with the accuracy at 87.0%.

In this multiclass classification problem, accuracy is not enough to show details in each class, so we calculate the precision, recall and f1-score for each class.

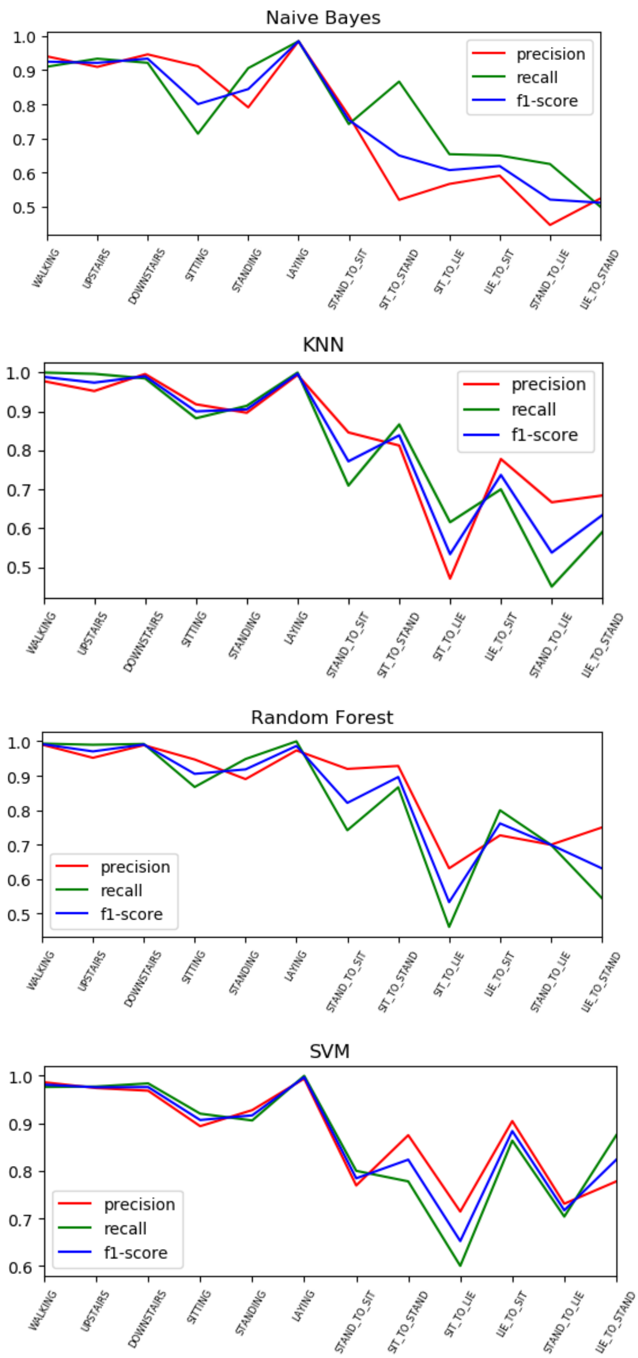

Fig. 5. Precision, recall and f1-score for each class

According to Fig. 5, it is clear that there is a significant difference between the performance of 6 basic activities (walking, upstairs, downstairs, sitting, standing, laying) and 6 postural transition activities (stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, lie-to-stand). The f1-scores for the former 6 classes share a similar pattern between 4 classifiers at above 80%. However, the f1-scores of the latter 6 classes have a higher deviation at a much lower level, from around 50% to 90%.

One explanation for this phenomenon is the distribution of training data. As shown in Fig. 6, the number of postural transition samples are much fewer than basic activities. They only have about 100 samples in each class on average. So the classifiers might not have enough data to get well trained. Although the performance for postural transition classes is poor, the overall performance is not heavily affected due to their small sample number.

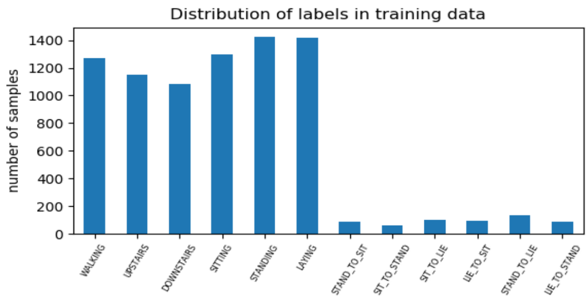

Fig. 6. Distribution of labels in training data

B. *Impact of parameters*

We have many parameters in the whole process of this project, and some of them can greatly affect the final results. In this section, we will discuss some important parameters. To simplify our comparison, we use the SVM model as the only classifier.

1) *Gravitational acceleration*

As mentioned before, we do not remove the gravitational acceleration with a high-pass filter, because gravity provides useful information in this case. We neither separate it due to the cost of complexity and the gain in accuracy.

To show the importance of gravity component, we remove the gravity by applying a high-pass Butterworth filter with 0.2Hz [12] [13] cut-off frequency in the pre-processing stage, after smoothing the signal. The signal shapes of original and filtered signals are shown in Fig. 7.

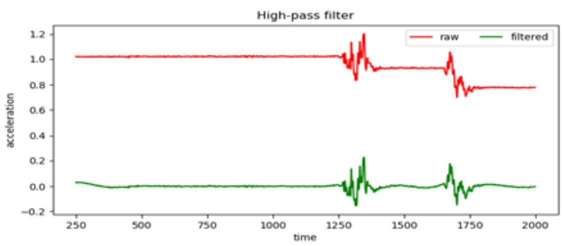

Fig. 7. Remove the gravitational acceleration with a high-pass filter

Then we do the following steps with the same parameters. Table IV shows the change of overall accuracy and Fig. 10 shows the comparison of f1-score for each class.

TABLE IV. IMPACT OF GRAVITY

	Accuracy
Gravity retained	94.8%
Gravity removed by high-pass filter	88.2%

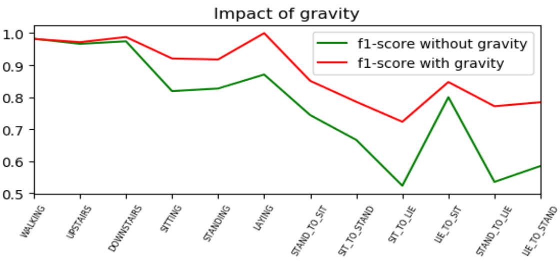

Fig. 8. Impact of gravity

If we remove the gravity by a high-pass filter, the overall accuracy will drop from 94.8% to 88.2%, and the f1-score for almost every class will decrease notably.

In this case, the smartphone is fixed to the subject's wrist, it can be considered as a smartwatch. As a result, the amplitudes of gravitational acceleration must follow certain patterns, and provide important information for classification.

2) *Window size and overlapping for feature extraction*

If we use smaller window size to extract features, we can get more samples and faster recognition speed, but the features may not reflect the activity patterns if the window size is too small. However, if the window size is large, it will reduce the number of samples and cause more lagging and computing in real world application [14]. Table V and Fig. 9 show the results of different window sizes used for feature extraction.

TABLE V. IMPACT OF WINDOW SIZE

	Number of samples	Accuracy
1-second window size	31944	90.2%
3-second window size	10251	94.8%
5-second window size	5952	95.0%

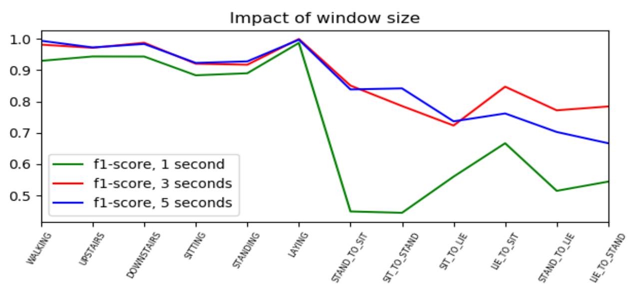

Fig. 9. Impact of window size

As shown above, the features extracted by 3 and 5 seconds windows result in very close f1-score, while 1-second window size features perform much weaker. In this case, 1 second might not be enough to demonstrate clear pattern for some activities.

Overlapping also plays an important role in feature extraction. It can increase number of data points and reduce the information loss at the boundaries. At the same time, it consumes more computing power. Table VI and Fig. 10 compare the results of different overlapping.

TABLE VI. IMPACT OF OVERLAPPING

	Number of samples	Accuracy
20% overlapping	7076	91.2%
50% overlapping	10251	94.8%
80% overlapping	22902	96.6%

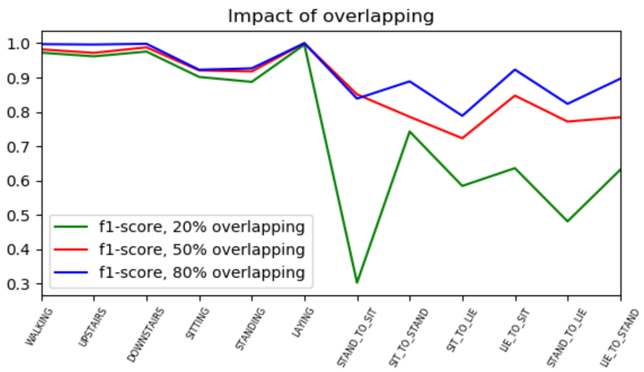

Fig. 10. Impact of overlapping

In this case, more overlapping can improve the performance, especially for the 6 postural transition activities. As we mentioned before, these classes have very small number of samples, which causes underfitting problem. Increasing the overlapping produces more samples and mitigates this problem to some extent.

3) *Dimensionality of Features*

We reduced the dimensionality from 60 to 30 in section II and captured more than 95% of the variance. To evaluate the influence of dimensionality, now we use 10-dimensional and 60-dimensional feature vectors for training and testing. The 60-dimensional features are the raw features without undergoing PCA.

TABLE VII. IMPACT OF DIMENSIONALITY

	Accuracy
10-dimensional features	85.5%
30-dimensional features	94.8%
60-dimensional features (no PCA)	95.4%

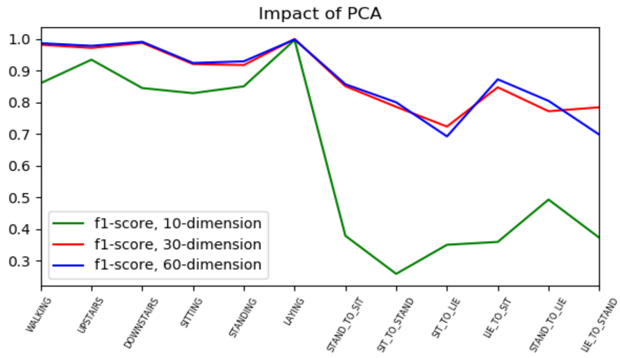

Fig. 11. Impact of feature dimensionality

As shown in Table VII and Fig. 11, using higher dimensional feature vectors (above 30) can hardly improve the performance. However, if the dimensionality number is reduced to 10, the accuracy will drop significantly from about 95% to 85%, and the 6 postural transition activities are more sensitive to this dimensionality reduction.

C. *Comparison with other researchers' work*

The HAPT dataset we used in this project is a public dataset, other researchers also proposed different methods for this dataset to improve the classification performance. Table VIII shows the comparison between our work and other papers. The result has proven the effectiveness of our approaches.

TABLE VIII. COMPARISON WITH OTHER PAPERS

Reference	Dimensionality Reduction Method	Classifier	Accuracy
Our project	PCA	SVM	94.8%
[15]	PCA	KNN+SVM	94.9%
[15]	LDA	KNN+SVM	95.7%
[16]	-	MLP with Artificial Bee Colony algorithm	95.5%
[16]	-	MLP with Genetic Algorithm	93.5%
[17]	-	LSTM	90.7%

IV. CONCLUSION

In this project, we present a Human Activity Recognition project using the HAPT dataset. The process of the project comprises signal pre-processing, feature extraction, dimensionality reduction, and classification with different models.

For this 12-class classification problem, we have reached a good result, with the overall accuracy above 90%. This result shows the effectiveness of our methods. Furthermore, we conduct a research on the influence of parameters in different stages of the project, to get a comprehensive understanding of HAR.

Our work can provide useful suggestion to other HAR systems. Besides, it is easier to apply the project into real world application, because it is smartphone-sensor-based and use raw signal data as the input.

REFERENCES

[1] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore, "Activity Recognition using Cell Phone Accelerometers," ACM SIGKDD Explorations Newsletter, Volume 12 Issue 2, pp. 74-82, December 2010.

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz, "Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine," Ambient Assisted Living and Home Care, pp. 216-223, 2012.

[3] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul J. M. Havinga, "Fusion of Smartphone Motion Sensors for Physical Activity Recognition," Sensors 2014, 14, 10146-10176.

[4] J.-L. Reyes-Ortiz, L. Oneto, A. Ghio, A. Sam´a, D. Anguita, and X. Parra, "Human activity recognition on smartphones with awareness of basic activities and postural transitions," Artificial Neural Networks and Machine Learning–ICANN 2014, pp. 177–184, 2014.

[5] Mostafa Elhoushi, Jacques Georgy, Aboelmagd Noureldin, and Michael J. Korenberg, "A Survey on Approaches of Motion Mode Recognition Using Sensors," IEEE Transactions on Intelligent Transportation Systems, Volume 18, Issue 7, July 2017.

[6] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul J.M. Havinga, "A Survey of Online Activity Recognition Using Mobile Phones," Sensors 2015, 15, 2059-2085.

[7] 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.

[8] A M Khan, Young-Koo Lee, S Y Lee, Tae-Seong Kim, "A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer," IEEE Transactions on Information Technology in Biomedicine, Volume 14, Issue5, pp. 1166-1172, September 2010.

[9] Yu-Liang Hsu, Shih-Chin Yang, Hsing-Cheng Chang, Hung-Che Lai, "Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network," IEEE Access, Volume 6, pp. 31715-31728, 23 May, 2018.

[10] Jeen-Shing Wang and Fang-Chen Chuang, "An Accelerometer-Based Digital Pen With a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition," IEEE Transactions on Industrial Electronics, Volume 59, pp. 2998-3007, Jul 2012.

[11] API document of GridSearchCV method in scikitlearn, https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

[12] Jafet Morales, David Akopian, "Physical activity recognition by smartphones, a survey," Biocybernetics and Biomedical Engineering, Volume 37, Issue 3, 2017, pp. 388-400.

[13] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," IEEE Transactions on Information Technology in Biomedicine, Volume 10, pp.156-167, January 2006.

[14] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares, and Ignacio Rojas, "Window Size Impact in Human Activity Recognition," Sensors 2014, Volume 14, pp. 6474-6499, 2014.

[15] Ihssene Menhour, M'hamed Bilal Abidine, and Belkacem Fergani, "A New Framework Using PCA, LDA and KNN-SVM to Activity Recognition Based SmartPhone's Sensors," 6th International Conference on Multimedia Computing and Systems (ICMCS), May 2018.

[16] Murat Taşkıran, Zehra Gülru Çam, and Nihan Kahraman, "An Efficient Method to Optimize Multi-Layer Perceptron for Classification of Human Activities," Communications & Instrumentation Engg. (IJCCIE) Volume 2, Issue 2, 2015.

[17] Toan H. Vu, An Dang, Le Dung, and Jia-Ching Wang, "Self-Gated Recurrent Neural Networks for Human Activity Recognition on Wearable Devices," Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pp. 179-185.