An LSTM-based Descriptor for Human Activities Recognition using IMU Sensors

Sara Ashry[1,2], Reda Elbasiony[1,3] and Walid Gomaa[1,4]
[1]*Cyber-Physical Systems Lab (CPS), Computer Science and Engineering Department (CSE), Egypt-Japan University of Science and Technology (E-JUST), Alexandria, Egypt*
[2]*Computers and Systems Department, Electronic Research Institute (ERI), Giza, Egypt*
[3]*Faculty of Engineering, Tanta University, Tanta, Egypt*
[4]*Faculty of Engineering, Alexandria University, Alexandria, Egypt*

Keywords: Human Activity Recognition, Auto Correlation, Median, Entropy, LSTM, Smart Watch, IMU Sensors.

Abstract: In this article, we present a public human activity dataset called 'HAD-AW'. It consists of four types of 3D sensory signals: acceleration, angular velocity, rotation displacement, and gravity for 31 activities of daily living ADL measured by a wearable smart watch. It is created as a benchmark for algorithms comparison. We succinctly survey some existing datasets and compare them to 'HAD-AW'. The goal is to make the dataset usable and extendible by others. We introduce a framework of ADL recognition by making various pre-processing steps based on statistical and physical features which we call AMED. These features are then classified using an LSTM recurrent network. The proposed approach is compared to a random-forest algorithm. Finally, our experiments show that the joint use of all four sensors has achieved the best prediction accuracy reaching 95.3% for all activities. It also achieves savings from 88% to 98% in the training and testing time; compared to the random forest classifier. To show the effectiveness of the proposed method, it is evaluated on other four public datasets: CMU-MMAC, USC-HAD, REALDISP, and Gomaa datasets.

1 INTRODUCTION

Researchers are continuously thinking in making everyday environment intelligent. For this purpose, human activity modeling and recognition is the basis of this research trend. Supporting accurate information on people's activities and conducts is one of the most key tasks in a widespread computing application; especially in the healthcare sector.

Activities of daily living (ADL) are the activities ordinary people have the ability for doing on a daily basis like eating, moving, individual hygiene, and dressing. For recognizing these activities, some smart homes employ various types of sensors (Bruno et al., 2012) such as microphones, various types of cameras, and motion sensors. However, those sensors have many restrictions concerning its fixed nature. For example, if the user wants to leave the place, he will not be observed from the fixed sensors and his activities won't be detectable. The other approach depends on wearable mobile-sensors which can be worn on different parts of the human body like wrists, legs, waist, and chest. Wearable accelerometers are generally used in this area, because of being small-sized, cheap, and embedded in a lot of smart-phones, watches, shoes, sensory gloves, and hand straps.

Seeing that wearable sensors are proper for continuous monitoring, they open the door to a world of novel healthcare applications like physical fitness monitoring, elder care support and etc. These applications foster recognizing human activities research by using wearable sensors. That is why; researchers focus their efforts for prototyping new wearable sensor systems, building human activity datasets, and developing machine learning techniques to recognize various types of human activities. In the current work, we focus on collecting and creating HAD-AW dataset for human activity recognition. It is clear how important the datasets in facilitating scientific research. We also introduce the ADL framework by making various pre-processing steps based on physical and statistical features which we call AMED.

Regarding wearable sensor datasets, most researchers develop activities recognition models based on their own datasets. But, some of these datasets have small size or contain a small number of volun-

Ashry, S., Elbasiony, R. and Gomaa, W.
An LSTM-based Descriptor for Human Activities Recognition using IMU Sensors.
DOI: 10.5220/0006902404940501
In *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2018) - Volume 1*, pages 494-501
ISBN: 978-989-758-321-6

Table 1: The selected ADLs and their motion primitives.

ADL	Motion Primitives (s)
Eating	Eat Sandwich with Hand.
Driving	Driving Car.
Individual Hygiene	Showering, Washing hands.
Sporting & Hobbies	Cycling, Rowing, Running, GYM weight back, Weight biceps, Weight chest, Weight shoulders, Weight triceps, Weight workout, Dancing, Drawing, Reading, Playing on Piano, Playing on Guitar, Playing on a violin.
Working	Writing on paper, Typing on keyboard.
Cooking	Cutting components, Flipping.
House Cleaning	Washing dishes, Sweeping, Wiping, Shaking the dust, Bed-making.
Others	Wearing Clothes, Put off clothes, Praying.

teers or focus on specific activities (e.g. cooking). Moreover, most of them are not accessible for public usage.

In this paper, we describe how we constructed the "HAD-AW" dataset (human activity dataset using Apple watch). We also compare it to a selection of similar existing datasets. It is created to include many of the common activities in daily life from a large and diverse group of subjects. The data is captured by an accurate inertial sensing device embedded at the smart watch. We have included 31 activities as shown in Table 1. The entire dataset are publicly available in [1]. We aim to enlarge the activities and the number of subjects in the future, and all updates will be supported on this website.

The paper is organized as follows. Section 1 is an introduction. Section 2 presents a survey of related work. Section 3 presents our collected dataset and the proposed descriptor that is fed to LSTM network and comparing the recognition performance with random forest algorithm as applied in (Gomaa et al., 2017). The descriptor is evaluated on other four public datasets as shown at the experimentation results in Section 4. Section 5 concludes the paper.

2 RELATED WORK

2.1 Used Sensors

At designing any sensor-based activity recognition system, the number of sensors and their locations are critical parameters. Regarding sensors locations, different body parts have been chosen from feet to chest. The selected locations are chosen according to the relevant activities. For example, ambulation activities such as walking, running, etc. were detected using waist sensor. Whereas, non-ambulation activities such as brushing teeth, eating, etc. can be classified effectively using a wrist sensor (Bruno et al., 2013).

Most of the related work systems require obtrusive sensors on the throat, chest, wrist, thigh, and ankle connected via wired links. It restricts the human movement, but, in healthcare applications involving elderly people or patients with heart disease, obtrusive sensors are not convenient. Furthermore, some datasets were collected under controlled conditions, and they classified a small number of activities. These drawbacks are solved in our approach by using only a smart Watch at the wrist and building a large dataset in realistic conditions without any supervision.

2.2 Major Existing Datasets

There is a limited number of the free datasets of human activities which are publicly available. We mention some of them in this section. Though each one has its own strengths, they don't meet our objective because they concentrate only on few types of activities. That is why; we have been motivated to collect and create our own dataset. A full comparison of these datasets and HAD-AW is shown in Table 2.

The problem of USC-HAD dataset that it contains only ambulation activities and the sensor placed at the hip that restrict the human movement. The REALDISP dataset (Baños et al., 2012) focused on evaluating sensor displacement in activity recognition by drawing scenarios. The first scenario is an ideal-placement: the sensors are fixed by the researcher. Secondly, self-placement scenario: the volunteer should put sensors himself on body parts. This dataset drawback is that it only focuses on fitness activities by using 9 obtrusive sensors. Regarding the CMU-MMAC (De la Torre et al., 2008) dataset; although it contains higher population-size and more plentiful modalities than any other dataset, it focuses only on cooking activities.

3 PROPOSED DATASET AND DESCRIPTOR FOR LSTM

3.1 HAD-AW Dataset

To overcome the limitations of the other existing datasets which is mentioned in section2.2, our dataset

[1]The HAD-AW dataset and its description are available on: https://www.researchgate.net/publication/324136132_HAD-AW_Dataset_Benchmark_For_Human_Activity_Recognition_Using_Apple_Watch

Table 2: Comparison between some of the existing datasets and HAD-AW dataset where F refers to female and M to male.

Dataset	Number of Subjects	Activities	Sensor Locations	Sensors	Comments
CMU-MMAC (De la Torre et al., 2008)	32	Food preparation, Cook five recipes: Pizza, Sandwich, Brownies, Scrambled eggs, Salad	Left and right forearm, arms, calves, thighs, wrists, abdomen and forehead.	Camera, Microphone, RFID, 3D gyroscopes, 3D accelerometers, 3D magnetometer, Ambient light, Heat flux sensor, Galvanic skin response, temperature, Motion Capture.	The dataset focuses on cooking activities and it is obtrusive due to many sensors on human body.
USC-HAD (Zhang and Sawchuk, 2012)	14 (7 F, 7 M)	Walk forward, walk left, walk right, walk up-stairs, walk down-stairs, run forward, jump, sit on chair, stand, sleep, elevator up, and elevator down	Front right hip	3D accelerometers, 3D gyroscopes	Data taken from one sensor location but it focused on ambulation activities.
REALDISP dataset (Baños et al., 2012)	12	Walking, Jogging, Running, Jump , Trunk twist, Waist/Lateral bends, forward stretching, Arms elevation/ crossing, Cycling, hand claps, Knees bending, Shoulders rotation and Rowing.	Left and Right calf, Left and Right thigh, Left and Right lower/upper arm, Back	3D acceleration, gyroscope, magnetic field, and 4D orientation.	The Dataset focused only on warm up, fitness and cool down exercises and sensors placed on 9 obtrusive body parts.
(Gomaa et al., 2017) dataset	3	Use telephone, Drink from glass, Pour water, Eat with knife/ fork, Eat with spoon, Climb/ Descend stairs, Walk, Get up/Lie down bed, Stand up/ Sit down chair, Brush teeth, Comb hair	Right wrist only.	3D accelerometers, 3D angular velocity, 3D rotation, 3D gravity.	Fourteen activities are collected by Apple watch on right wrist.
HAD-AW	16 (9 F, 7 M).	Thirty-one activities are mentioned in table 1.	Right wrist only.	3D accelerometers, 3D angular velocity, 3D rotation, 3D gravity.	Large dataset is collected by only Apple watch on right wrist. 31 different activities are recorded.

has been carefully designed with the following aims: Firstly, many subjects with variation in gender, age, weight, and height are considered in the dataset. Secondly, the recorded activities match the most common and basic human activities in their daily lives as shown in Table 1. It is useful for elder care, and personal fitness monitoring applications. Finally, the used smart watch captures most of the human activity signals accurately and robustly.

3.1.1 Data Collection

Sixteen subjects (9 females and 7 males) are volunteered to record activities by wearing an Apple watch on the right wrist. Each subject repeats the same action about ten times on average. So, we have 160 sample for each activity on average. Volunteers ages range from 20 to 55 years old with weights range from 55 kg to 95 kg. The raw signals contain information about 3D angular velocity, 3D orientation (roll, pitch, and yaw), 3D gravity components, and 3D user acceleration where the sampling frequency is 50 Hz. Figure 1 shows snapshots of the data collection process. Figure 2 shows sample of activities data from the z-axis of the tri-axial accelerometer. We collected our dataset using Apple watch series one. It weighs only about thirty grams and it has a Dual-Core processor, eight GB internal storage, and 512 MB RAM. The newest version of it has more storage capacities.

Figure 1: Snapshots of performing different ADL activities.

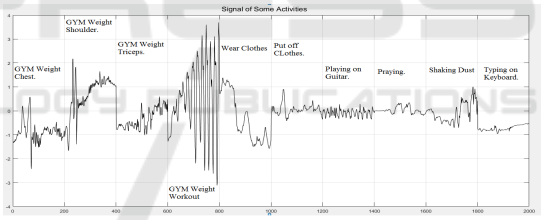

Figure 2: Sample of activities data from the z-axis of the tri-axial accelerometer.

3.2 Descriptor for LSTM

3.2.1 LSTM Recurrent Neural Network

An LSTM layer is a recurrent neural network (RNN) layer which supports time and data series in the network. The layer does extra interactions in the training stage to help improving gradient flow over sequences. The greatest advantage of the recurrent neural networks is their capability to take the contextual information into consideration when mapping between input and output sequences through hidden layer-units. It can automatically detect general features and capture the temporal dependencies between the features sequence (Donahue et al., 2015).

The descriptor features, which works as a classifier for the whole feature sequence. Each LSTM unit takes the updated network state from the previous unit as shown in Fig. 3 and Fig. 4. However, LSTM has a disadvantage of tuning a lot of parameters that need to be chosen carefully as mentioned in section 4.3.

Equations 1, 2 define the cell state (c) and the output or hidden state (h) at time step t while i, f, g, and

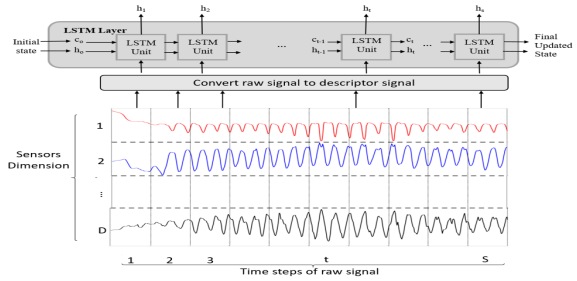

Figure 3: The flow of a time series X with D sensors signals through an LSTM layer where h symbolizes the hidden state and c is the cell state.

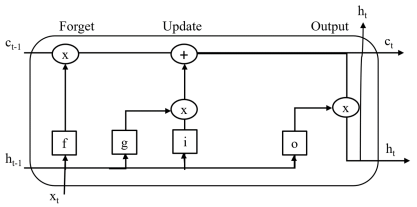

Figure 4: The LSTM unit where blocks f, i, o are sigmoid function and block g is tanh function as illustrated in Equation 3, 4, 5, 6.

o in Equations 3, 4, 5, 6 [2] show the input gate, forget gate, layer input, and output gate respectively. $\odot$ refers to the element-wise vectors multiplication and σ symbolizes the sigmoid function. The learnable parameters of the LSTM are the input weights (W), the recurrent weights (R), and the bias (b).

$$CellState(c_t) = f_t \odot c_{t-1} + i_t \odot g_t \quad (1)$$
$$HiddenState(h_t) = o_t \odot \tanh(c_t) \quad (2)$$
$$InputGate(i_t) = \sigma(W_i X_t + R_i h_{t-1} + b_i) \quad (3)$$
$$ForgetGate(f_t) = \sigma(W_f X_t + R_f h_{t-1} + b_f) \quad (4)$$
$$LayerInput(g_t) = \tanh(W_g X_t + R_g h_{t-1} + b_g) \quad (5)$$
$$OutputGate(o_t) = \sigma(W_o X_t + R_o h_{t-1} + b_o) \quad (6)$$

3.2.2 Proposed AMED Descriptor for LSTM

In this paper, we aim at creating a robust and discriminative descriptor to classify human activities with high accuracy. Although LSTM automatically extracts some general features, it doesn't achieve the required accuracy as will be shown in the experiment section. In contrast, providing specific engineered features facilitates the training process and increases the classification accuracy. It also helps in reducing the training and testing time. Figure 5 shows the proposed framework at the training phase for learning the network while the testing phase has the same preprocessing step. I used Matlab for coding.

[2]Permission is taken from the MathWorks company for using the equations and imitating figure 3, 4 for illustration.

First, the system divides each training raw signal into multiple parts with lengths 10 second for each part or (500 readings as sampling rate equals 50 HZ). By drawing the activities signals, we have observed that 10 sec has enough representation for most of activities pattern. Then, the autocorrelation function with a certain lag is applied to the data as shown in Equation 7, 8 where h is the function lag and $\bar{x}$ is the sample mean. Following that, the median and entropy features are concatenated with the output of the autocorrelation function to form a complete descriptor. We call the descriptor 'AMED' (Autocorrelation Median Entropy Descriptor). Figure 6 shows the descriptor applied on different activities which produces discriminative pattern for each activity.

$$acf(h) = \frac{\gamma(h)}{\gamma(0)} \quad (7)$$
$$\gamma(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}) \quad (8)$$

The lag parameter is a tuning parameter which is chosen based on trial and error experiments. When it equals to 20, the highest accuracy is achieved. Without using any descriptor, the normal LSTM deal with the input signal as raw data with length 500 in our data. While using the proposed descriptor; which convert the raw single input data with length 500 to the AMED signal with length 23; reduces the feature space by 95.4% for each sensor signal. It doesn't only reduce the features space, but significantly minimizes the execution time, the storage space, and maximizes the accuracy as illustrated in details in section 4.1.

4 EXPERIMENTS

In this section, we evaluate our proposed algorithm and HAD-AW dataset via two groups of experiments. The first group studies the change in classification accuracy which results from changing the number of activities, as well as the type and number of the selected sensory data. It also focus on studying the effect of using different LSTM descriptors on both the accuracy and time of the classification process.

The second group of experiments makes a comparison between our proposed algorithm and one of the recent RF-based approach which has been proposed in (Gomaa et al., 2017) for the same purpose of recognizing human activities. The two algorithms are applied on five different datasets.

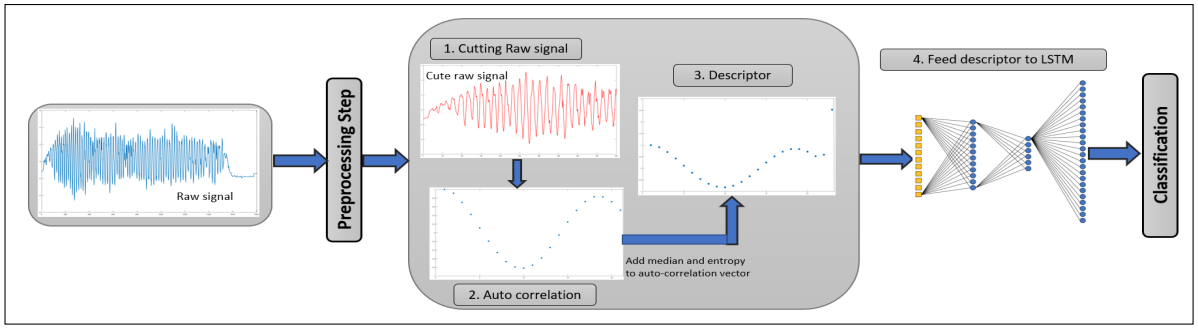

Figure 5: A simple visualization for the proposed framework in the training phase. Note that in the real experiment the input of the preprocessing step is a matrix with dimension N*M where N is the horizontal length of 1D sensor data and M is the length of all training samples of the 31 activities, where each sample has vertical length with all 3D sensors signals.

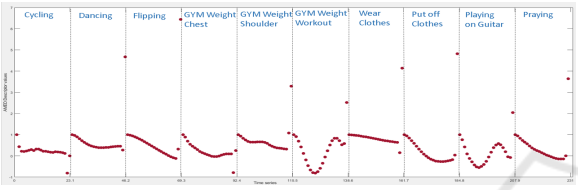

Figure 6: Descriptors signal series of different activities.

4.1 Experiments on HAD-AW Dataset

First, in order to study the effect of changing the number of activities and relevant sensory data on the accuracy, we performed a total of 49 evaluation processes grouped into a set of 7 main tests, each test contains 7 experiments. In each test, various sensors number are used as illustrated in Table 3.

Each test contains 7 experiments, where, in each experiment we try to classify different number of motion primitives as illustrated in Table 4. This is done by adding four more activities randomly to each experiment until reaching the last experiment. The last experiment includes the whole set of activities. Figure 7 shows the accuracy of the AMED descriptor of LSTM on seven main tests containing the seven experiments of different activities counts. Figure 8 shows the accuracy of the selected RF algorithm (Gomaa et al., 2017) when applying the same experimental setup described in Tables 3, 4.

Next, we performed other 8 experiments using different input features for LSTM as illustrated in Table 5, where all the experiments try to classify all activities in HAD-AW dataset using all sensory data. Table 5 shows the resulted accuracy and the total time of the training and testing.

Table 3: Experiments setup.

Test	Sensors type used in Experiment
Test 1	3D Rotation
Test 2	3D Angular Velocity
Test 3	3D Gravity
Test 4	3D Acceleration
Test 5	3D Angular Velocity, 3D Acceleration
Test 6	3D Rotation, 3D Angular Velocity, 3D Acceleration
Test 7	3D Rotation, 3D Angular Velocity, 3D Gravity, 3D Acceleration

Table 4: Motion primitives in each experiment.

Experiments	Activities on Each Experiment
Exp. 1	Flipping, drawing, cycling and cutting components.
Exp. 2	Exp.1 activities plus running, driving, eat with hand and playing on Guitar.
Exp. 3	All activities on Exp.2 plus GYM weight back, biceps, chest and shoulder.
Exp. 4	All activities on Exp.3 plus GYM weight triceps, workout, washing dishes and bed making.
Exp. 5	All activities on Exp.4 plus wear clothes, sweeping, typing on keyboard and washing hands.
Exp. 6	All activities on Exp.5 plus reading, rowing, writing on paper and wiping.
Exp. 7	All activities on Exp.6 plus dancing, put off clothes, praying, shaking dust, showering, playing on piano and playing on a violin.

4.2 A Comparison with Random Forest using Different Datasets

To evaluate the proposed algorithm, we compared it with the RF-based algorithm at(Gomaa et al., 2017) by applying both algorithms to five different datasets including HAD-AW as illustrated in Table 6.

The experiments setup for testing all activities of each dataset is as follows: Activities on CMU-

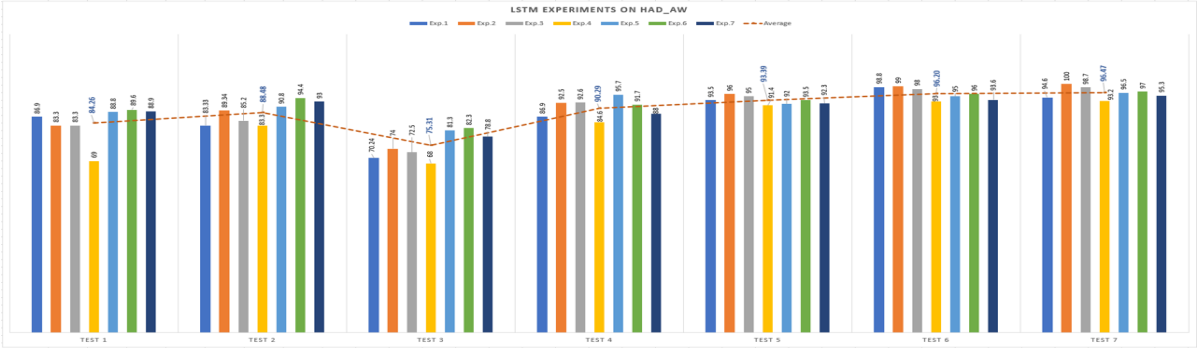

Figure 7: Accuracy of the proposed LSTM descriptor on HAD-AW with different setup categories as shown in tables 3, 4.

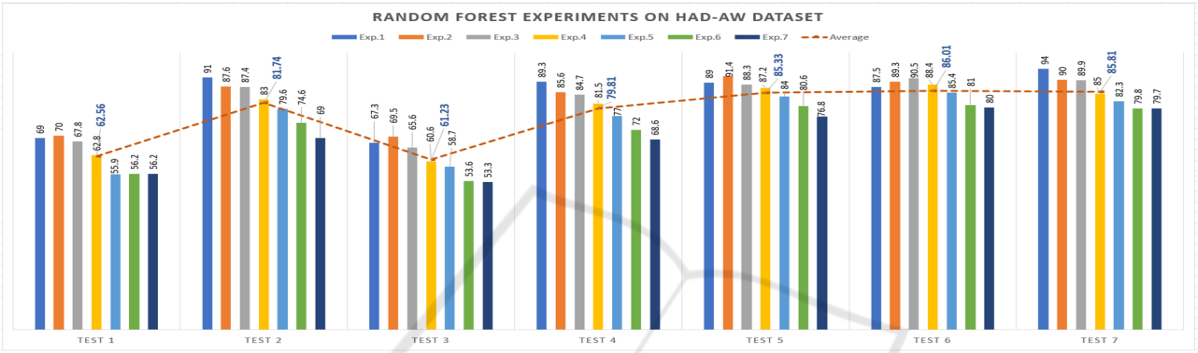

Figure 8: Accuracy of the RF-based method used in (Gomaa et al., 2017) on HAD-AW dataset applied on different setup categories as described in tables 3, 4.

Table 5: A comparison between different Experiments with respect to accuracy and computational time of training and testing. Experiments include all activities and all sensors where we test the HAD-AW dataset on different LSTM configuration regards to number of epochs and using various input feature vector for feeding to LSTM.

Experiment	Description	Input	Epochs	Accuracy	Time (sec)
Exp. 1	Raw data.	500	500	55%	780
Exp. 2	Normalized raw data.	500	500	50%	740
Exp. 3	Autocorrelation with lag 20 for normalized data	21	500	65%	200
Exp. 4	Autocorrelation with lag 20 for raw data.	21	500	75%	220
Exp. 5	Autocorrelation with lag 20 concatenated to median at raw data.	22	500	83%	228
Exp. 6	Autocorrelation with lag 20 concatenated with median, and entropy for raw data (AMED Descriptor).	23	150	89%	180
Exp. 7	AMED Descriptor.	23	250	92%	220
Exp. 8	AMED Descriptor.	23	500	95.3%	350

MMAC (De la Torre et al., 2008); which are recorded by random four subjects; are tested using 3D accelerometer and gyroscope sensors placed on the right wrist. The USC-HAD dataset (Zhang and Sawchuk, 2012) are tested using 3D accelerometer and gyroscope sensors recorded by all subjects. In REALDISP Dataset (Baños et al., 2012), we tested all activities recorded by all volunteers using 3D accelerometer and gyroscope and we used the data recorded for the self-placement scenario as illustrated in section 2.5. At Gomaa dataset; (Gomaa et al., 2017); 3D accelerometer, 3D rotation and 3D angular velocity sensors are used in recognition experiments. For HAD-AW dataset, we used the data recorded by all subjects and all IMU sensor of Apple watch.

4.3 Discussion

From figures 7 and 8, by calculating the average accuracy for the all experiments of each test, we notice that test 3 which used 3D gravity only achieved the lowest average accuracy (75.31%, 61.23% for LSTM and RF respectively). Tests 1 and 2 achieved reasonable average accuracy 84.26%, 88.48% for LSTM

Table 6: A comparison between RF (Gomaa et al., 2017) and the proposed LSTM-based method using different public datasets. We use some measuring metrics like accuracy, sensitivity, specificity, precision, and F-measure by getting the average of all these measurement over all classes.

Dataset	Method	Accuracy	Sensitivity	Specificity	Precision	F-Measure	Time (Sec)
CMU-MMAC	RF	61.23%	43.71%	66.08%	46.84%	43.77%	20734
	LSTM	84.44%	61.50%	90.19%	63.68%	61.57%	479.6
USC-HAD	RF	78.5%	70.45%	98.2%	60.5%	65.1%	35460
	LSTM	96.31%	88.32%	85.19%	90.50%	90.8%	600
REALDISP Dataset	RF	89.58%	71.21%	84.6%	67.58%	69.34%	474.67
	LSTM	94.5%	88.38%	99.58%	93.16%	90.7%	50
(Gomaa et al., 2017)	RF	81.64%	82.47%	98.67%	84.6%	83.53%	8515
	LSTM	93.2%	88.29%	99.52%	91.80%	89.22%	128
HAD-AW	RF	79.74%	81.78%	99.25%	81.87%	81.82%	18734
	LSTM	95.3%	92.9%	90.3%	93.7%	93.3%	350

and 62.56%, 81.74% for RF. However, Tests 4 and 5 achieved higher average accuracy 90.29%, 93.39% for LSTM and 79.81%, 85.33% for RF. Finally tests 6 and 7 achieved the highest average accuracy 96.2%, 96.47% for LSTM and 86%, 85.81% for RF. Thus, there is an evidence showing that the accuracy of the classification process is strongly affected by the selected sensory information.

We noticed that using combinations of all sensory signals generally improve the accuracy to reach 95.3% for all activities on experiment 7 of test 7 in figure 7. However, using both the acceleration and angular velocity only also achieved an acceptable high accuracy of 92.5% for LSTM over the all activities on experiment 7 of test 5.

In addition, Table 5 shows the accuracy and the processing time while trying different features on LSTM where the experiments include all activities and all sensory data of the HAD-AW dataset. The AMED descriptor in experiment 8 achieves the highest accuracy of 95.3% and save 55.12% from the execution time of using raw data without any preprocessing in experiment 1 and it enhanced the accuracy between two experiments by 40.3%. Additionally, in the AMED descriptor, the size of the feature vector for each sensor signal is only 23 while the raw signal input length is 500. Thus, it saves the storage space by 95.4%. This shows the importance of using effective features rather than using raw data. The proposed descriptor also save 98.3% of the execution time of training and testing achieved by the RF method in (Gomaa et al., 2017) as RF execution time is 18734 sec for experiment 7 and test 7 at figure 8.

When we have tried to normalize the raw data in experiment 2, the accuracy is not enhanced as the normalization reduces the band between features value and make them between the same range which reduces the discrimivatiy between them and affects negatively on the accuracy. We have tested different features as in experiment 4 and 5, but they didn't achieve a high accuracy. AMED descriptor in experiment 6, 7, 8 achieved an accuracy of 89%, 92%, 95.3% for the epochs number equals to 150, 250, and 500 respectively with the best accuracy at epochs 500. When we tried to increase the epochs up to 500, the accuracy value became almost constant.

Table 6 shows some evaluation metrics for testing the proposed approach of LSTM on different public datasets compared with the RF-based algorithm proposed in (Gomaa et al., 2017). Obviously, the total accuracy enhancement for all different datasets ranges from 5% to 23.3% as shown in Table 6 in addition to the execution time of training and testing data saved by around 88% to 98%. This enhancement can be interpreted by taking into consideration that the random forest computational cost is $O(M(m*nlogn))$ where n refers to the number of training instances, m refers to the dimension of the feature space and M is the size of the forest. The number of trees is typically on the order of thousands. Hence, random forest consumes more execution time. The most challenging step in the proposed method is tunning the LSTM parameters such as the number of epochs, batch size, and weights size. We have empirically; using trial-and-error; chosen the batch size to be 50 and the weights size to be 70. The epochs that achieved the best accuracy on testing different datasets is chosen to be 500.

The CMU-MMAC dataset is tested by other researchers algorithms. For instance, the performance of testing IMU data alone of one subject by using 1-NN is 56.8% (Spriggs et al., 2009). Additionally, the authors in (Zhang and Piccardi, 2015) presented partial-ranking structural SVM (PR-SSVM) approach with an accuracy equals to 69.8% on this dataset. Our proposed method accuracy is 84.44% for recognizing activities recorded by random four subjects.

Regarding the USC-HAD dataset, the authors in (Politi et al., 2014) proposed a feature extraction approach for SVM algorithm which achieved accuracy of 93.52%. They compare their proposed method

of SVM with other three algorithms like MLP algorithm, IBK approach and J48 method that achieved an accuracy equals to 90.37%, 88.72% and 89.33% respectively. Our approach outperform all the previous work with an accuracy reaches 96.3%.

Regarding the REALDISP dataset, authors in (Banos et al., 2014) tested the activities after extracting the best features on the decision trees, K-Nearest Neighbor, Nave Bayes algorithms. The accuracies were 90%, 96%, and 72% respectively for the ideal-location setting data. It was 78%, 89%, 65% for the self-placement data. The difference between the ideal-location setting and the self placement data are illustrated in section 2.2. Our approach achieves 94.5% for the the ideal-placement data.

5 CONCLUSION AND FUTURE WORK

This paper introduced the HAD-AW dataset; which includes 31 human activities; as a reference source for human activity recognition research by using a smart watch. Additionally, we presented a human motion recognition framework based on tri-axial sensory data of IMU sensors. The framework exploits a feature reduction as a preprocessing step where the raw signals are parameterized by a combination of some statistical and physical features.

The experimental results indicated that the recognition accuracy reaches 95.3% for HAD-AW dataset, 96.3% for USC-HAD dataset (Zhang and Sawchuk, 2012), 84.44% for CMU-MMAC dataset (De la Torre et al., 2008), 94.5% for the REALDISP dataset (Baños et al., 2012), and 93.2% for the dataset in (Gomaa et al., 2017). Moreover, it saved the executing time of training and testing by 88% to 98% compared to the RF-based method for different datasets. It is worth mentioning that when we used the proposed approach to test the combination of all 31 activities of HAD-AW dataset and the 14 activities of (Gomaa et al., 2017) , the total accuracy reaches 90.2% for the whole 45 combined activities.

In the future, We aim to increase the number of activities and collect new dataset using Myo device which contains data from IMU sensors, electromyographic (EMG) sensors and magnetometer. We will compare the recognition accuracy between both Apple watch and Myo device by using different algorithms. We also plan to collect another dataset of daily human activities in a full continuous stream scenarios and developing approaches for recognizing them.

REFERENCES

Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M. A., and Amft, O. (2012). A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1026–1035. ACM.

Banos, O., Toth, M. A., Damas, M., Pomares, H., and Rojas, I. (2014). Dealing with the effects of sensor displacement in wearable activity recognition. *Sensors*, 14(6):9995–10023.

Bruno, B., Mastrogiovanni, F., Sgorbissa, A., Vernazza, T., and Zaccaria, R. (2012). Human motion modelling and recognition: A computational approach. In *Automation Science and Engineering (CASE), 2012 IEEE International Conference on*, pages 156–161. IEEE.

Bruno, B., Mastrogiovanni, F., Sgorbissa, A., Vernazza, T., and Zaccaria, R. (2013). Analysis of human behavior recognition algorithms based on acceleration data. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1602–1607. IEEE.

De la Torre, F., Hodgins, J., Bargteil, A., Martin, X., Macey, J., Collado, A., and Beltran, P. (2008). Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. *Robotics Institute*, page 135.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., and Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634.

Gomaa, W., Elbasiony, R., and Ashry, S. (2017). Adl classification based on autocorrelation function of inertial signals. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on*, pages 833–837. IEEE.

Politi, O., Mporas, I., and Megalooikonomou, V. (2014). Human motion detection in daily activity tasks using wearable sensors. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 2315–2319. IEEE.

Spriggs, E. H., De La Torre, F., and Hebert, M. (2009). Temporal segmentation and activity classification from first-person sensing. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference On*, pages 17–24. IEEE.

Zhang, G. and Piccardi, M. (2015). Structural svm with partial ranking for activity segmentation and classification. *IEEE Signal Processing Letters*, 22(12):2344–2348.

Zhang, M. and Sawchuk, A. A. (2012). Usc-had: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1036–1043. ACM.