Experiences with Publicly Open Human Activity Data Sets
Studying the Generalizability of the Recognition Models

Pekka Siirtola, Heli Koskimäki and Juha Röning
Biomimetics and Intelligent Systems Group, P.O. BOX 4500, FI-90014, University of Oulu, Oulu, Finland

Keywords: Human Activity Recognition, Accelerometer, Open Data Sets, Cross-validation.

Abstract: In this article, it is studied how well inertial sensor-based human activity recognition models work when training and testing data sets are collected in different environments. Comparison is done using publicly open human activity data sets. This article has four objectives. Firstly, survey about publicly available data sets is presented. Secondly, one previously not shared human activity data set used in our earlier work is opened for public use. Thirdly, the genaralizability of the recognition models trained using publicly open data sets are experimented by testing them with data from another publicly open data set to get knowledge to how models work when they are used in different environment, with different study subjects and hardware. Finally, the challenges encountered using publicly open data sets are discussed. The results show that data gathering protocol can have a statistically significant effect to the recognition rates. In addition, it was noted that often publicly open human activity data sets are not as easy to apply as they should be.

1 INTRODUCTION

Human activity recognition using inertial sensors, especially using accelerometers, has become one of the most studied area of pattern recognition. One reason for this is that activity recognition can be applied to many different types of application, including health and fitness monitoring; personalized advertising; smarthomes that anticipates the user's needs; and self-managing system that adapts to user's activities (Lockhart et al., 2012).

Often, the models for activity recognition are user-independent. This is challenging as people are different, and therefore, a user-independent recognition model that provides accurate recognition results for one person does not necessary provide as high results for other person (Albert et al., 2012). In the article the reason for this was that recognition models were trained with healthy study subjects and tested with subjects with difficulties to move. Differences between persons are not the only thing that cause variation to the recognition results. In fact, it has been shown that a model that works in one environment may provide totally different results when used in other environment. In Ermes *et. al.* (Ermes et al., 2008) it was studied how the recognition rate decreases when data is collected with and without guidance. In the study, a activity recognition model was trained to recognize nine different sports/everyday activities and when it was tested using data collected with guidance, the recognition rate was 91%. When the same model was tested using data collected without guidance, including four out of nine trained activities, the average recognition rate was only 64%. There are many possible reasons for this phenomena as several things in the environment can have an effect to the recognition rates, these include changes in weather, terrain and location that can cause problems to recognition models. In addition, in the real-life many other unseen contingencies can happen and the data set used to train the recognition models cannot include all of these. Most importantly, the study shows that data gathering protocol has an effect to the collected data, and therefore, to the trained models and recognition rates. Thus, to get better knowledge how well the trained models actually work in different situations, it is important to validate them using data gathered by different study subjects but also collected in different environment and sensors.

In some of the studies, the validation of the accuracy of the recognition models in different environments has been done by implementing the trained models to mobile phone application and then used real-time real-life situations (Shoaib et al., 2014; Siirtola and Röning, 2013). In these cases, the validation data may include contingencies that are not included in the training data but the environment where the data are gathered does not change a lot as for instance it is

Siirtola, P., Koskimäki, H. and Röning, J.
Experiences with Publicly Open Human Activity Data Sets - Studying the Generalizability of the Recognition Models.
DOI: 10.5220/0006553302910299
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 291-299
ISBN: 978-989-758-276-9

normally collected in the same country as the training data. However, open data sets can be seen as a solution to this problem: by testing the trained models using them, it is possible to get knowledge how models work in different environment, with different study subjects and hardware. This approach obviously has advantages compared to implementing the trained models to a physical device as it saves time because a separate validation data session is not needed.

In this study inertial sensor-based human activity recognition is studied using publicly open data sets. Study has following aims:

1. to survey what kind of open data sets are available
2. one previously not shared human activity data set used in our earlier work is opened for public use
3. experiment how accurately the models trained using a one open data set work when they are validated with other open data set
4. the challenges encountered using publicly open data sets are discussed.

The paper is organized as follows: Section 2 gives a brief survey about publicly open data sets and Section 3 explain which are the ones selected for this study. Section 4 introduces the used methods and describes the experimental protocol while Section 5 shows the results and discusses about them. Finally, the conclusions are in Section 6.

2 OPEN HUMAN ACTIVITY DATA SETS

There are various open human activity data sets collected using inertial sensors available. In fact, we found 15 different open data sets. In addition, we made one of our own data set open, this data set was used in Siirtola & Röning (Siirtola and Röning, 2012) and can now been found from our research units webpage (Biomimetics and Intelligent Systems Group, 2017). Most of the other data sets can be found at the UCI Machine Learning Repository (Lichman, 2013). Open data sets and some of their characteristics are listed in Table 1. Some of these included also data from other sensors than inertial sensors, but they are not listed in the table.

While several open data sets are available, comparing and cross-validating them is not straightforward as data sets differ from each other in many ways as it can be seen from Table 1. For instance, sensor position is not the same in all data sets. The most common sensor positions are wrist, chest and hip but also sensor positions are used such as trouser's pocket, foot and back. The problem is that from each of these body position the gathered data is different meaning that if the recognition model is trained using data from one body position, it cannot recognize activities if the sensor is positioned in some other position (Siirtola, 2015). Moreover, one of the challenges of the open data sets listed in Table 1 is that the used sampling rate is not always the same. The sampling rate varies from 20Hz to 200Hz and the most commonly used frequency is 50Hz. Again, this makes the cross-validation of the data sets more challenging as the data sets used in model training, testing and validation needs to have the same sampling rate to obtain reliable model and results. What also limits the number of data sets that can be in cross-validation is that almost all data sets are collected from different activities. Most of the data sets are collected from daily activities like walking, running, sitting and standing while some concentrate for instance on sports activities. This limits the using of data sets as in order to use cross-validate data sets, they need to contain same activities. Easily the most common activity is walking, which is included in almost all data sets. However, the definition of walking differs from data set to data set: in some of the data sets walking, walking downstairs and walking downstairs are considered as three different activities as in some data sets they all have the same label. One more difference in data sets is that the number of study subjects varies from one person to 30 persons. Therefore, it is clear that some data sets have more variation than others.

However, based on the information shown in the Table 1 it is not possible to conclude which data set is the best as for instance the numbers do not tell how much data per activity these data sets include and how much variability they include. This means that model trained and tested using one data set and proving highly accurate results, does not necessarily provide as high results when it is tested using another data set which is collected using different sensors, in different environment, and by giving different instructions given to study subjects. In fact, the purpose of this study is to experiment how the model accuracy changes when it is tested using the data from the same data set compared to when it is tested using different open data set.

3 EXPERIMENTAL DATA SETS

All the open data sets listed in Table 1 were not used in cross-validation process as in order to cross-validate data sets they need to have the same sam-

Table 1: List of publicly open human activity data sets collected using inertial sensors.

Author	Body position	Frequency	Activities	Inertial sensors	Study subjects
(Shoaib et al., 2014)	trouser's pocket, arm, wrist, belt	50Hz	walking, running, standing, sitting, cycling and walking upstairs and downstairs	accelerometer, gyroscope, magnetometer	7
(Anguita et al., 2013)	waist	50Hz	walking, walking upstairs, walking downstairs, sitting, standing, laying	accelerometer, gyroscope	30
(Siirtola and Röning, 2012)	trouser's pocket	40Hz	walking, running, cycling, idling, driving car	accelerometer	7
(Banos et al., 2015)	chest, wrist, ankle	50Hz	12 activities, including walking, running, cycling and sitting	accelerometer, gyroscope, magnetometer	10
(Casale et al., 2012)	chest	52Hz	7 activities including walking, standing and walking at stairs	accelerometer	15
(Stisen et al., 2015)	waist, arm	50-200Hz	biking, sitting, standing, walking, stair up and stair down	accelerometer	9
(Koskimäki and Siirtola, 2014)	arm	100Hz	36 gym activities	accelerometer	1
(Barshan and Yüksek, 2014)	chest	25Hz	19 activities including sitting, standing, walking, running, cycling	accelerometer, gyroscope, magnetometer	8
(Bruno et al., 2013)	wrist	32Hz	14 activities including brush teeth, climb stairs, comb hair, descend stairs and walking	accelerometer	16
(Reiss and Stricker, 2012)	wrist, chest, ankle	100Hz	18 activities including lying, sitting, standing, walking, running, cycling	accelerometer, gyroscope, magnetometer	9
(Kwapisz et al., 2011)	trouser's pocket	20Hz	walking, jogging, upstairs, downstairs, sitting, standing	accelerometer	29
(Zhang and Sawchuk, 2012)	hip	100Hz	12 activities including walking, running, upstairs, downstairs, sitting and standing	accelerometer, gyroscope, magnetometer	14
(Chavarriaga et al., 2013)	wrist, chest, limb, shoulder, foot	30Hz	groom room, prepare and drink coffee, prepare and drink sandwich, cleanup	accelerometer, gyroscope, magnetometer	12
(Baños et al., 2012)	left and right calf and thigh, back, 4 on arms	50Hz	33 activities including walking, running, jumping and cycling	accelerometer, gyroscope, magnetometer	17
(Micucci et al., 2017)	trouser's pocket	50Hz	9 activities including walking, running, upstairs and downstairs	accelerometer	30
(Ugulino et al., 2012)	waist, tight, arm, ankle	50Hz	walking, standing, sitting, sitting down, standing up	accelerometer	4

pling rate and activities and they need to be collected from the same body position. However, the number of study subjects does not need to be the same. It was decided to cross-validate data sets from two body positions: trouser's pocket and wrist. Trouser's pocket was chosen as one of the studied position as according to some studies it is the most common position for a phone (Ichikawa et al., 2005). In addition, wrist was chosen as other position as smartwatches and wrist-worn activity monitors are probably the most obvious devices to implement activity recognition algorithms.

From the both chosen body position two data sets were selected for this study. Three of the data sets introduced in Table 1 contain data gathered from trouser's pocket: Kwapisz *et. al.* (Kwapisz et al., 2011), Siirtola & Röning (Siirtola and Röning, 2012) and Shoaib *et. al.* (Shoaib et al., 2014). The activities we are aiming to recognize are *walking, running, idling (=sitting and standing) and cycling*. As Kwapisz *et. al.* does not include cycling -activity, cross-validation is performed using two data sets: Siirtola & Röning (Siirtola and Röning, 2012) and Shoaib *et. al.* (Shoaib et al., 2014). In turn, one is used for training and other for validation. The problem with the selected data sets is that they are collected using different sampling rates, 40Hz and 50Hz, respectively. As the greatest common factor of 40 and 50 is 10, the data sets are down sampled to 10Hz. With Siirtola & Röning (Siirtola and Röning, 2012) this is done by taking only every fourth observation into consideration, and in the case of Shoaib *et. al.* (Shoaib et al., 2014) by taking only every fifth observation into consideration.

Several data sets contained data gathered from wrist, and the ones chosen for this study were Shoaib *et. al.* (Shoaib et al., 2014), Banos *et. al.*(Banos et al., 2015). These were selected as they have the same sampling rate and also several common activities: *walking, running, cycling, sitting, standing, and walking at stairs*. Again, in the cross-validation process one data sets in turn is used for training and other for validation.

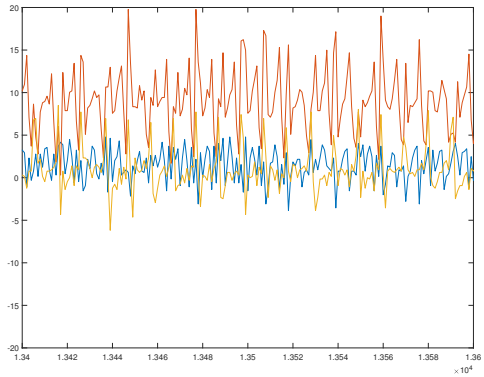
Examples from the selected data sets are shown in Figures 1, where 10 seconds of acceleration data from walking signal is presented. From these figures it can be seen that the scale of acceleration signal is approximately the same in all data sets. However, the comparison of signals collected from trouser's pocket (Figures 1(a) & 1(b)) show that signals are on different level. This is most likely due to a fact that sensor can lay on the pocket in a numerous different orientations. Therefore, in the pre-processing stage the effect of orientation was eliminated by square summing acceleration channels to obtain the magnitude acceleration signal, which is orientation independent. Wrist-position is more stable than pocket, as there is basically only one possible way to wear the sensor, and therefore the orientation of the sensor should be approximately same for each study subject wearing the sensor in the same wrist. However, visualization (Figures 1(c) & 1(d)) of the data sets collected from wrist and selected for this study show that sensors used in Shoaib *et. al.* (Shoaib et al., 2014) and Banos *et. al.* (Banos et al., 2015) have been different. Apparently the order of acceleration sensors has been different in these sensors, and therefore, x-axis acceleration of Banos *et. al.* behaves like z-axis acceleration of Shoaib *et. al.*. This was simply fixed by changing the same order for both data sets but this difference shows that it is not always that straightforward to use publicly open data sets.

4 ACTIVITY RECOGNITION PROCESS

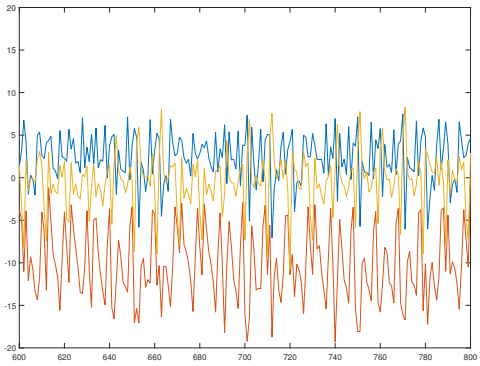
Activity recognition process is divided into three main phases (data collection, training and activity recognition), and each of these can be divided into subphases (Incel et al., 2013), see Figure 2. In this section, it is explained what methods this study uses in different stages.

Activity recognition was done using a sliding window technique. The signals from the sensors were divided into equal-sized smaller sequences, also called windows. Window size of 2.0 seconds was used with every data set.From these windows, features were extracted and finally the classification of the sequences was done using models trained based on these features. The features used in this study included for instance standard deviation, minimum, maximum, median, and different percentiles (10, 25, 75, and 90). Moreover, the sum of values above or below percentile (10, 25, 75, and 90), square sum of values above of below percentile (10, 25, 75, and 90), and number of crossings above or below percentile (10, 25, 75, and 90) were extracted and used as features. Altogether 61 features were extracted. These same features are used for instance in Siirtola *et. al.* (Siirtola et al., 2016).

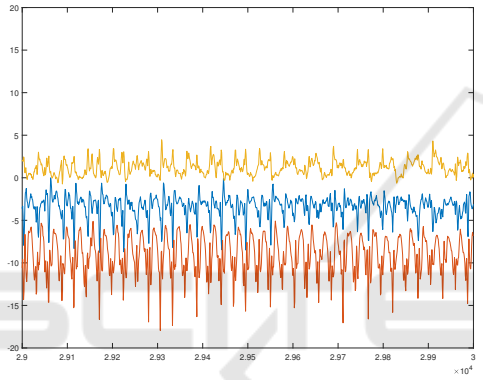
In order to achieve the highest possible recognition rates, the most descriptive features for each model were selected using a sequential forward selection (SFS) method (Devijver and Kittler, 1982). Moreover, to reduce the number of misclassified windows, the final classification was done based on the majority voting of the classification results of three adjacent windows. Therefore, when an activity changes, a new activity can be detected when two adjacent win-



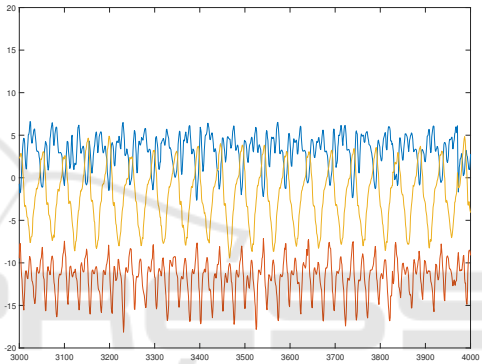
(a) Example from data set Siirtola & Röning (Siirtola and Röning, 2012) where the position of the sensor is trouser's pocket.



(b) Example from data set Shoaib *et. al.* (Shoaib et al., 2014) where the position of the sensor is trouser's pocket.



(c) Example from data set Banos *et. al.* (Banos et al., 2015) where the position of the sensor is wrist.



(d) Example from data set Shoaib *et. al.* (Shoaib et al., 2014) where the position of the sensor is wrist.

Figure 1: 3D acceleration data (blue = x-axis, red = y-axis, and yellow = z-axis) from the each selected data set describing 20 seconds of walking signal.

dows are classified as a new activity.

It was decided to do the experiments using LDA (linear discriminant analysis) and QDA (quadratic discriminant analysis (Hand et al., 2001)) classifier as in our previous studies (Siirtola and Röning, 2012; Siirtola and Röning, 2013) we have noticed that they are not only accurate but also computationally light, and therefore, sufficient to be implemented to smartphones and used 24/7. In addition, they are fast to train. LDA is used to find a linear combination of features that separate the classes best. The resulting combination may be employed as a linear classifier. QDA is a similar method, but it uses quadratic surfaces to separate classes (Hand et al., 2001).

In the last stage of activity recognition process, using the trained recognition model, an unknown streaming signal can be classified. Before its class label can be defined, new signal must be processed in the same way as training data was processed when recognition models were trained, see Figure 2. Therefore, at first, streaming data is pre-processed and windowed. Then, the featured used to train the model are extracted from the window and these are given as input to the trained recognition model to obtain the predicted activity class. Note that when new data is classified, in each stage, the same parameters must be used that were used to train the models (Bishop, 2006).

5 EXPERIMENTS AND DISCUSSION

For the experiments model training is performed using protocols presented in Figure 3 using separate training, validation and testing data sets. Figure 3(a) shows the protocol used and how the data is divided when the same publicly open data set is used for training and testing, one person's data in turn is used for

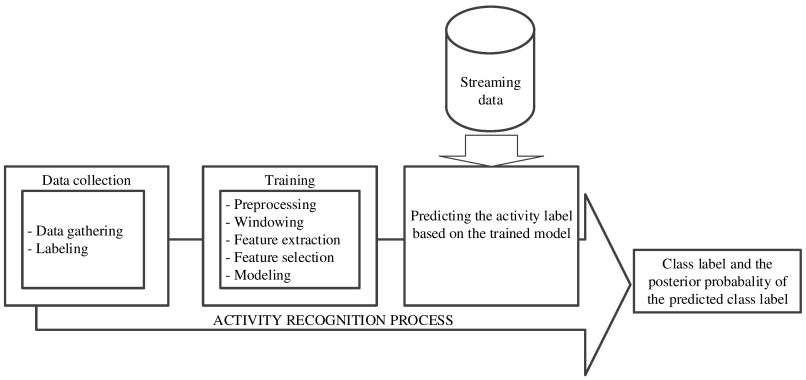


Figure 2: Activity recognition process is divided into three main phases, and each of these can be divided into subphases (Incel et al., 2013).

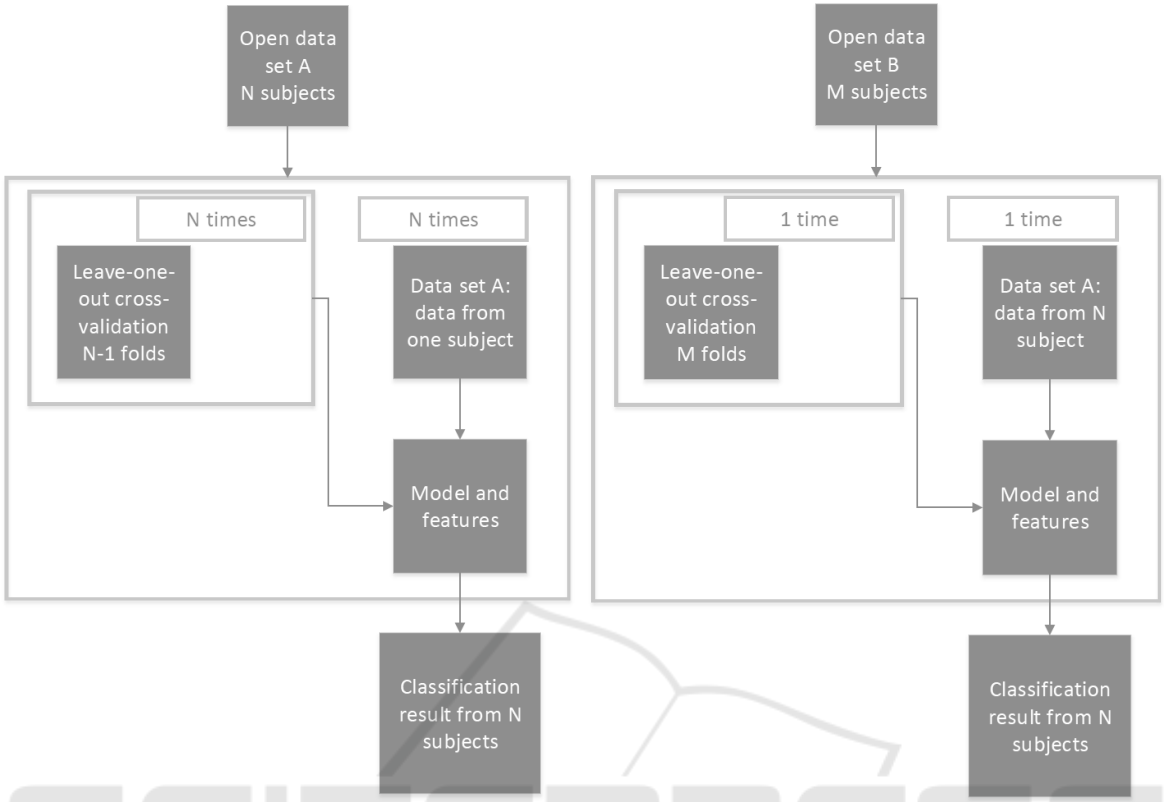
testing and other for training. Therefore, the same data is never used for training and testing. Figure 3(b) shows the protocol used when the one publicly open data set is used for model training and other for testing.

The results of the experiments using trouser's pocket and wrist as a sensor position are shown separately for each study subject in Table 2 and 3, respectively. The accuracies were obtained by calculating true positive rate of each class and by calculating average of these. The results show that in general QDA produces better results than LDA. Therefore, it was concentrated on analyzing QDA results.

When the results of data from trouser's pocket (Table 2) are studied, it can be seen that the recognition accuracy is a bit higher when the recognition model is trained and tested using a data from the same data set compared to training model with one data set and testing it with other. While this result came as no surprise, the difference was smaller than expected. The reason for this can be the small number of activities (4) studied in this sensor position. Therefore, the samples from different activities can locate in the very different parts of the feature space making task quite easy and differences between data gathering protocols, sensors, study subjects and other factors do not have that big of an effect to the recognition rates. Though the difference does not seem big, still according to the paired t-test the difference is statistically significant in three cases out of four, only when Siirtola data is used for testing and classification is based on QDA classifier the difference is not statistically significant.

The data from a wrist-worn sensor included more activities (6) than data from trouser's pocket, making it more difficult to classify. Therefore, also the results had more variance. This scenario was experimented using data sets from Banos *et. al.* (Banos et al., 2015) and Shoaib *et. al.* (Shoaib et al., 2014). When Shoaib data set was classified using a model trained using Banos data, the accuracy using QDA was 5 percentage units lower (86.6% vs. 81.2%) compared to training and testing model using data from one data set. However, when the experiment was performed other way around, meaning that Banos data set was classified using models trained with Shoaib data, the difference was over 13 percentage units (95.4% vs. 82.3%). Therefore, while using Banos data is cross-validated using leave-one-out method the recognition accuracy is almost perfect (95.4%), and the rate is noticeable lower when a recognition model trained using Banos data is tested with Shoaib data (81.2%). This is not the case with Shoaib data (86.6% vs. 82.3%), and therefore, the recognition rates and models obtained using Shoaib data for training are more predictable and their generalizability is better than the ones trained using Banos data. Also this time, according to the paired t-test the difference is statistically significant in three cases out of four. Only when Banos data is used for testing and the classification process is performed using LDA classifier, the different is not statistically significant.

The results of Table 3 show that the data collection protocol has been different between the data sets and the end result is that Banos data set has less variation than Shoaib. On the other hand, this was not necessarily the only reason for the obtained results as these data sets had other differences as well, as shown in Figure 1. It was noted that the order of the three acceleration signals was different in these data sets and it was fixed in the pre-processing stage, as explained in Section 3. This shows a challenge regarding using open data sets, they are not always usable out-of-the-box. In fact, it would be really beneficial to the research area to further study publicly open data sets to find differences like this and report them. In addition, a brief study of other data sets showed also other differences between data sets which makes validation



(a) The model training protocol when the same data set is used for training and validation.

(b) The model training protocol when the different data set is used for training and validation.

Figure 3: Protocols used in the experiments.

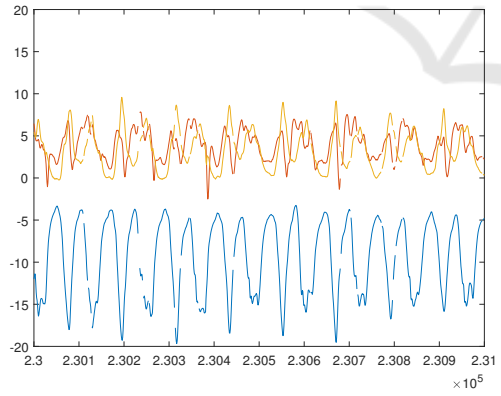


Figure 4: Walking signal from Reiss & Stricker (Reiss and Stricker, 2012) is different to signals from other studied data sets.

of them burdensome. For instance, Anguita data set (Anguita et al., 2013) had different scale for accelerometer values than the data sets used in this study. With Anguita *et. al.* (Anguita et al., 2013) the unit for accelerometer values is gravity g when for the data sets used in this study it is meters in second m/s^2. We also had problems with Reiss data (Reiss and Stricker, 2012) when we tried to use it in wrist-sensor scenario. When it was cross-validated using other data sets, the results were really poor. The models did not work at all. In Figure 4, ten seconds of walking signal from this data set is visualized. The visualization shows that this data is totally different to those shown in Figure 1. Most likely Reiss data is filtered, and therefore, it can not be used with other data sets. In addition, there are missing values which may have an effect to recognition rates.

Therefore, as a summary: publicly open human activity data sets have differences which makes their usage difficult. Some of these differences are easy to fix, some hard or even inflexible. The worst scenario is of course that there are differences that user does not find out at all. Therefore, to make data set usage more easy, guidelines for data gathering should be made for the research area to obtain reusable data sets as currently the data gathering protocols can vary so much between data sets that they cannot be used together.

While in this study, the focus of the study are publicly open human activity data sets, the observations made here are also valid in other application fields as

Table 2: The results of the experiments using trouser's pocket as a sensor position.

Test data: Shoaib data set, Classifier QDA											
Training	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**Mean**
Shoaib	96.3%	97.4%	92.7%	95.5%	93.1%	97.2%	95.6%	98.0%	96.5%	92.2%	95.4%
Siirtola	91.1%	94.6%	89.6%	93.2%	91.4%	91.9%	88.3%	91.3%	93.2%	89.1%	91.4%
Test data: Shoaib data set, Classifier LDA											
Shoaib	90.4%	97.7%	92.0%	95.2%	92.1%	96.2%	93.6%	96.3%	96.0%	92.5%	94.2%
Siirtola	93.5%	94.7%	87.1%	92.0%	93.5%	91.2%	88.3%	90.8%	92.6%	86.2%	91.0%
Test data: Siirtola data set, Classifier QDA											
Training	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	-	-	**Mean**
Siirtola	92.8%	91.0%	93.2%	90.3%	85.4%	94.7%	98.1%	92.6%	-	-	92.3%
Shoaib	89.7%	92.0%	94.8%	86.5%	85.7%	90.9%	93.7%	90.9%	-	-	90.5%
Test data: Siirtola data set, Classifier LDA											
Siirtola	93.7%	90.9%	90.3%	95.0%	90.5%	99.1%	98.4%	96.1%	-	-	94.2%
Shoaib	86.8%	87.5%	87.0%	70.4%	82.7%	94.3%	90.0%	92.2%	-	-	86.4%

Table 3: The results of the experiments using wrist as a sensor position.

Test data: Shoaib data set, Classifier QDA											
Training	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**Mean**
Shoaib	79.6%	82.1%	91.3%	85.4%	82.2%	84.1%	97.2%	91.7%	88.2%	83.9%	86.6%
Banos	87.3%	81.8%	83.4%	83.3%	75.6%	79.0%	86.2%	82.4%	81.5%	71.7%	81.2%
Test data: Shoaib data set, Classifier LDA											
Shoaib	85.1%	90.2%	91.8%	94.7%	92.3%	96.2%	97.3%	92.4%	91.0%	80.3%	91.1%
Banos	58.2%	73.5%	69.8%	61.7%	62.1%	70.8%	66.8%	52.4%	63.2%	56.5%	63.5%
Test data: Banos data set, Classifier QDA											
Training	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**Mean**
Banos	96.3%	97.4%	92.7%	95.5%	93.1%	97.2%	95.6%	98.0%	96.5%	92.2%	95.4%
Shoaib	73.8%	96.1%	75.0%	77.5%	79.6%	81.0%	83.8%	87.3%	85.9%	83.3%	82.3%
Test data: Banos data set, Classifier LDA											
Banos	72.7%	68.6%	79.3%	84.2%	76.8%	80.5%	87.7%	93.4%	72.5%	85.6%	80.1%
Shoaib	66.2%	91.3%	65.8%	60.1%	75.9%	92.0%	95.2%	70.5%	82.5%	75.4%	77.4%

well: the recognition models should be tested with data sets collected from different environments and publicly open data sets should be easily reusable.

6 CONCLUSION

The original aim of this article was to survey and cross-validate publicly open inertial sensor-based human activity data sets, and release one new for public use, to see how accurately a model which is trained using a data gathered in one environment and location works when it is tested using a data gathered in a totally different environment. This was experimented in two different scenarios and the results of this article show that data gathering environment, and protocol has an effect to the results. The accuracies obtained in a new environment are lower than from original environment, and according to the paired t-test this difference is statistically significant in six cases out of eight. In addition, it seems that the difference is the bigger the more difficult the classification task is. However, eventually this was not the main outcome of this article. Instead, the main outcome of this article is that publicly open data sets are not easy and straightforward enough to be used for validation purposes. Several open inertial-based data sets are publicly available. However, cross-validation of them is problematic as most of the data sets have different activities, sampling rate, and body position. While these differences are quite straightforward to find out, it was noticed that data sets have other differences that are not as easy to detect. For instance it was noticed that there were differences in sensor orientation and scale of y-axis. To avoid this, our future work includes a guide and code to get non-burdensome access to publicly open data sets. In addition, guidelines for data gathering should be made for the research area to obtain more easily reusable data sets.

ACKNOWLEDGMENT

The authors would like to thank Infotech Oulu for funding this work.

We acknowledge collaboration with Tianjin Normal University related multisensor fusion technology, wearable computing, machine learing and health care.

REFERENCES

Albert, M., Toledo, S., Shapiro, M., and Kording, K. (2012). Using mobile phones for activity recognition in parkinsons patients. *Frontiers in neurology*, 3:1–7.

Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. In *ESANN*.

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

Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado-Terriza, J. A., Lee, S., Pomares, H., and Rojas, I. (2015). Design, implementation and validation of a novel open framework for agile development of mobile health applications. *Biomedical engineering online*, 14(2):S6.

Barshan, B. and Yüksek, M. C. (2014). Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *The Computer Journal*, 57(11):1649–1667.

Biomimetics and Intelligent Systems Group (2017). http://www.oulu.fi/bisg/node/40364. Accessed: 2017-10-24.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

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

Casale, P., Pujol, O., and Radeva, P. (2012). Personalization and user verification in wearable systems using biometric walking patterns. *Personal and Ubiquitous Computing*, 16(5):563–580.

Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., Millán, J. d. R., and Roggen, D. (2013). The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042.

Devijver, P. A. and Kittler, J. (1982). *Pattern recognition: A statistical approach*. Prentice Hall.

Ermes, M., Pärkkä, J., Mäntyjärvi, J., and Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE transactions on information technology in biomedicine*, 12(1):20–26.

Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of data mining*. MIT Press, Cambridge, MA, USA.

Ichikawa, F., Chipchase, J., and Grignani, R. (2005). Where's the phone? a study of mobile phone location in public spaces. In *2nd Asia Pacific Conference on Mobile Technology, Applications and Systems*, pages 1–8. IET.

Incel, O., Kose, M., and Ersoy, C. (2013). A review and taxonomy of activity recognition on mobile phones. *BioNanoScience*, 3(2):145–171.

Koskimäki, H. and Siirtola, P. (2014). Recognizing gym exercises using acceleration data from wearable sensors. In *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*, pages 321–328. IEEE.

Kwapisz, J. R., Weiss, G. M., and Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82.

Lichman, M. (2013). UCI machine learning repository. http://archive.ics.uci.edu/ml. University of California, Irvine, School of Information and Computer Sciences.

Lockhart, J. W., Pulickal, T., and Weiss, G. M. (2012). Applications of mobile activity recognition. In *2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 1054–1058, New York, NY, USA.

Micucci, D., Mobilio, M., and Napoletano, P. (2017). Unimib shar: a new dataset for human activity recognition using acceleration data from smartphones. *arXiv preprint arXiv:1611.07688v2*.

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

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

Siirtola, P. (2015). Recognizing human activities based on wearable inertial measurements: methods and applications. *Doctoral dissertation, Department of Computer Science and Engineering, University of Oulu*, (Acta Univ Oul C 524).

Siirtola, P., Koskimäki, H., and Röning, J. (2016). Personal models for ehealth-improving user-dependent human activity recognition models using noise injection. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, pages 1–7. IEEE.

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

Siirtola, P. and Röning, J. (2013). Ready-to-use activity recognition for smartphones. In *Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on*, pages 59–64. IEEE.

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

Ugulino, W., Cardador, D., Vega, K., Velloso, E., Milidiú, R., and Fuks, H. (2012). Wearable computing: Accelerometers data classification of body postures and movements. In *Advances in Artificial Intelligence-SBIA 2012*, pages 52–61. Springer.

Zhang, M. and Sawchuk, A. A. (2012). 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.