A Recommendation Specific Human Activity Recognition Dataset with Mobile Device's Sensor Data

Alexandros Vrochidis, Vasileios G. Vasilopoulos, Konstantinos Peppas, Valia Dimaridou, Iordanis Makaratzis, Apostolos C. Tsolakis, Stelios Krinidis, Dimitrios Tzovaras

A Recommendation Specific Human Activity Recognition Dataset with Mobile Device's Sensor Data*

Alexandros Vrochidis [0000-0002-9738-1553] and Vasileios G. Vasilopoulos [0000-0003-0542-2467] and Konstantinos Peppas [0000-0001-6341-1294] and Valia Dimaridou and Iordanis Makaratzis and Apostolos C. Tsolakis [0000-0003-2606-1402] and Stelios Krinidis [0000-0002-9666-7023] and Dimitrios Tzovaras [0000-0001-6915-6722]

Information Technologies Institute, Centre for Research and Technology Hellas, 57001, Thessaloniki, Greece
`krinidis@iti.gr`

Abstract. Human activity recognition is a challenging field that grabbed considerable research attention in the last decade. Two types of models can be used for such predictions, those which use visual data and those which use data from inertial sensors. To improve the classification algorithms in the sensor category, a new dataset has been created, targeting more realistic activities, during which the user may be more prompt to receive and act upon a recommendation. Contrary to previous similar datasets, which were collected with the device in the user's pockets or strapped to their waist, the introduced dataset presents activities during which the user is looking on the screen, and thus most likely interacts with the device. The dataset from an initial sample of 31 participants was gathered using a mobile application that prompted users to do 10 different activities following specific guidelines. Finally, towards evaluating the resulting data, a brief classification benchmarking was performed with two other datasets (i.e., WISDM and Actitracker datasets) by employing a Convolutional Neural Network model. The results acquired demonstrate a promising performance of the model tested, as well as a high quality of the dataset created, which is available online on Zenodo.

Keywords: Mobile Inference, Sensor Mining, Human Activity Recognition, Deep Neural Network.

1 Introduction

Human Activity Recognition (HAR) is a research topic with applications found in a wide variety of fields. It provides information about someone's activity automatically and unobtrusively, by using sophisticated technologies, such as computer vision and machine learning [1]. In general, depending on the application, the human activities examined can be either simple like jogging or walking, or more complex like peeling a potato. A characteristic example of HAR can be found in the sports field [2], where it

*This work has been co-funded by the European Union and the General Secretariat of Research and Technology , Ministry of Development \& Investments, under the project (SIT4Energy - T2EDK-0911) of the Bilateral S\&T Cooperation Program Greece – Germany 2017.

can help recognize the sport-related activities over common domestic activities. According to Chen and Shen [3], each individual has its own specific and discriminative movement patterns and should be treated accordingly. Resulting such complexities, an apparent challenge arises, which can be addressed only if there is an adequate amount of data to properly explore the extreme heterogeneity identified. Hence, the need of creating and making available HAR datasets of high quantity and quality becomes apparent.

Quite a lot of such datasets have been collected over the years, not always in the most optimal way. There are datasets for video methods like the 20BN-something-something V2 [4], the VLOG dataset [5], and the EPIC-KITCHENS [6]. This kind of dataset requires camera setups, which introduces a greater risk of violating the privacy of participants, which is one of the reasons that it is more difficult to create this kind of datasets. To tackle this problem, Ryoo et al. [15], proposed a fundamental approach of HAR, that uses low-resolution anonymized videos. In this way, privacy is protected and computing costs are reduced. However, Beyond privacy, there are also additional issues identified and highlighted in the literature. Zhang et al. [12], state what difficulties appeared, on this kind of research, like long-distance and low-quality videos, which are common in these cases. To mitigate such problems, more expensive approaches are used, either in terms of equipment or processing. Pradhan et al. [13] proposed a system based on event-based camera data while facing the very low latency and data sparsity coming from event-based vision sensors. Depth cameras can also be used for classifying human activities. In their research Jalal et al. [14], created a robust HAR model from analyzing continuous sequences of depth map.

On the other hand, sensor-type datasets are not so heavy to manipulate and they do not usually face privacy problems. Some datasets in this category are the Physical Activity Monitoring for Aging People (PAMAP2) [7], which contains 18 different physical activities, the University of Southern California Human Activity Dataset (USC-HAD) [8], which is consisted of low-level daily activities like cleaning the house and the Wearable Human Activity Recognition Folder (WHARF) [9], which includes accelerometer data recordings that are used for the recognition of simple daily life activities with wearable sensing systems. These datasets, used wearable sensors like heart rate monitors, watches, and devices that include IMUs sensors. On the contrary, the Wireless Sensor Data Mining (WISDM) [10] and Actitracker [11] datasets, which are quite extensively being used in the literature and are further analyzed in the next section, are two datasets containing data coming only for mobile devices inertial sensors. Such datasets become more and more important, as mobile devices have become a necessity and are a data-rich environment for various applications.

The dataset presented in the present study belongs to the sensor type of dataset, as it uses triaxial data, coming from inertial sensors of mobile phones. The challenge that the introduce dataset aims to address is that of identifying the optimal timing for delivering a recommendation or a nudge to the user, with the highest probability of acting upon it. Following the notion of micro-moments [16], the ideal moment to deliver a recommendation to the user, with the maximum impact, can be found during or in between other activities, and not when the user is not actively interacting with the mobile device. The ultimate goal is to employ this dataset towards training an activity tracking model that will be able to recognize activities while the mobile screen is activated, meaning that most likely the user is looking at it and interacts with the device.

To create this dataset, an android mobile application was developed and was afterwards circulated to a wide audience throughout Europe. Users were called to perform 10 different activities (i.e., variations of standing, siting, walking, running, lying down, ascending or descending stairs) while they were looking at the mobile screen. Following the authors' re-definition of Micro-moments [16], the activity tracking recognition is only used when the mobile device's screen is turned on because these moments produce the highest probability of the user looking at the device and being alert. Previously used datasets like WISDM [10] and Actitracker [11], were collected through devices that were inside the user's pockets or strapped to a belt in their waist. In this case, for all activities (except the one that mobile is left on a table surface), the dataset is gathered while users look at the phone screen.

To properly assess the created dataset, a benchmarking is also delivered over these three datasets, using a state-of-the-art Convolution Neural Network (CNN), towards evaluating not only the performance of such an approach over the datasets, but also the quality of the dataset itself.

The rest of this paper is organized as follows. In Section 2, related work on both the challenge addressed and other related datasets is presented. In Section 3, the methodology for creating the dataset is described in detail, followed by brief documentation of the CNN employed for the evaluation, including the relevant metrics in Section 4. Then, Section 5 presents the evaluation results of the dataset. Finally, Section 6 concludes the manuscript, along with future improvements.

2 Related Work

2.1 Human Activity Recognition

Human activity recognition (HAR) is the problem of identifying the specific action of a human, based on sensor data and it is a time-series classification task that is challenging. Actions like these can be, specific movements when someone is indoors or they can also be activities like walking, jogging, and ascending stairs. Sensor data can remotely be recorded and recognition tasks like that seek a profound high-level knowledge of human activities from sensor data. Many of the proposed methods are including deep learning. Wang et al. [17], surveyed the advance of deep learning-based sensor activity recognition. They tried to find which models are the best ones for each activity recognition challenge. Each model is proposed for different activity recognition, depending on its length and type.

Hassan et al. [18], proposed a deep learning method for human activity recognition, that uses smartphone inertial sensors. They first extracted features, using Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA) to make their model robust. Then they trained a Deep Belief Network (DBN) for activity recognition. Comparing their method with Support Vector Machine (SVM) and Artificial Neural Networks (ANN) models, they found that their model has better accuracy in both traditional and non-traditional activities. The overall accuracy of their model was 95.85%.

Two years later, Peppas et al. [19], approached the problem by using Convolutional Neural Networks, proposing a model, which can make real-time physical activity recognition on smart mobile devices. Tri-axial accelerometer data were used, taken from mobile devices of humans when they were in the activities. Then by using a two-layer Convolutional Neural Network (CNN), they achieved 94.18% accuracy on the WISDM dataset, while they reduced storage space by 5-8 times.

HAR can be used for environmental purposes too. In [16], a Google Activity Recognition API was used, for recognizing the user's physical activity, while possible detected activities in this API are in a vehicle, on a bicycle, on foot, still, tilting, and unknown. Researchers used this, for recognizing activities like when someone's device is still and when it is tilting. Tilting is recognized when a device's angle is changed significantly. When someone picks his phone from the table, or when he is sitting and then standing up, the API classifies the activity as tilting. After recognizing the activities, they tried to find some moments, which were redefined as micro-moments, in which energy-related recommendations will be sent to users, to maximize their receptiveness. Their purpose was to create a novel approach to changing energy behavior by using a mobile application for exploiting user attributes of micro-moments.

2.2 Previous HAR Datasets

The Wireless Sensor Data Mining project [10] was developed by Fordham University. Its goal was to explore the research issues related to mining data from mobile device's inertial sensors. An Android-based application was built to collect data from users. Phones that contained tri-axial accelerometers were used to produce data. Their users were 29 volunteers which were asked to do some specific activities while they had their mobile phone in their pocket. These activities were six in total, including jogging, walking, ascending and descending stairs, sitting, and standing.

From raw accelerometer data, 43 features were extracted and they were collected every 50 milliseconds, so for each second, there were 20 samples. After collecting data, some machine learning algorithms were used for evaluating the dataset. Results show that high levels of accuracy can be achieved, for two of the most common activities, which are walking and jogging. For walking the accuracy achieved was 93.6%, by using a logistic regression algorithm and for jogging, the score was 98.3% by using a multilayer perceptron algorithm. High accuracy was also achieved, in other activities. Walking upstairs and walking downstairs were two of the lowest accuracies achieved. In walking upstairs class, the score was 61.5% with multilayer perceptron algorithm and for walking downstairs the accuracy was 55.5 % with the J48 algorithm.

Actitracker [11] was a similar project by Fordham University and its difference with WISDM is that this dataset is a real-world one, while the other is a controlled testing dataset. This dataset also consists of tri-axial accelerometer data samples while it is bigger than the previous one. In this project, there were 563 volunteers. In all cases, data were collected in total 20 times for each second.

The activities that were tracked in this dataset were the same as in WISDM. They include walking upstairs and downstairs, standing, sitting, jogging, and walking. The difference here is that these activities take place in an uncontrolled environment. The total number of samples is also much higher due to more volunteers. In this dataset, there were 2.980.765 samples while in WISDM there were 1.098.207.

The main difference with the CHARM dataset is that these datasets, contain data originating from mobile inertial sensors, while the device is either in the user's pocket or mounted on a belt. In the CHARM case, those micro-moments that the user looks at the screen, should be found. This is the reason that a new dataset had to be created with data collected when the user looks at the screen. Moreover, knowing the android model of the device and some demographic information about each user can help in better analysis.

3 CERTH Human Activity Recognition Mobile (CHARM) Dataset

To have a well-formed dataset, a custom android mobile application was created and distributed to various users around Europe. This application was given to each user, alongside some instructions, and each user should give some general demographic information about him. The number of users was restricted to 31 and was distributed remotely due to coronavirus restrictions. After doing this, he should follow the instructions, and he should do activities like walking, running, ascending stairs, etc., while he had his phone on hand. Users could skip an activity in case it was difficult for them to complete. Since there were various mobile devices among the users, the accelerometer calibrations were different and the only common setting was the sampling rate, which contributes to the in the wild nature of the dataset.

Table 1. Information about the activities done by users.

Activity	Number of Repetitions	Duration
Sitting on a Chair	2 Repetitions	30 Seconds
Sitting on a Couch	2 Repetitions	30 Seconds
Standing	2 Repetitions	30 Seconds
Lying Up	2 Repetitions	30 Seconds
Lying by Side	2 Repetitions	30 Seconds
Device on Surface	2 Repetitions	30 Seconds
Walking	2 Repetitions	30 Seconds
Running	2 Repetitions	30 Seconds
Walking Upstairs	6 Repetitions	10 Seconds
Walking Downstairs	6 Repetitions	10 seconds

Each activity had a start button, which defines its start and end after 35 seconds. Walking upstairs and downstairs last 90 seconds and they are segmented in 15-second intervals, which means that each user has to walk stairs 6 times upwards and 6 downwards. The frequency of the accelerometer data collection is 20 Hz, which means that data were collected every 50ms. There was a total of 20 samples for each second and after ending, each user was asked if everything went according to the rules, or not. If not, the user should repeat the process. The android application used tri-axial accelerometer data to record each activity. After completing the activities, these data are uploaded to a server, where they can be retrieved afterward. All activities can be seen in Table 1.

Except knowing the device's models of users participated in the CHARM dataset, some demographic information was held, respectively to GDPR law. Most of the users were between the ages 25-50 and between the height of 171-180. 4 participants were less than 25 years old, 21 were between 25-50, and 6 were 51-75, which shows a variety in measurements. As for the height, there were 6 between 150-160 cm, 7 between 161-170 cm, 10 between 171-180 cm, 7 between 181-190 cm, and 1 between 191-200cm.

The database used is InfluxDB and it is an open-source time-series database. This allows the overall size reduction and easier data handling, while each type of data is stored in separate tables. These tables are the registration, the accelerometer data, and the gyro data. The connection of the mobile application to the database is established through the InfluxDB-java client library. The application requires the SDK28 to run on Android 8.0 or newer versions. Its structure enables the registration of multiple users from the same device. Each user can append measurements to the database, after following specific instructions and confirming the proper completion of the activity. All users could delete personal information from the database and could also delete their measurements, according to GDPR laws. Furthermore, the data of each user is anonymized, so that a user cannot be identified from its data.

Table 2. Number of the CERTH HAR data points per activity.

Activity	Number of Data Points	Percentage (%)
Sitting on a Chair	37.144	12.55
Sitting on a Couch	35.344	11.94
Standing	31.190	10.54
Lying Up	32.389	10.94
Lying by Side	29.984	10.13
Device on Surface	30.000	10.13
Walking	27.600	9.32
Running	22.792	7.70
Walking Upstairs	24.396	8.24
Walking Downstairs	25.188	8.51

Accelerometer data were divided into time windows, in order to exploit the temporal information and periodicity of the signals. Each window had 50 data points and its time duration was 2.5 seconds. There are 3 vectors $a_{x,y,z}$ in each data point, one for each axis.

Then 40 statistical features for each window were created, to encode the global characteristics of the time series. The feature types created were the average (3 features), the standard deviation (3 features), the average absolute difference (3 features), the average resultant acceleration (1 feature), and the binned distribution (30 features). First, each time window of each channel is centered around its average, and then it is being fed to the network. These features are the same as the ones in the WISDM and Actitracker datasets [10] and an analysis of them was conducted in [19].

After completing the dataset retrieval procedure, some preprocessing methods were applied, to clear the dataset from incorrect measurements. Specifically, any measurement with a sampling frequency lower than 98% of the target one, which was 20Hz, was discarded, along with any null values. The number of data points per activity is presented in Table 2. Some actions have more data points than others because more users completed them and it has to do with how difficult each action was. Some people were not able to run or climb stairs and that's why those actions have fewer data points. Sitting on a chair has most of the data points, while activities like running and walking stairs have the least, due to their difficulty.

4 Experimental Setup

To evaluate the produced dataset, a comparison was made with a state-of-the-art methodology, compared it with the two other datasets identified (i.e., the WISDM and Actitracker datasets).

4.1 Convolutional Neural Network

A convolutional neural network is a hierarchical Feed-Forward Neural Network (FFNN), in which each neuron in one layer is connected to all neurons in the next layer. They are inspired by the biological visual system. Apart from fully connected layers, it consists of convolutional layers. In those layers, the network learns filters that are sliding along the input data and they are applied to its sub-regions. Its architecture is presented in Fig. 1.

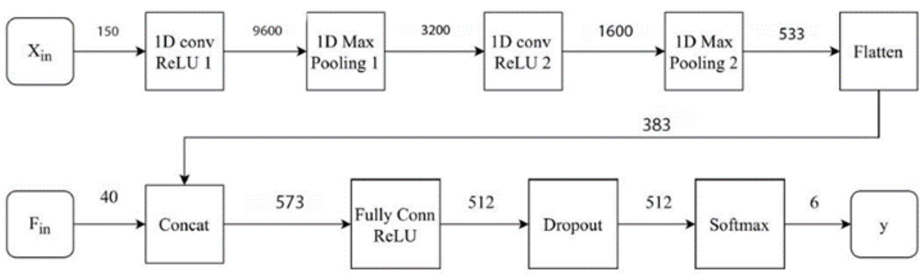

Fig. 1. Representation of the developed CNN architecture.

As explained in detail in [19], CHARM system architecture consists of 7 steps. Firstly, the accelerometer data which are sized 50 x 3 axis, are fed into the first layer with 192 convolutional filters. The kernel1 size is 12, and the stride of the convolution is 1. To its output, applied the ReLU function was applied. After that, a max-pooling layer follows. Its kernel size is 3 x 1 and its stride is 3. This way, feature representation is reduced by 3. Then, another convolutional layer is added, to learn more abstract and hierarchical features. It has 96 convolutional filters and a kernel size of 12. Step convolution is 1 and ReLU function is applied to its output. A final max-pooling layer with kernel size 3 x 1 and a stride of 3 is then used to further reduce the feature representation by 3. The output max-pooling layer is then flattened with the statistical features in 3. The joint vector is given to a fully connected layer that has 512 neurons and the ReLU function is applied to its output. A dropout layer was then added with a rate of 0.5 in order to avoid overfitting. Finally, the output of the layer is passed to a softmax layer, to compute a probability distribution over 10 activity classes. The optimizer used for the training model is a stochastic gradient descent with momentum and a constant learning rate.

4.2 Evaluation Metrics

To evaluate the results of this dataset some of the most common metrics, described in [20] were used to compute the performance of the neural network, with the CHARM dataset given. Firstly, the accuracy metric was calculated, which is the ratio of correct predictions over the total predictions, the precision metric, which represents the rate of correct predictions of a class over the total number of predictions for this class. Also, the recall metric was used, to compute the fraction of correct predictions of a class to the total real data points of it. Finally, using recall and precision, the f1 metric was computed, which is a combination of them and is described among the previously mentioned metrics.

4.3 Evaluation Scenarios

The scenarios used for evaluation were 3 in total, with scenario 1 being the training and testing with the CHARM dataset. In that case, cross-validation was used in order to not have the same data in the training and test split. Scenario 2 included a model trained with the CHARM dataset and tested with the WISDM and Actitracker datasets as input. Cross-validation was also used for computing evaluation metrics. Scenario 3 included two models in total. The first model was trained with the WISDM dataset and tested with the CHARM and the second was trained with the Actitracker dataset and tested with the CHARM dataset. To be able to make comparisons with the other models, CHARM labels were adapted to other dataset's number of labels. The CHARM dataset has 10 labels, while WISDM and Actitracker have 6. This adaptation was necessary, to make comparisons.

5 Results

To test the created dataset (which currently has 31 subjects in total) in real life use cases, we used the CNN model described in Section 4.1 and executed the experiments described above. Then the classification quality according to specific metrics, presented in Section 4.2 was calculated. Each experiment ran with the previously mentioned CNN in Section 4.1. Next, the training parameters used are described. Window size is described as Nw, while epochs are described as e. Optimizer momentum is described as β and learning rate as λ. For all of experiments, the window size had 50 data points. The model was trained for 100 epochs, while optimizer momentum was 0.9 and the learning rate was 0.01. This means that there were Nw = 50, e = 100, $\beta = 0.9$ and $\lambda = 0.01$. The datasets used for experiments are described in the appropriate section 3.

Firstly, the CNN model was trained on the new dataset, and then, using the appropriate metrics, the model's accuracy, recall, precision, and f1 on the CHARM dataset, were counted. To keep the evaluation user-independent, a 10-fold cross-validation method was implemented and then, one model was trained with an Actitracker dataset and one with WISDM. Then, both of them were tested on the CHARM dataset. The results of CHARM metrics are presented in Table 3.

Table 3. 10-fold cross-validation results for CHARM dataset.

Metric	Average	Std. Deviation
Accuracy	69.13%	7.33%
Precision	69.98%	17.81%
Recall	66.64%	18.24%
F1	66.45%	17.69%

Having examined the 10-fold cross-validation results, the confusion matrix of the CHARM dataset was created. In Table 4, the number of predictions for each class is presented and by this, activities that are confused can be seen. First, the average number of predictions for each class is calculated, followed by a percentage, that shows its relationship with the other classes. From this table, it can be seen that both static and dynamic activities are well classified. The activity that has the highest percentage is having the mobile on a surface with 98.8% and it is followed by Lying up with 85.6% and Lying Side reaching 80.5%. There is naturally a common misclassification between sitting on a chair and sitting on a couch. Couch had 33.6% misclassified as a chair, while chair had 27.1% misclassified as a couch. These two were the biggest percentages of misclassification. As for the dynamic activities like walking, running, upstairs, and downstairs, all activities were noticed having accuracy over 60% and the most confused one was upstairs with walking. Surprisingly it is not confused with downstairs. Similar confusion between walking and stairs had been mentioned in [19], however using the WISDM dataset. Walking is also most confused with upstairs and not with running.

After examining the confusion matrix of the CHARM model, comparisons with WISDM and Actitracker datasets were conducted. For scenario 1, the accuracy from cross-validation was 69.13%, while for scenario 2, when testing with the WISDM dataset the accuracy reached 79.73%. When the testing dataset was the Actitracker, the accuracy was 78.30%, which shows that the CHARM model can perform well with other input datasets. For scenario 3 when the training model was the WISDM the accuracy was 46.23% and when it was the Actitracker it was 30.74% which shows that accuracy is increased significantly when using the created dataset for training. This shows that the existing datasets are not accurate enough for our intended use case, so a new dataset such as CHARM is necessary. More detailed results could be seen in Fig. 2. In the left figure, the accuracy of CHARM models is presented in Y-axis, when it has as input the WISDM, Actitracker, and CHARM datasets respectively. In the right figure, scenario 3 is presented.

Table 4. Confusion matrix of the CHARM dataset.

	Chair	Couch	Standing	Surface	Lying Up	Lying Side	Walking	Running	Upstairs	Downstairs
Chair	**142.5 (45.4 %)**	85 (27.1 %)	72.9 (23.2 %)	2.9 (0.9 %)	3 (1.0 %)	0.8 (0.3 %)	3.3 (1.1 %)	0.7 (0.2 %)	0.6 (0.2 %)	2.5 (0.8 %)
Couch	99.7 (33.6 %)	**128.3 (43.2 %)**	33.4 (11.2 %)	8.1 (2.7 %)	11.4 (3.8 %)	4.5 (1.5 %)	4.6 (1.5 %)	2.2 (0.7 %)	1.9 (0.6 %)	2.9 (1.0 %)
Stand-ing	69.1 (24.7 %)	32.5 (11.6 %)	**148 (52.9 %)**	11 (3.9 %)	8.4 (3.0 %)	0.4 (0.1 %)	7.6 (2.7 %)	0.1 (0.0 %)	0.4 (0.1 %)	2.3 (0.8 %)
Sur-face	2.4 (0.9 %)	0 (0.0 %)	0 (0.0 %)	**276.7 (98.8 %)**	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)	0.4 (0.1 %)	0 (0.0 %)	0.5 (0.2 %)
Lying Up	1.2 (0.4 %)	13.9 (4.6 %)	8.7 (2.9 %)	12.5 (4.1 %)	**259.5 (85.6 %)**	4.6 (1.5 %)	0.7 (0.2 %)	0.2 (0.1 %)	0.4 (0.1 %)	1.5 (0.5 %)
Lying Side	11 (3.9 %)	19 (6.8 %)	0.8 (0.3 %)	5.5 (2.0 %)	11.5 (4.1 %)	**225.2 (80.5 %)**	2.3 (0.8 %)	1.3 (0.5 %)	1.1 (0.4 %)	2.1 (0.8 %)
Walk-ing	2.6 (1.0 %)	2.2 (0.9 %)	8.7 (3.4 %)	11.2 (4.3 %)	0 (0.0 %)	5 (1.9 %)	**168.7 (65.5 %)**	11 (4.3 %)	32.2 (12.5 %)	16 (6.2 %)
Run-ning	0 (0.0 %)	1.7 (0.8 %)	0.5 (0.2 %)	9.4 (4.4 %)	0 (0.0 %)	0.2 (0.1 %)	17.3 (8.1 %)	**128.8 (60.6 %)**	36.5 (17.2 %)	18.3 (8.6 %)
Up-stairs	1.4 (0.7 %)	1.7 (0.9 %)	0.7 (0.4 %)	6.6 (3.4 %)	0 (0.0 %)	0.3 (0.2 %)	33.5 (17.1 %)	8.6 (4.4 %)	**125.3 (63.9 %)**	17.9 (9.1 %)
Down stairs	1.6 (0.8 %)	1.1 (0.5 %)	3.4 (1.7 %)	6.4 (3.2 %)	0 (0.0 %)	0.5 (0.2 %)	28.3 (14.1 %)	11 (5.5 %)	24.2 (12.0 %)	**124.9 (62.0 %)**

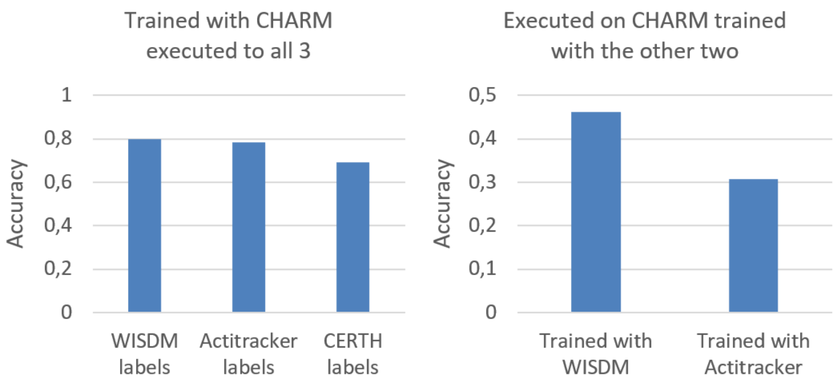

Fig. 2. Evaluation charts presenting CHARM, WISDM and Actitracker datasets.

6 Conclusions

In this paper, a more realistic new dataset has been presented in terms of Human Activity Recognition during activities through which the end-user might be more prompt to receive a recommendation towards a certain action. The presented dataset contains 31 participants and 10 activities, throughout which data from the mobile accelerometer and gyroscope have been collected, including also certain metadata (e.g., android version). After collecting and preprocessing the data, its added value was demonstrated by applying a classification model, based on a state-of-the art CNN model. Based on a benchmarking analysis performed with two other datasets, a 10-fold Cross-validation showed that the explored model works effectively in all three datasets, while the CHARM dataset is of the same high quality with its counterparts. It has also been proved that the model presented can perform well with other input datasets. In future work, the CHARM dataset (openly available in Zenodo [21]) is expected to grow even more, with more users throughout Europe, introducing significant scientific value to the research community.

References

1. Vrigkas M., Nikou C., Kakadiaris I.A.: A review of Human Activity Recognition Methods. Frontiers in Robotics and Ai, vol. 2 (2015).
2. Hsu Y. L., Yang S. C., Chang H. C., Lai H. C.: Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network. IEEE vol. 6, pp. 31715-31728 (2018).
3. Chen Y., Shen C.: Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access vol. 5, pp. 3095-3110 (2017).
4. Goyal R., Kahou S., Michalski V., Materzynska J., Westphal S., Kim H., Haenel V., Fruend I., Yianilos P., Mueller M., et al.: The "Something Something" Video Database for Learning and Evaluating Visual Common Sense. Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 5842-5850 (2017).

5. Fouhey D. F., Kuo W., Efros A. A., Malik J.: From Lifestyle Vlogs to Everyday Interactions. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, pp. 4991-5000 (2018).
6. Damen D., Doughty H., Farinella G.M., Fidler S., Furnari A., Kazakos E., Moltisanti D., Munro J., Perrett T, Price W. et al.: Scaling Egocentric Vision: The EPIC-KITCHENS dataset. ArXiv:1804.02748, (2018)
7. Reiss A., Stricker D.: Introducing a New Benchmarked Dataset for Activity Monitoring. In Proccedings of the 2012 16th International Symposium on Wearable Computers, Newcastle UK, pp. 108-109 (2012).
8. Zhang M., Sawchuk A.A.: USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors. In Proceedings of the ACM International Conference on Ubi. Comp. Workshop on Situation, Activity and Goal Awareness pp. 1036-1043 (2012).
9. Bruno B., Mastrogiovanni F., Sgorbissa A.: Wearable inertial sensors: Applications, challenges, and public test benches. Robotics & Au-tomation Magazine, IEEE, vol 22, pp. 116–124, (2015).
10. Kwapisz J.R., Weiss G.M., Moore S.A.: Activity Recognition Using Cell Phone Accelerometers. SIGKDD Explor. Newsl. 12, 74-82 (2011).
11. Lockhart J.W, Weiss G.M., Xue J.C., Gallagher S.T., Grosner A.B., Pulickal T.T.: Design Considerations for the WISDM Smart Phone Sensor Mining Architecture. In Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data, San Diego, CA, USA, 21-24 August, pp. 25-33 (2011).
12. Zhang S., Wei Z., Nie J., Huang L., Wang S., Li Z.: A Review on Human Activity Recognition Using Vision-Based Method. Journal of Healthcare Engineering, (2017).
13. Pradhan B. R., Bethi Y., Narayanan S., Chakraborty A., Thakur C. S.: n-HAR: A Neuromorphic Event-Based Human Activity Recognition System Using Memory Surfaces. IEEE International Symposium on Circuits and Systems, Japan, pp. 1-5 (2019).
14. Jalal A., Kim Y. H., Kim Y.J., Kamal S., Kim D.: Robust human activity recognition from depth video using spatiotemporal multi-fused features. Pattern Recognition, vol. 61, pp. 295-308, (2017).
15. Ryoo M., Rothrock B., Fleming C., Yang H. J.: Privacy-Preserving Human Activity Recognition from Extreme Low Resolution. Proceedings of the AAAI Conference on Artificial Intelligence, (2017).
16. Peppas K., Chouliara A., Tsolakis A., Krinidis S., Tzovaras D.: Redefining Micro-moments for improving energy behaviour: The SIT4Energy approach. IEEE SmartWorld, Ubiquitous Intelligence & Comp., Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Comp., Internet of People and Smart City Innovation (2019).
17. Wang J.,Yiqiang C., Shuji H., Xiaohui P., Lisha H.: Deep Learning for Sensor-based Activity Recognition: A Survey. Pattern Recognition Letters, vol. 119, pp. 3-11 (2019).
18. Hassan M. M., Uddin MD. Z., Mohamed A., Almogren A.: A robust human activity recognition system using smartphone sensors and deep learning. Future Generation Computer Systems, vol.81, pp. 307-313 (2018).
19. Peppas K., Tsolakis A., Krinidis S., Tzovaras D.: Real-Time Physical Activity Recognition on Smart Mobile Devices Using Conv. Neural Networks. Appl. Sci. 10, no. 23: 8482 (2020).
20. Sasaki Y.: The truth of the F-measure. Teach Tutor Master (2007).
21. Vasilopoulos G.V. et al.: CERTH Human Activity Recognition Mobile (CHARM) Dataset. Available Online: https://doi.org/10.5281/zenodo.4642560.