OpenHAR: A Matlab Toolbox for Easy Access to Publicly Open Human Activity Data Sets - introduction and experimental results

Pekka Siirtola, Heli Koskimäki and Juha Röning

Abstract This study introduces OpenHAR, a free Matlab toolbox to combine and unify publicly open data sets. It provides an easy access to accelerometer signals of ten publicly open human activity data sets. Data sets are easy to access as OpenHAR provides all the data sets in the same format. In addition, units, measurement range and labels are unified, as well as, body position IDs. Moreover, data sets with different sampling rates are unified using downsampling. What is more, data sets have been visually inspected to find visible errors, such as sensor in wrong orientation. OpenHAR improves re-usability of data sets by fixing these errors. In addition, OpenHAR provides Matlab code to easily select only desired parts of this data. In fact, also users without Matlab have an access to OpenHAR data but not for this code. For them, the whole OpenHAR data is provided as a one .txt-file. Altogether OpenHAR contains over 65 million labeled data samples. This is equivalent to over 280 hours of data from 3D accelerometers. This includes data from 211 study subjects performing 17 daily human activities and wearing sensors in 14 different body positions. A small experiment was performed to show the potential of OpenHAR. The experiment was done using three classifiers: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and classification and regression tree (CART). The experiment showed that using LDA and QDA classifiers and OpenHAR data, as high recognition rates can be achieved in a previously unseen test data than by using a data set specially collected for this purpose. With CART the results obtained using OpenHAR data were slightly lower.

Biomimetics and Intelligent Systems Group, University of Oulu, FI-90014 Oulu, Finland
e-mail: pekka.siirtola@oulu.fi, heli.koskimaki@oulu.fi, jjr@oulu.fi

1 Introduction

Inertial sensor based human activity recognition using wearable and smartphone sensors is a widely studied area of pattern recognition. It can be applied to several different types of applications. Mostly it is used for health and fitness monitoring, but also for personalized advertising; smarthomes that anticipates the user's needs; self-managing system that adapts to user's activities, and for context-aware applications [9, 13].

The first step to build a reliable human activity recognition model is to collect an extensive data set from the studied problem. Unfortunately, this can be very time consuming as this data not just needs to be collected but labeled as well. Luckily, it is not always necessary to collect a new data set as it is more and more common that data sets used in the previous activity recognition studies are made publicly available. Moreover, by combining multiple publicly open data sets, bigger data set can be build and bigger data set normally means more general and accurate recognition model. The problem is that it is not always that easy to combine data sets.

In our previous article [15], different publicly open human activity data sets were cross-validated. The recognition models were trained using one data set and tested using another to see how well models work when data for training and testing are collected in different environments and using different sensors. It was noted in the article that publicly open data sets are not always that easy to use and combine. For instance, it was noted that data sets are often stored in different formats, sensor orientation varies, units are not always the same, etc. Moreover, in [14] personalized human activity recognition models were studied and the experiments were based on publicly open data set containing data from ten study subjects. However, in the study, it was decided that data from one study subject was not used as apparently one subject had worn sensor in different orientation than others making this data non-uniform with other subjects data.

This study is an extension to [16], which introduced OpenHAR - a free Matlab toolbox combining publicly open data sets. It provides an easy access to accelerometer signals of ten publicly open human activity data sets. Data sets are easy to access as OpenHAR provides all the data sets in the same format, units, measurement range and labels are unified, as well as, body position IDs. Moreover, data sets with different sampling rates are unified using downsampling. What is more, data sets have been visually inspected to find visible errors, such as sensors in wrong orientation. OpenHAR improves re-usability of data sets by fixing these errors. In addition, OpenHAR provides Matlab code to easily select only desired parts of this data. In fact, also users without Matlab have an access to OpenHAR data but not for this code. For them, the whole OpenHAR data is provided as a one .txt-file. The study most similar to this is presented in [3], where a data set called AcctionNet collating six publicly open data sets was introduced. This data contains over 10 million labeled accelerometer samples samples from 13 activities. The data

sets used in AcctionNet are partly the same as the ones used in OpenHAR. The main difference between AcctionNet and OpenHAR is that OpenHAR is not just a data set, it also provides tools to select only that part of data that is important to certain application.

2 OpenHAR

2.1 Collated data sets

OpenHAR combines ten publicly open data sets, which are listed in Table 1. Common with these data sets is that they all contain raw accelerometer data collected with reasonable sampling rate from activities of daily living. There is also other publicly open data sets available, but they were not included to this study as they do not fulfill our requirements: for instance, Reiss *et. al.* [11] was not included to OpenHAR as data of it filtered and not raw. Moreover, SHL data set which is an excellent and extensive activity data set by Gjoreski *et. al.* [6] was not included to OpenHAR as it is so huge compared to selected ten data sets that it would have a too dominant role in the combined data set.

According to Table 1, data sets of OpenHAR are not consistent which means that combining of the data sets is not as straightforward as it could be. Data sets are in multiple format and also data files are grouped in several different ways. In some cases, the whole data set is stored in one single file but often data are divided into multiple folders and files. In addition, when it comes to activity labels, both integers and strings were used as labels in the original data sets. Moreover, numerical labels did not have the same response in different data sets. In some cases, labels also had different meaning, for instance depending on the studied data set, *walking*, *walking upstairs* and *walking downstairs* had own labels but it was also possible that label *walking* included walking at flat level and walking at stairs. Another difference in the data sets is the used sampling frequency of acceleration data which varied from 40Hz to 200Hz. Moreover, accelerometer values differed in the provided value range and units. Visual mining of the data sets also showed some errors and non-uniformities from data sets. For instance, there are cases where sensor orientation of one study subject is not the same than for others.

2.2 Unifying data sets

The aim of the study was to unify the selected data sets, and therefore, provide an easy access to these data sets and combine them to get access to a bigger data set. The data sets presented in Table 1 comes in multiple file

Table 1 OpenHAR contains ten publicly open human activity data sets.

Data set ID	Author	File format	Frequency	Labels	Range and unit
1	Banos *et. al.*[2]	.log	50Hz	numeral	$\pm 24m/s^2$
2	Ortiz *et. al.* [1]	.txt	50Hz	numeral	$\pm 2g$
3	Shoaib *et. al.* [12]	.csv	50Hz	strings	$\pm 20m/s^2$
4	Siirtola & Röning [18]	.txt	40Hz	numeral	$\pm 20m/s^2$
5	Stisen *et. al.* [19]	.csv	50 - 200Hz	strings	$\pm 40m/s^2$
6	USC-HAD [23]	.mat	100Hz	numeral	$\pm 6g$
7	UniMib-SHAR [10]	.mat	50Hz	numeral	$\pm 20m/s^2$
8	HuGaDB [5]	.txt	60Hz	numeral	± 32767
9	RealworldHAR [21]	.csv	50Hz	strings	$\pm 20m/s^2$
10	MobiAct [22]	.csv	200Hz	strings	$\pm 20m/s^2$

types and formats. In some cases, the whole data set is in one single file but often data are divided into multiple folders and files based on study subject ID, body position ID or activity labels. In fact, one of the main benefits of OpenHAR is that it provides code to load these without taking care of file formats, and the resulting data set has only one format.

Another problem in combining open data sets is that currently labels can be numeral or strings, and a number or string can have different meaning in different data sets. OpenHAR unifies these activity labels. It provides all the labels in numerical format and these labels have only one meaning. Activity labels used in OpenHAR are presented in Table 3. However, some of these activities are overlapping, which needs to be noted when OpenHAR data are used. For instance, [18] contained activity *idling*, which is a combination of sitting and standing, while in other data sets *sitting* and *standing* were considered as two separate activities. Similarly some data sets consider *walking*, *walking upstairs* and *walking downstairs* as separate activities and in some data sets these all three are considered as one activity called *walking*. The same goes to *elevator up* and *elevator down* activities, in some cases they are combined as *elevator*-activity (direction not defined). These subjects performed altogether 17 daily human activities, Table 3 although most data is from walking (19.9%), standing (15.6%) and sitting (13.1%) activities.

Depending on the purpose of the original article, sensor position differs between data sets. As the measured sensor values are greatly dependent on the body position of the sensor, each observation of OpenHAR has a cell defining from which body position the value has been measured. These positions and the position ID's are listed in Table 4. It is worth noting that some of these are overlapping. For instance, in some studies sensor position was defined as trouser's pocket, meaning that it can either left or right, while in some cases position was explicitly defined as left of right pocket. In addition, in some cases study subjects were allowed to decide the orientation of the sensor while in some studies orientation was fixed. Moreover, some of the body positions were combined as they are so similar, for instance hip, waist and

Table 2 Fixes needed for the datasets.

Data set ID	Fixes
1	timestamp added
3	subj. 8, belt: orientation fixed
4	timestamp added
5	sampling rates unified
6	timestamp added
7	timestamp added
8	timestamp added
9	subj.8, chest: orientation fixed, subj.15, thigh: orientation fixed, subj.3, upperarm: orientation fixed, subj.3, waist: orientation fixed
10	-

belt positions were combined as one. Most of the data is from hip (22.5%) and trouser's pocket (22.2%, including also data from thigh).

Another difference in the data sets is the used sampling frequency of acceleration data which varies from 40Hz to 200Hz. One feature of OpenHAR is that it unifies sampling rates using depending on which data sets are combined. Unifying is based on down-sampling by finding the greatest common divider of the sampling rates of the selected data sets. If all ten data sets are combined, the sampling rate of the resulting data set is 10Hz, which has been shown to be enough to reliably recognize activities [17, 8, 20].

Table 1 shows that unit of the measurement is either g (gravity) or m/s^2. Moreover, accelerometer differed in provided value range and units. OpenHAR converts all the units as m/s^2 to enable the joint usage of the data sets. In addition, the range of measurements is unified.

Visual mining of the data sets also showed some errors and non-uniformities from data sets, see Table 2. For instance, there are cases where sensor orientation of one study subject is not the same than for others. These were corrected by changing the common coordinate system to all data files within a data set. However, it should be noted that the orientation of a sensor is the same only within original data sets but when two original data sets are compared, the orientation of a sensor can differ. Nevertheless, this orientation issue can be solved for example using features extracted from magnitude signal, which is automatically calculated by OpenHAR using formula $\sqrt{(\mathbf{x}^2 + \mathbf{y}^2 + \mathbf{z}^2)}$, where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are the raw acceleration measurements from 3D accelerometer. Moreover, only a few original data set has timestamps, these were added to all data sets based on the sampling rate and by considering that it remains constant.

Altogether, OpenHAR contains over 65 million labeled data samples. This is equivalent to over 280 hours of data from 3D accelerometers. This includes data from 211 study subjects, see Figure 1. While the amount of data from each study subject varies a lot between data set (minimum 2 minutes, maximum 710 minutes), on average there is 80 minutes of data from each subject.
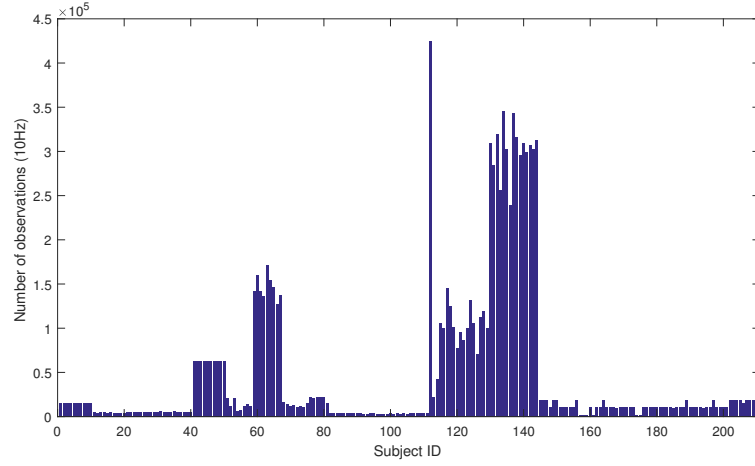
Fig. 1 OpenHAR contains data from 211 study subjects.

Table 3 Data set includes accelerometer data from 17 activities. However, some of these are overlapping.

Activity ID	Activity	Amount of data
1	Standing	15.6%
2	Sitting	13.1%
3	Lying	8.0%
4	Idling (= sitting + standing)	0.4%
5	Walking	19.9%
6	Walking (inc. walking at stairs)	0.2%
7	Walking stairs up	10.3%
8	Walking stairs down	8.9%
9	Walking at stairs (inc. up and down)	0.2%
10	Running (inc. jogging)	10.4%
11	Biking	4.8%
12	Jumping	1.8%
13	Sitting in car	1.9%
14	Elevator up	1.1%
15	Elevator down	0.9%
16	Falling	0.7%
99	Null	1.9%

3 Using OpenHAR

To use OpenHAR, the first step is to download the toolbox from our research group's web page[1]. Next step is to download and unpack all ten original data sets (see ReadMe.txt -file to find from where to download them and where

---

[1] OpenHAR is available at: https://www.oulu.fi/bisg/datasets

Table 4 OpenHAR includes data from 14 body positions. However, some of these are overlapping.

Position ID	Position	Amount of data
1	Hip (inc. belt and waist)	22.5%
2	Trouser's pocket, left (fixed orientation)	1.3%
3	Trouser's pocket, right (fixed orientation)	1.3%
4	Trouser's pocket, any (inc. thigh)	22.2%
5	Chest	7.0%
6	Wrist, any (inc. forearm)	10.1%
7	Upper arm	7.7%
8	Head	6.4%
9	Shin (inc. leg)	13.3%
10	Ankle	0.5%
11	Trouser's pocket, left (free orientation)	0.5%
12	Trouser's pocket, right (free orientation)	0.5%
13	Foot, left	3.4%
14	Foot, right	3.4%

to unpack them). After this, everything is set. To get the whole OpenHAR experience, Matlab is required. However, users without Matlab can also use OpenHAR. Unified data from all the ten data sets is available as a single .txt-file. Therefore, the sampling rate of this data set is 10Hz. However, users without Matlab cannot have the whole OpenHAR experience as they have access to the data, but cannot use the commands presented in this section. This data can also be downloaded from our research group's web page.

Using OpenHAR is easy, only one command is needed to download all the data to *data*-file. ([`data, sampling_rate`] = `getOpenHAR()`). In addition, this command returns the sampling rate of data. In this case, all the data are downloaded and so the sampling rate would be 10Hz. In fact, this is the data that is provided as a single .txt-file for users without Matlab.

User can also load only wanted parts of the whole data set by three name-value -pairs arguments (`'datasets'`, `'activities'`, and `'positions'`). User can specify several or only one name and value pair argument in any order.

If the purpose is not to use data from all ten original data sets, wanted data sets can be specified as the comma-separated pair consisting of `'datasets'` and a vector containing th IDs of the wanted data sets, see Table 1 for dataset IDs. For example, the command [`data, sampling_rate`] = `getOpenHAR('datasets', [1 3 5])` returns only measurements from original data sets corresponding to data set IDs 1 ([2]), 3 ([12]), and 5 ([19]). If all the selected data sets do not have the same sampling frequency, OpenHAR unifies them and return the sampling rate of the combined data set.

Similarly, if only some activities or body positions are of interest, wanted activities and body positions can be specified as the comma-separated pair consisting of `'activities'` and `'positions'`, and a vector containing the

IDs of the wanted activities or body positions. IDs for these are listed in Tables 3 and 4. This means that user can select only data from some activities or body positions, or only data from some activities from selected body positions.

Each case, the code returns *data*-file. This file has nine columns of data in the following order: *data set ID, position ID, user ID, activity ID, timestamp, x-axis acceleration, y-axis acceleration, z-axis acceleration* and *magnitude acceleration*. File does not have a header.

When OpenHAR is used, we encourage users to cite to this article, and in addition, to cite to the publications were the original data sets were introduced. BibTeX-information for OpenHAR and original articles are provided in the ReadMe.txt -file of OpenHAR Matlab-package.

4 Experiments

This section shows how OpenHAR can be used to build reliable models for human activity recognition. Models are tested using one data set of OpenHAR and trained using the rest nine data sets. Experiments are done using three different classifiers: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and classification and regression tree (CART). LDA is used to find a linear combination of features that separate the classes in an optimal way. The resulting combination and the the hyperplane separating the classes can then be employed as a linear classifier. QDA is a similar method but uses quadratic hyperplanes to separate classes [7]. CART is a standard decision tree, which uses certain criterion to partition the space spanned by the input variables to maximize the score of class purity [4]. Moreover, as sensor orientation within the ten data sets of OpenHAR can differ, it was decided to use only features extracted from the magnitude acceleration signal in the machine learning process. Altogether 17 features were extracted from the magnitude acceleration signal. These included standard deviation, mean, maximum, minimum, different percentiles, and frequency domain features.

In the experiment, data from Shoaib *el. al.* (data set ID 3) was use for testing, and the rest of the OpenHAR data for training. Therefore, the original testing data includes eight activities: walking, running, sitting, standing, jogging, biking, walking upstairs and walking downstairs. However, for the experiment walking, walking upstairs, and walking downstairs were combined as one walking-activity, and sitting and standing were combined as idling. The same was done for training data, and this way it was possible to use the rest nine data sets for training. Moreover, in OpenHAR jogging is labeled as running. Thus, the testing data sets contained four activities: walking, running, idling, and biking. In addition, Shoaib *el. al.* contains data from left and right trouser's pocket, wrist, upper arm, and belt positions. Therefore,

body position IDs 1, 2, 3, 4, 5, 6, 7, 8, 11, and 12 were selected for training data set.

Training was performed so that in the first place only one out of nine training data sets is used for training, and this model is tested using our test data set. Then incrementally more data sets, one at the time, are added to the training data set. Moreover, as a comparison, Shoaib *el. al.* is classified using leave-one-subject-out -method. This means that in turn one persons data is used for testing and data from other persons for training.
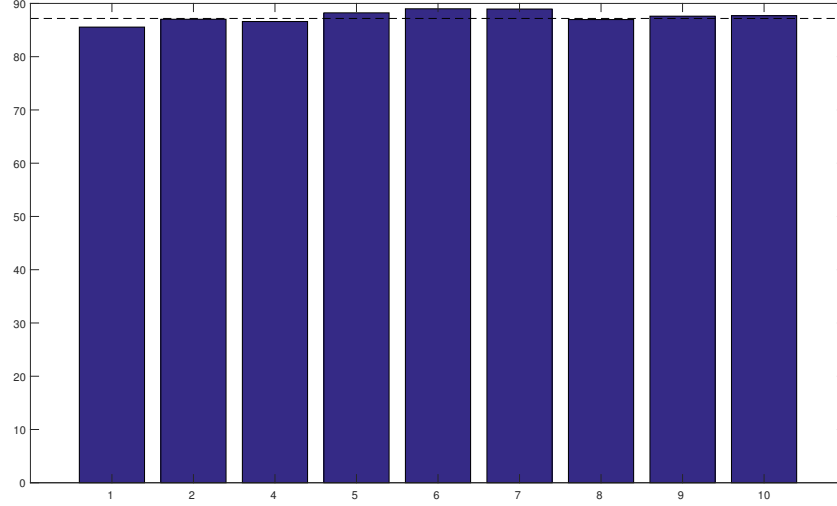
Fig. 2 Classification accuracies for Shoaib *el. al.* data set using LDA classifier, when new OpenHAR data sets are incrementally added to training data. Bar number shows the data set ID of the added data set. Horizontal line shows the mean accuracy obtained by classifying Shoaib *el. al.* using leave-one-subject-out -method.

Shoaib *el. al.* contains data from 10 persons, and the recognition results were calculated separately to each person. The average accuracy from these was calculated and the results are presented in Figures 2, 3, and 4. In addition, variance from these accuracies was calculated but variance remained practically the same no matter which training data was used. The results presented in the figures show that the best classification results are obtained using more than one OpenHAR data set. Therefore, it is beneficial to incrementally add more and more data sets to obtain the highest possible recognition accuracy. Moreover, it can be noted that with using LDA and QDA classifiers with OpenHAR data it is possible to obtain as high or even higher recognition rates in a previously unseen test data than by dividing this unseen data for test and training data. This means that instead of using time for collecting and labeling accelerometer data set, one can save time and just download OpenHAR and use the data provided by it. On the other hand, it
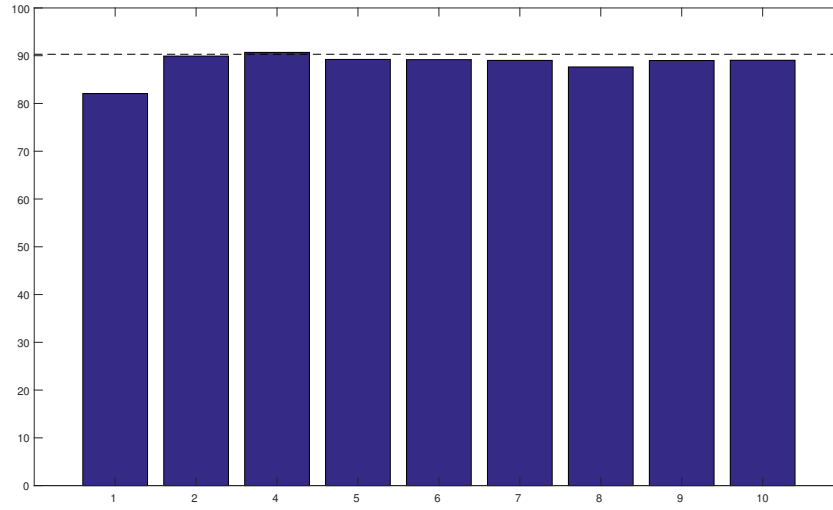
Fig. 3 Classification accuracies for Shoaib *el. al.* data set using QDA classifier, when new OpenHAR data sets are incrementally added to training data. Bar number shows the data set ID of the added data set. Horizontal line shows the mean accuracy obtained by classifying Shoaib *el. al.* using leave-one-subject-out -method.
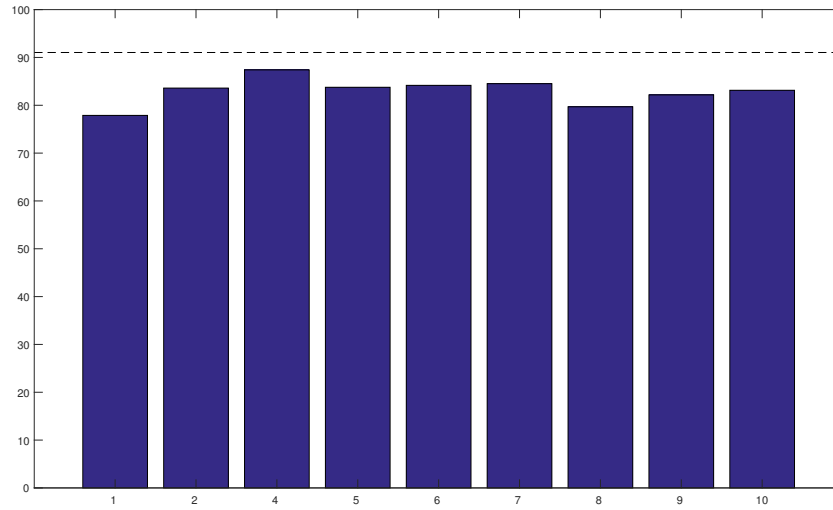

Fig. 4 Classification accuracies for Shoaib *el. al.* data set using CART classifier, when new OpenHAR data sets are incrementally added to training data. Bar number shows the data set ID of the added data set. Horizontal line shows the mean accuracy obtained by classifying Shoaib *el. al.* using leave-one-subject-out -method.

should be noted that with CART the results obtained using OpenHAR data were slightly lower than with other two classifiers. However, also when using CART, in the best case (using data set IDs 1, 2, and 4 as training data) the difference in accuracies was not big when comparing results obtained, 91.0% vs. 87.4%, respectively.

The experiment showed in this section was limited, but still it shows the potential of OpenHAR. However, to show the full potential of OpenHAR, more experiment should be done with multiple classifiers, including deep learning methods.

5 Discussion and conclusion

This article presents OpenHAR, a free Matlab toolbox combining ten publicly available human activity data sets. The extra value provided by OpenHAR is that it provides easy access to these ten data sets. In fact, OpenHAR provides all the data sets in the same format In addition, units, measurement range and labels are unified, as well as, body position IDs. Moreover, data sets with different sampling rates are unified using downsampling. What is more, data sets have been visually inspected to find visible errors, such as sensors in wrong orientation. OpenHAR improves re-usability of data sets by fixing these errors. Matlab code is provided to enable an easy way to select only desired parts of these data sets. In fact, also users without Matlab have an access to OpenHAR data but not for this code. With over 65 million labeled observation, 211 study subjects, 17 activities and 14 body position, OpenHAR is the most comprehensive accelerometer based human activity data set to date. OpenHAR can be downloaded from https://www.oulu.fi/bisg/datasets.

OpenHAR opens a lot of new possibilities to researchers and application developers. For instance, OpenHAR provides a great testbed to study deep learning methods and other data hungry classifiers. In addition, OpenHAR contains data from ten different data gathering protocols, which means that using OpenHAR is it possible to experiment how a model that is trained using data from one location and environment works when it is tested in other location. Thus, it can be used to test methods of transfer learning. In addition, OpenHAR contains data from 211 study subjects, which is more than any other data sets. The high number of study subjects is especially important when user-independent models are trained. Moreover, when used with traditional classifiers, more data usually means better and more accurate models. Most importantly, OpenHAR is publicly open, and therefore, studies based on it are easy to replicate.

A small experiment was done to show the potential on OpenHAR. The experiment was done using three classifiers: LDA, QDA, and CART. The experiment showed that using LDA and QDA classifiers and OpenHAR data, as high recognition rates can be achieved in a previously unseen test data

than by using a data set specially collected for this purpose. With CART the results obtained using OpenHAR data were slightly lower. The experiment was limited, but still it shows the potential of OpenHAR. However, to show the full potential of OpenHAR, the future work includes more comprehensive experiments and experimenting with multiple classifiers, including deep learning methods. Moreover, new data sets can be added to OpenHAR in the future when suitable data set are published. In addition, currently OpenHAR focuses on daily activities, however, it could be extended to include other types of activities as well.

6 Acknowledgements

References

1. Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., and Reyes Ortiz, J. L. A public domain dataset for human activity recognition using smartphones. In *Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (2013), 437–442.
2. Banos, O., Garcia, R., Holgado-Terriza, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A., and Villalonga, C. mhealthdroid: A novel framework for agile development of mobile health applications. In *Ambient Assisted Living and Daily Activities*, L. Pecchia, L. L. Chen, C. Nugent, and J. Bravo, Eds., Springer International Publishing (Cham, 2014), 91–98.
3. Bartlett, J., Prabhu, V., and Whaley, J. Acctionnet: A dataset of human activity recognition using on-phone motion sensors. In *Proceedings of the 34th International Conference on Machine Learning* (Sydney, Australia, 2017).
4. Breiman, L. *Classification and regression trees*. Routledge, 2017.
5. Chereshnev, R., and Kertész-Farkas, A. Hugadb: Human gait database for activity recognition from wearable inertial sensor networks. In *International Conference on Analysis of Images, Social Networks and Texts*, Springer (2017), 131–141.
6. Gjoreski, H., Ciliberto, M., Wang, L., Morales, F. J. O., Mekki, S., Valentin, S., and Roggen, D. The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access* (2018).
7. Hand, D. J., Mannila, H., and Smyth, P. *Principles of data mining*. MIT Press, Cambridge, MA, USA, 2001.
8. Kose, M., Incel, O. D., and Ersoy, C. Online human activity recognition on smart phones. In *Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data*, vol. 16 (2012), 11–15.
9. Lockhart, J. W., Pulickal, T., and Weiss, G. M. Applications of mobile activity recognition. In *2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, ACM (New York, NY, USA, 2012), 1054–1058.

10. Micucci, D., Mobilio, M., and Napoletano, P. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences 7*, 10 (2017), 1101.

11. Reiss, A., and Stricker, D. Introducing a new benchmarked dataset for activity monitoring. In *Wearable Computers (ISWC), 2012 16th International Symposium on*, IEEE (2012), 108–109.

12. Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., and Havinga, P. Fusion of smartphone motion sensors for physical activity recognition. *Sensors 14*, 6 (2014), 10146–10176.

13. Siirtola, P., Komulainen, J., Kellokumpu, and V. Effect of context in swipe gesture-based continuous authentication on smartphones. In *26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2018.* (April 25-27 2018), 639—-644.

14. Siirtola, P., Koskimäki, H., Röning, and J. Personalizing human activity recognition models using incremental learning. In *26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2018.* (April 25-27 2018), 627—-632.

15. Siirtola, P., Koskimäki, H., and Röning, J. Experiences with publicly open human activity data sets -studying the generalizability of the recognition models. In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, SCITEPRESS (2018), 291—-299.

16. Siirtola, P., Koskimäki, H., and Röning, J. Openhar: A matlab toolbox for easy access to publicly open human activity data sets. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, ACM (2018), 1396–1403.

17. Siirtola, P., Laurinen, P., Röning, J., and Kinnunen, H. Efficient accelerometer-based swimming exercise tracking. In *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on*, IEEE (2011), 156–161.

18. Siirtola, P., and Röning, J. Recognizing human activities user-independently on smartphones based on accelerometer data. *International Journal of Interactive Multimedia and Artificial Intelligence 1*, 5 (June 2012), 38–45.

19. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjærgaard, M. B., Dey, A., Sonne, T., and Jensen, M. M. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, ACM (2015), 127–140.

20. Suto, J., Oniga, S., Lung, C., and Orha, I. Comparison of offline and real-time human activity recognition results using machine learning techniques. *Neural Computing and Applications* (2018), 1–14.

21. Sztyler, T., and Stuckenschmidt, H. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on*, IEEE (2016), 1–9.

22. Vavoulas, G., Chatzaki, C., Malliotakis, T., Pediaditis, M., and Tsiknakis, M. The mobiact dataset: Recognition of activities of daily living using smartphones. In *ICT4AgeingWell* (2016), 143–151.

23. Zhang, M., and Sawchuk, A. A. Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *ACM International Conference on Ubiquitous Computing (Ubicomp) Workshop on Situation, Activity and Goal Awareness (SAGAware)* (Pittsburgh, Pennsylvania, USA, September 2012).