Patrick Lieser, Alaa Alhamoud, Hosam Nima, Björn Richerzhagen, Sanja Huhle, Doreen Böhnstedt, Ralf Steinmetz. *Situation Detection based on Activity Recognition in Disaster Scenarios.* In Proc. of the International Conference on Information Systems for Crisis Response and Management (ISCRAM), 2018.

Situation Detection based on Activity Recognition in Disaster Scenarios

Patrick Lieser
KOM, Technical University Darmstadt,
Darmstadt, Germany
patrick.lieser@kom.tu-darmstadt.de

Alaa Alhamoud
KOM, Technical University Darmstadt,
Darmstadt, Germany
alaa.alhamoud@kom.tu-darmstadt.de

Hosam Nima
KOM, Technical University Darmstadt,
Darmstadt, Germany
hosam.nima@kom.tu-darmstadt.de

Björn Richerzhagen
KOM, Technical University Darmstadt,
Darmstadt, Germany
bjoern.richerzhagen@kom.tu-darmstadt.de

Sanja Huhle
KOM, Technical University Darmstadt,
Darmstadt, Germany
sanja.huhle@kom.tu-darmstadt.de

Doreen Böhnstedt
KOM, Technical University Darmstadt,
Darmstadt, Germany
doreen.boehnstedt@kom.tu-darmstadt.de

Ralf Steinmetz
KOM, Technical University Darmstadt,
Darmstadt, Germany
ralf.steinmetz@kom.tu-darmstadt.de

ABSTRACT

In disaster situations like earthquakes and hurricanes, people have difficulties accessing shelter and requesting help. Many smartphone applications provide behavioral advice or means to communicate during such situations. However, to what extent a person is affected by a disaster is often unclear, as these applications rely on the user's subjective assessment. Therefore, detecting a user's situation is key to provide more meaningful information in such applications and to allows first responders to better assess incoming messages.

We propose a predictive model that recognizes four normal and ten disaster-related activities achieving an average f1-score of up to 90.1%, solely based on sensor readings of the subject's mobile device. We conduct an extensive measurement-based evaluation to assess the impact of individual model parameters on the prediction accuracy. Our model is orientation-independent, position-independent, and subject-independent, making it an ideal foundation for future context-aware emergency applications.

Keywords

Disaster Relief, Activity Recognition, Prediction Model, Wearables

INTRODUCTION

Activity recognition based on smart devices has attracted remarkable attention from researchers in recent years (Attal et al. 2015), largely for monitoring fitness and health activities. Up until now such context-aware systems are used in several specific scenarios such as elder care (Alwan et al. 2006), health-care and health recovery (Guo et al. 2016; Garcia et al. 2013), stress detection (MacLean et al. 2013; Webb et al. 2013) and to enhance self reflection (Kefalidou et al. 2014). To the best of our knowledge, activity recognition systems were not specifically used before to detect disaster-related activities. In the following, we sketch a context-aware system with the main aim of helping people in disaster situations by recognizing their activity based on their physical movement.

Such a system can be divided into two main subsystems. The first subsystem is responsible for recognizing the current activity of a subject and decide whether it is related to a disaster or not. The second subsystem is responsible for using and integrating the information of a subject's activity in applications or communication systems for disaster relief. In this work, we address the first subsystem: detecting the situation of a subject based on activity recognition.

To recognize the activities of a subject, activity recognition systems depend on analyzing a series of observations that can be obtained from different sources. An example of such sources are wearable sensors that the subject wears in different positions of her/his body (Mantyjarvi et al. 2001; Gyorbiro et al. 2009).

Another source of observations are mobile devices like smartphones and smartwatches (Chetty et al. 2015; Otebolaku and Andrade 2016). In addition, several research works depend on smart environments for activity recognition (Cook 2010) where several sensors, for example motion, temperature, humidity, and power sensors, are integrated into the environment. Todays smartphones and smartwatches (Cecchinato et al. 2015) are equipped with a rich set of accurate sensors and come with sufficient battery capacity, storage and processing capabilities. In addition, they are widely deployed, making them an ideal source for observations in a disaster scenario.

The remainder of the paper is structured as follows: First, we present use-cases for the activity recognition in disaster scenarios. Afterwards, the approach for the activity prediction models, the resulting research objectives, and the necessary data collection process is described, followed by the development concept of the models. Then, we provide an in-depth description and performance evaluation of the developed impersonal and position-independent smartwatch and smartphone prediction models, as well as the combined smartphone-smartwatch prediction model. Finally, we discuss relevant related work and conclude the paper.

USECASES

Earthquakes, volcanic eruptions, and tsunamis are examples of spontaneous occurring natural disasters that cause huge destruction and jeopardize the life of many people. In such a case the affected population suffers from the difficulties of accessing safe places and contacting aid organizations to get help.

Knowing a person's current activity in exceptional situations is important for different use-cases. For emergency response and disaster relief there exist a variety of useful smartphone-based applications and communications services that help people to cope with the aftermath of a disaster (Lieser et al. 2017; Reuter et al. 2017; Hossmann et al. 2011). If the current activity of a user is known to the system (as additional meta-data), it could enable activity-aware support mechanisms. For example, if the activity *running downstairs* is detected, the user's interface (font and button size) of an emergency application could be enlarged to support the service's ease of use. Additionally, an emergency response center can use the user's activity information to filter incoming emergency messages according to their importance, to prioritize assistance efforts of those who seem to be injured or immobile.

Going one step further, an application can utilize information about the current activity of a person to proactively call for help or to provide situation relevant information such as evacuation maps or behavioral instructions. For example, badly wounded people need help immediately but are often not able to contact medical organizations or surrounding people by themselves. Automatically sending out SOS-messages containing the location of the user could save lives. Additionally, when people are hurrying to shelter places, protecting themselves from falling debris, or crawl, their hands are not free to operate a smartphone. By detecting this activity the respective application could switch to voice in- and output instead of requiring people to operate the phone with their hands.

The aforementioned use-cases are only a few examples of how activity recognition of a subject can be used to enhance and improve existing disaster related services.

PROPOSED APPROACH AND RESEARCH OBJECTIVES

In this work, we develop a system to recognize the physical activities of humans and decide whether this activity is related to the behavior typically performed in disaster situations. We build this system on top of sensor readings collected from the subject's smart devices. Table 2 shows the list of all the activities we consider in this work. These disaster related activities were specified after studying international guidelines for emergency preparedness from different governmental organisations[1,2], NGO's[3] and national disaster management authorities[4] and also during discussions with local fire departments and different emergency operation centers. We chose these specific activities since we consider them as the most relevant activities in disaster situations.

Some of the studied activities are normal activities that are daily performed by most of the people such as normal walking, sitting, going upstairs or going downstairs. The reason behind involving such general activities in our research work is to increase the capability of our model to distinguish between normal activities and the activities that are usually performed in disaster situations.

As will be clarified in the next section, we conduct our work by utilizing four smart devices, namely two smartphones and two smartwatches which have to be worn by the subjects. Our ultimate goal is to build a system which is position-independent, subject-independent, and orientation-independent. A position-independent model is characterized by being able to correctly recognize the activities regardless of the position in which a subject wears the smart device, for example the hand on which the subject wears her/his smartwatch. A model is said to be orientation-independent if it correctly predicts the target activity regardless of the orientation of the smart device. As the subject wears her/his smartwatch in only one specific orientation, a smartwatch-based model is implicitly orientation-independent. However, this is not the case for smartphone-based models. A model is said to be subject-independent if it is able to predict the activities of subjects for whom no data was available during the training phase.

We achieve the aforementioned goal by building a predictive model that can recognize subjects' activities solely based on the sensor readings collected from their smartwatches. By conducting a comprehensive evaluation study in which we utilize different combinations of our four smart devices as observation sources, we show that the smartwatch-based model can recognize all activities with very good predictive performance.

DATA COLLECTION PROCESS

In this work, we study a combination of activities that, to the best of our knowledge, has not been studied before. We collected our dataset using four smart devices, namely two Nexus 5X smartphones and two Samsung gear live S2 galaxy smartwatches. Another Nexus 5X smartphone is used as a master device to coordinate the data collection process conducted by the other four devices. Twenty-two contributors participated in the data collection process. Their characteristics are shown in Table 1. As stated before, we study fourteen different activities. Nine of the contributors performed all of the studied activities while the other contributors performed only subsets of those activities. Table 2 presents detailed information about each activity with regard to the number of contributors who performed it as well as the collection duration.

Table 1. Contributor characteristics

Age	18-60 years
Weight	70-110 kg
height	155-185 cm
Gender	20 males, 2 females

While performing the studied activities, each contributor wore a smartwatch on each of her/his hands and put a smartphone in each of her/his front trouser pockets. We call these four devices the slave devices and we use them to capture sensor readings. A third smartphone, namely the master device is used to control the collection process. To collect the data, we developed two Android applications, namely the

[1]www.ready.gov/
[2]www.fema.gov/
[3]www.redcross.org/
[4]www.ndma.gov.in/en/

Table 2. Activities List

Activities	Shortcut	Contributors	Duration (min)
Crawling backward	CB	12	11.6
Crawling like kids	CLK	12	18.3
Going upstairs	GU	12	16.3
Going upstairs fast	GUF	10	14.6
Going downstairs	GD	10	14.4
Going downstairs fast	GDF	9	12.7
Going downstairs with injured right leg	GDWIRL	10	21.2
Normal sitting	NS	4	35
Sitting with hands on head	SWHOH	19	40
Kneeling	K	16	23.6
Kneeling with hands on head	KWHOH	18	44
Normal walking	NW	18	38.7
Walking with injured right leg	WWIRL	12	20.4
Walking with injured left leg	WWILL	12	21.3

master application and the slave application. We chose Android as an operating system to build our applications on top of it as it provides several important features with regard to security, portability, and efficiency in networking, power, and memory management. In addition to that, Android is widely used by a wide spectrum of famous smartphone makers such as Samsung, Huawei, LG, and HTC which gives us the potential to run our application on top of several hardware platforms.

Through the master application, the user can start and stop the collection process. Moreover, she/he can log the user and activity names, specify the list of sensors to collect their readings, the collection frequency, and the collection duration. Each of the slave devices runs the slave application whose main role is to respond to the data collection commands sent from the master application over Bluetooth.

Table 3 shows the list of sensors we monitor during the data collection process. We did not use all of those sensors in our research work. Nevertheless, we captured their readings for future studies. We use a sampling frequency of 40 Hz.

Table 3. The sampled sensors

Smartphone sensors	Smartwatch sensors
Accelerometer	Accelerometer
Gyroscope	Gyroscope
Gravity	Gravity
Magnetometer	Magnetometer
Uncalibrated magnetometer	Uncalibrated magnetometer
Game rotation vector	Game rotation vector
Rotation vector	Rotation vector
Linear accelerometer	Linear accelerometer
Barometer	Barometer
Geomagnetic Rotation Vector	-

CONCEPT AND DEVELOPED MODELS

Based on the chosen combinations of smart devices, we develop three main types of predictive models, namely:

- The smartwatch model which is built based on the observations obtained from the left-hand smartwatch.
- The smartphone model which is built based on the observations obtained from the left-pocket smartphone.

- The hybrid smartphone-smartwatch model which is built based on the observations obtained from the left-hand smartwatch and right-pocket smartphone.

For each of the aforementioned models, we can distinguish between two different model types, namely personal and impersonal models. As its name implies, a personal model is designed to recognize the activities of a specific person. Therefore, a personal predictive model is trained and evaluated using the data of only one subject. In contrast to that, an impersonal model is supposed to recognize the activities of any subject even when no data belonging to this subject was a part of the training phase. Therefore, an impersonal model is trained using the data of a certain set of subjects and evaluated using the data of a different set for which no training phase has been conducted. In this work, we only focus on impersonal models as our goal is to build an activity recognition model which is able to generalize beyond its training data.

To extract the features required for building our predictive models from the raw sensor data, we split this data into time windows such that each window represents a training/testing instance for which we extract the designed set of features. We present the impact of window size on the selected set of features as well as on the performance of the predictive models by using four different window size values, namely 2, 4, 6, and 8 seconds. We also present the impact of the sensor sampling frequency on the performance of our models by considering four sampling frequencies, namely 1, 10, 20, and 40 Hz. As a classification technique, we use the algorithm of random forests (Breiman 2001) as it has proven to achieve very good results in building activity recognition models (Weiss et al. 2016)(Maurer et al. 2006).

IMPERSONAL SMARTWATCH MODEL

In this section, we introduce our impersonal smartwatch predictive model. We build this model based on the sensor readings obtained from the left-hand smartwatch. To create an impersonal smartwatch model, we apply the leave-one-out principle. As mentioned before, nine subjects performed all the studied activities while each of the other 13 subjects performed a certain subset of activities. As a first step, the data needs to be divided into training and testing sets. By following the leave-one-out principle, we select one of the nine subjects who performed all activities to be the testing set of our model. As a training set, we use that data of all other 21 subjects. This process is repeated for all the nine subjects so that we test our impersonal smart watch model on each one of them. As a result of this procedure, we obtain nine different impersonal smartwatch models based on the subject chosen for the testing set. The final predictive performance of our model is calculated by averaging the results obtained from all of these nine models.

We consider four different values for the window size, namely 2, 4, 6, and 8 seconds. As mentioned before, the used set of features depends on the selected window size. In the next sections, we introduce the steps we followed to build and evaluate the impersonal models corresponding to each of the window sizes.

Two-Second Impersonal Smartwatch Model

The first step in building a predictive model is to design a set of features that can efficiently discriminate the different classes from each other and thereby achieves a very good predictive performance. We choose summary statistics, namely minimum, mean, median, maximum, standard deviation (SD), quartile25 (Qu_1), quartile75 (Qu_3) to build our set of features. After dividing the raw sensor data into windows, we calculate these values during each window for the different axis combinations of the accelerometer sensor, linear accelerometer, gyroscope as well as the gravity sensor. Furthermore, we calculate the correlation between the time and the readings of the barometer sensor. In order to identify the important features out of the whole set of features, we utilize the varimp method provided by the R caret package (Kuhn 2016).

The third column in Table 4 shows the final set of features we got for the two-second impersonal smartwatch model after applying the algorithm of variable importance on the total set of features. As can be seen from the table, different sets of summary statistics have been identified as the most important for different axis combinations.

By applying the leave-one-out principle, we build and evaluate nine different models based on the subject chosen for the testing set. For the different models we use the average f1-score as performance indicator (Goutte and Gaussier 2005). The f1-score is the harmonic average of precision and recall. Precision denotes the proportion of predicted positive cases that are correctly real positives and recall is the

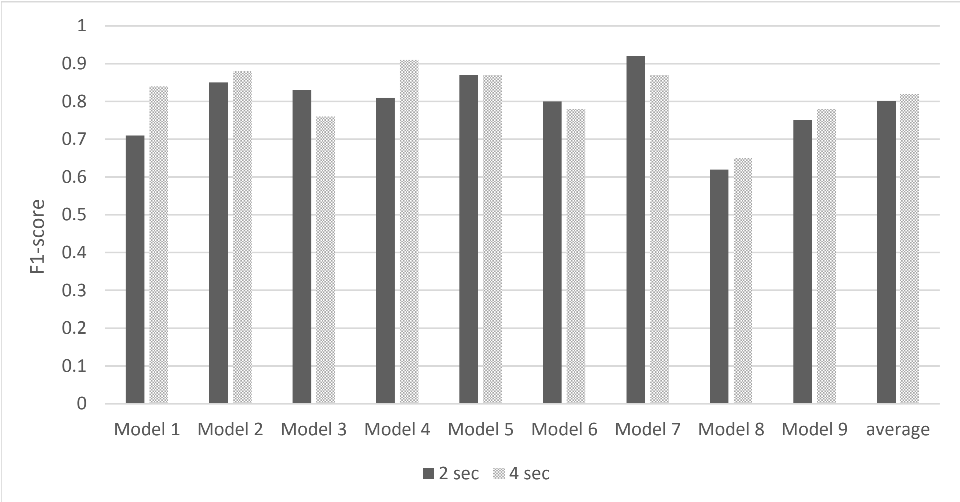

Figure 1. Performance evaluation of the two-second and four-second impersonal smartwatch model

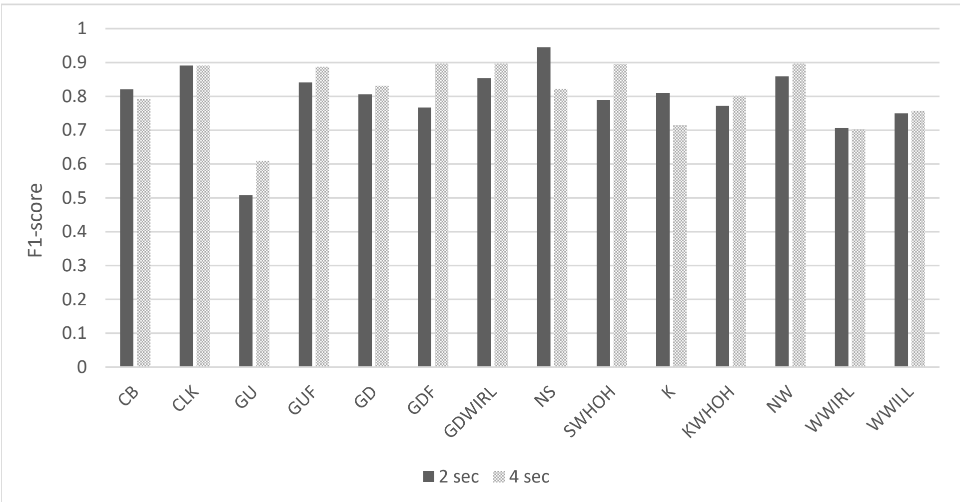

Figure 2. Performance evaluation of the two-second and four-second impersonal smartwatch model with regard to each of the studied activities

proportion of real positive cases that are correctly predicted positive. The f1-score reaches its best value at 1 when there is perfect precision and perfect recall.

The black bars in Figure 1 show the evaluation results for each of those models in terms of the average f1-score achieved in recognizing the designed set of activities. This figure shows that the two-second impersonal smartwatch model achieves an average f1-score value of 80% with regard to the nine testing subjects.

The black bars in Figure 2 show the achieved predictive performance with regard to each of the studied activities in terms of the average f1-score value that has been achieved by the nine models. We notice from this figure that the models achieve a low predictive performance with an average f1-score of 50.6% in recognizing the activity of going upstairs. In most cases, the models predict the examples belonging to this activity as normal walking or going downstairs. The reason behind that lies in the fact that the motion of the hands while performing those activities is almost the same which makes it difficult for a predictive model to separate them well.

Four-Second Impersonal Smartwatch Model

We build this model by following the same steps used in building the two-second model in the previous section. As a first step, we identify the set of important features we use to build this model. By applying the variable importance algorithm on the summary statistics, we get the set of features for the four-second model shown in fourth column of Table 4. Comparing with the features of the two-second model, we notice that some features are deleted and others are added. This can be explained by the fact that changing window size renders certain features more or less important.

By applying the leave-one-out principle on the data of the nine subjects, we build and evaluate nine different models as we did for the two-second model. The evaluation results corresponding to each of those models are shown by the gray bars in Figure 1. By comparing this model with the two-second model, we notice that the performance has slightly improved with an average f1-score value of 82%. The average performance of the nine models corresponding to each of the studied activities are represented by the gray bars shown in Figure 2. We notice a remarkable improvement over the two-second model regarding the activities of going upstairs, going downstairs fast, and sitting with hands on head with 10% to 14% increase in the average f1-score value. However, a remarkable decrease of 10% to 12% in the average f1-score can be noticed for the activities of normal sitting, and kneeling. Moreover, we still notice that the lowest predictive performance has been achieved in recognizing the activity of going upstairs.

Table 4. Two-second, four-second, six-second and eight-second impersonal smartwatch model: used set of features

Sensor	Axis/Axes-relation	Features			
		2 sec	**4 sec**	**6 sec**	**8 sec**
Accelerometer	Magnitude(x,y,z)	Mean, max, SD, Q_1, Q_3	Mean, SD, Q_1, Q_3	Mean, Q_1, Q_3	Mean, Q_1, Q_3
	x	-	Max, SD, Q_1	SD	-
	y	-	SD, Q_1	SD, Q_1	SD, Q_1
	z	-	SD, Q_1	SD	SD
	Magnitude(x,z)	SD	SD	-	-
	Magnitude(y,z)	Min, Q_1	Min, Q_1	Min, Q_1	Min
	Magnitude(x,y)	-	Mean, SD, Q_3	Q_3	Q_3
	x/z	Mean, max, Q_3	Mean, Q_3	Mean, Q_3	Mean, Q_3
	z/x	Mean, max, SD, Q_1	Mean, max, min, SD, median, Q_1	Mean, max, SD, Q_1	Mean, max, SD, Q_1
Linear Accelerometer	Magnitude(x,y,z)	Q_3	Q_1	Q_1	Q_1
	x	SD	Mean, SD, Q_3	Mean, SD, max, Q_3	Mean, SD, Q_3
	y	Mean, SD, Q_1	Mean	Mean	Mean
	z	SD	-	-	-
	Magnitude(x,y)	Median, Q_1	Q_1	Q_1	Q_1
	Magnitude(x,z)	-	SD	SD	SD
	z/y	Q_3	-	-	-
Gyroscope	x	-	-	Mean	Mean
	y	Mean, median	Mean, median	Mean, median	Median
	Magnitude(x,z)	-	Q_1	Q_1	-
	x/y	Median	Median, Q_3	Median, Q_3	Median
	y/x	Median	Median, Q_1	Median, Q_1	Median
Gravity	x	Mean, min, SD, median, Q_1	Mean, min, SD, median, Q_1, Q_3	Mean, min, SD, median, Q_1, Q_3	Mean, min, SD, median, Q_1
	y	-	SD	-	-
	z	Max, min	Max, min, Q_3	Max, min	Max
	Magnitude(x,z)	SD	-	-	-
	Magnitude(y,z)	-	SD	SD	-
	x/z	-	Max	Max	-
	z/x	-	Min	Min	-
Barometer	Readings, time	Correlation	Correlation	Correlation	Correlation

Six-Second Impersonal Smartwatch Model

Using the variable importance method, we identify the set of features used to build this model as shown in the fith column of Table 4. As we notice from this table, changing the window size to six seconds renders certain features as less or more important as it was the case for the four-second window.

Applying the method of variable importance on the six-second model generates a very important observation. With a six-second window size, the correlation between the time and barometer readings is identified as the most important feature for all nine models. This can be explained by the fact that a longer window size gives enough time for the barometer readings to have meaningful correlations with the different activities. Based on the accompanied changes in the sensed air pressure, we can categorize our set of activities into three categories, namely:

- Activities that lead to an increasing sensed air pressure such as going upstairs and going upstairs fast.
- Activities that lead to a decreasing sensed air pressure such as going downstairs, going downstairs fast, and going downstairs with injured leg.
- Activities that do not lead to a significant change in the sensed air pressure such as normal walking and normal sitting.

Figure 3 shows the time series representing the barometer readings corresponding to six different activities, namely going downstairs, going downstairs fast, going upstairs, going upstairs fast, normal walking, and normal sitting. Theses readings were captured for six seconds and with a 40 Hz sampling frequency. Under each sub-figure, we mention the respective activity along with the computed correlation between the time and its barometer readings. We observe from this figure that the correlation has a high positive value for the activities going downstairs and going downstairs fast. It has a high negative value in case of going upstairs and going upstairs fast. Therefore, using this correlation as a feature simplifies the process of distinguishing between going upstairs and downstairs. Regarding the activities of normal walking and normal sitting, the readings of barometer sensor do not remarkably increase or decrease along the time. Therefore, a very low correlation value can be noticed for those both activities.

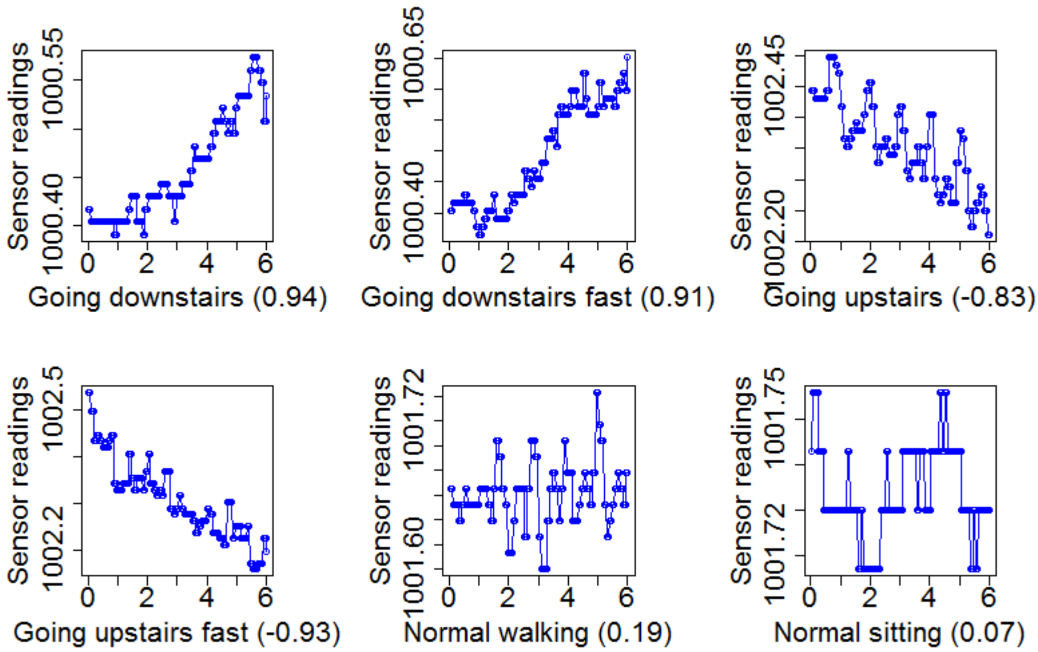

Figure 3. Correlation between the time and barometer readings

As a result, we conclude that increasing the window size improves the role of barometer sensor in recognizing the studied activities. To show its impact, we evaluate the performance of the six-second nine models with and without the barometer-time correlation as a feature. Figure 4 shows the predictive performance of the nine models in both cases. It is clear from this figure that the correlation feature positively affects the predictive performance of eight models. Moreover, we notice that the six-second model is more stable than the two-second and four-second models. Apart from the result corresponding to the eighth subject, all the f1-score values are above 80%. The stability in performance the model achieves when tested with data coming from different subjects proves that it generalizes well beyond the training data.

Figure 5 presents the predictive performance of the nine models for both cases with regard to each of the studied activities. This figure proves the importance of the correlation feature in improving the prediction results for most of the studied activities. We notice a significant increase in the average f1-score values achieved in recognizing the activity of going upstairs. As we have seen, all the previous models suffered from low f1-score values in recognizing this activity. Moreover, all of them confused between the activities

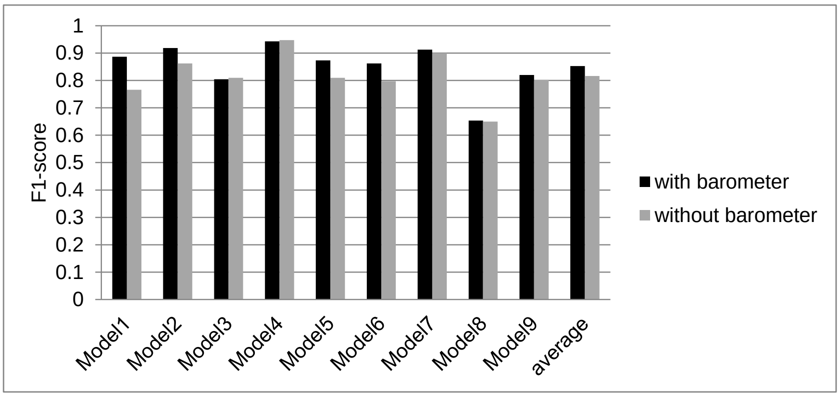

Figure 4. Performance evaluation of the six-second impersonal smartwatch model

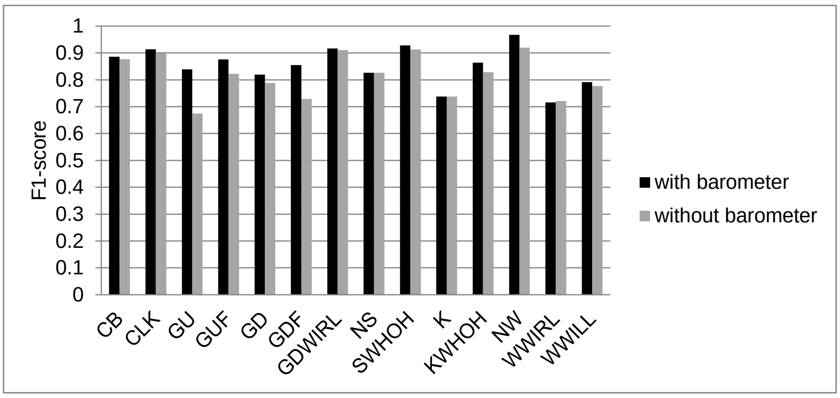

Figure 5. Performance evaluation of the six-second impersonal smartwatch model with regard to each of the studied activities

that are performed on stairs. We solved this problem so that no observable confusion between such activities can occur by utilizing the barometer-time correlation feature with the suitable window size.

Eight-Second Impersonal Smartwatch Model

To further inspect the effect of window size on the predictive performance of our impersonal smartwatch model, we increase it to 8 seconds and build the corresponding model. As done for all the previous models, we obtain the set of features shown in the last column of Table 4 and use it to build this model based on the method of variable importance. We build and evaluate nine different models by applying the leave-one-out principle. The black bars in Figure 6 show the evaluation results corresponding to each of those nine models. With an 82% average f1-score value, the eight-second impersonal smartwatch model achieves lower performance than the six-second model. Figure 7 shows the average performance of the nine models with regard to each of the studied activities. This figure clearly shows that the nine models are not able to achieve a good performance in recognizing the activity of going upstairs. With an average f1-score of 67.3%, the eight-second impersonal smartwatch model does not reach the performance achieved by the six-second model. As a result, we conclude that increasing the window size from 6 to 8 seconds has a negative impact on the performance of our impersonal smartwatch model.

Discussion and Comparison Between the Impersonal Smartwatch Models

The evaluation results of the impersonal smartwatch models with regard to the different window sizes clearly show that the six-second model outperforms all the other models. This can be seen in Figure 8 as it achieves the best f1-score value. By studying four window sizes, we have seen that the features corresponding to each window size should be independently extracted as the features have different importance levels based on the used window size. To prove that, we use the features of the two-second model to build a model with a window size of 6 seconds. This model achieves an average f1-score value of

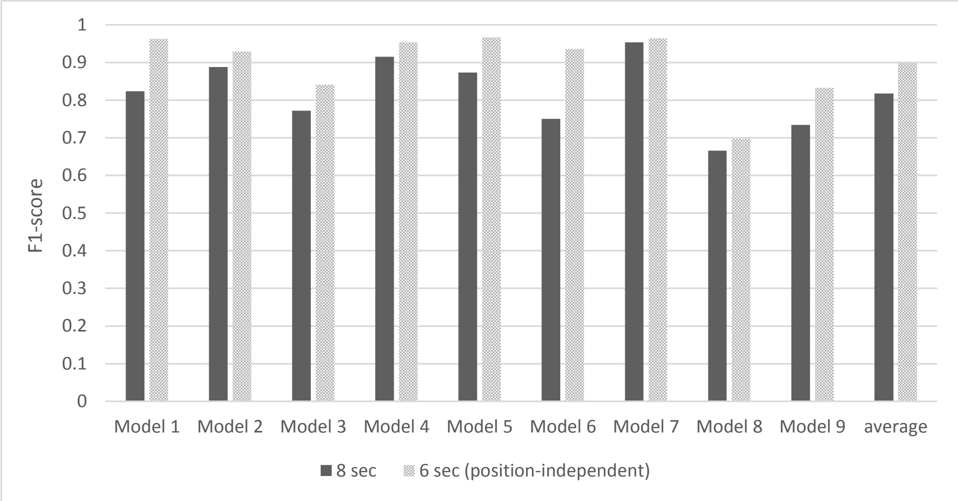

Figure 6. Performance evaluation of the eight-second impersonal smartwatch model and the position independent six-second impersonal smartwatch model

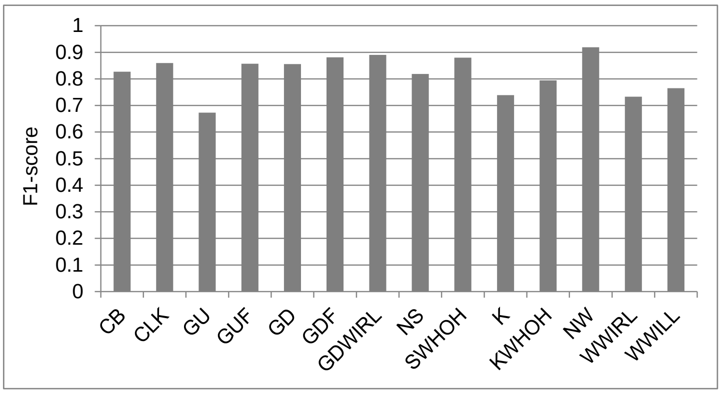

Figure 7. Performance evaluation of the eight-second impersonal smartwatch model with regard to each of the studied activities

83.9%. By comparing this value to the 85.2% f1-score achieved by our six-second impersonal smartwatch model, we notice that identifying a new set of features enhances the predictive performance.

As mentioned before, we used a sampling frequency of 40 Hz to capture the sensor readings. Using such a high frequency consumes the battery of the smart device and requires high processing capabilities. To check whether it is possible to reduce this frequency and get the same level of performance, we used four different sampling frequencies, namely 1, 10, 20, and 40 Hz. Figure 9 shows the results achieved by using those different frequencies with the six-second impersonal smartwatch model. The model gives its best predictive performance with a frequency of 40 Hz. This performance decreases to reach an f1-score value of 52% using a frequency of 1 Hz. By using a sampling frequency of 10 Hz, the model achieves an f1-score value of 84.6% which is 0.6% less than the value achieved with a sampling frequency of 40 Hz. These results indicate that a sampling frequency of 10 Hz can be used to reduce the required power and processing capabilities while keeping the predictive performance almost at the same level.

Position-independent Impersonal Smartwatch Model

Based on the previous evaluation results, the six-second impersonal smartwatch model proved to achieve the best predictive performance with an average f1-score of 85.2%. This model is orientation-independent as it only uses the smartwatch as an observation source. Moreover, it is subject-independent as it achieved a stable performance for eight testing subjects. Our goal in this section is to make this model position-independent. We achieve this goal by training it using the sensor readings of both left-hand and right-hand smartwatches. We use the same set of features listed in the fifth column of Table 4. Before building this model, we need to perform the following steps:

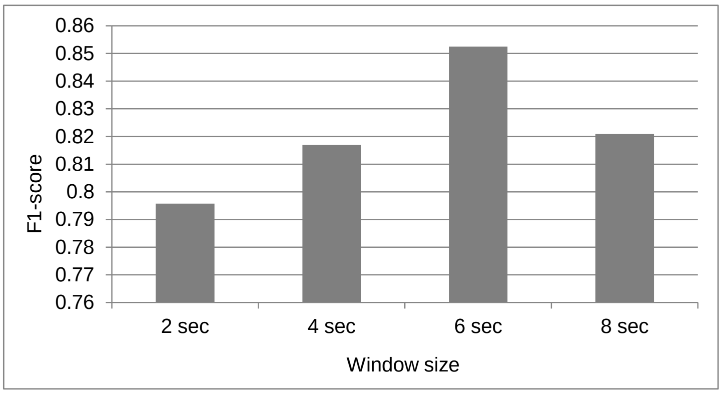

Figure 8. Evaluation summery of the impersonal smartwatch models

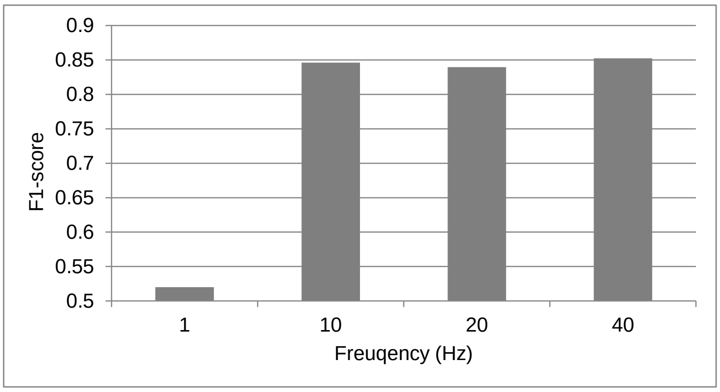

Figure 9. Performance evaluation of the six-second impersonal smartwatch model with different sampling frequencies

- We delete the activity of going downstairs with injured right leg from the dataset as it has been only performed where the handrails are on the left side of the stairs.
- We combine the activities of walking with injured right leg and walking with injured left leg into one activity with the name walking with injured leg (WWIL).
- We combine the activities of kneeling with hands on head and sitting with hands on head into one activity with the name hands on head (HOH).
- We combine the activities of kneeling and normal sitting into one activity with the name sitting (S).

By following the same previous leave-one-out principle, we evaluate the predictive performance of this model using the same previous nine subjects as testing sets. For evaluation purposes, we only use the sensor readings obtained from the left-hand smartwatch as we assume the subject to be wearing only one smartwatch in real-world applications. The gray bars in Figure 6 illustrate the average f1-score achieved for each of the nine subjects as well as their average. As we notice from this figure, this model outperforms all the previous models with an average f1-score of 90.1%. Figure 10 shows the average f1-score values achieved by this model with regard to each of the studied activities.

IMPERSONAL SMARTPHONE MODEL

To build our impersonal smartphone model, we use the sensor readings obtained from the left-pocket smartphone. We consider four different window size values, namely 2, 4, 6, and 8 seconds. Based on the method of variable importance, we identify the set of important features for each window size. We build and evaluate the models by following the same previous leave-one-out principle. Table 6 shows the evaluation results corresponding to each window size. We notice from this table that the six-second impersonal smartphone model outperforms all other models. Table 5 shows the set of features used in building this model. This table shows that we only use orientation independent features. With regard to

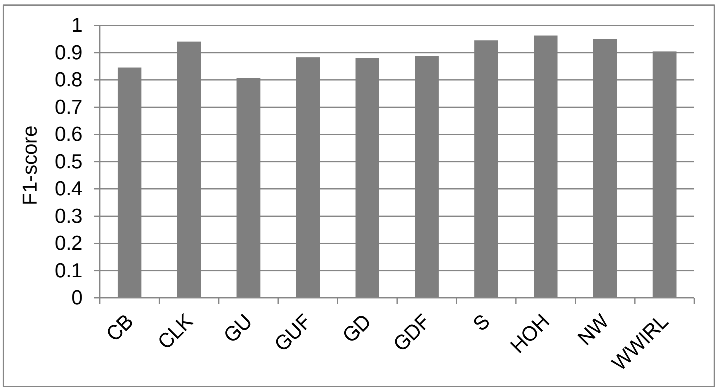

Figure 10. Performance evaluation of the position-independent six-second impersonal smartwatch model with regard to each of the studied activities

x- and y-axes, we only use the magnitude over both of them as it is orientation-independent. All the readings captured over the z-axis are implicitly orientation-independent.

Table 5. Six-second impersonal smartphone model: used set of features

Sensor	Axis/Axes relation	Features (6 sec)
Accelerometer	Magnitude(x,y,z)	Mean, max, SD, Q_1
	Magnitude(x,y)	Max, Q_3
	z/Magnitude(x,y)	SD, median, Q_1
	z	SD, median
Linear Accelerometer	Magnitude(x,y,z)	Median, Q_1, SD
	Magnitude(x,y)	SD, median, Q_1, Q_3,
Gyroscope	Magnitude(x,y,z)	Mean, Q_1, median, Q_3
	Magnitude(x,y)	Mean, median, Q_1, Q_3
Gravity	Magnitude(x,y)	SD
	Magnitude(x,y)/z	Median, SD, Q_1
	z/Magnitude(x,y)	SD, median, Q_3
	z	SD, Q_3
Barometer	Readings, time	Correlation

Table 6. Performance evaluation of the impersonal smartphone model with different window size values

Window size (Min)	2	4	6	8
F1-score (%)	59.6	61.7	64	63.7

As shown by the black bars in Figure 11, the average f1-score value corresponding to the nine models is 64%. It represents a low value in comparison to the performance of the six-second impersonal smartwatch model. The f1-score values corresponding to each of the studied activities are represented by the black bars in Figure 12. This figure illustrates the reason behind such a low predictive performance. The f1-score values corresponding to the activities of kneeling, normal sitting, sitting with hands on head, and kneeling with hands on head are between 27.9% and 43.8%. The position of smartphone in the trouser pocket results in similar sensor readings corresponding to these activities. This is because the subject's legs while performing those activities are still and have the same direction. Although the other activities are considered as hand- and leg-based activities, the impersonal smartphone model does not achieve a good predictive performance in recognizing them. The reason behind that lies in using only orientation-independent features to make the system applicable in real-world scenarios and comparable to the impersonal smartwatch model.

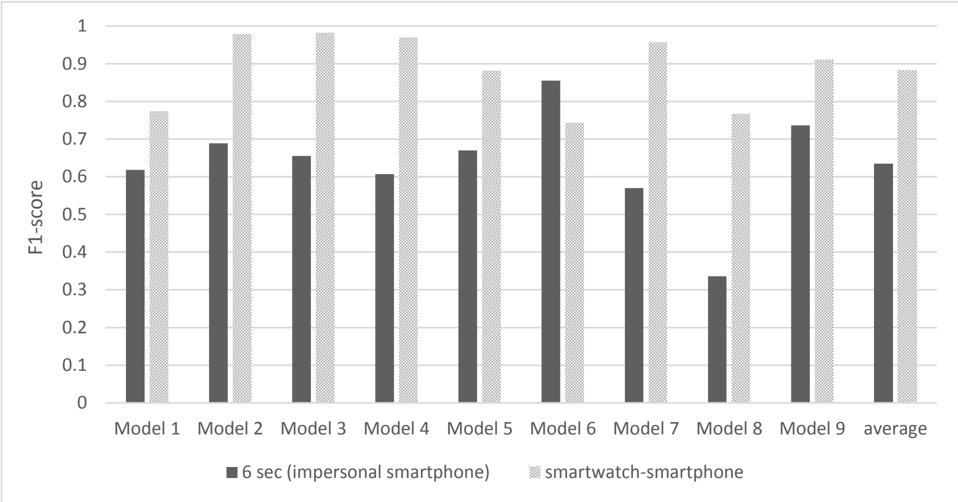

Figure 11. Performance evaluation of the six-second impersonal smartphone model and the smartwatch-smartphone model

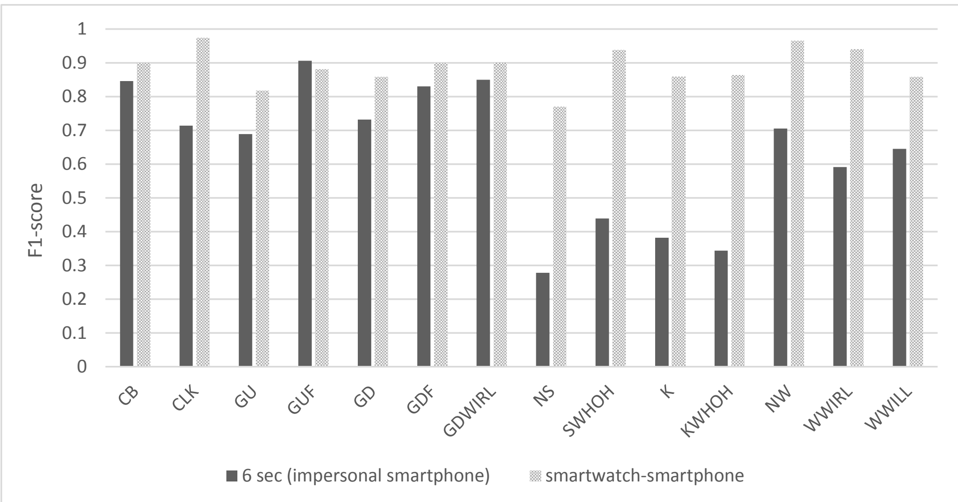

Figure 12. Performance evaluation of the six-second impersonal smartphone model and the smartwatch-smartphone model with regard to each of the studied activities

HYBRID SMARTPHONE-SMARTWATCH MODEL

In this section, we discuss the effect of combining two smart devices, namely a smartphone and a smartwatch on the predictive performance of our activity recognition model. In this setting, we assume the subject to wear a smartwatch on her/his left hand while carrying a smartphone in her/his right pocket. The goal of building such a model is to show the impact of capturing sensor readings from different positions of subject's body on the predictive performance of activity recognition models. Several of our activities are hand-based, namely normal sitting, sitting with hands on head, kneeling, and kneeling with hands on head. All smartphone models achieved low predictive performance in recognizing those activities. The remaining set activities is considered as hand- and leg-based activities.

By capturing the readings of sensors that are positioned on the hand and leg of a subject, the model should be able to achieve a better predictive performance in recognizing our set of activities. To build this model, we use a six-second window size with a 40 Hz sampling frequency. As an evaluation methodology, we utilize the previous leave-one-out principle with a feature set which combines the features listed in the fifth column in Table 4 and in Table 5.

The gray bars in Figure 11 show the predictive performance achieved for each of the nine subjects in terms of f1-score values. The average f1-score values with regard to the single activities are represented by the gray bars in Figure 12. These results show that using a combination of smart devices leads to an enhancement in the predictive performance of activity recognition models. Such results prove that capturing the motions of subject's hand and leg simplifies the process of body-based activity recognition. However, smartwatch-based models are more applicable in real-world scenarios as it is more realistic in disaster situations to assume the subject to be wearing a smartwatch than to be carrying a smartphone while being affected by the disaster.

RELATED WORK

Several research projects have been presented in the domain of activity recognition using smart devices. In this section, we introduce a group of these projects with the main focus of identifying the factors that affect the predictive performance of activity recognition models using smart devices. Examples of such factors are window size, model type i.e. personal or impersonal, and sensor placement. Kwapisz et al. (Kwapisz et al. 2011) studied the effect of window size on the performance of their activity recognition model. They came out with the result that a 10-second window size leads to a better predictive performance when compared to a 20-second window. However, they did not provide a detailed discussion regarding the reasons behind such a result.

Shoaib et al. (Shoaib et al. 2013) showed that the performance of activity recognition models is highly affected by the position and type of used sensors. They studied the effect of using accelerometer and gyroscope sensors together and individually where they found out that using both sensors leads to an improvement in the performance of their final model. Moreover, they studied the effect of incorporating the magnetometer sensor data into the feature space. Based on their evaluation study, they found out that the magnetometer, when used individually or in combination with the accelerometer and gyroscope sensors, does not provide any remarkable improvement in recognizing their set of activities. Rasekh et al. (Rasekh et al. 2014) studied the effect of high dimensionality on the performance of their activity recognition models. They showed that reducing the dimensionality of the feature set improves the predictive performance of their models.

Lockhart et al. (Lockhart and Weiss 2014) studied the impact of model type on the performance of activity recognition models. In their work, they compared between personal, hybrid, and impersonal models. Their evaluation results showed that personal models achieve the best predictive performance. Moreover, they found out that hybrid models tend to outperform the impersonal models. Weiss et al. (Weiss et al. 2016) compared the performance of smartwatch- and smartphone-based models in recognizing hand-based activities. Their results showed that a remarkable improvement in recognizing hand-based activities such as eating and drinking can be achieved by utilizing smartwatches as observation sources.

CONCLUSION, DISCUSSION & FUTURE WORK

In this paper, we presented our system for detecting people in disaster related situations by recognizing their physical activities. We built this system using two smartwatches and three smartphones. We designed a set of activities that are typically performed by people in disaster scenarios. We created a dataset containing 22 contributors performing the disaster activity set as well as everyday activities. Based on this dataset we conducted a comprehensive evaluation study to achieve our goal of building an orientation-independent, position-independent, and subject-independent model which is applicable in real-world scenarios. We built and evaluated a diverse set of models based on different combinations of smart devices. Our optimal model achieves a high average f1-score value of 90.1% in recognizing ten different activities related to disaster scenarios solely based on one smartwatch as an observation source. This model is characterized by being orientation-independent, position-independent, and subject-independent.

This work discussed the factors affecting the predictive performance of activity recognition models. We studied the impact of window size on designing the optimal set of features used to build the classification models. Moreover, we proved that the predictive performance is highly dependent on the chosen window size and sampling frequency. Our work shows the possibility of achieving a trade-off between an accurate recognition model and a moderate sampling frequency which preserves the resources of smart devices.

Since the created dataset contains only activities which mainly involve persons younger than 40 years old, the question arises how well the created models can reliably detect activities of persons of other age groups. In future work, we want to focus especially on the elderly who are often limited in their mobility and often need special help in the event of a disaster. A further uncertainty of the data set lies in the realistic execution of disaster activities. After best efforts, during the data collection, we tried to ensure that the volunteers carry out the activities as realistically as possible. However, it is assumed that the real catastrophic situation, including various stress factors, can lead to different movement patterns during the execution of the activities. To minimize this potential source of error, we want to include more contributors of different age groups and health conditions. Utilizing our data recording procedure in planned emergency drills, for example earthquake and tsunami drills (Johnson et al. 2014), more situation aware and more realistic data could be recorded and used for the training. Even though we proposed a pre-learned subject-independent prediction model, each model could also be improved by an individual learning phase performed by the smartphone/smartwatch owner. An interesting research question arises, if the mentioned individual learning phase, containing only everyday activities, could lead to higher prediction accuracy of the disaster activity set. This way the additional training could be executed during the day without any disturbance of the user.

ACKNOWLEDGMENT

This work has been co-funded by the following projects: the LOEWE initiative (Hesse, Germany) within the NICER[5] project, the German Federal Ministry of Education and Research within the SMARTER[6] and the Software Campus project "DisasterSense" (Support Code: 13N13406 & 01IS12054).

REFERENCES

Alwan, M., Rajendran, P. J., Kell, S., Mack, D., Dalal, S., Wolfe, M., and Felder, R. (2006). “A smart and passive floor-vibration based fall detector for elderly”. In: *Information and Communication Technologies, 2006. ICTTA’06. 2nd*. Vol. 1. IEEE, pp. 1003–1007.

Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., and Amirat, Y. (2015). “Physical Human Activity Recognition Using Wearable Sensors”. In: *Sensors* 15.12, pp. 31314–31338.

Breiman, L. (2001). “Random Forests”. In: *Machine learning* 45.1, pp. 5–32.

Cecchinato, M. E., Cox, A. L., and Bird, J. (2015). “Smartwatches: the Good, the Bad and the Ugly?” In: *Proceedings of the 33rd Annual ACM Conference extended abstracts on human factors in computing systems*. ACM, pp. 2133–2138.

Chetty, G., White, M., and Akther, F. (2015). “Smart Phone based Data Mining for Human Activity Recognition”. In: *Procedia Computer Science* 46, pp. 1181–1187.

Cook, D. J. (2010). “Learning Setting-Generalized Activity Models for Smart Spaces”. In: *IEEE intelligent systems* 2010.99, p. 1.

Garcia, J. J., Bruyckere, H. de, Keyson, D. V., and Romero, N. (2013). “Designing personal informatics for self-reflection and self-awareness: the case of children with attention deficit hyperactivity disorder”. In: *International Joint Conference on Ambient Intelligence*. Springer, pp. 109–123.

[5]Networked Infrastructureless Cooperation for Emergency Response
[6]Smartphone-based Communication Networks for Emergency Response

Goutte, C. and Gaussier, E. (2005). "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation." In: *ECIR*. Vol. 5. Springer, pp. 345–359.

Guo, J., Zhou, X., Sun, Y., Ping, G., Zhao, G., and Li, Z. (2016). "Smartphone-Based Patients' Activity Recognition by Using a Self-Learning Scheme for Medical Monitoring". In: *Journal of medical systems* 40.6, pp. 1–14.

Gyorbiro, N., Fabian, A., and Homanyi, G. (2009). "An Activity Recognition System For Mobile Phones". In: *Mobile Networks and Applications* 14.1, pp. 82–91.

Hossmann, T., Legendre, F., Carta, P., Gunningberg, P., and Rohner, C. (2011). "Twitter in disaster mode: Opportunistic communication and distribution of sensor data in emergencies". In: *Proceedings of the 3rd Extreme Conference on Communication: The Amazon Expedition*. ACM, p. 1.

Johnson, V. A., Johnston, D. M., Ronan, K. R., and Peace, R. (2014). "Evaluating children's learning of adaptive response capacities from ShakeOut, an earthquake and tsunami drill in two Washington State school districts". In: *Journal of Homeland Security and Emergency Management* 11.3, pp. 347–373.

Kefalidou, G., Skatova, A., Brown, M., Shipp, V., Pinchin, J., Kelly, P., Dix, A., and Sun, X. (2014). "Enhancing self-reflection with wearable sensors". In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, pp. 577–580.

Kuhn, M. (2016). *caret: Classification and Regression Training*.

Kwapisz, J. R., Weiss, G. M., and Moore, S. A. (2011). "Activity Recognition Using Cell Phone Accelerometers". In: *ACM SigKDD Explorations Newsletter* 12.2, pp. 74–82.

Lieser, P., Alvarez, F., Gardner-Stephen, P., Hollick, M., and Boehnstedt, D. (2017). "Architecture for Responsive Emergency Communications Networks". In: *Global Humanitarian Technology Conference (GHTC), IEEE*.

Lockhart, J. W. and Weiss, G. M. (2014). "The Benefits of Personalized Smartphone-Based Activity Recognition Models." In: *SDM*. SIAM, pp. 614–622.

MacLean, D., Roseway, A., and Czerwinski, M. (2013). "MoodWings: a wearable biofeedback device for real-time stress intervention". In: *Proceedings of the 6th international conference on PErvasive Technologies Related to Assistive Environments*. ACM, p. 66.

Mantyjarvi, J., Himberg, J., and Seppanen, T. (2001). "Recognizing Human Motion With Multiple Acceleration Sensors". In: *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*. Vol. 2. IEEE, pp. 747–752.

Maurer, U., Smailagic, A., Siewiorek, D. P., and Deisher, M. (2006). "Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions". In: *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*. IEEE, 4–pp.

Otebolaku, A. M. and Andrade, M. T. (2016). "User Context Recognition Using Smartphone Sensors and Classification Models". In: *Journal of Network and Computer Applications* 66, pp. 33–51.

Rasekh, A., Chen, C.-A., and Lu, Y. (2014). "Human Activity Recognition Using Smartphone". In: *arXiv preprint arXiv:1401.8212*.

Reuter, C., Kaufhold, M.-A., Leopold, I., and Knipp, H. (2017). "Katwarn, NINA or FEMA? Multi-Method Study on Distribution, Use and Public Views on Crisis Apps". In: *European Conference on Information Systems (ECIS)*.

Shoaib, M., Scholten, H., and Havinga, P. J. (2013). "Towards Physical Activity Recognition Using Smartphone Sensors". In: *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)*. IEEE, pp. 80–87.

Webb, A. K., Vincent, A. L., Jin, A., and Pollack, M. H. (2013). "Wearable sensors can assist in PTSD diagnosis". In: *Body Sensor Networks (BSN), 2013 IEEE International Conference on*. IEEE, pp. 1–6.

Weiss, G. M., Timko, J. L., Gallagher, C. M., Yoneda, K., and Schreiber, A. J. (2016). "Smartwatch-based Activity Recognition: A Machine Learning Approach". In: *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, pp. 426–429.