Journal of Ambient Intelligence and Smart Environments 12 (2020) 239–261
DOI 10.3233/AIS-200558
IOS Press

Smoking recognition with smartwatch sensors in different postures and impact of user's height

Sumeyye Agac [a,*], Muhammad Shoaib [b] and Ozlem Durmaz Incel [a]
[a] *Department of Computer Engineering, Galatasaray University, Ciragan Cd. No:36, 34349, Istanbul, Turkey*
E-mails: konaksumeyye@gmail.com, odincel@gsu.edu.tr
[b] *Pervasive Systems Research Group, University of Twente, Enschede, The Netherlands*
E-mail: m.shoaib@utwente.nl

Abstract. Currently, smartwatches are mainly used as an extension of smartphones. However, equipped with various motion sensors, they are also effective devices for human activity recognition, particularly for those involving hand and arm movements. In this paper, we investigate the smoking recognition problem with motion sensors on smartwatches using supervised learning algorithms. For this purpose, we collected a dataset from 11 participants including ten different activities. The dataset includes different smoking variations in four different postures, such as smoking while standing, as well as similar activities, such as eating, and other activities, such as walking. Instead of approaching the problem as a binary classification problem, such as smoking and other, we are interested in differentiating smoking in different postures. Our aim is to explore the parameter space that may affect the recognition process on a large and complex dataset, considering 4 different window sizes and overlaps, 63 different features extracted from each sensor, 4 different sensors, 2 different sensor combinations, 3 classifiers and 10 different activities. Additionally, we analyze the impact of participants' height on the recognition performance. The results show that, simple time-domain features and the combination of accelerometer and gyroscope sensors perform the best. When we consider the impact of height on the recognition performance, the results show that it does not have a significant effect when all activities are considered, however, it does have an effect on smoking while standing, particularly for participants with a significant height difference than others.

Keywords: Activity recognition, wearable computing, motion sensors, feature engineering

1. Introduction

Smartwatches integrated with various sensors are emerging as ideal platforms for human activity recognition, particularly for sports and well-being applications [15]. Compared to smartphones, they have advantages, such as ease of carrying, being attached to the wrist instead of being carried in a pocket or bag. Moreover, they make it easier to recognize more complex activities, especially those involving hand and arm movements, such as eating, typing, and drinking. Nowadays, the most common uses of smartwatches include getting the notifications on the watch rather than on the phone, watching the time and following the steps taken by the user, as a step-tracker. However, with the variety of the sensors included in the smartwatches, they can be used to recognize more complex activities and assist the user to track his or her routines and patterns [10].

One of the behavioral patterns that the users may be interested to track is the smoking pattern, such as the number of cigarettes smoked, periods, and time of smoking. Such tracking can be useful for the user to get an insight into his or her smoking behavior and it can be useful in an effective intervention for behavior change, such as quitting smoking or reducing the

*Corresponding author. E-mail: konaksumeyye@gmail.com.

number of cigarettes smoked per day. Particularly, for smoking cessation programs, self reporting puts a burden on the user, but smartwatches can enable automated self reporting and provide more context to the smoking activity [26]. However, it is challenging to detect smoking compared to simpler locomotion activities, such as walking, running, because it is not a periodic activity unlike these simple activities. Moreover, it can be performed in various postures (sitting, walking, standing) and in combination with different activities (in a group while chatting, alone, while drinking coffee). It can also be confused with similar activities, such as eating, drinking, that involve similar hand gestures. Height of the users may also impact the recognition performance.

In this paper, we investigate the recognition of smoking activity with the motion sensors available on smartwatches. For this purpose, we collected a dataset from 11 participants including ten different activities. The data includes four different smoking variations: smoking while sitting (smokeST), smoking while standing (smokeSD), smoking in a group (smokeGroup), and smoking while walking (smokeWalk), to address the challenge of smoking recognition in different postures. Moreover, it would be interesting to detect the posture/context while smoking for a behavioral change. For example, if a smoker often smokes within a group, this may show that he smokes for socializing and recognizing this may increase his awareness about his behaviour. The dataset is not only composed of smoking but also includes activities with similar hand/arm gestures: eating while sitting (eat), drinking while sitting (drinkST), drinking while standing (drinkSD). Standing (stand), sitting (sit) and walking (walk) activities are also performed alone to differentiate these activities in combination with smoking.The dataset contains 45 hours of activity and 17 hours of smoking in different postures and to the best of our knowledge, this is the largest smoking dataset compared to previous studies. Our dataset is publicly available for reproducibility which can be found at [7].

Our aim is to analyze the recognition of smoking activity in detail, with a focus particularly on using different sensors, different and comprehensive set of features. Seventeen features from four dimensions (x, y, z and magnitude) of accelerometer, gyroscope, and linear acceleration sensors are extracted. The linear acceleration sensor can be considered as a virtual sensor. It is calculated from the raw accelerometer readings by subtracting the gravity effect on x, y and z dimensions [27]. Another set of seventeen features from pitch and roll values computed from the accelerometer readings [4] is also used in the analysis. Pitch and roll values are calculated using Equation (1) and Equation (2), respectively. Variable x, y and z represent the accelerometer readings in three dimensions and g represents the gravity of Earth, i.e., 9.81 m/s^2.

$$\beta = \frac{180}{\pi}.\tan^{-1}(y/g, z/g) \tag{1}$$

$$\alpha = \frac{180}{\pi}.\tan^{-1}(x/g, z/g) \tag{2}$$

As the final parameter, we investigate the impact of height on smoking recognition performance. To the best of our knowledge, this has not been investigated in previous studies. We analyze the impact of height both using regression analysis and by grouping the users with dissimilar heights. We use three different classifiers, namely support vector machine, random forest and multilayer perceptron which are commonly used for activity recognition [21,24,29]. Scikit-learn (Version 0.18.1) is used for the analysis of our dataset. Moreover, these classifiers are shown to perform well in terms of resource consumption, such as battery, memory and CPU cycles when implemented on smart watches [22]. Although we do not study the analysis of resource consumption in this paper, in a previous study [22], we analyzed the impact of using different classifiers, sensors, sampling rates and window sizes and here we investigate the recognition performance by focusing on resource-aware parameters.

Compared to previous works on smoking recognition [19,26], we do not study the recognition of puffs (hand-to-mouth gesture) to identify smoking periods. In the data collection phase, this makes it easy to label only the start and end of smoking sessions instead of a fine-grain gesture labeling and enables continuous recognition. The main contributions and highlights of the paper are summarized as follows:

– We explore the parameter space of activity recognition that may affect the performance on a large and complex dataset in the context of smoking, considering 4 different postures, 4 different window sizes and overlaps, 63 different features extracted from each sensor, 4 different sensors and 2 different sensor combinations and 3 classifiers. We show that average recognition performance considering 10 activities with accelerometer using simple features is around 77% and when it is combined with gyroscope it is around 83% (in terms of F1-score).

- Instead of recognizing only smoking or not, we explore the performance of recognition in different smoking postures, such as smoking while sitting, while standing, while walking and while in a group. We show that smoking while sitting and standing are recognized with 70% F1-score, while walking is with 94% and while in a group is with 63% F1-score. Smoking in a group is found to be more difficult to recognize due to different patterns exhibited by the participants.
- We also investigate the impact of subject's height on the smoking recognition performance and the results show that, the recognition performances of smoking while standing and sitting are reduced when tested on a participant with a significantly different height than those participants in the training data, however impact of height is not very clear in other activities.

The rest of the paper is organized as follows: In Section 2, we present the related studies that focus on smoking recognition with wearable devices. In Section 3, we present our methodology for analysis, including the dataset, feature sets, classification methods. In Section 4, we present the results of our analysis in terms of feature set, classifier performance, sensor or sensor-combination performance and user height, together with a discussion on the presented results. Finally, in Section 5, we present the conclusions and future work.

2. Related work

The common use of smartwatches provides an opportunity to realize human activity recognition on these devices and for the activity recognition process, there are many parameters to explore, such as the types of sensors, sampling rates, window sizes, features, and classifiers.

The authors in [19] focused on the detection of only the smoking activity using a wrist-worn accelerometer-based device. Smoking sessions were performed by four participants. The authors reported a precision of 51.2% and a person-dependent recall of 70%. Similarly, the authors in [28] used a sensor device attached to the wrist of three participants. The data was collected in a controlled environment and subjects were asked to repeatedly perform six different activities including smoking. Particularly for smoking, they utilized movement detection, including arm moving up followed by the arm moving down after taking the puff. In [2], a respiratory inductive plethysmography (RIP) sensor, which collects respiration data with inhalation and exhalation of smoking and is worn around the chest area, was used rather than a wrist-worn device. They use a dataset collected from 10 participants over 13 individual smoking sessions. In another study [8], authors introduced the PACT2.0 system that consists of an instrumented lighter, hand module, and chest module with an embedded data log capability. The lighter records the time and duration of lighting, the hand module includes inertial sensors for tracking hand-to-mouth gestures and the chest module monitors breathing patterns and cardiac activity (ECG sensor).

In [26], the authors use a dataset of 11.8 hours where six participants performed smoking activities, such as smoking while sitting, standing, eating, walking, using a phone, and talking in a group and they used two accelerometers at wrist position. They propose a two-layer model for automatic detection of puffs and smoking activities. They achieve an F1-score of 70% for puffing and 79% for smoking detection using person-dependent cross validation.

In [14], the smoking detection problem is studied using a wristband, containing three sensors, including accelerometer, gyroscope and magnetometer. They used a dataset collected from 15 participants for a total of 17 smoking, 10 eating, 6 drinking sessions. For smoking, they reported a precision of 91% and a recall of 81%.

In a recent study [3], the focus is on the successful detection of smoking events with an accelerometer sensor on a smartwatch. 120 hours of data is collected from ten participants. Using artificial neural networks, they aim to classify the raw data into two groups: smoking and non-smoking. They achieve an accuracy of approximately 90% for the smoking activity.

In [25], the authors evaluated a smartwatch-based system to detect smoking activity using accelerometer and gyroscope. Unlike other studies, this system does not require a connected smartphone and runs on a low-cost smartwatch. They performed a preliminary validation in a laboratory setting with 13 participants and free-living conditions. In this system, an instance of smoking is comprised of three movements which are hand raising to mouth, hand stationary at mouth, hand moving away from mouth. They report 86% precision and 71% recall in free-living conditions.

In another recent study [1], it is proposed to apply a convolutional neural network (CNN) on the same dataset as in our study. The data is divided into 3 groups according to the activities and as a result, they obtained an F1-score of 92–96%. Although the main focus of this paper is not on deep learning methods, in Section 4.3, we briefly discuss our findings when deep learning algorithms are applied on this dataset.

In [24], we proposed a two-layer hierarchical smoking detection algorithm (HLSDA) and analyzed its performance on the same dataset. Our aim was to see the impact of using a lazy context rule-based correction method that utilizes neighboring data segments on the performance of activity recognition. In that study, we only used 6 features, including mean, standard deviation, minimum, maximum, kurtosis, and skewness. We showed that using HLSDA increases the performance up to 11% in terms of F1-measure. In a recent study [23], we implemented the HLSDA algorithm on a smartwatch and analyzed its performance for online smoking recognition on the watch and the phone. Compared to[23,24], in this paper we explore all the parameters of the whole parameter space that may affect the recognition process, identify smoking in different postures rather than only detecting smoking and we also investigate the impact of smoker's height on the smoking recognition performance.

A detailed comparison of studies that mainly use wrist-worn sensors is summarized in terms of these parameters in Table 1 to show the differences of this study than those in the literature.[1] In some studies, very few features have been used. For example, in [19] only mean and variance, in [23] only min, max, mean and standard deviation are used as a feature set. However, in this paper, we extract 17 features from each four dimensions (x, y, z and, magnitude) of each of the sensors, as mentioned. There are also studies that investigate the use of different features extracted from each sensor [2,18,26]. But these studies use all the features in a single feature set and they did not create different feature combinations to better observe their effect on the performance of activity recognition. Most of the studies approached the problem as a binary classification problem, such as smoking and not, or few classes are targeted, however we are interested in classifying smoking, eating, drinking in different postures as well as other activities. We evaluated the performance of recognition with four sensors (the accelerometer, the linear acceleration, the gyroscope and the pitch-roll) individually and some of them in combination.

3. Methodology for smoking recognition

In this section the methodology for smoking recognition is explained, together with the characteristics of the collected dataset.

3.1. Dataset and preprocessing

45 hours of sensor data is collected with eleven participants including ten different activities. The data was collected mainly for the detection of smoking activity, however, as mentioned, it contains variations of smoking activity, other similar activities, and other activities. More explicitly, the activities are smoking while standing (smokeSD), smoking while sitting (smokeST), eating (eat), drinking while standing (drinkSD), drinking while sitting (drinkST), standing (stand), sitting (sitting), smoking while walking (smokeWalk), walking (walk) and smoking in a group conversation (smokeGroup). Each participant repeated the activities five times. All activities were performed for 5.23 hours except smoking while walking, walking and smoking in a group conversation were performed for 2.31, 2.31 and 4.17 hours, respectively. More details about the duration of activities can be found in [24].

During the collection of data, one smartphone and one smartwatch were used by every participant. The smartphones (Samsung Galaxy S2 or S3) were placed

[1] Sensors: A: Accelerometer, G: Gyroscope, LA: Linear Acceleration, M: Magnetometer, BMP:bioimpedance, PEDO:pedometer, PROX:proximity, RIP:respiratory inductive plethysmograph. Features: max: maximum, min: minimum, std: standard deviation, snr: signal to noise ratio, rms: root mean square, corr: correlation coefficient, mse: mean squared error, absdiff: absolute difference, erd: euclidean related distance, levens: levenstein distance. Classifiers: HLSDA: Hierarchical Smoking Detection Algorithm, RF: Random Forest, TB: Threshold Based, GMM: Gaussian Mixture Model, DT: Decision Tree, KNN: K Nearest Neighbors, SVM: Support Vector Machine, CRF: Conditional Random Field, DTa: Decision Table, NB: Naïve Bayes, MLP: Multi-Layer Perceptron, ANN: Artificial Neural Networks, RBAI: Rule-Based Artificial Intelligence. Activities: S: smoking, S_{ST}: smoking while sitting, S_{SD}: while standing, S_G: while in a group, S_W: while walking, S_E: while eating, S_D: while drinking, S_{UP}: while using phone, D: drinking, D_{ST}: drinking while sitting, D_{SD}: while standing, E: eating, ST: sitting, SD: standing, W: walking, T: giving a talk, WR: writing, TY: typing, J: jogging, B: biking, W_U: walking upstairs, W_D: walking downstairs, R: running, STR: stretching, SC: scrubbing, FL: folding laundry, BT: brushing teeth, R_{EL}: riding elevator, R_{ES}: riding escalator, W_{CI}: walking carrying items, WOC: working on computer, ED: eating or drinking, RD: reading, STT: strength-training, V: vacuuming, LD: lying down, CS: climbing stairs, TOC: typing on computer, TS: tying shoes, JA: jacks. NP: not provided.

Table 1
Comparative analysis of related work and our study

Ref.	Sensors	Window size-overlap	Features	Classifiers	Activities	Classes
[26]	A (2)	1, 3, 5, 7, 9, 15, 20, 25, 35 sec-50%	mean, std, max, min, median, kurtosis, skewness, percentile, snr, rms, peak-peak amplitude, peak rate, corr, crossing rate between axes, slope, mse, r-squared	RF, TB	S_{SD}, S_{ST}, S_G, S_W, S_E, S_D, S_{UP}	S, puffing
[24]	A, G	30 sec-0%	max, min, kurtosis, skewness	HLSDA	S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W	S, E, D, inactive
[19]	A	5.4 sec-NP	mean, variance	GMM	S_{SD}, *others*	S/Not
[28]	A, G	32*0:05 sec-50%	mean, max, std, peak-peak, RMS, corr	SVM	S_{SD}, W, SD, WR, JA, J	S_{SD}, W, SD, WR, JA, J
[8]	A, G, PED, ECG, GPS, RIP, PROX	13.63 sec-NP	NP	NP	S	S
[14]	A, G, M	20 sec-NP	duration, speed, distZ, distXY, dist, roll velocity, roll, pitch	RF, CRF	S_{SD}, S_{SG}, S_W, E, D, others	S, E
[3]	A	5 sec-NP	NP	ANN, RBAI	S, E, D, W, TOC, TS	S/not, abnormal, improper use
[25]	A, G	NP	NP	DT	S, E, D	S/Not
[1]	A, G	30 sec	max, min, skewness, kurtosis	CNN	S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, S_G, S_W	S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, S_G, S_W
[23]	A, G	30 sec-0%	mean, min, max, std	DT, RF, SVM, MLP	S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W	S/Not
This study	A, G, LA	20, 30 sec-0%, 50%	mean, std, skewness, kurtosis, min, max, range, integration, corr, rms, absdiff, spectral energy, entropy, coefficient sum, erd, levens	SVM, RF, MLP	S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W	S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W

in the pocket of their right pants and the smartwatches (LG Watch R, LG Watch Urbane or Sony Watch 3) on the wrist of the their dominant hand. During the treatment of this data, we used only the data collected from the smartwatch considering that purpose is to analyze the performance of smartwatch for activities where the hand movements are more significant. The utilized sensors are accelerometer, gyroscope and linear acceleration. The data was sampled at 50 Hz from all sensors. More details of the dataset are presented in Table 2.

We segment raw data into different time windows, then we compute different features for each segment. For the feature extraction phase, we use a sliding window approach. Since the window size and the overlap are important factors on continuous activity recognition, we consider four different cases as follows:

- Case 1: Window size of 20 seconds with a 0% overlap.
- Case 2: Window size of 30 seconds with a 0% overlap.
- Case 3: Window size of 20 seconds with a 50% overlap.
- Case 4: Window size of 30 seconds with a 50% overlap.

We tested smoking recognition with smaller or larger window sizes as well, however, recognition success was optimal for these window sizes (20 and 30 seconds).

Sensors provide information from three axes, x, y and z. We also added a fourth dimension called magnitude, which is the sum of the square root of the readings of the three axes. Magnitude is commonly used in activity recognition studies, to prevent the impact

Table 2
Participants and details of the collected dataset

Participant no	Activities performed	Activity duration (minutes)	Total duration (minutes)	Gender	Height (cm)	Age (years)	Cigarette usage
1	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W$	43	430	male	180	25	8–10 per day
2	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W$	47	470	male	172	30	0–10 per week
3	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G, S_W, W$	48	480	male	175	25	2–6 per day
4	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G$	37	296	male	156	28	0–10 per week
5	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G$	18	144	male	174	23	18–20 per day
6	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G$	20	160	female	164	20	3–7 per week
7	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G$	16.8	134.4	male	181	20	9–11 per day
8	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD, S_G$	20	160	female	172	29	4–6 per day
9	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD$	24	168	male	167	35	0–10 per week
10	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD$	19	133	male	181	27	7–12 per day
11	$S_{ST}, S_{SD}, D_{ST}, D_{SD}, E, ST, SD$	18.6	130.2	male	170	45	15–20 per day

of orientation on the sensor readings. We calculate the features from these four components. In addition, to better utilize the rotation information from smartwatches, we create a fourth sensor called pitch and roll. Raw accelerometer readings are used for computing pitch and roll values using the method in [4].

3.2. *Feature sets*

We calculate 17 features from each four dimensions (x, y, z and magnitude) of accelerometer (ACC), linear acceleration (LACC), gyroscope (GYR) and pitch and roll (PR) sensors from the segmented raw data. List of features that we compute is as follows:

– **Mean**: The average value of samples over a time window. It gives us a central value for a time window.
– **Standard deviation (std)**: The square root of the variance. It shows how much data sample is spread out around the mean and thus it gives an indication about the stability of sample [5].
– **Median**: Middle number of a sample. It divides the data sample in two parts: high half and lower half [5].
– **Skewness**: The measure calculated by lack of symmetry of data sample around its mean [24]. A sample is not symmetric if its distribution does not same to the left and right of the mean.
– **Kurtosis**: The measure of whether the data in the sample has a lot of or less data in its tails compared to normal distribution [24].
– **Min**: The minimum value of samples over a time window.
– **Max**: The maximum value of samples over a time window.
– **Range**: The difference between the maximum and the minimum of samples over a time window.
– **Integration**: The measure used to estimate the speed and distance of the signal under the data curve and this is commonly applied to accelerometer data [12].
– **Correlation**: Pearson's product-moment coefficient is the most commonly used correlation coefficient [17]. It measures the relationship between each pair of axis and can be applied for accelerometer or gyroscope readings [13]. It is effective to discriminate one dimensional activities such as walking and climbing stairs [30].
– **Root mean square (rms)**: The square root of the mean of the squares of data over a time window.
– **Absolute difference (absdiff)**: The sum of the differences between each data sample and the av-

Table 3
Feature sets

Set	Features	Domain
F1	min, max, skewness, kurtosis	Time
F2	mean, std, min, max	Time
F3	median, std, min, max, range, mean	Time
F4	mean, std, integration, correlation, rms, absdiff	Time
F5	spectral energy, entropy, coefficient sum	Frequency
F6	erd	String
F7	levenshtein	String

erage of sample divided by the number of data points [9].

- **Spectral energy**: The squared sum of spectral coefficients of signal over the length of the sample window[9].
- **Entropy**: The entropy metric can be roughly considered as frequency distribution which is high if the distribution is flat and low if peaky [11]. It helps to differentiate activities which have similar energy values but different activity patterns [5].
- **Sum of coefficients (coeffsum)**: The sum of the first five FFT coefficients.
- **Euclidean related distance (erd)**: The square root of the sum of the squares of the differences between corresponding data over a time window.
- **Levenshtein distance**: The measurement of similarity between two strings. It determines the smallest number of insertions, deletions, and substitutions needed to transform the first to the second [6].

Features are comprised of time, frequency and string domain features. These individual features have been reported to be suitable for running on mobile phones and wearables and have been extensively used in previous studies [5] on context recognition from motion data. Using all features together may be inefficient in terms of computation, particularly when running on a smart watch. Instead of using all the features, we organized seven different feature groups to better observe their effects. More details about feature sets and their domains are presented in Table 3. One may argue that, instead of grouping the features, feature selection methods could be applied. In another preliminary study, we worked with different feature selection algorithms and while grouping the features we used its findigs, keeping in mind to use a small number of features for reduced complexity, as also discussed in [23].

3.3. *Classifiers and validation*

There are several algorithms for classification that have been applied to activity recognition. Particularly, we used Support Vector Machine (SVM), Random Forest (RF) and Multilayer Perceptron(MLP) which are commonly used for activity recognition [21,24,29]. Scikit-learn implementations of the classifiers are used with the default settings and parameters. As the parameters of the RF algorithm, we used 11 trees (large number may increase the memory consumption), gini split, maximum depth none and two splits. For SVM, rbf kernel, 1.0 penalty parameter, 3 as the degree of the kernel function are used. For the MLP classifier, 1 hidden layer, constant learning rate, 200 as the maximum number of iterations are used.

In the validation phase, we realize an evaluation with 10-fold cross-validation without shuffling. In this method, the mechanism consists of dividing the dataset into ten equal parts; use nine of these parts for training and one part for testing. In each iteration, the part used for testing is different, thus, all data is used for testing and for training. By using stratified classification, every part has nearly the same length. We also used a person-independent evaluation method when we explore the impact of height on the recognition performance in Section 4.2.2.

4. Performance analysis

In this section, we present the results of our recognition analysis. As mentioned, we explore a large set of parameters: 4 different window sizes/overlaps, 7 feature sets extracted from each sensor, 4 different sensors and 2 different sensor combinations and 3 classifiers. First, in Section 4.1, we analyze the recognition performance when all activities are considered. We investigate the impact of window size and overlap, feature set and sensors. In Section 4.2, we exclude the activities of smoking while walking, walking and smoking in a group since they were not performed by all participants. Similarly, we investigate the effect of mentioned parameters. Moreover, we investigate the impact of users' height in the recognition phase by training with different user groups and using regression. As the performance metric, we report F-measure (F1-score) values which is the harmonic mean of precision and recall. We choose F-measure because it is a considered a balance performance metric by taking into account both recall and precision.

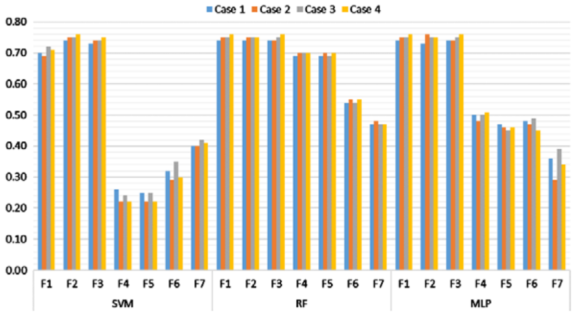

Fig. 1. Impact of window size and overlap using accelerometer.

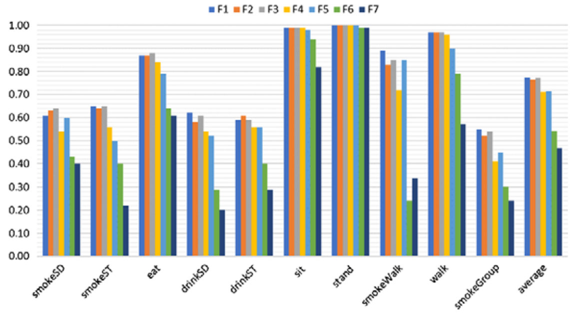

Fig. 2. Impact of feature set using accelerometer with RF and Case 4.

4.1. Scenario 1: All activities

In this section, we present the results obtained by following the methodology explained in Section 3 considering all ten activities. In the presented results, the F1-score values range between zero and one, but in the text, they are discussed in terms of percentages, for the ease of reading.

4.1.1. Impact of window size and overlap

In this section, we explore the impact of window size on the performance of classifiers. We change the window sizes and the ratio of window overlaps, according to the four cases explained in Section 3.1. We present and discuss the results obtained using accelerometer only in this section, however, the results with other sensors are also presented in the Appendix, Table 15.

In Fig. 1, the results using SVM, RF and MLP are presented to compare different cases which were introduced in Section 3. The y-axis shows the F1-score values obtained with different feature sets, whereas the x-axis shows these feature sets (shown in Table 3). When the results of different cases are compared, using Case 4 achieves the highest F1-score for all three classifiers, which is 76% for SVM using F2, for MLP using F1 and F3, for RF using F1 and F3. For these mentioned feature set and classifier combinations, there is only a small difference in other cases. The F1-scores obtained with Case 1, Case 2, and Case 3 are only 1–2% smaller than the results with Case 4. With other feature sets, namely F4 to F7, results obtained with all cases (1 to 4) are much lower, differing between 22% to 70%.

Furthermore, when we compare the performance of classifiers, RF is the best classifier, considering all feature sets. Particularly, the F1-scores of F4, F5 and F6 are 22%, 22% and 30% using SVM whereas it is 70%, 70% and 55% using RF, which results in 48%, 48% and 25% higher F1-score. Whereas, the results with other feature sets, particularly with F2 and F3, are similar with all the classifiers, ranging between 75 and 76% F1-score.

As mentioned, these results were obtained using only accelerometer. The F1-scores for the other sensors are presented in Appendix Table 15. Considering all cases, we observe that RF with Case 4 achieves the highest F1-score, followed by RF with Case 3. As mentioned, we tested smaller or larger window sizes as well, however recognition performance was lower. In the next sections, while evaluating the impact of feature set and sensors, we will continue to present and discuss the results obtained with Case 4 and the random forest classifier. However, all results are presented in the Appendix.

4.1.2. Impact of feature set

In this section, Case 4 is fixed using accelerometer with the random forest classifier. The F1-score results of the other classifiers with each feature set using accelerometer are presented in Appendix Table 16.

In Fig. 2, we present the F1-scores achieved per activity as well as the average F1-score considering all the activities. The highest performance for four activities (smokeST: 65% with F1 and F3; sit: 99% with F1 to F4; stand: 100% with F1 to F5; walk: 97% with F1 to F3) is obtained with more than one feature set. For the rest of activities, the best performance is achieved using F1 for drinkSD which is 62%, for smokeWalk (89%) and smokeGroup (55%), F2 for drinkST (61%) and F3 for smokeSD (64%) and for eat (88%). With other feature sets, relatively lower performances are achieved. The worst feature set for all activities is F7 which contains only a string domain feature (levenshtein distance). The average differences between first three feature sets and the rest is relatively high (16%). Although F1 and F3 results are very close to each other, F3 is slightly higher than F1. When we also

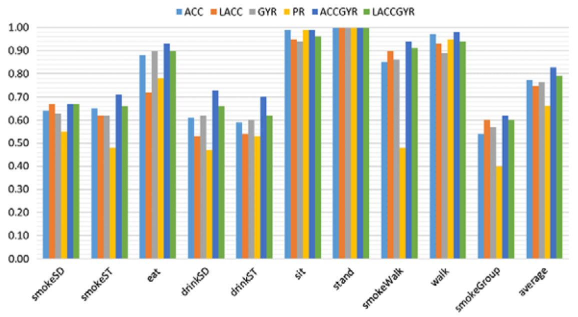

Fig. 3. Impact of sensors using F3 with RF and Case 4.

consider the performance with other classifiers in Table 16, by ranking the feature sets in terms of F1-score, F3 again ranks as the best feature set, which is followed by F1 and F2. Since, F3 includes the features in F2, but also has median and range, this was an expected result. Moreover, F3 and F1 include min and max in common but F3 has more features. If computing more features is a concern, then F1 or F2 can also be used with a small tradeoff in F1-score.

For the smoking variations, the performance of the smokeWalk is the highest (89%) using F1 and the smoke Group is the lowest 55% with F1. This is due to the fact that, smokeGroup is very similar to smokeSD. If we use both of these variations as one type of activity, the recognition performance improves as shown in one of our previous works where all these smoking variations were considered as one smoking activity [24]. In smokeGroup sessions, all smokers didn't talk or move their hand too much. This makes the smokeGroup activity very similar to smokeSD.

4.1.3. *Impact of sensors and fusion of sensors*

In this section, impact of individual sensors (ACC, LACC, GYR and PR), combination of accelerometer and gyroscope (ACCGYR) and combination of linear acceleration and gyroscope (LACCGYR) are analyzed. For this evaluation, RF, Case 4 and F3 are fixed. In Appendix Table 17, the F1-score results for all feature sets with all classifiers are presented.

In Fig. 3, we present the results per activity and average of all activities. The highest smoking performance is 94%, obtained for the smokeWalk using ACCGYR whereas it is 100% for stand and sit activities. In general, the best performances are obtained with the fusion of accelerometer and gyroscope for all activities. Considering all activities, on average, using only accelerometer exhibits a performance of (77%) and only gyroscope (76%). Combining accelerometer and gyroscope improves the average performance of activity recognition (83%). Similarly, using fusion of linear acceleration and gyroscope increases performance compared to only using linear acceleration and only gyroscope. The worst performance is obtained with pitch and roll features which is 66% on average. As shown in Appendix Table 17, using accelerometer gyroscope combination achieves the highest score considering the F1, F2 and F3 cases and the RF and MLP classifiers, which is followed by the combination of linear acceleration and gyroscope combination and this is followed by using only the accelerometer.

Example confusion matrices for only accelerometer sensor is given in Table 4 and for the fusion of accelerometer and gyroscope is given in Table 5. We use F3 and Case 4 in both these tables. We observe that smoking and drinking activities are confused with each other. However, two activities that are mostly confused with each other are smokeSD and smokeGroup. This may be due to the fact that people do not talk much in the group while collecting smokeGroup data and do not actively use hand movements. This may cause the activity to be confused with smoking while standing. We have previously observed higher recognition performance for smoking when its different variations were considered as one smoking activity [24]. By comparing Table 4 and Table 5, as mentioned, we observe that the use of fusion of accelerometer and gyroscope reports an increase, compared to use of only accelerometer except stand and sit activities which are already well recognized.

As a conclusion, fusion of sensors improves the recognition performance, however, resource consumption due to additional sensors increases [22]. Hence, the trade-off between high recognition rate and high resource consumption should be further investigated since battery lifetime is a limitation with smartwatches.

4.2. *Scenario 2: Less activities*

As mentioned, not all activities were performed by all eleven participants. Activity smokeWalk, walk and smoke Group were performed in total only by 3, 3 and 8 participants, respectively. In order to look at the common smoking variations of smoking while sitting and standing, we removed the smoking in a group and smoking while walking and created a second scenario. This also allowed us to create a balanced dataset where all these activities were performed by all participants. Thus, in Scenario 2, we consider all participants, but we do not include smoking while walking, smoking

Table 4
Confusion matrix using ACC with F3, RF and Case 4

		Predicted									
		smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand	smokeWalk	walk	smokeGroup
actual	smokeSD	**836**	43	0	36	4	0	1	18	0	318
	smokeST	79	**792**	15	59	239	3	0	1	0	66
	eat	0	11	**1132**	80	28	0	0	0	2	2
	drinkSD	50	88	120	**744**	213	0	4	3	3	29
	drinkST	9	206	48	233	**732**	19	0	1	0	7
	sit	0	1	0	1	0	**1253**	0	0	0	0
	stand	0	0	0	0	0	0	**1253**	1	0	1
	smokeWalk	29	1	0	8	0	0	0	**458**	7	51
	walk	0	0	0	10	0	0	0	8	**534**	1
	smokeGroup	339	54	5	12	5	0	0	35	1	**550**

Table 5
Confusion matrix using ACCGYR with F3, RF and Case 4

		Predicted									
		smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand	smokeWalk	walk	smokeGroup
actual	smokeSD	**860**	54	0	42	4	0	1	0	0	295
	smokeST	102	**856**	9	32	206	4	0	0	0	45
	eat	0	5	**1164**	79	5	0	0	0	0	2
	drinkSD	57	45	45	**917**	150	0	3	1	1	35
	drinkST	2	175	18	163	**871**	24	0	0	0	2
	sit	0	0	0	0	1	**1254**	0	0	0	0
	stand	1	0	0	0	0	0	**1253**	0	0	1
	smokeWalk	4	0	0	5	1	0	0	**520**	9	15
	walk	0	0	0	4	1	0	0	10	**537**	1
	smokeGroup	292	30	8	17	1	0	0	19	0	**634**

while in group conversation, and only walking. We evaluate this scenario by choosing the best case (Case 4 and F3) that is explored in Section 4.1.

4.2.1. *Impact of sensor fusion and classifiers*

In this section, our aim is to analyze the effect of sensors and classifiers using Case 4 and F3 to better understand if focusing on less activities improves the performance of recognition.

We present the results in Fig. 4 for all sensors. This scenario brings significant improvement for the performance of smokeSD which is an increase of 23%, 22% and 21% using ACC, PR and ACCGYR and an increase of 12%, 13% and 17% using LACC, GYR and LACCGYR, compared to the previous scenario whose results are shown in Fig. 3. Removing some variations of smoking activities such as smokeGroup improves the performance of smokeSD. For the smokeST, eat and drinkSD, it performs equally or slightly better in a range of 5% and 1%.

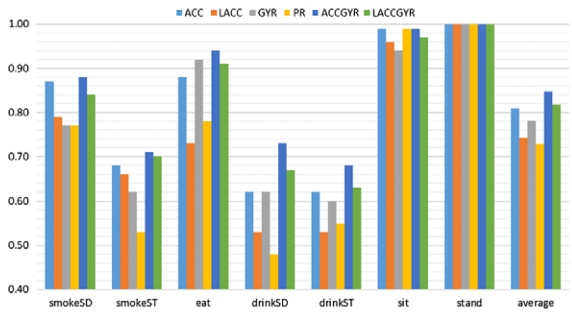

Fig. 4. Impact of sensors using F3 with RF and Case 4.

On average, the best performances are obtained with the fusion of accelerometer and gyroscope for all activities. Similar to Scenario 1, the worst performance on average of all activities is obtained again using pitch and roll sensors (73%). Using only accelerometer gives a performance of 81% and only gyroscope 78%, which were 77% and %76 in Scenario 1. Combining accelerometer and gyroscope improves the av-

Table 6
Confusion matrix using ACC with F3, RF and Case 4

		Predicted class						
		smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
actual class	smokeSD	**1131**	80	0	44	0	0	1
	smokeST	123	**843**	14	65	207	2	0
	eat	0	16	**1116**	87	35	1	0
	drinkSD	87	86	109	**767**	200	1	4
	drinkST	6	197	39	239	**759**	15	0
	sit	0	0	0	0	3	**1252**	0
	stand	0	0	0	0	0	0	**1254**

Table 7
Confusion matrix using ACCGYR with F3, RF and Case 4

		Predicted class						
		smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
actual class	smokeSD	**1137**	80	1	33	4	0	1
	smokeST	120	**873**	4	37	215	5	0
	eat	3	4	**1167**	75	6	0	0
	drinkSD	68	63	53	**903**	163	0	4
	drinkST	3	184	14	185	**849**	20	0
	sit	1	0	0	0	2	**1252**	0
	stand	0	0	0	0	0	0	**1254**

erage performance of activity recognition (85%). Similarly, using fusion of linear acceleration and gyroscope (82%) increases performance compared only linear acceleration (74%) and only gyroscope (78%).

In Table 6 and Table 7, we show the confusion matrices of accelerometer and fusion of accelerometer and gyroscope respectively. Once again, it can be seen that mainly smoking and drinking are confused with each other. We observe less confusion for eat, sit and stand activities. Compared to Scenario 1, we see a significant decrease in the number of confused smokeSD activities. As mentioned in Scenario 1, smokeSD was mainly confused with smokeGroup activity. In this scenario, we removed smokeGroup, so this change positively affects the recognition of smokeSD. This shows that if smoking while standing and smoking while in group conversation are considered as one variation, it can easily be recognized. We had observed a similar result previously in [24].

To understand the best classifier among SVM, RF and MLP, in Table 8, we show the results of three classifiers as well as their improvements compared to Scenario 1. We observe that RF is the best classifier with an average performance of 76%. For the best sensor which is ACCGYR, the performance of corresponding classifier is 83%. There is not so much difference observed between RF and MLP. Performance of SVM is slightly lower compared to the other two. Our approach improves the recognition performance and the highest improvement is observed for PR sensor which is around 7%.

4.2.2. *Impact of height*

After collecting the dataset, some extra questions were asked to the participants. These were about how often they smoke, their height and their age, as presented in Table 2.

Depending on the participants' height, the frequency and patterns of the performed activities may vary. For example, while a shorter person can take his hand to his mouth faster, it may take longer for someone who is taller. Starting from this point of view, we tried to determine whether the height of participants really affects the activity recognition process. Weight may affect activities related to locomotion, but in our case the activities are related to arm movements, and hence the height of the arm. Gender may affect the results but according to our observations while data collection, there were not significant differences between male and female participants in their smoking patterns. Therefore, in this section, we explore whether the height of participants impacts the performance of smoking recognition or not. For this purpose, among

Table 8
Increase in F1-scores compared to Scenario 1

	SVM	Increase	RF	Increase	MLP	Increase
ACC	0,76	0,03	0,79	0,03	0,79	0,02
LACC	0,66	0,01	0,70	0,01	0,69	0,00
GYR	0,72	0,02	0,76	0,03	0,72	0,01
PR	0,45	0,09	0,71	0,06	0,65	0,06
ACCGYR	0,75	0,03	0,83	0,03	0,82	0,02
LACCGYR	0,70	0,02	0,78	0,05	0,77	0,00
average	0,67	0,03	0,76	0,03	0,74	0,02

the eleven participants, we have created two separate groups consisting of participants with closer heights. Besides this, to balance the amount of data in groups, we determined that each group should contain an equal number of participants. The heights of participants in the first group (G1) are 180 cm, 181 cm and 181 cm for participant 1 (P1), P7 and P10, respectively. In the second group (G2), the heights are 164 cm, 167 cm and 170 cm for P6, P9 and P11, respectively. As mentioned in Section 4.2, all activities are not performed by every participant, we analyze the activities that were common to these six. Based on the results shown in Section 4.1 and 4.2, we performed our analysis with Case 4, feature set 3, accelerometer and gyroscope combination (ACCGYR) and RF classifier.

In the first phase, we perform in-group analysis which means the aim is to find the performance of each participant in his own group. More clearly, we train the data of G1 and G2 separately to create two models, then we test each participant of G1 with G1's model and same for G2.

In Table 9, only the results with the fusion of accelerometer and gyroscope are presented for ease of presentation and to focus on the impact of the height. If we compare the groups, all participants have over 90% performance for all activities except drinkSD and drinkST. The performance of G2 (P6, 9 and 11) is better than G1 (P1, 7 and 10), particularly for participant 6 and 9. Generally, in in-group analysis, there is no participant with a performance less than 80% for any activity. F1-score results of all sensors for six participants are presented in Appendix Table 18.

In the second phase, we perform the out-group analysis. We test each participant with the training data from the other group rather than their own group. For example, P1 belongs to group 1, we test his performance by using group 2 (G2) as the training set. In Table 10, the results of ACCGYR (fusion accelerometer and gyroscope) are presented. We observe that the results are slightly different or not at all for some participants. Again, the participant 11 has the worst results for all activities except for drinkST. As observed in the first phase, the performance of G2 is better than, especially for participant 6 and 9 which have an average performance of 95% for both. All the sensor results are presented in Appendix Table 19.

The average performance difference between in and out group analysis does not exceed 2% for all participants. Even if we look at the activities of individual participants, the performance difference never exceeds 5%, but usually is either 0% or 1%. Based on this, we find that using different height groups does not significantly affect the performance results.

Additionally, we decided to analyze an extreme case particularly using G1 and G2 as the training set, and we chose participant 4 as the outlier, with height of 156 cm, to test. Recognition performances with P4 are presented in Table 11. Comparing training set G1 and G2, we did not observe a difference over than 1% for all activities except for the smoking activities. On average, considering in-group and out-group analysis, the smoking performances were between 91% and 96%. However, in this analysis, for smokeSD and smokeST, we observe a performance of 86% and 83% using G1, and 83% and 78% using G2. As the height differences are too much between P4 and the two groups, it might affect the performance of smokeSD. Because in these two cases, usually, participants take their hand really down and then it comes to the mouth where the angular distance might make a difference. However, smokeST is a way complex activity. Because the hand to mouth angular distance can take many shapes depending how someone smokes. We can report that height does affect the smokeSD if it is significantly different as the case of P4. However, smokeST can be affected by the different variations of activity.

4.2.3. Per-participant analysis

In the last phase, we did a per participant analysis. The aim is to explore the performance of leave-out-one-subject method for a specific individual. The model is trained using all participants' data except a participant X and then, it is tested on the data of participant X. Moreover, we investigate whether the height of participants has an impact on the recognition performance using statistical analysis. For this purpose, for each of the eleven participants, we train the data of the rest ten participants together for creating a classification model. Then, we test each participant with a particular height, separately, using this model. For example, for P1, we create training model using data of P2

Table 9
Impact of height in-group using ACCGYR, F3, RF and Case 4

	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
P1	0.97	0.93	0.96	0.81	0.84	0.96	1
P6	0.94	0.9	0.96	0.95	0.94	0.95	0.99
P7	0.92	0.86	0.96	0.84	0.82	0.94	0.99
P9	0.99	0.93	0.97	0.94	0.92	0.97	1
P10	0.95	0.91	0.96	0.88	0.84	0.93	1
P11	0.93	0.9	0.89	0.8	0.81	0.91	0.99
avg	0.95	0.91	0.95	0.87	0.86	0.94	1

Table 10
Impact of height out-group using ACCGYR, F3, RF and Case 4

	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
P1	0.96	0.9	0.96	0.83	0.85	0.96	1
P6	0.96	0.95	0.92	0.93	0.92	0.95	0.99
P7	0.95	0.91	0.95	0.82	0.81	0.94	1
P9	0.99	0.92	0.97	0.93	0.9	0.96	0.99
P10	0.95	0.94	0.96	0.89	0.88	0.95	0.99
P11	0.93	0.9	0.9	0.82	0.85	0.94	1
avg	0.96	0.92	0.94	0.87	0.87	0.95	1

Table 11
Impact of height for P4 using G1 and G2 separately as training data with ACCGYR, F3, RF and Case 4

	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
Gr1	0.86	0.83	0.95	0.85	0.85	0.97	1
Gr2	0.83	0.78	0.96	0.86	0.86	0.98	1

Table 12
Significance-F results

Activity	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
Significance	0.29	0.09	0.61	0.04	0.24	0.09	0.95

to P11 and we test this model on P1's data. Firstly, F1-score results of all sensors were obtained for all participants (see Appendix Table 20). Then, we explore whether the participants' activity performances change significantly based on the height parameter. To examine this, we used regression analysis with the default confidence level which is 95%. After creating the regression model for each activity, particularly, we focused on the significance-F value. In Table 12, we presented the significance-F results of regression model for all activities based on height parameter. It shows that, all generated regression results are not significant for all the considered activities except drinkSD. The significance-F value of drinkSD is lower than 0.05, which means that statistically, there is a relation between the participants' height and the recognition performance of drinking while standing activity.

In Table 13, we also present the detailed analysis of the performance differences between per participant analysis (Table 20) and cross validation analysis (Section 4.2) considering Scenario 2. In Scenario 2, cross validation results do not contain performance per participant, we use average F1-scores of this analysis to make comparisons. It is clear that, cross validation analysis performs better for all activities except sit activity with LACC, GYR and LACCGYR.

Table 13
Performance difference between per participant validation and cross validation results

	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
ACC	−0.07	−0.14	−0.05	−0.10	−0.08	−0.01	0.00
LACC	−0.05	−0.16	−0.10	−0.09	−0.09	0.01	0.00
GYR	−0.03	−0.04	−0.03	−0.07	−0.05	0.01	−0.01
ACCGYR	−0.08	−0.07	−0.05	−0.12	−0.03	−0.01	0.00
LACCGYR	−0.07	−0.12	−0.02	−0.10	−0.08	0.00	0.00

Particularly for smokeST and drinkSD activities, this approach achieves an important decrease in the performance which is in a range of 4% and 16%. As the performance of stand was very high which is almost 100%, it is not possible to achieve a change for this activity. However, we observed a decrease of 1% for gyroscope sensor.

4.3. Discussion

In this section, we summarize our findings and identify the open issues for further investigation.

- *Simple features perform sufficiently well*: When we analyze the performance of recognition with different feature sets, we observe that F1 (min, max, skewness, kurtosis) and F3 (median, std, min, max, range, mean) perform the best. These features are all time domain features and can be computed easily. If smoking recognition is performed online on watches or on other wearables [16], this will be an advantage.
- *Larger windows perform better*: Compared to shorter window sizes, 1 to 5 seconds, used in the recognition of simple activities, such as walking, sitting, larger window sizes perform better for smoking recognition. In our analysis, Case 4 (30 second window size with 50% overlap) performed slightly better than the other cases. While simpler locomotion activites follow shorter periods, such as 1–2 seconds, in smoking, hand-to-mouth gesture, inhalation, mouth-to-hand and breaks between these patterns take longer periods. This is also true for eating and drinkin activities.
- *When efficient features are used, performance of classifiers is similar*: In our analysis, RF classifier is the best performing classifier in most cases, which is followed by MLP and then SVM. However, when efficient feature sets are used, such as F1 or F3, their performances are very similar. Although we did not change or optimize the parameters of the classifiers, their performances are acceptable, for example in Scenario 2, when only accelerometer is used average F1-scores for SVM, RF and MLP are 80%, 81% and 80%, respectively. Again, if online recognition is to be performed, in another study [22], we show that these classifiers perform well in terms of resource consumption, such as battery, memory and CPU cycles.
- *Combination of gyroscope with accelerometer improves the results*: The best performances are obtained with the fusion of accelerometer and gyroscope for all activities. Considering all activities, using only accelerometer exhibits a performance of (77%) on average and only gyroscope (76%). Combining accelerometer and gyroscope improves the average performance of activity recognition to 83% on average, considering all the activities. However, when a single sensor is to be used, then the accelerometer is the best performing one as shown in the literature [21]. Fusion of linear acceleration and gyroscope increases performance compared to only using linear acceleration and only gyroscope as well, however the results are not as high as the combination of accelerometer and gyroscope. We should also note that linear acceleration sensor consumes more battery than accelerometer.
- *Puff Detection versus smoking session detection*: In this paper, we focus on smoking session detection, however, there are some studies in the literature using wrist sensors for puff detection [3,20,26,28]. The sensor data collected in our study using smartwatch devices may be sufficient to perform this puff analysis but we need also to have puff labels (such as the number or start-end time of puffs) to test the performance of the analysis and these were not logged during the data collection phase.
- *User's height may affect the performance of smoking and drinking while standing*: While there is no significant difference between in group and

Table 14
Preliminary results for LSTM with accelerometer and gyroscope

Epoch	Batch size	Final accuracy
30	512	0.91509914
30	1024	0.9142906
30	2048	0.9030981

out group analysis, the recognition performances of smoking while standing and sitting are reduced when tested on a participant with a different height than those participants in the training data. Besides, significance-F values of per participant analysis showed that drinking while standing activity also may be affected by height. On the other hand, impact of height is not very clear in other activities.

- *Smoking in group is more difficult to recognize*: Compared to the other variants, smoking in a group is more difficult to recognize due to different patterns exhibited by the participants and it is mostly confused with smoking while standing. This can be further investigated with the use of both phone and watch data to obtain better performance.
- *Use of deep learning algorithms may improve the results*: In a recent study [1], a convolutional neural network (CNN) was applied on the same dataset as in our study and 92–96% F1-score was achieved. Although the main focus of this paper is not on deep learning methods, we also applied a deep learning method, namely, Long Short Term Memory (LSTM) on the dataset, we show the preliminary results in Table 14 using accelerometer and gyroscope readings. More detailed results will be presented in a paper which is under preparation. We briefly discuss our findings when deep learning algorithms are applied on this dataset. 70% of the dataset was used for training, 15% for validation and 15% for testing. The model was constructed using Tensorflow library[2] and Keras Framework[3] and number of hidden layers was 32 and learning rate was 0.0025. Unlike in [1], we did not extract any features and did not divide the data into groups. Our preliminary results presented in Table 14 suggest that, use of deep learning algorithms may improve the results, however training and running such models is costly on resource-limited devices, such as smart watches, hence a cloud-supported model could be a viable solution.

[2] https://www.tensorflow.org/
[3] https://keras.io/

5. Conclusion and future work

In this paper, we studied the recognition of smoking activity in different postures by using the motion sensors available on smartwatches using a challenging dataset which includes different variations of smoking as well as activities including similar hand gestures to smoking. We followed a detailed analysis: with a focus on using different classifiers, different and comprehensive set of features, different window sizes and different window overlap ratios. As the final parameter, we investigated the impact of height in the training phase on smoking recognition performance. The results show that smoking activities can be recognized with simple features, such as median, std, min, max, range, mean. Compared to smaller window sizes used in the recognition of simpler activities, larger window sizes, such as 30 seconds, perform better. When we compare the performance of different classifiers, when efficient features are used, their performances are similar. Smoking in a group is more difficult to recognize compared to other variants of smoking due to different patterns exhibited by different participants. When we analyze the impact of height on smoking recognition, it does not have a significant effect when all activities are considered. However, it does have an effect on drinking while standing activity based on statistical results. Currently, we are working on the analysis of online smoking recognition where we utilize the findings from this paper and analyze the resource consumption of different parameter sets on the wearable devices besides the recognition performance.

Acknowledgements

This work is supported by Tubitak under Grant number 117E761 and by the Galatasaray University Research Fund under Grant Number 17.401.004.

Appendix

In Tables 15, 16, 17, 18, 19, 20, each cell shows the F1-score performance of a specific feature-set with a specific classifier in a specific case, using a specific sensor.

Table 15
Scenario 1: F1 scores considering all cases

		Case 1			Case 2			Case 3			Case 4		
		svm	rf	mlp	svm	rf	mlp	svm	rf	mlp	svm	rf	mlp
ACC	F1	0.7	0.74	0.74	0.69	0.75	0.75	0.72	0.75	0.75	0.71	0.76	0.76
	F2	0.74	0.74	0.73	0.75	0.75	0.76	0.75	0.75	0.75	0.76	0.75	0.75
	F3	0.73	0.74	0.74	0.74	0.74	0.74	0.74	0.75	0.75	0.75	0.76	0.76
	F4	0.26	0.69	0.5	0.22	0.7	0.48	0.24	0.7	0.5	0.22	0.7	0.51
	F5	0.25	0.69	0.47	0.22	0.7	0.46	0.25	0.69	0.45	0.22	0.7	0.46
	F6	0.32	0.54	0.48	0.29	0.55	0.47	0.35	0.54	0.49	0.3	0.55	0.45
	F7	0.4	0.47	0.36	0.4	0.48	0.29	0.42	0.47	0.39	0.41	0.47	0.34
LACC	F1	0.6	0.65	0.65	0.6	0.66	0.65	0.63	0.67	0.67	0.62	0.68	0.67
	F2	0.57	0.7	0.67	0.59	0.71	0.69	0.6	0.71	0.7	0.6	0.72	0.7
	F3	0.59	0.71	0.67	0.58	0.71	0.67	0.61	0.71	0.69	0.6	0.73	0.68
	F4	0.3	0.69	0.5	0.27	0.7	0.48	0.32	0.7	0.54	0.28	0.71	0.51
	F5	0.27	0.64	0.47	0.24	0.65	0.44	0.27	0.65	0.47	0.24	0.66	0.45
	F6	0.46	0.52	0.51	0.45	0.52	0.52	0.49	0.52	0.5	0.46	0.53	0.52
	F7	0.38	0.43	0.22	0.4	0.45	0.17	0.38	0.44	0.31	0.41	0.45	0.22
GYR	F1	0.66	0.71	0.69	0.66	0.72	0.7	0.69	0.72	0.7	0.68	0.74	0.71
	F2	0.58	0.72	0.68	0.61	0.73	0.69	0.6	0.73	0.7	0.62	0.74	0.71
	F3	0.58	0.72	0.7	0.61	0.73	0.71	0.6	0.73	0.72	0.62	0.75	0.71
	F4	0.34	0.68	0.54	0.32	0.7	0.43	0.35	0.69	0.57	0.33	0.71	0.49
	F5	0.29	0.68	0.4	0.25	0.69	0.42	0.29	0.7	0.44	0.26	0.7	0.4
	F6	0.51	0.56	0.52	0.5	0.57	0.55	0.52	0.56	0.53	0.53	0.57	0.54
	F7	0.35	0.42	0.25	0.37	0.41	0.24	0.35	0.42	0.27	0.37	0.42	0.25
PR	F1	0.39	0.63	0.6	0.36	0.64	0.6	0.39	0.64	0.6	0.37	0.64	0.6
	F2	0.28	0.64	0.61	0.27	0.65	0.6	0.3	0.65	0.63	0.27	0.65	0.62
	F3	0.27	0.65	0.61	0.26	0.66	0.6	0.29	0.66	0.62	0.27	0.66	0.61
	F4	0.26	0.6	0.24	0.22	0.6	0.22	0.27	0.6	0.27	0.22	0.61	0.23
	F5	0.24	0.6	0.22	0.21	0.61	0.24	0.25	0.6	0.26	0.21	0.61	0.24
	F6	0.36	0.37	0.11	0.34	0.35	0.08	0.37	0.36	0.1	0.36	0.35	0.1
	F7	0.17	0.29	0.12	0.17	0.29	0.09	0.18	0.29	0.14	0.17	0.28	0.11

Table 16
Scenario 1: F1 scores of each feature set using ACC and Case 4

		F1	F2	F3	F4	F5	F6	F7
SVM	smokeSD	0.60	0.63	0.62	0.24	0.24	0.15	0.42
	smokeST	0.55	0.65	0.63	0.00	0.00	0.26	0.22
	eat	0.84	0.90	0.89	0.00	0.12	0.10	0.63
	drinkSD	0.60	0.66	0.65	0.00	0.00	0.04	0.09
	drinkST	0.48	0.62	0.61	0.00	0.00	0.07	0.21
	sit	0.93	0.99	0.99	0.64	0.56	0.90	0.66
	stand	1.00	1.00	1.00	0.99	0.99	1.00	0.97
	smokeWalk	0.86	0.75	0.74	0.00	0.00	0.00	0.08
	walk	0.88	0.93	0.91	0.00	0.00	0.08	0.60
	smokeGroup	0.46	0.49	0.48	0.00	0.00	0.04	0.12
RF	smokeSD	0.61	0.63	0.64	0.54	0.60	0.43	0.40
	smokeST	0.65	0.64	0.65	0.56	0.50	0.40	0.22
	eat	0.87	0.87	0.88	0.84	0.79	0.64	0.61
	drinkSD	0.62	0.58	0.61	0.54	0.52	0.29	0.20
	drinkST	0.59	0.61	0.59	0.56	0.56	0.40	0.29
	sit	0.99	0.99	0.99	0.99	0.98	0.94	0.82
	stand	1.00	1.00	1.00	1.00	1.00	0.99	0.99
	smokeWalk	0.89	0.83	0.85	0.72	0.85	0.24	0.34
	walk	0.97	0.97	0.97	0.96	0.90	0.79	0.57
	smokeGroup	0.55	0.52	0.54	0.41	0.45	0.30	0.24
MLP	smokeSD	0.65	0.63	0.65	0.37	0.29	0.39	0.34
	smokeST	0.63	0.62	0.66	0.41	0.28	0.34	0.17
	eat	0.88	0.91	0.89	0.57	0.49	0.37	0.58
	drinkSD	0.60	0.61	0.61	0.25	0.21	0.14	0.10
	drinkST	0.57	0.57	0.59	0.40	0.42	0.34	0.16
	sit	0.96	0.97	0.97	0.91	0.88	0.87	0.53
	stand	0.99	0.99	0.99	0.92	0.90	0.94	0.82
	smokeWalk	0.91	0.83	0.85	0.16	0.22	0.15	0.08
	walk	0.97	0.96	0.96	0.72	0.58	0.63	0.43
	smokeGroup	0.60	0.49	0.55	0.26	0.22	0.26	0.01

Table 17
F1 score of each sensor and fusion of sensors using Case 4

		ACC	LACC	GYR	PR	ACCGYR	LACCGYR
SVM	F1	0.71	0.62	0.68	0.37	0.69	0.65
	F2	0.76	0.6	0.62	0.27	0.78	0.68
	F3	0.75	0.6	0.62	0.27	0.78	0.68
	F4	0.22	0.28	0.33	0.22	0.22	0.26
	F5	0.22	0.24	0.26	0.21	0.22	0.24
	F6	0.3	0.46	0.53	0.36	0.27	0.47
	F7	0.41	0.41	0.37	0.17	0.54	0.51
RF	F1	0.76	0.68	0.74	0.64	0.8	0.77
	F2	0.75	0.72	0.74	0.65	0.8	0.77
	F3	0.76	0.73	0.75	0.66	0.81	0.78
	F4	0.7	0.71	0.71	0.61	0.78	0.77
	F5	0.7	0.66	0.7	0.61	0.78	0.75
	F6	0.55	0.53	0.57	0.35	0.71	0.68
	F7	0.47	0.45	0.42	0.28	0.6	0.55
MLP	F1	0.76	0.67	0.71	0.6	0.8	0.75
	F2	0.75	0.7	0.71	0.62	0.8	0.77
	F3	0.76	0.68	0.71	0.61	0.81	0.76
	F4	0.51	0.51	0.49	0.63	0.6	0.56
	F5	0.46	0.45	0.4	0.24	0.51	0.52
	F6	0.45	0.52	0.54	0.1	0.29	0.66
	F7	0.34	0.22	0.25	0.11	0.44	0.33

Table 18
F1-scores of in-group analysis using F3, RF and Case 4

	Sensor	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
P1	ACC	0.95	0.89	0.90	0.73	0.85	0.96	1.00
	LACC	0.88	0.82	0.81	0.70	0.65	0.89	1.00
	GYR	0.84	0.78	0.90	0.67	0.61	0.82	1.00
	ACCGYR	0.97	0.93	0.96	0.81	0.84	0.96	1.00
	LACCGYR	0.87	0.80	0.92	0.74	0.66	0.87	0.99
P6	ACC	0.98	0.96	0.92	0.90	0.87	0.95	0.99
	LACC	0.92	0.86	0.86	0.94	0.92	0.87	0.99
	GYR	0.82	0.69	0.88	0.84	0.84	0.87	0.99
	ACCGYR	0.94	0.90	0.96	0.95	0.94	0.95	0.99
	LACCGYR	0.86	0.83	0.89	0.88	0.89	0.93	1.00
P7	ACC	0.94	0.91	0.88	0.82	0.81	0.93	1.00
	LACC	0.89	0.73	0.71	0.76	0.74	0.89	0.98
	GYR	0.83	0.79	0.92	0.72	0.70	0.78	0.95
	ACCGYR	0.92	0.86	0.96	0.84	0.82	0.94	0.99
	LACCGYR	0.84	0.83	0.96	0.87	0.82	0.92	0.99
P9	ACC	0.99	0.87	0.98	0.96	0.89	0.96	1.00
	LACC	0.92	0.85	0.85	0.65	0.71	0.80	1.00
	GYR	0.88	0.86	0.97	0.79	0.71	0.71	0.98
	ACCGYR	0.99	0.93	0.97	0.94	0.92	0.97	1.00
	LACCGYR	0.90	0.85	0.95	0.79	0.71	0.74	1.00
P10	ACC	0.94	0.85	0.96	0.84	0.80	0.94	1.00
	LACC	0.90	0.83	0.91	0.70	0.77	0.87	1.00
	GYR	0.93	0.87	0.97	0.87	0.78	0.83	0.97
	ACCGYR	0.95	0.91	0.96	0.88	0.84	0.93	1.00
	LACCGYR	0.93	0.90	0.97	0.87	0.79	0.86	0.99
P11	ACC	0.93	0.95	0.92	0.84	0.90	0.94	1.00
	LACC	0.80	0.62	0.77	0.62	0.69	0.70	0.99
	GYR	0.84	0.72	0.76	0.55	0.63	0.69	0.97
	ACCGYR	0.93	0.90	0.89	0.80	0.81	0.91	0.99
	LACCGYR	0.82	0.72	0.85	0.68	0.68	0.76	1.00

Table 19
F1-scores of out-group analysis using F3, RF and Case 4

	Sensor	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
P1	ACC	0.97	0.90	0.88	0.72	0.83	0.94	1.00
	LACC	0.91	0.84	0.79	0.70	0.63	0.87	1.00
	GYR	0.84	0.77	0.93	0.68	0.62	0.85	0.99
	ACCGYR	0.96	0.90	0.96	0.83	0.85	0.96	1.00
	LACCGYR	0.88	0.82	0.93	0.77	0.70	0.86	1.00
P6	ACC	0.97	0.93	0.94	0.94	0.93	0.95	0.99
	LACC	0.88	0.80	0.89	0.90	0.92	0.88	0.99
	GYR	0.84	0.75	0.88	0.86	0.88	0.83	0.97
	ACCGYR	0.96	0.95	0.92	0.93	0.92	0.95	0.99
	LACCGYR	0.87	0.84	0.93	0.92	0.91	0.95	1.00
P7	ACC	0.95	0.87	0.88	0.83	0.82	0.95	1.00
	LACC	0.89	0.91	0.75	0.77	0.74	0.92	0.99
	GYR	0.79	0.75	0.92	0.78	0.71	0.77	0.95
	ACCGYR	0.95	0.91	0.95	0.82	0.81	0.94	1.00
	LACCGYR	0.90	0.88	0.94	0.88	0.83	0.91	0.99
P9	ACC	0.99	0.88	0.96	0.95	0.89	0.96	1.00
	LACC	0.92	0.84	0.91	0.74	0.73	0.72	1.00
	GYR	0.92	0.86	0.96	0.80	0.75	0.72	0.97
	ACCGYR	0.99	0.92	0.97	0.93	0.90	0.96	0.99
	LACCGYR	0.89	0.82	0.96	0.79	0.75	0.78	1.00
P10	ACC	0.96	0.87	0.95	0.83	0.82	0.93	0.99
	LACC	0.90	0.85	0.91	0.71	0.78	0.87	0.99
	GYR	0.94	0.89	0.96	0.85	0.76	0.84	0.99
	ACCGYR	0.95	0.94	0.96	0.89	0.88	0.95	0.99
	LACCGYR	0.90	0.90	0.97	0.87	0.83	0.88	1.00
P11	ACC	0.94	0.93	0.89	0.90	0.86	0.94	0.99
	LACC	0.81	0.65	0.83	0.70	0.69	0.75	1.00
	GYR	0.78	0.72	0.75	0.58	0.72	0.68	0.97
	ACCGYR	0.93	0.90	0.90	0.82	0.85	0.94	1.00
	LACCGYR	0.77	0.67	0.83	0.66	0.66	0.69	1.00

Table 20
F1-scores of per participant analysis using F3, RF and Case 4

	Sensor	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
P1	ACC	0.76	0.72	0.87	0.56	0.68	1.00	1.00
	LACC	0.81	0.66	0.70	0.43	0.41	0.96	1.00
	GYR	0.72	0.60	0.88	0.60	0.47	0.91	0.96
	ACCGYR	0.83	0.75	0.96	0.70	0.71	1.00	1.00
	LACCGYR	0.80	0.62	0.90	0.65	0.51	0.96	1.00
P2	ACC	0.98	0.87	0.74	0.62	0.62	1.00	1.00
	LACC	0.93	0.77	0.89	0.65	0.71	0.99	1.00
	GYR	0.95	0.87	0.87	0.76	0.75	0.91	1.00
	ACCGYR	0.98	0.92	0.94	0.82	0.83	0.99	1.00
	LACCGYR	0.96	0.88	0.90	0.80	0.79	0.98	1.00
P3	ACC	0.72	0.64	0.84	0.61	0.50	0.97	1.00
	LACC	0.57	0.51	0.63	0.44	0.25	1.00	1.00
	GYR	0.56	0.43	0.89	0.54	0.35	0.99	1.00
	ACCGYR	0.68	0.43	0.93	0.64	0.49	0.98	1.00
	LACCGYR	0.61	0.54	0.89	0.55	0.41	1.00	1.00
P4	ACC	0.62	0.24	0.96	0.53	0.64	1.00	1.00
	LACC	0.74	0.61	0.77	0.48	0.60	0.94	1.00
	GYR	0.70	0.39	0.96	0.68	0.57	0.90	0.98
	ACCGYR	0.65	0.33	0.99	0.62	0.70	1.00	1.00
	LACCGYR	0.73	0.42	0.94	0.55	0.56	0.94	1.00
P5	ACC	0.49	0.20	0.92	0.32	0.67	0.99	1.00
	LACC	0.58	0.19	0.46	0.17	0.32	1.00	1.00
	GYR	0.59	0.46	0.88	0.40	0.25	0.99	1.00
	ACCGYR	0.46	0.43	0.89	0.51	0.68	0.99	1.00
	LACCGYR	0.46	0.38	0.85	0.30	0.33	1.00	1.00
P6	ACC	0.95	0.80	0.67	0.31	0.10	1.00	1.00
	LACC	0.78	0.59	0.60	0.59	0.33	1.00	1.00
	GYR	0.84	0.73	0.80	0.65	0.79	0.99	1.00
	ACCGYR	0.95	0.87	0.74	0.56	0.68	1.00	1.00
	LACCGYR	0.88	0.75	0.80	0.76	0.77	1.00	1.00
P7	ACC	0.85	0.57	0.55	0.44	0.36	1.00	1.00
	LACC	0.67	0.27	0.46	0.41	0.28	1.00	1.00
	GYR	0.75	0.67	0.93	0.49	0.61	1.00	1.00
	ACCGYR	0.83	0.74	0.75	0.56	0.55	1.00	1.00
	LACCGYR	0.82	0.63	0.94	0.61	0.40	1.00	1.00
P8	ACC	0.91	0.61	0.87	0.76	0.60	0.95	0.99
	LACC	0.77	0.35	0.54	0.40	0.48	0.98	1.00
	GYR	0.81	0.48	0.87	0.50	0.56	0.99	1.00
	ACCGYR	0.88	0.72	0.90	0.72	0.64	0.94	0.99
	LACCGYR	0.92	0.59	0.83	0.56	0.53	0.98	1.00
P9	ACC	0.97	0.31	0.88	0.57	0.62	1.00	1.00
	LACC	0.93	0.51	0.59	0.45	0.44	0.95	1.00
	GYR	0.92	0.66	0.91	0.53	0.56	0.92	1.00
	ACCGYR	0.96	0.66	0.95	0.62	0.67	1.00	1.00
	LACCGYR	0.92	0.53	0.89	0.57	0.54	0.97	1.00

Table 20

(Continued.)

	Sensor	smokeSD	smokeST	eat	drinkSD	drinkST	sit	stand
P10	ACC	0.73	0.69	0.93	0.29	0.54	1.00	1.00
	LACC	0.72	0.74	0.68	0.22	0.56	0.98	1.00
	GYR	0.65	0.69	0.95	0.38	0.61	0.99	1.00
	ACCGYR	0.73	0.78	0.96	0.36	0.68	1.00	1.00
	LACCGYR	0.69	0.75	0.98	0.35	0.68	0.99	1.00
P11	ACC	0.87	0.25	0.88	0.75	0.57	0.87	1.00
	LACC	0.68	0.27	0.65	0.59	0.50	0.86	1.00
	GYR	0.63	0.37	0.82	0.54	0.57	0.89	0.99
	ACCGYR	0.80	0.37	0.81	0.63	0.55	0.93	1.00
	LACCGYR	0.66	0.30	0.86	0.61	0.51	0.88	1.00

References

[1] F. Alharbi and K. Farrahi, A convolutional neural network for smoking activity recognition, in: *2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom)*, IEEE, 2018, pp. 1–6.

[2] A.A. Ali, S.M. Hossain, K. Hovsepian, M.M. Rahman, K. Plarre and S. Kumar, mPuff: Automated detection of cigarette smoking puffs from respiration measurements, in: *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*, ACM, 2012, pp. 269–280.

[3] A.C. Cole, D. Anshari, V. Lambert, F.J. Thrasher and H. Valafar, Detecting smoking events using accelerometer data collected via smartwatch technology: Validation study, *JMIR Mhealth Uhealth* **5**(12) (2017), 189. doi:10.2196/mhealth.9035.

[4] D. Coskun, O.D. Incel and A. Ozgovde, Phone position/placement detection using accelerometer: Impact on activity recognition, in: *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2015, pp. 1–6. doi:10.1109/ISSNIP.2015.7106915.

[5] D. Figo, P.C. Diniz, D.R. Ferreira and J.M. Cardoso, Preprocessing techniques for context recognition from accelerometer data, *Personal and Ubiquitous Computing* **14**(7) (2010), 645–662. doi:10.1007/s00779-010-0293-9.

[6] J.G. Fiscus, J. Ajot, N. Radde and C. Laprun, Multiple dimension levenshtein edit distance calculations for evaluating automatic speech recognition systems during simultaneous speech, in: *The International Conference on Language Resources and Evaluation (LERC)*, 2006.

[7] P.S.R. Group, Our dataset available at, in: *Pervasive Systems Research Group*, 2017.

[8] M. Imtiaz, R. Ramos-Garcia, V. Senyurek, S. Tiffany and E. Sazonov, Development of a multisensory wearable system for monitoring cigarette smoking behavior in free-living conditions, *Electronics* **6**(4) (2017), 2079–9292. doi:10.3390/electronics6040104.

[9] Ö.D. Incel, Analysis of movement, orientation and rotation-based sensing for phone placement recognition, *Sensors* **15**(10) (2015), 25474–25506. doi:10.3390/s151025474.

[10] V. Janko, B. Cvetkovic, A. Gradisek, M. Lustrek, B. Strumbelj and T. Kajtna, e-Gibalec: Mobile application to monitor and encourage physical activity in schoolchildren, *Journal of Ambient Intelligence and Smart Environments* **9**(5) (2017), 595–609. doi:10.3233/AIS-170453.

[11] H. Lu, J. Yang, Z. Liu, N.D. Lane, T. Choudhury and A.T. Campbell, The jigsaw continuous sensing engine for mobile phone applications, in: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, SenSys'10*, ACM, New York, NY, USA, 2010, pp. 71–84. ISBN 978-1-4503-0344-6. doi:10.1145/1869983.1869992.

[12] M. Nambu, Body surface mounted biomedical monitoring system using bluetooth, in: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 1824–1825. doi:10.1109/IEMBS.2007.4352668.

[13] J.L.R. Ortiz, Human activity dataset generation, in: *Smartphone-Based Human Activity Recognition*, Springer, 2015, pp. 59–78.

[14] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan and E. Kalogerakis, Risq: Recognizing smoking gestures with inertial sensors on a wristband, in: *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ACM, 2014, pp. 149–161.

[15] I. Pires, N.M. Garcia, V. Felizardo, N. Pombo and F. Florez-Revulta, Validation of a method for the estimation of energy expenditure during physical activity using a mobile device accelerometer, *Journal of Ambient Intelligence and Smart Environments* **10**(3) (2018).

[16] H. Rezaie and M. Ghassemian, An adaptive algorithm to improve energy efficiency in wearable activity recognition systems, *IEEE Sensors Journal* **17**(16) (2017), 5315–5323. doi:10.1109/JSEN.2017.2720725.

[17] J.L. Rodgers and W.A. Nicewander, Thirteen ways to look at the correlation coefficient, *The American Statistician* **42**(1) (1988), 59–66. doi:10.2307/2685263.

[18] N. Saleheen, A.A. Ali, S.M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. al'Absi and S. Kumar, puffMarker: A multi-sensor approach for pinpointing the timing of first lapse in smoking cessation, in: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 2015, pp. 999–1010. doi:10.1145/2750858.2806897.

[19] P.M. Scholl and K. Van Laerhoven, A feasibility study of wrist-worn accelerometer based detection of smoking habits, in: *Innovative Mobile and Internet Services in Ubiquitous Comput-*

ing (IMIS), 2012 Sixth International Conference on, IEEE, 2012, pp. 886–891.

[20] V.Y. Senyurek, M.H. Imtiaz, P. Belsare, S. Tiffany and E. Sazonov, Smoking detection based on regularity analysis of hand to mouth gestures, *Biomedical Signal Processing and Control* **51** (2019), 106–112, http://www.sciencedirect.com/science/article/pii/S1746809419300242. doi:10.1016/j.bspc.2019.01.026.

[21] M. Shoaib, S. Bosch, O.D. Incel, H. Scholten and P.J. Havinga, Fusion of smartphone motion sensors for physical activity recognition, *Sensors* **14**(6) (2014), 10146–10176. doi:10.3390/s140610146.

[22] M. Shoaib, O.D. Incel, H. Scholten and P. Havinga, Resource consumption analysis of online activity recognition on mobile phones and smartwatches, in: *36th IEEE – International Performance Computing and Communications Conference*, 2017.

[23] M. Shoaib, O.D. Incel, H. Scholten and P. Havinga, Smoke-Sense: Online activity recognition framework on smartwatches, in: *Mobile Computing, Applications, and Services*, K. Murao, R. Ohmura, S. Inoue and Y. Gotoh, eds, Springer International Publishing, Cham, 2018, pp. 106–124. ISBN 978-3-319-90740-6. doi:10.1007/978-3-319-90740-6_7.

[24] M. Shoaib, H. Scholten, P.J. Havinga and O.D. Incel, A hierarchical lazy smoking detection algorithm using smartwatch sensors, in: *E-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on*, IEEE, 2016, pp. 1–6.

[25] A. Skinner, C. Stone, H. Doughty and M. Munafo, StopWatch: The preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking, *Nicotine and Tobacco Research* (2018).

[26] Q. Tang, D.J. Vidrine, E. Crowder and S.S. Intille, Automated detection of puffing and smoking with wrist accelerometers, in: *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, ICST*, Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2014, pp. 80–87.

[27] Y.E. Ustev, O. Durmaz Incel and C. Ersoy, User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal, in: *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, ACM, 2013, pp. 1427–1436. doi:10.1145/2494091.2496039.

[28] J.P. Varkey, D. Pompili and T.A. Walls, Human motion recognition using a wireless sensor-based wearable system, *Personal and Ubiquitous Computing* **16**(7) (2012), 897–910. doi:10.1007/s00779-011-0455-4.

[29] S. Wang, J. Yang, N. Chen, X. Chen and Q. Zhang, Human activity recognition with user-free accelerometers in the sensor networks, in: *2005 International Conference on Neural Networks and Brain*, Vol. 2, 2005, pp. 1212–1217. doi:10.1109/ICNNB.2005.1614831.

[30] J.-Y. Yang, J.-S. Wang and Y.-P. Chen, Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers, *Pattern recognition letters* **29**(16) (2008), 2213–2220. doi:10.1016/j.patrec.2008.08.002.