HTAD: A Home-Tasks Activities Dataset with Wrist-accelerometer and Audio Features

Enrique Garcia-Ceja[1], Vajira Thambawita[2,3], Steven A. Hicks[2,3], Debesh Jha[2,4], Petter Jakobsen[5], Hugo L. Hammer[3], Pål Halvorsen[2], and Michael A. Riegler[2]

[1] SINTEF Digital, Norway
[2] SimulaMet, Norway
[3] Oslo Metropolitan University, Norway
[4] UIT The Arctic University of Norway
[5] Haukeland University Hospital, Norway
michael@simula.no

Abstract. In this paper, we present HTAD: A Home Tasks Activities Dataset. The dataset contains wrist-accelerometer and audio data from people performing at-home tasks such as sweeping, brushing teeth, washing hands, or watching TV. These activities represent a subset of activities that are needed to be able to live independently. Being able to detect activities with wearable devices in real-time has the potential for the realization of assistive technologies with applications in different domains such as elderly care and mental health monitoring. Preliminary results show that using machine learning with the dataset leads to promising results, but also that there is still improvement potential. By making this dataset public, researchers can test different machine learning algorithms for activity recognition, especially, sensor data fusion methods.

Keywords: Activity recognition · Dataset · Accelerometer · Audio · Sensor fusion.

1 Introduction

Automatic monitoring of human physical activities has become of great interest in the last years since it provides contextual and behavioral information about a user without explicit user feedback. Being able to automatically detect human activities in a continuous unobtrusive manner is of special interest for applications in sports [17], recommendation systems, and elderly care, to name a few. For example, appropriate music playlists can be recommended based on the user's current activity (exercising, working, studying, etc.) [22]. Elderly people at an early stage of dementia could also benefit from these systems, like by monitoring their hygiene-related activities (showering, washing hands, or brush teeth) and sending reminding messages when appropriate [20]. Human activity recognition (HAR) also has the potential for mental health care applications [12] since it can be used to detect sedentary behaviors [4], and it has been shown that there is

an important association between depression and sedentarism [5]. Recently, the use of wearable sensors has become the most common approach to recognizing physical activities because of its unobtrusiveness and ubiquity, specifically, the use of accelerometers [9, 16, 18], because they are already embedded in several commonly used devices like smartphones, smart-watches, fitness bracelets, etc.

In this paper, we present HTAD: a Home Tasks Activities Dataset. The dataset was collected using a wrist accelerometer and audio recordings. The dataset contains data for common home tasks activities like *sweeping, brushing teeth, watching TV, washing hands*, etc. To protect users' privacy, we only include audio data after feature extraction. For accelerometer data, we include the raw data and the extracted features. The dataset can be downloaded via: `https://osf.io/4dnh8/`

There are already several related datasets in the literature. For example, the epic-kitchens dataset includes several hours of first-person videos of activities performed in kitchens [6]. Another dataset, presented by Bruno et al., has 14 activities of daily living collected with a wrist-worn accelerometer [3]. Despite the fact that there are many activity datasets, it is still difficult to find one with both: wrist-acceleration and audio. The authors in [21] developed an application capable of collecting and labeling data from smartphones and wrist-watches. Their app can collect data from several sensors, including inertial and audio. The authors released a dataset[6] that includes 2 participants and point to another website (http://extrasensory.ucsd.edu) that contains data from 60 participants. However, the link to the website was not working in the present date (August-10-2020). Even though the present dataset was collected by 3 volunteers, and thus, is a small one compared to others, we think that it is useful for the activity recognition community and other researchers interested in wearable sensor data processing. The dataset can be used for machine learning classification problems, especially those that involve the fusion of different modalities such as sensor and audio data. The dataset was previously used in [11]. This dataset can be used to test data fusion methods [14] and used as a starting point towards detecting more types of activities in home settings. Furthermore, the dataset can potentially be combined with other public datasets to test the effect of using heterogeneous types of devices and sensors.

This paper is organized as following: In section 2, we describe the data collection process. Section 3 details the feature extraction process, both, for accelerometer and audio data. In section 4, the structure of the dataset is explained. Section 5 presents baseline experiments with the dataset, and finally in section 6, we present the conclusions.

2 Data collection

The home-tasks data were collected by 3 individuals. They were 1 female and 2 males with ages ranging from 25 to 30. The subjects were asked to perform

[6] https://www.kaggle.com/yvaizman/the-extrasensory-dataset

7 home-task activities including: *mop floor*, *sweep floor*, *type on computer keyboard*, *brush teeth*, *wash hands*, *eat chips* and *watch TV*. The *eat chips* activity was conducted with a bag of chips. Each individual performed each activity for approximately 3 minutes. If the activity lasted less than 3 minutes, an additional trial was conducted until the 3 minutes were completed. The volunteers used a wrist-band (Microsoft Band 2) and a smartphone (Sony XPERIA) to collect the data.

The subjects wore the wrist-band in their dominant hand. The accelerometer data was collected using the wrist-band internal accelerometer. Figure 1 shows the actual device used. The inertial sensor captures motion from the x, y, and z axes, and the sampling rate was set to 31 Hz. Moreover, the environmental sound was captured using the microphone of a smartphone. The audio sampling rate was set at 8000 Hz. The smartphone was placed on a table in the same room where the activity was taking place.

An in-house developed app was programmed to collect the data. The app runs on the Android operating system. The user interface consists of a dropdown list from which the subject can select the home-task. The wrist-band transfers the captured sensor data and timestamps over Bluetooth to the smartphone. All the inertial data is stored in a plain text format.

Fig. 1. Wrist-band watch.

3 Feature extraction

In order to extract the accelerometer and audio features, the original raw signals were divided into segments of three seconds long. The segments are not overlapped. The three seconds were chosen because, according to Banos *et al.* [2], this is a typical value for activity recognition systems. They did comprehensive tests by trying different segments sizes and they concluded that small segments produce better results compared to longer ones. From each segment, a set of features were calculated which are known as *feature vectors* or *instances*. Each *instance* is characterized by the audio and accelerometer features. In the following section, we provide details about how the features were extracted.

3.1 Accelerometer features

From the inertial sensor readings, 16 measurements were computed including: The *mean*, *standard deviation*, *max* value for all the x, y and z axes, *pearson correlation* among pairs of axes (xy, xz, and yz), *mean magnitude*, *standard deviation of the magnitude*, the *magnitude area under the curve* (AUC, Eq. 1) , and *magnitude mean differences* between consecutive readings (Eq. 2). The *magnitude* of the signal characterizes the overall contribution of acceleration of x, y and z. (Eq. 3). Those features were selected based on previous related works [7, 10, 24].

$$AUC = \sum_{t=1}^{T} magnitude(t) \tag{1}$$

$$meandif = \frac{1}{T-1} \sum_{t=2}^{T} magnitude(t) - magnitude(t-1) \tag{2}$$

$$Magnitude(x, y, z, t) = \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2}, \tag{3}$$

where $a_x(t)^2$, $a_y(t)^2$ and $a_z(t)^2$ are the squared accelerations at time t.

Figure 2 shows violin plots for two of the accelerometer features: mean of the x-axis and mean of the y-axis. Here, we can see that overall, the mean acceleration in x was higher for the *brush teeth* and *eat chips* activities. On the other hand, the mean acceleration in the y-axis was higher for the *mop floor* and *sweep* activities.

3.2 Audio features

The features extracted from the sound source were the Mel Frequency Cepstral Coefficients (MFCCs). These features have shown to be suitable for activity classification tasks [1, 8, 13, 19]. The 3 second sound signals were further split into 1 second windows. Then, 12 MFCCs were extracted from each of the 1 second windows. In total, each instance has 36 MFCCs. In total, this process resulted in the generation of 1,386 instances. The tuneR R package [15] was used to extract the audio features. Table 1 shows the percentage of instances per class. More or less, all classes are balanced in number.

4 Dataset structure

The main folder contains directories for each user and a *features.csv* file. Within each users' directory, the accelerometer files can be found (*.txt* files). The file names are comprised of three parts with the following format: *timestamp-acc-label.txt*. *timestamp* is the timestamp in Unix format. *acc* stands for accelerometer and *label* is the activity's label. Each *.txt* file has four columns: timestamp and the acceleration for each of the x, y, and z axes. Figure 3 shows an example

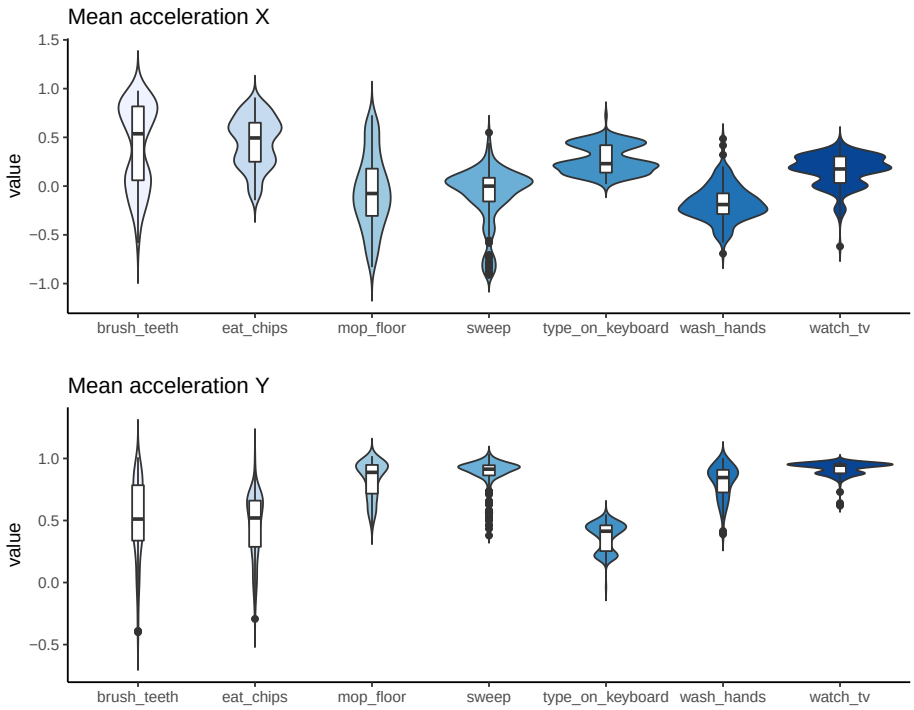

Fig. 2. Violin plots of mean acceleration of the x and y axes.

Table 1. Distribution of activities by class.

Class	Proportion
Brush teeth	12.98%
Eat chips	20.34%
Mop floor	13.05%
Sweep	12.84%
Type on keyboard	12.91%
Wash hands	12.98%
Watch TV	14.90%

of the first rows of one of the files. The *features.csv* file contains the extracted features as described in section 3. It contains 54 columns. *userid* is the user id. *label* represents the activity label and the remaining columns are the features. Columns with a prefix of *v1_* correspond to audio features whereas columns with a prefix of *v2_* correspond to accelerometer features. In total, there are 36 audio

features that correspond to the 12 MFCCs for each second, with a total of 3 seconds and 16 accelerometer features.

```
1468360517664,-0.12915039,0.9797363,-0.21191406
1468360517693,-0.13500977,0.98168945,-0.21118164
1468360517743,-0.1496582,0.9819336,-0.20336914
1468360517763,-0.16894531,0.9892578,-0.21606445
1468360517788,-0.18847656,0.99658203,-0.20581055
1468360517818,-0.1850586,0.97998047,-0.21362305
1468360517857,-0.19140625,0.97216797,-0.21533203
1468360517904,-0.18066406,0.9692383,-0.21411133
1468360517921,-0.1730957,0.9560547,-0.21435547
1468360517962,-0.17871094,0.9626465,-0.2163086
```

Fig. 3. First rows of one of the accelerometer files.

5 Baseline experiments

In this section, we present a series of baseline experiments that can serve as a starting point to develop more advanced methods and sensor fusion techniques. In total, 3 classification experiments were conducted with the present dataset. For each experiment, different classifiers were employed, including ZeroR (baseline), a J48 tree, Naive Bayes, Support Vector Machine (SVM), a K-nearest neighbors (KNN) classifier with $k = 3$, logistic regression, and a multilayer perceptron. We used the WEKA software [23] version 3.8 to train the classifiers. In each experiment, we used different sets of features. For experiment 1, we trained the models using only *audio features*, that is, the MFCCs. The second experiment consisted of training the models with only the 16 *accelerometer features* described earlier. Finally, in experiment 3, we combined the *audio and accelerometer* features by aggregating them. 10-fold cross-validation was used to train and assess the classifier's performance.

Table 2. Classification performance (weighted average) with audio features. The best performing classifier was KNN.

Classifier	False-Positive Rate	Precision	Recall	F1-Score	MCC
ZeroR	0.203	0.041	0.203	0.069	0.000
J48	0.065	0.625	0.623	0.624	0.559
Naive Bayes	0.049	0.720	0.714	0.713	0.667
SVM	0.054	0.699	0.686	0.686	0.637
KNN	0.037	0.812	0.788	0.793	0.761
Logistic regression	0.062	0.654	0.652	0.649	0.591
Multilayer perceptron	0.041	0.776	0.769	0.767	0.731

Table 3. Classification performance (weighted average) with accelerometer features. The best performing classifier was KNN.

Classifier	False-Positive Rate	Precision	Recall	F1-Score	MCC
ZeroR	0.203	0.041	0.203	0.069	0.000
J48	0.036	0.778	0.780	0.779	0.743
Naive Bayes	0.080	0.452	0.442	0.447	0.365
SVM	0.042	0.743	0.740	0.740	0.698
KNN	0.030	0.820	0.820	0.818	0.790
Logistic regression	0.031	0.800	0.802	0.800	0.769
Multilayer perceptron	0.031	0.815	0.812	0.812	0.782

Table 4. Classification performance (weighted average) when combining all features. The best performing classifier was Multilayer perceptron.

Classifier	False-Positive Rate	Precision	Recall	F1-Score	MCC
ZeroR	0.203	0.041	0.203	0.069	0.000
J48	0.035	0.785	0.785	0.785	0.750
Naive Bayes	0.028	0.826	0.823	0.823	0.796
SVM	0.020	0.876	0.874	0.875	0.855
KNN	0.014	0.917	0.911	0.912	0.899
Logistic regression	0.022	0.859	0.859	0.859	0.837
Multilayer perceptron	0.014	0.915	0.914	0.914	0.901

Tables 2, 3 and 4 show the final results. When using only audio features (Table 2), the best performing model was the KNN in terms of all performance metrics with a Mathews correlation coefficient (MCC) of 0.761. In the case when using only accelerometer features (Table 3), the best model was again KNN in terms of all performance metrics with an MCC of 0.790. From these tables, we observe that most classifiers performed better when using accelerometer features with the exception of Naive Bayes. Next, we trained the models using all features (accelerometer and audio). Table 4 shows the final results. In this case, the best model was the multilayer perceptron followed by KNN. Overall, all models benefited from the combination of features, of which some increased their performance by up to $\approx 15\%$, like the SVM which went from an MCC of 0.698 to 0.855.

All in all, combining data sources provided enhanced performance. Here, we just aggregated the features from both data sources. However, other techniques can be used such as late fusion which consists of training independent models using each data source and then combining the results. Thus, the experiments show that machine learning systems can perform this type of automatic activity detection, but also that there is a large potential for improvements - where the HTAD dataset can play an important role, not only as an enabling factor, but also for reprocucibility.

6 Conclusions

Reproducibility and comparability of results is an important factor of high-quality research. In this paper, we presented a dataset in the field of activity recognition supporting reproducibility in the field. The dataset was collected using a wrist accelerometer and captured audio from a smartphone. We provided baseline experiments and showed that combining the two sources of information produced better results. Nowadays, there exist several datasets, however, most of them focus on a single data source and on the traditional *walking, jogging, standing, etc.* activities. Here, we employed two different sources (accelerometer and audio) for home task activities. Our vision is that this dataset will allow researchers to test different sensor data fusion methods to improve activity recognition performance in home-task settings.

References

1. Al Masum Shaikh, M., Molla, M., Hirose, K.: Automatic Life-Logging: A novel approach to sense real-world activities by environmental sound cues and common sense. In: Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on. pp. 294–299 (Dec 2008). https://doi.org/10.1109/ICCITECHN.2008.4803018
2. Banos, O., Galvez, J.M., Damas, M., Pomares, H., Rojas, I.: Window Size Impact in Human Activity Recognition. Sensors **14**(4), 6474–6499 (2014). https://doi.org/10.3390/s140406474, `http://www.mdpi.com/1424-8220/14/4/6474`
3. Bruno, B., Mastrogiovanni, F., Sgorbissa, A., Vernazza, T., Zaccaria, R.: Analysis of human behavior recognition algorithms based on acceleration data. In: 2013 IEEE International Conference on Robotics and Automation. pp. 1602–1607. IEEE (2013)
4. Ceron, J.D., Lopez, D.M., Ramirez, G.A.: A mobile system for sedentary behaviors classification based on accelerometer and location data. Computers in Industry **92**, 25–31 (2017)
5. Ciucurel, C., Iconaru, E.I.: The importance of sedentarism in the development of depression in elderly people. Procedia - Social and Behavioral Sciences **33**(Supplement C), 722 – 726 (2012). https://doi.org/https://doi.org/10.1016/j.sbspro.2012.01.216, `http://www.sciencedirect.com/science/article/pii/S1877042812002248`, pSIWORLD 2011
6. Damen, D., Doughty, H., Farinella, G.M., Fidler, S., Furnari, A., Kazakos, E., Moltisanti, D., Munro, J., Perrett, T., Price, W., Wray, M.: Scaling egocentric vision: The epic-kitchens dataset. In: European Conference on Computer Vision (ECCV) (2018)
7. Dernbach, S., Das, B., Krishnan, N.C., Thomas, B.L., Cook, D.J.: Simple and Complex Activity Recognition through Smart Phones. In: Intelligent Environments (IE), 2012 8th International Conference on. pp. 214 –221 (Jun 2012). https://doi.org/10.1109/IE.2012.39
8. Galván-Tejada, C.E., Galván-Tejada, J.I., Celaya-Padilla, J.M., Delgado Contreras, J.R., Magallanes-Quintanar, R., Martinez-Fierro, M.L., Garza-Veloz, I.,

López-Hernández, Y., Gamboa-Rosales, H.: An analysis of audio features to develop a human activity recognition model using genetic algorithms, random forests, and neural networks. Mobile Information Systems **2016**, 1–10 (2016)
9. Garcia, E.A., Brena, R.F.: Real time activity recognition using a cell phone's accelerometer and wi-fi. In: Workshop Proceedings of the 8th International Conference on Intelligent Environments. Ambient Intelligence and Smart Environments, vol. 13, pp. 94–103. IOS Press (2012). https://doi.org/10.3233/978-1-61499-080-2-94
10. Garcia-Ceja, E., Brena, R.: Building personalized activity recognition models with scarce labeled data based on class similarities. In: García-Chamizo, J.M., Fortino, G., Ochoa, S.F. (eds.) Ubiquitous Computing and Ambient Intelligence. Sensing, Processing, and Using Environmental Information, Lecture Notes in Computer Science, vol. 9454, pp. 265–276. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-26401-1_25, http://dx.doi.org/10.1007/978-3-319-26401-1_25
11. Garcia-Ceja, E., Galván-Tejada, C.E., Brena, R.: Multi-view stacking for activity recognition with sound and accelerometer data. Information Fusion **40**, 45 – 56 (2018). https://doi.org/http://dx.doi.org/10.1016/j.inffus.2017.06.004, http://www.sciencedirect.com/science/article/pii/S1566253516301932
12. Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K.J., Tørresen, J.: Mental health monitoring with multimodal sensing and machine learning: A survey. Pervasive and Mobile Computing **51**, 1 – 26 (2018). https://doi.org/https://doi.org/10.1016/j.pmcj.2018.09.003, http://www.sciencedirect.com/science/article/pii/S1574119217305692
13. Hayashi, T., Nishida, M., Kitaoka, N., Takeda, K.: Daily activity recognition based on DNN using environmental sound and acceleration signals. In: Signal Processing Conference (EUSIPCO), 2015 23rd European. pp. 2306–2310 (Aug 2015). https://doi.org/10.1109/EUSIPCO.2015.7362796
14. Khaleghi, B., Khamis, A., Karray, F.O., Razavi, S.N.: Multisensor data fusion: A review of the state-of-the-art. Information Fusion **14**(1), 28 – 44 (2013). https://doi.org/https://doi.org/10.1016/j.inffus.2011.08.001, http://www.sciencedirect.com/science/article/pii/S1566253511000558
15. Ligges, U., Krey, S., Mersmann, O., Schnackenberg, S.: tuneR: Analysis of music (2014), http://r-forge.r-project.org/projects/tuner/
16. Mannini, A., Sabatini, A.M.: Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors **10**(2), 1154–1175 (2010). https://doi.org/10.3390/s100201154, http://www.mdpi.com/1424-8220/10/2/1154
17. Margarito, J., Helaoui, R., Bianchi, A.M., Sartor, F., Bonomi, A.G.: User-independent recognition of sports activities from a single wrist-worn accelerometer: A template-matching-based approach. IEEE Transactions on Biomedical Engineering **63**(4), 788–796 (2016)
18. Mitchell, E., Monaghan, D., O'Connor, N.E.: Classification of sporting activities using smartphone accelerometers. Sensors **13**(4), 5317–5337 (2013)
19. Nishida, M., Kitaoka, N., Takeda, K.: Development and preliminary analysis of sensor signal database of continuous daily living activity over the long term. In: Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. pp. 1–6. IEEE (2014)
20. Richter, J., Wiede, C., Dayangac, E., Shahenshah, A., Hirtz, G.: Activity recognition for elderly care by evaluating proximity to objects and human skeleton data.

In: International Conference on Pattern Recognition Applications and Methods. pp. 139–155. Springer (2016)
21. Vaizman, Y., Ellis, K., Lanckriet, G., Weibel, N.: Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2018)
22. Wang, X., Rosenblum, D., Wang, Y.: Context-aware mobile music recommendation for daily activities. In: Proceedings of the 20th ACM international conference on Multimedia. pp. 99–108. ACM (2012)
23. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 3 edn. (2011)
24. Zhang, M., Sawchuk, A.A.: Motion Primitive-Based Human Activity Recognition Using a Bag-of-Features Approach. In: ACM SIGHIT International Health Informatics Symposium (IHI). pp. 631–640. Miami, Florida, USA (Jan 2012)