

Daily Living Activity Recognition Using Wearable Devices: A Features-rich Dataset and a Novel Approach

Maurizio Leotta, Andrea Fasciglione, Alessandro Verri

Abstract:

Automated daily living activity recognition is a relevant task since it allows to assess the health status of a subject both objectively and remotely. Having a reliable measure is important since it gives precise indications to doctors and researchers interested in evaluating the effectiveness of treatments or drugs (e.g., in the context of clinical studies). The possibility to perform this task remotely is more convenient for the patients and acquired increasing importance not only due to the current pandemic, but also because of the regularly growing population of elderly people that could benefit from remote monitoring.

In this paper, first, we describe a novel wearable-device-based dataset that contains data (1) of a high number of daily life activities, coming from a real-life scenario, (2) recorded by applying multiple devices on different parts of the body, and (3) recorded with medical-grade devices at a high sampling frequency. Then, second, we describe a machine learning-based method for activity recognition.

Our approach takes in input a dataset and through multiple phases allows to recognise the activities performed by the subjects with a good degree of accuracy (up to 0.92 expressed as F1 score depending on the location).

Digital Object Identifier (DOI):

https://doi.org/10.1007/978-3-030-68790-8_15

Associated Dataset available on  HARVARD
Dataverse

<https://doi.org/10.7910/DVN/G23QTS>

Copyright:

© 2021 Springer Nature Switzerland

The final authenticated version is available online at:

https://doi.org/10.1007/978-3-030-68790-8_15

Daily Living Activity Recognition Using Wearable Devices: A Features-rich Dataset and a Novel Approach

Maurizio Leotta^[0000-0001-5267-0602], Andrea Fasciglione^[0000-0003-4453-9369],
Alessandro Verri^[0000-0001-9777-9986]

Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi (DIBRIS)

Università di Genova, Italy

maurizio.leotta@unige.it

Abstract. Automated daily living activity recognition is a relevant task since it allows to assess the health status of a subject both objectively and remotely. Having a reliable measure is important since it gives precise indications to doctors and researchers interested in evaluating the effectiveness of treatments or drugs (e.g., in the context of clinical studies). The possibility to perform this task remotely is more convenient for the patients and acquired increasing importance not only due to the current pandemic, but also because of the regularly growing population of elderly people that could benefit from remote monitoring.

In this paper, first, we describe a novel wearable-device-based dataset that contains data (1) of a high number of daily life activities, coming from a real-life scenario, (2) recorded by applying multiple devices on different parts of the body, and (3) recorded with medical-grade devices at a high sampling frequency. Then, second, we describe a machine learning-based method for activity recognition. Our approach takes in input a dataset and through multiple phases allows to recognise the activities performed by the subjects with a good degree of accuracy (up to 0.92 expressed as F1 score depending on the location).

Keywords: Activity Recognition · Wearable Devices · Actigraph.

1 Introduction

Recently an increasing interest in finding reliable methods for monitoring patients suffering from different diseases (or simply elderly) emerged, in particular using remote and non-intrusive methods (e.g. [7, 11, 15]). Many diseases, in fact, strongly impact daily life activities because of their effects (e.g., Multiple Sclerosis, Pulmonary Arterial Hypertension, or Parkinson). A possible way to assess the physical condition of a patient is based on assessing how much time is spent on specific activities and how the amount of time dedicated on each activity changes over time (e.g., a subject can decide to stop vacuuming because of the feeling of fatigue, or could need more time to eat because of hand tremors).

While working on a possible method to create an activity recognition classifier, starting from data recorded by wearable devices and in the context of a collaboration with a major pharmaceutical company (Janssen), we found the need of creating a novel dataset regarding daily life activities. Indeed, by analysing the state of the art concerning the public available datasets recorded with wearable devices, we noticed that in spite of the large interest on this topic, there is a lack of datasets having, at the same time,

the following characteristics: (1) containing data of numerous and different daily life activities, (2) containing data recorded using high-quality sensors (both concerning frequency and accuracy), and (3) containing data from different synchronised devices positioned on different parts of the body. From the point of view of a researcher, this lack could become an obstacle to perform more in-depth investigations and to conceive more advanced approaches to the problem of the Activity Recognition using wearable devices.

For this reason, in this work we present, as *first contribution*, the ongoing effort of creating a novel dataset that meets the three characteristics mentioned above.

Concerning the first characteristic, our dataset includes 17 different daily life activities performed in real-life scenarios (e.g., eating, using laptop, handwriting, vacuuming, walking, going upstairs/downstairs). On the contrary, available datasets often include a few activities (often 8 or less) or activities that are limited to a particular context (e.g., cooking, breakfast morning-routine, or walking at different speeds only) [4, 17, 1, 13].

Concerning the second characteristic, we recorded our dataset using professional devices produced by Actigraph¹. These devices are medical-grade activity monitors that thanks to their characteristics and to their reliability, have been broadly considered in different studies and for different purposes (e.g. [6], [12]). More in detail, we have used two Actigraph GT9X Link, with a sampling frequency of 100 Hz and one Actigraph Centrepoint, with a sampling frequency of 256 Hz, all of them relying on high-quality internal sensors (e.g., accelerometers, gyroscopes, magnetometers). This is an important characteristic of our dataset since, very often, the currently available datasets contain data recorded with low cost (and precision) sensors like the ones included in Android smartphones, with a sampling frequency of 50 Hz (or lower) and non-certified accuracy of the values provided [10, 16, 13]. The high frequency of these devices can be useful for researchers in order to perform analysis by re-sampling data to different frequencies.

Another strength of our dataset (third characteristic) lies in the fact that we have placed the three aforementioned devices, synchronised together, on three different parts of the participants' bodies: dominant wrist, right side of the hip, and right ankle (on the right side). Having synchronised data coming from different parts of the body, like in our dataset, would allow researchers to find methods based on the correlation of these data, and thus creating more accurate activity analysis or recognition approaches. On the contrary, existing datasets are usually focused on only one part of the body (e.g. pockets or the hip) [10, 20]. The fact of wearing more devices could be quite annoying for a subject in real settings: the trade-off between having more data and creating a discomfort for the patient should be carefully analysed.

As a *second contribution*, in this paper we describe a machine learning (ML) based method for the Activity Recognition. Our approach takes in input a dataset and through multiple phases allows to recognise the activities performed by the subjects with a good degree of accuracy (up to 0.92 expressed as F1 score depending on the location).

This paper is organised as follows: Section 2 describes the dataset collection procedure and the structure of the obtained raw data. Section 3 briefly describes the proposed approach that we have used to evaluate the predictability of the activities using our dataset. Section 4 reports on the empirical evaluation of the approach, while Section 5 reports related works and Section 6 concludes the paper.

¹ ActiGraph, LLC (Pensacola, FL, USA), <https://actigraphcorp.com/>

2 Dataset of Daily Living Activities Creation

The creation of the *Daily Living Activities* dataset has been performed in three main phases: (1) Data Recording, (2) Data Extraction, and (3) Data Labelling and Cleaning. The final output is a labelled dataset containing the raw data of 17 daily living activities ready to be used by researchers for a variety of possible studies. In the following of this section, we describe the three phases in detail.

2.1 Data Recording

To record our dataset, we followed a precise protocol that we defined and that had been reviewed and approved by the ethical committee of the Department in which the data recordings took place. Apart from the operational details on the procedures to follow during data recordings, our protocol includes also a step where each participant is asked to sign an informed consent: this allow us to share, with the research community, all the data recorded and the physical bio-metric characteristic of each participant. The dataset is currently available online on our department website² but we are working to share it on relevant dataset repositories such as the UC Irvine Machine Learning Repository³ or on the Harvard Dataverse⁴ so that it could also be indexed by specific search engines (e.g. Google Dataset Search).

Participant inclusion/exclusion criteria. The former includes the following: (1) to be able to perform the request actions, (2) age of 18 and over, and (3) to understand the purpose of the study and willing to participate in the study. On the contrary, exclusion criteria include any planned surgery or procedures that would interfere with the conduct of the study and any major mobility difficulties.

Participants Characteristics. The preliminary version of our dataset currently includes data of 8 volunteers: males aged between 23-37, with a weight between 52-90 kg and height between 172-186 cm. Regarding the dominant hand, two subjects out of 8 were left-handed, while the other ones (6 out of 8) were right-handed. In the next months we expect to record the data of 25-30 additional participants. Detailed information (i.e. age, height, weight, dominant hand) for each subject are reported within the dataset itself.

Activity Recording Procedure. During the data recording, we asked to all the participants to perform the 17 different activities listed in Table 1. We split the list of activities into two different sets, as in the Table 1: Set A and Set B; the former have been performed for a fixed time, while the latter not. The differences between the two sets lie mainly in the fact that activities in Set B were constrained to a particular path or to a flight of stairs, while activities in Set A were quite stationary and did not require the subjects to move along a path. By depending on a fixed path, moreover, it is also possible to measure the different walking speed (e.g. in term of meters/second or of steps/second) of the subjects as an additional information. More in details, Set A activities have been performed for more than 120 seconds (in general for about 150 seconds), and we included in our dataset the central 120 seconds of each execution in order to obtain cleaner data. On the other hand, Set B includes: *Walking* performed for 160 meters (in at least 110 seconds); *Walking Fast* performed for 205 meters (in at least 110 seconds); and *Going Downstairs*,

² <https://sepl.dibris.unige.it/2020-DailyActivityDataset.php>

³ <https://archive.ics.uci.edu/ml/index.php> ⁴ <https://dataverse.harvard.edu/>

Set A	Set A	Set B
1. Relaxing on a chair	7. Brushing Teeth	13. Walking
2. Keyboard Typing	8. Sweeping	14. Walking Fast
3. Using the Laptop	9. Vacuuming	15. Going Downstairs
4. Handwriting	10. Eating (a soup)	16. Going Upstairs
5. Washing Hands	11. Dusting a surface	17. Going Upstairs Fast
6. Washing Face	12. Rubbing a surface	

Table 1. List of Activities performed

Going Upstairs, *Going Upstairs Fast* performed using a single flight of stairs with no intermediate floors between the steps for an average time of 40 seconds.

The participants have been followed and instructed during the data recording: they have been told what activity should have been performed and some details on it, but it has not been imposed to move exactly in a particular way⁵. We have shown a WHO video⁶ on how to wash the hands and we have better explained the difference between *dusting* and *rubbing*: respectively, dust a surface, or rub to clean a really dirty surface. Even if the chosen activities are characterised by relatively standard movements, as expected, we noticed that different subjects had their own way to execute movements related to the activity. This is a positive characteristic since it can help to understand the natural differences that can occur when analysing and comparing different subjects. The only constraint established during the data recording has been to always use the dominant hand (i.e. the one where the device was placed) when performing those actions that mostly involve a single arm movements (e.g. handling the vacuum/broom with the dominant hand while vacuuming/sweeping). Otherwise, the signal recorded from the wrist would have had no information regarding the pattern of the dominant hand.

Devices and their Positioning. In the last years, Actigraph, a leading provider of wearable physical activity and sleep monitoring solutions for the global scientific community, proposed several actigraphy devices. In this work, we used two Actigraph GT9X Link and one Actigraph Centrepoint Insight Watch. They are activity monitors equipped with high precision and fast reading sensors. In detail, both are equipped with a three-axial accelerometer while the GT9X also includes a complete IMU (Inertial Measurement Unit⁷). The IMU is an electronic chip capable of capturing position and rotation data for advanced analyses. It contains a secondary accelerometer, a gyroscope, a magnetometer, and a temperature sensor. Additional information on the two devices could be found in the official web site⁸. Table 2 shows the kinds of sensors available in two devices and the corresponding measurement units.

⁵ Regarding the possible variation of the subjects behaviour while performing activities by knowing they were participating in an experiment (also known as Hawthorne effect), it is important to note that (1) there have not been judgements on how well subjects were performing activities, therefore they could behave in any way they preferred, (2) due to the short amount of time spent in recording data, possible variations in subjects behaviour happened in the entire recording (i.e. no effect on possible train and test data sets). ⁶ "WHO: How to handwash? With soap and water", <https://www.youtube.com/watch?v=3PmVJQUCm4E>

⁷ https://en.wikipedia.org/wiki/Inertial_measurement_unit ⁸ Actigraph GT9X Link - <https://actigraphcorp.com/actigraph-link/>, Actigraph Centrepoint Insight Watch - <https://actigraphcorp.com/cpiw/>

Device	Sensor Type	Units of Measure
Both	3 axis primary accelerometer	g
GT9X	3 axis secondary accelerometer	g
GT9X	3 axis gyroscope	degrees/s
GT9X	3 axis magnetometer	microTesla (μT)
GT9X	temperature sensor	Celsius

Table 2. Characteristics of the devices sensors

The three wearable devices were worn by the participants as follows and with the following settings:

- 1 Actigraph Centrepoint at the dominant wrist. Accelerometer recording at a sampling rate of 256 Hz.
- 1 Actigraph GT9X Link at the right hip at the height of the iliac crest (using the device belt clip). IMU (i.e., accelerometer, magnetometer, and gyroscope) recording at a sampling rate of 100 Hz.
- 1 Actigraph GT9X Link at the height of the right ankle placed, with the help of the belt clip, on the subject’s right side of the shoe, over the malleolus. IMU recording at a sampling rate of 100 Hz.

Regarding the calibration of the devices, they have been precisely calibrated (using the automated procedure of the device) at the beginning of each data recording session. **Ground Truth Definition.** The ground truth annotation has been performed by the two first authors of this paper, in parallel, by following the subjects performing activities, using a chronometer took note of the starting and ending time of each activity. Moreover, while recording walking data researchers ensured that the subjects were following a specific walking path so that we could retrieve the average walking speed of the subjects for optional and additional tests.

2.2 Raw Data Extraction

After recording data with the subjects, we extracted the raw data from the devices using the proprietary software system developed for Actigraph devices. Then we exported the data as .csv files. The two kinds of devices that we used were equipped with different sets of sensors, so the output of each kind of device will be different. The .csv produced for the Actigraph GT9X Link, will contain 11 columns:

- ‘*Timestamp*’: timestamp of the sampled values
- ‘*Accelerometer X*’, ‘*Accelerometer Y*’, ‘*Accelerometer Z*’: instantaneous accelerations for each axis, measured in units of gravity (G)
- ‘*Temperature*’: IMU temperature, in Celsius degree
- ‘*Gyroscope X*’, ‘*Gyroscope Y*’, ‘*Gyroscope Z*’: instantaneous measure of the gyroscope for each axis, measured in degrees/sec
- ‘*Magnetometer X*’, ‘*Magnetometer Y*’, ‘*Magnetometer Z*’: instantaneous measured magnetic field for each axis, measured in microTesla (mT)

For each row of the file, it is possible to find the sampled value at the specified timestamp from each of the sensors and axis. The .csv file produced using data recorded with the Actigraph Centrepoint, instead, will only have these columns: ‘*Timestamp*’, ‘*Accelerometer X*’, ‘*Accelerometer Y*’, ‘*Accelerometer Z*’.

2.3 Data Labelling and Cleaning

Thanks to the ground truth, we were able to label the data precisely. Labels were associated with each row of the recorded data indicating which activity is carried out in such instant. Basically, a new column has been attached to data, where for each row we had a number corresponding to the activity performed in that instant (e.g., 1 is *Relaxing*, 2 is *Keyboard Typing*, ...).

During this data processing step, we also used a label to identify data that had to be removed because it was not useful or that could lead to misleading results (e.g., data recorded in between two different actions).

3 A Possible Approach to Daily Living Activity Recognition

In this section, we describe in detail the steps composing our approach aimed at recognising the activities performed by a subject wearing an actigraphy device. The approach is based on the usage of Support Vector Machine (SVM). More in detail, starting from the labelled *raw data* (e.g., from our dataset), the first activity consists in a (1) *features extraction* phase, after which data will be split into *training* and *test data*. Training data will be used for (2) *tuning the hyperparameter* and (3) *training the SVM model* with the correct parameters. At this point, the SVM model can be used for (4) *recognising daily living activities* on novel unseen data. Thus we use *test data* to evaluate the accuracy of the trained SVM model and, in general, of our approach. In the following subsections, we will describe in detail the first three steps (1,2,3) while the fourth will be described in Section 4. Our approach has been implemented using Python and with the help of the Jupyter platform⁹; we relied on the *Scikit-learn* library¹⁰, also known as *sklearn*, since it provides several instruments for data analysis that were useful in our study.

3.1 Features Extraction

As done in other similar studies like the one of Staudenmayer et al. [14], we have extracted the feature set made up of feature vectors and associated labels. To do so, we have used the *sliding window* approach to compute the features, using only the accelerometer data (however the approach can be simply extended to include gyroscope and magnetometer data). In this phase, a sliding window passes over the data and for each axis (X, Y, Z) we extract some measures on the data contained in the window: mean, variance, standard deviation, median absolute deviation, percentiles (10Th, 25Th, 75Th, 90Th). Having eight measures per axis allows to compute 24 features for each window that composed the feature set used in the evaluation, described in Section 4.

Data at the end of each activity recording is discarded when not enough for building a window (i.e., the remaining data covers less time than the length of the sliding window).

⁹ <https://jupyter.org/> ¹⁰ <https://scikit-learn.org/>

Regarding the sliding window, its length represents an important parameter on which results could potentially highly depend. For this reason, in previous experiments, we have performed some analysis to understand how the length of the window and the overlap between subsequent windows could affect accuracy. After these tests, we have decided to use windows that were 2.0 seconds long, with 95% of overlap each other. This value is motivated by the fact that typical human periodic movements have a period of no more than two seconds (e.g., each step during walking or hand movement during toothbrushing).

After the features extraction step, data is ready to be used in any ML algorithm. For our approach, we have chosen to rely on Support Vector Machine (SVM). SVM, in fact, has been already used to estimate physical activity from accelerometers in the literature, showing good performances in this kind of task (e.g., [21, 5]). When using SVM data need to be *standardised* in order to obtain better results. This is needed since SVM is based on the idea of finding the hyperplane that best divides different classes by maximising the distance between the hyperplane and the data (i.e. Support Vectors), if one feature (i.e. one dimension) has larger values than the others, it will prevail on the others when computing distances. This will not be a problem if we standardise data: we did so by removing the mean and scaling to unit variance. Finally, to further prepare our data to feed the algorithm, we have also split our data into *training data* and *test data*: 75% and 25% of data of each activity, respectively.

3.2 Hyperparameter Tuning

SVM needs some parameters to be tuned in order to achieve the best result: C and γ , in combination with the different used *kernels* (*Radial Basis Function* and *Polynomial* kernels). Focusing on the hyperparameter tuning, we know that while constructing a machine learning model, a general goal is to choose parameters such that we obtain a model that is able to learn, in the best way, all information from the training data, while, at the same time, it should be able to generalise well to new data. This problem of balancing these properties is known as the *Bias Variance Trade-off* problem [18]. One possible way to find the best model is to use the Cross Validation method [3].

Cross-Validation is a frequently used procedure for evaluating a model. The basic idea is that training data are divided into complementary subsets; one subset is used to train the model and we validate the results using the other subset. To do so, we have decided to use the Grid Search method [2] for choosing the best parameters for the algorithm. For each parameter of the algorithm, a list of possible values is given in input to Grid Search. Basically, each combination of the selected values generates a model that is then evaluated. The output of Grid Search is then the list of chosen parameters that performed the best.

3.3 Training Model to Predict Data

After computing the best parameters for the Support Vector Machine model, the next step has been to train the model with the training data and using the previously found parameters, in order to conclude the process. Once the model has been created, it was ready to be fed with new unseen data in order to output its predictions. As explained

before we have split, at the beginning, our whole processed data in training data and test data; the latter have been used for this last step to evaluate the accuracy of the created model. In Section 4 we will analyse the obtained results considering, separately, data of each body location.

4 Empirical Evaluation of the Approach

As a case study for showing one of the possible results achievable with our dataset we report in this section the evaluation of our approach using the data we collected. The research question we investigated is the following:

RQ: What is the accuracy of the proposed approach in classifying the activities performed by a subject?

Note that in this preliminary study, we independently consider the three devices/body locations. Moreover, we analyse only the performances of a person-dependent model where the training is on a subject and the test on the same subject.

4.1 Procedure

To answer our research question, first, we computed three confusion matrices for each subject in our dataset (one matrix for each of the three devices employed). More in detail, the values in each confusion matrix refer to the percentage of processed data of a specific class C_a that have been predicted to belong to the class C_b . More precisely, let us assume that we are reading the confusion matrix starting from the first row, representing the class C_a : each value we see in this row represents the percentage of data belonging to C_a that has been labelled as belonging to the class of the corresponding column. A flawless result would be represented as a matrix in which all the values on the diagonal are 100.0%, and the other values are 0.0% meaning that all the unseen data have been classified with the correct corresponding label.

Second, we averaged the eight confusion matrices (one per subjects) creating a single confusion matrix for each considered body location. In this way, we can answer our RQ by providing the average for each activity considered in our dataset.

An important aspect to consider when judging the quality of the results in the confusion matrices is that a baseline model that would randomly recognise the activity could have an accuracy equal to the probability of assigning the correct label that is:

$$\frac{1}{\#(classes)} = \frac{1}{17} = 5.9\%.$$

4.2 Results

In Figure 1 it is possible to see the confusion matrix obtained using **wrist** data, averaged over all eight subjects' results. The overall mean F1 score obtained is 0.92 ± 0.03 (mean \pm standard deviation). In general, we can say that the recognition of most of the activities achieved good results (values on the diagonal of the matrix are always greater than 0.74). In this case, the most noticeable outliers are in the wrong classifications of *keyboard typing in using laptop*, that are indeed very similar activities. The same analysis can be valid for the classifications of *sweeping* and *vacuuming*. The lower accuracy values are

obtained in most of those activities that mostly involve legs movements (*walking, going downstairs/upstairs*): those are indeed quite similar activities when "observed" from the wrist.

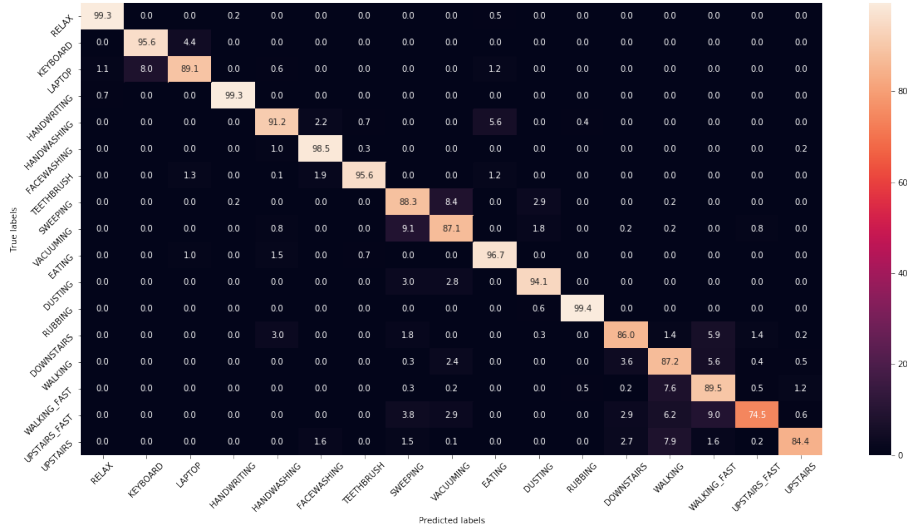


Fig. 1. Average Confusion Matrix obtained with Wrist Data

In Figure 2 and Figure 3 we present the confusion matrices obtained using, respectively, **hip** and **ankle** data, averaged over all eight subjects results, for which it is possible to perform similar considerations. The mean F1 score obtained with *hip* data is 0.81 ± 0.04 , while the mean F1 score obtained with *ankle* data is 0.75 ± 0.06 .

Analysing these confusion matrices (Figures 2 and 3, we noticed both expected and unexpected results. In fact, as expected, since many performed activities mostly involve peculiar movements of the arms (e.g. *brushing teeth, washing hands/face, sweeping*), results obtained using hip and ankle data have a lower mean accuracy than the results obtained using wrist data. For the same reason, we were expecting to obtain low accuracy for the activities performed while sitting or while not walking (e.g. *using laptop, relaxing, handwriting*) since the hip and ankles are not involved in any movements. On the contrary, we achieved quite high accuracies.

We further analysed our data in order to explain these results. By plotting the accelerometer data, we noticed that there was a perceptible difference in the values between such different activities even in the ankle and hip data. We interpreted this as the fact that subjects, during data recording, unintentionally changed the orientation of the devices (e.g. by slightly moving a leg while sitting). These involuntary movements were leading to noticeable changes in the accelerometer values because of the variation in the orientation with respect to the earth gravity g . We concluded that in some cases, the right classification of activities happens not because of the peculiar characteristics of the activity movements, but because of the particular orientation of the device.

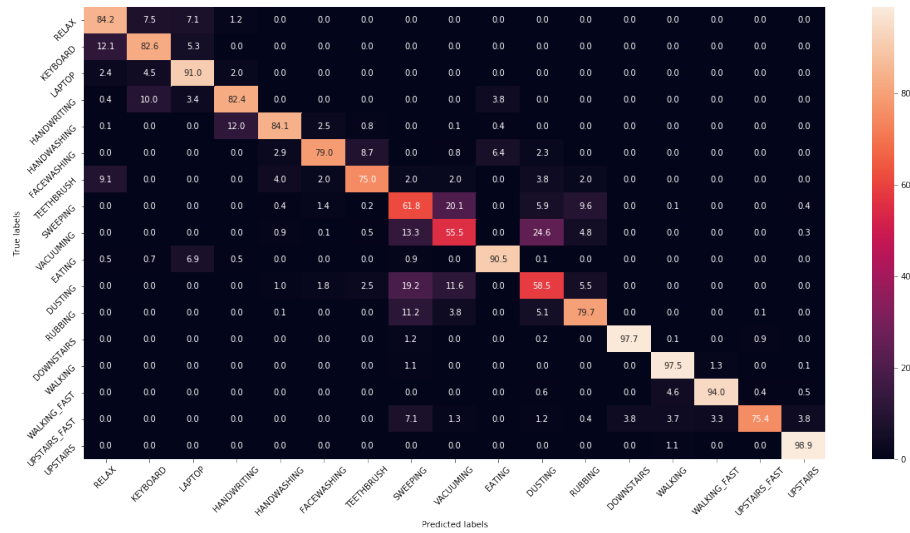


Fig. 2. Average Confusion Matrix obtained with Hip Data

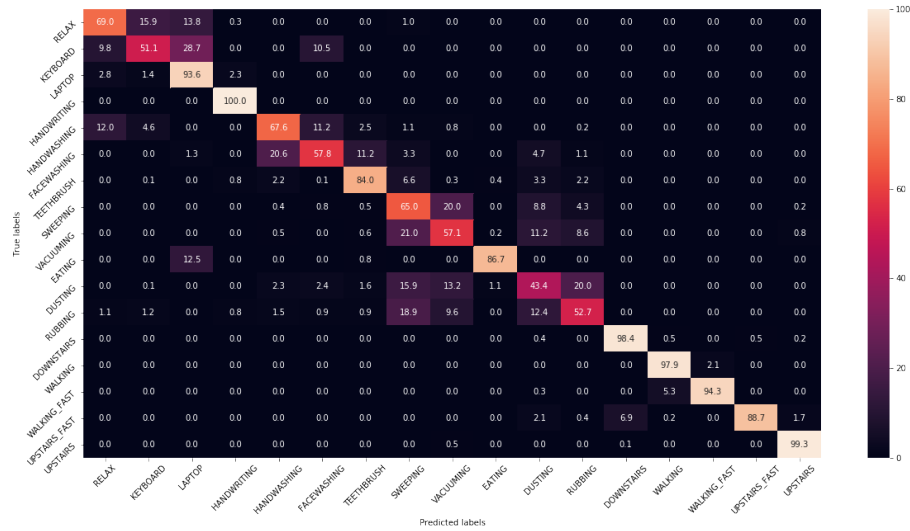


Fig. 3. Average Confusion Matrix obtained with Ankle Data

Therefore, in order to avoid this problem, when dealing with both hip and ankle data, we considered only activities that actively involve those parts of the body. In particular, we selected: *relaxing* (as a stationary activity), *sweeping*, *vacuuming*, *dusting*, *rubbing*, *going downstairs*, *walking*, *walking fast*, *going upstairs*, *going upstairs fast* and excluded *keyboard typing*, *using laptop*, *handwriting*, *hands washing*, *face washing*, *teeth brushing*, *eating*.

We show the confusion matrices obtained with the latest reduced activity set in Figure 4 (regarding hip data) and Figure 5 (regarding ankle data). In this case, the mean F1 score obtained with *hip* data is 0.48 ± 0.02 (mean \pm standard deviation), and the mean F1 score obtained with *ankle* data is 0.47 ± 0.03 .

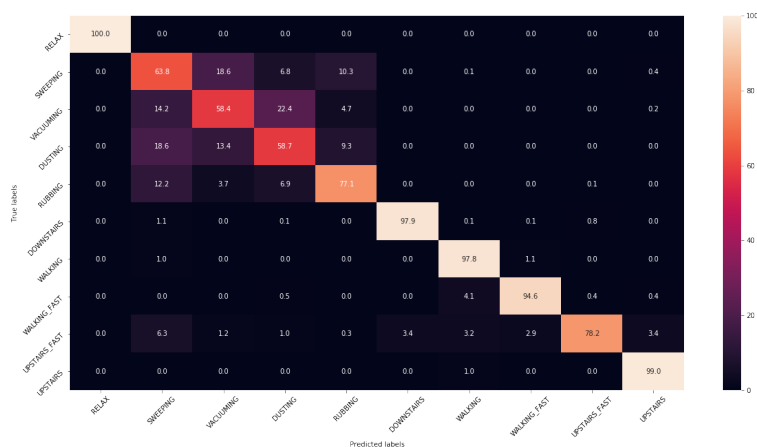


Fig. 4. Average Confusion Matrix obtained with Hip Data, limited on hip-related activities

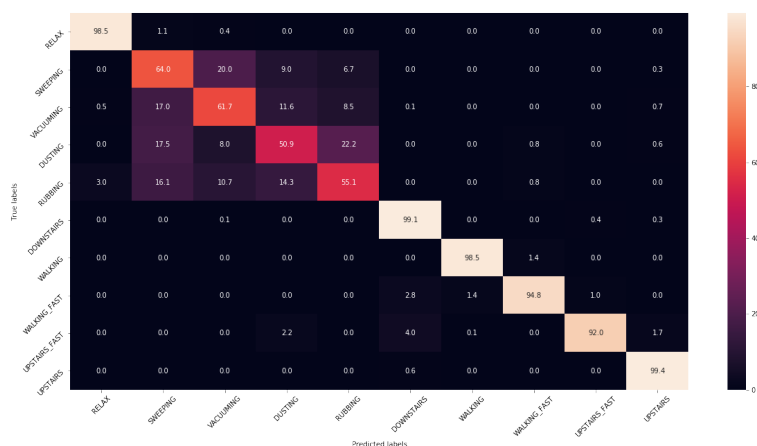


Fig. 5. Average Confusion Matrix obtained with Ankle Data, limited on ankle-related activities

In both confusion matrices (Figures 4 and 5) with a limited set of activities, it is clearly evident the scarce accuracy of the classifier in discerning from *sweeping*, *vacuuming*, *dusting* and *rubbing*. Indeed, all of the four listed activities have been performed

by doing small and slow steps around the room when recording data. Regarding the overall mean F1 scores it is clearly a consequence of the wrong classification of the four activities previously listed. On the same topic, we should also consider that by having fewer activities to be recognised, any wrong activity classification will have a significant impact on the F1 score.

On the other hand, the classifier is able to recognise with a quite high accuracy (always over 78%) all the other activities and, in particular, the ones that required the subjects to walk and use stairs (*walking, going downstairs, upstairs, upstairs fast*), in which both ankle and hip are more involved.

5 Related Works

In this section, we will briefly analyse related works, starting with the publicly available datasets on activities recorded with wearable devices and then presenting some approaches to activity recognition.

Datasets. As briefly explained in Section 1, when looking for a publicly available dataset, we have focused our analysis on three main criteria: (1) number and kind of recorded activities, (2) reliability of the recorded data according to the used device, and (3) which and how many parts of the body have been interested during data recording. To the best of our knowledge, a dataset satisfying the three aforementioned criteria is not currently available and this motivated our proposal.

About the first criterion, it is possible to find datasets focused on specific contexts of daily life: De la Torre et al. [17] presented a dataset on cooking activities, while Chavarriaga et al. [4] proposed a dataset on activities performed while preparing breakfast. On the other hand, it is also possible to find datasets related to a wider list of activities. Possible examples are the work of Anguita et al. [1], including more generic activities like *sitting, standing, walking, walking upstairs/downstairs* or the work of Leutheuser et al. [8], including activities from a daily life scenario (e.g., *walking, vacuuming, washing dishes, lying, sitting*). We have noticed that many available datasets include quite similar activities such as *walking* but at different speeds, or in different directions, *sitting, standing* or *lying*. Micucci et al. [10], in fact, with their brief literature review, have found out that the most frequent activities included in daily life activities dataset are: *walking, standing* and *walking downstairs/upstairs*.

Regarding the second criterion, a large number of the datasets that we have analysed used data recorded with an Android smartphone, with a requested sampling frequency of 50 Hz (e.g. [16, 1, 13]). Nevertheless, according to the work of Micucci et al., Android OS does not guarantee the consistency between the requested and the effective frequency sampling rate, therefore, the acquisition rate actually fluctuates during the acquisition [10]. This fact reduces, in our opinion, the reliability of the recorded data. On the contrary, some datasets use efficient devices with a high sampling frequency rate (> 100 Hz) as the work of Leutheuser et al. [8] or the work of Zhang et al. [20].

On the third criterion, during our investigation on existing datasets, we have seen that some were focused only on one part of the body, particularly on pockets or the hip. Possible examples are the works of Micucci et al. [10], Zhang et al. [20] or Anguita et al. [1]. On the other hand, other available datasets include data of multiple sensors on

different parts of our body, usually including waist, wrist, hip and ankle data. This is the case of the works of Sztyler et al. [16], Shoaib et al. [13] or Leutheuser et al. [8]. In our opinion, having data retrieved from different parts of our body would allow to achieve higher accuracy in activity recognition purposes.

Approaches for Activity Recognition. Different approaches for classifying daily-life activities using Machine Learning algorithms have been proposed in the last years. Here we will consider three works that have similar scenarios to ours. Indeed, all of the considered methods deal with a triaxial accelerometer worn on the wrist by participants of the experiments while performing some activities. All the devices used in the considered experiments recorded accelerations at a frequency of 80-100 Hz.

The work of Zhang et al. [21] tried to classify 4 main categories of activities: *sedentary* (lying, standing, PC working), *household* (window washing, sweeping, etc.), *walking* and *running* at different speeds. Mannini et al [9] tried to recognise as well 4 categories of activities: *ambulation, cycling, sedentary and other*. Yang et al. [19] has categories of activities more similar to our scenario: *walking, running, scrubbing, standing, working at a computer, vacuuming, brushing teeth* and *sitting*. For what concerns the features, the ones used in all the experiments are based on three different aspects: (1) time (mean, standard deviation, mean absolute deviation, etc. of acceleration over time); (2) frequency spectrum (first dominant frequencies and their power in some particular ranges - e.g. [0.6, 2.5] Hz) and, (3) wavelet, based on the Discrete Wavelet Transform, therefore obtaining features linking both the frequency and the time domain.

Regarding the windows length, in the aforementioned works this parameter was varying from 2.0 and 12.8 seconds, or, in term of number of samples, from 100 samples to 1152 (but taken at different frequencies). Those studies that compared the performances over the same data but using different length for the windows (e.g. [9]) have shown that longer windows would have meant higher performances, but also that 4.0 seconds windows were sufficient to obtain acceptable results.

Regarding the used ML algorithms, Zhang et al. [21] tested performances using different algorithms (Decision Trees, Naive Bayes, Linear Regression, Support Vector Machine, Neural Networks) showing that all the algorithms had good performances (>95.0%), with the DT and SVM being the ones with better results. Experiments in [9] used Support Vector Machine only, while in [19] a “neuro-fuzzy” classifier (classifiable as a Neural Network) has been used. All of the analysed algorithms in the documents were obtaining good and almost similar results, despite of the used algorithms (overall accuracy always over 86%).

The major differences w.r.t. our approach regard: (1) the windows length (we adopted shorter windows of 2.0 seconds) and, (2) the number of activities to be classified (higher in our case).

6 Conclusions and Future Work

In this paper, we have presented the current progress concerning the creation of a daily life activities dataset recorded while wearing multiple medical-grade wearable devices. With the help of the proposed approach for activity recognition, we have shown an example of prospective results that researchers could obtain using our dataset.

While being still incomplete (since we are working on recording the data of additional subjects), our dataset has interesting characteristics that could help researchers to perform deeper studies on the field of activity recognition. Differently from already available datasets, ours contains data: (1) of numerous daily life activities, (2) recorded using professional devices, and (3) of three synchronised devices on wrist, hip and ankle. Thanks to these characteristics, our dataset could help researchers to perform various kind of studies such as (1) work on subject-independent models, (2) find possible correlations between data of different sensors and different parts of the body while performing activities.

We have also described a preliminary approach to activity recognition and evaluated it using data from our dataset. By considering independently each device (i.e., body location), we have evaluated the predictions of a person-dependent model. From our results we have seen that it is possible to train a person-dependent model able to recognise the performed activities precisely, since using *wrist* data we achieved an average overall accuracy of 0.92 expressed as F1 score.

As future work, we are currently making progress on both the dataset and the proposed approach. Regarding the dataset, in the next months we plan to add data of 25-30 new subjects, in order to involve a more heterogeneous population (e.g., including female subjects or subjects with different ages). Concerning the proposed approach, we plan to evaluate and compare the accuracy of other classifiers (e.g. Random Forest, Neural Networks) or the influence derived from using different parameters for windows length and overlap. Additionally, we are studying the accuracy of a *subject independent* model, that would be able to recognise activities performed by a new unseen person with no need of subject-related training data. We will then compare the accuracy of our proposed approach with other already existing, by using our dataset as a benchmark. Moreover, in our future work we plan to include in our dataset also subjects suffering from physical impairments due to various diseases (e.g. Pulmonary Arterial Hypertension, Multiple Sclerosis, Parkinson’s disease) or the advanced age. This could also help researchers in better understanding and measuring the impact of such conditions on the daily life.

Acknowledgement. This work is an activity of the Software Engineering for Healthcare Lab, a joint laboratory between the University of Genova and Janssen Pharmaceuticals (Johnson & Johnson group). We want to show our gratitude to Massimo Raineri e Giacomo Ricca of Janssen Pharmaceuticals for the support provided.

Disclaimer. Responsibility for any information and result reported in this paper lies entirely with the Authors. Janssen Pharmaceuticals was not involved in any form of data collection or analysis.

References

1. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: Esann. vol. 3, p. 3 (2013)
2. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *The Journal of Machine Learning Research* **13**(1), 281–305 (2012)
3. Cawley, G.C., Talbot, N.L.: On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research* **11**, 2079–2107 (2010)

4. Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S.T., Tröster, G., Millán, J.d.R., Roggen, D.: The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* **34**(15), 2033–2042 (2013)
5. He, Z., Jin, L.: Activity recognition from acceleration data based on discrete cosine transform and svm. In: 2009 IEEE International Conference on Systems, Man and Cybernetics. pp. 5041–5044. IEEE (2009)
6. John, D., Freedson, P.: Actigraph and actual physical activity monitors: a peek under the hood. *Medicine and science in sports and exercise* **44**(1 Suppl 1), S86 (2012)
7. Kantoch, E., Augustyniak, P., Markiewicz, M., Prusak, D.: Monitoring activities of daily living based on wearable wireless body sensor network. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 586–589. IEEE (2014)
8. Leutheuser, H., Schulhaus, D., Eskofier, B.M.: Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. *PloS one* **8**(10), e75196 (2013)
9. Mannini, A., Intille, S.S., Rosenberger, M., Sabatini, A.M., Haskell, W.: Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise* **45**(11), 2193 (2013)
10. Micucci, D., Mobilio, M., Napolitano, P.: Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences* **7**(10), 1101 (2017)
11. Pitta, F., Troosters, T., Probst, V., Spruit, M., Decramer, M., Gosselink, R.: Quantifying physical activity in daily life with questionnaires and motion sensors in copd. *European respiratory journal* **27**(5), 1040–1055 (2006)
12. Robusto, K.M., Trost, S.G.: Comparison of three generations of actigraphTM activity monitors in children and adolescents. *Journal of sports sciences* **30**(13), 1429–1435 (2012)
13. Shoaib, M., Scholten, H., Havinga, P.J.: Towards physical activity recognition using smartphone sensors. In: 2013 IEEE 10th international conference on ubiquitous intelligence and computing and 2013 IEEE 10th international conference on autonomic and trusted computing. pp. 80–87. IEEE (2013)
14. Staudenmayer, J., He, S., Hickey, A., Sasaki, J., Freedson, P.: Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of applied physiology* **119**(4), 396–403 (2015)
15. Stikic, M., Larlus, D., Ebert, S., Schiele, B.: Weakly supervised recognition of daily life activities with wearable sensors. *IEEE transactions on pattern analysis and machine intelligence* **33**(12), 2521–2537 (2011)
16. Szytler, T., Stuckenschmidt, H.: On-body localization of wearable devices: An investigation of position-aware activity recognition. In: 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom). pp. 1–9. IEEE (2016)
17. De la Torre, F., Hodgins, J., Bargteil, A., Martin, X., Macey, J., Collado, A., Beltran, P.: Guide to the carnegie mellon university multimodal activity (cmu-mmact) database (2009)
18. Von Luxburg, U., Schölkopf, B.: Statistical learning theory: Models, concepts, and results. In: *Handbook of the History of Logic*, vol. 10, pp. 651–706. Elsevier (2011)
19. Yang, J.Y., Chen, Y.P., Lee, G.Y., Liou, S.N., Wang, J.S.: Activity recognition using one triaxial accelerometer: A neuro-fuzzy classifier with feature reduction. In: *International conference on entertainment computing*. pp. 395–400. Springer (2007)
20. Zhang, M., Sawchuk, A.A.: Usc-had: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. pp. 1036–1043 (2012)
21. Zhang, S., Rowlands, A.V., Murray, P., Hurst, T.L., et al.: Physical activity classification using the GENEa wrist-worn accelerometer. Ph.D. thesis, Lippincott Williams and Wilkins (2012)