University of Nebraska at Omaha
DigitalCommons@UNO

Computer Science Faculty Publications

Department of Computer Science

7-5-2011

Centinela: A human activity recognition system based on acceleration and vital sign data

Óscar D. lara

Alfredo J. Perez

Miguel A. Labrador

José D. Posada

Follow this and additional works at: https://digitalcommons.unomaha.edu/compscifacpub

Part of the Computer Sciences Commons

Please take our feedback survey at: https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE

UNO LIBRARIES
LIBRARY.UNOMAHA.EDU

Centinela: A human activity recognition system based on acceleration *and* vital sign data

Óscar D. Lara [a], Alfredo J. Pérez [a], Miguel A. Labrador [a], José D. Posada [b]

[a] *Department of Computer Science and Engineering. University of South Florida, Tampa, FL 33620, United States*

[b] *Departamento de Ingeniería Electrónica. Universidad Autónoma del Caribe, Barranquilla, Colombia*

***Keywords**:*

Feature extraction, Classification, Subject-independent analysis, Transient features

Abstract

This paper presents Centinela, a system that combines acceleration data with vital signs to achieve highly accurate activity recognition. Centinela recognizes five activities: walking, running, sitting, ascending, and descending. The system includes a portable and unobtrusive real-time data collection platform, which only requires a single sensing device and a mobile phone. To extract features, both statistical and structural detectors are applied, and two new features are proposed to discriminate among activities during periods of vital sign stabilization. After evaluating eight different classifiers and three different time window sizes, our results show that Centinela achieves up to 95.7% overall accuracy, which is higher than current approaches under similar conditions. Our results also indicate that vital signs are useful to discriminate between certain activities. Indeed, Centinela achieves 100% accuracy for activities such as running and sitting, and slightly improves the classification accuracy for ascending compared to the cases that utilize acceleration data only.

1. Introduction

In the past decade, there has been a significant advance of mobile devices and sensors in regards to size, cost, and power. This has enabled new sources of data to study people's daily activities and behaviors. Hence *human-centric sensing* came into picture as a promising research area in computer science [1]. Particularly, the recognition of human physical activities has become a task of high interest within the field, especially for medical, military, and security applications. For instance, patients with dementia and other mental pathologies could be monitored to detect abnormal activities and thereby prevent undesirable consequences [2]. An interactive game might also require information about which activity the user is performing in order to respond accordingly. In tactical scenarios, the soldiers' activities along with their location may be useful to send alerts in case of danger.

All these applications need to solve the activity recognition problem, which from a practical point of view, can be defined as follows: given a time window W, defined within time instants t_i and t_j, which contains a set of time series $S = \{S_0, . . . , S_k−1\}$, from each of the k measured attributes, the goal is to determine the activity performed during W from a predefined set of mutually exclusive activities (e.g., sitting, walking, eating, etc.). Now, recognizing human activities is not a trivial task. As a matter of fact, several challenges lie in this process, such as the selection of the attributes to be measured, the extraction of meaningful features, and the recognition of ambiguous activities. Energy consumption is also a critical issue in terms of deciding which sensors to turn on and off at any time, or setting the optimal sampling resolution [3].

Most of the previously proposed schemes in activity recognition collect data from either triaxial accelerometers, video sequences [4], or environmental variables. However, little work has been reported considering vital sign data. We believe there is a noticeable relationship between the behavior of the vital signs and the physical activity. When an individual begins running, for instance, it is expected that their heart rate and breath amplitude increase. Consequently, we hypothesize that higher human activity recognition accuracy can be achieved using both acceleration and vital sign data. To illustrate this, consider the situation in Fig. 1. Data from triaxial acceleration and vital signs were recorded while a subject was ascending after walking. Note that the acceleration signals within most time intervals are very similar for both activities. Instead, the heart rate time series exhibits a very clear pattern, as a person requires more physical effort to climb stairs than to walk. This might allow us to classify said activities more accurately.

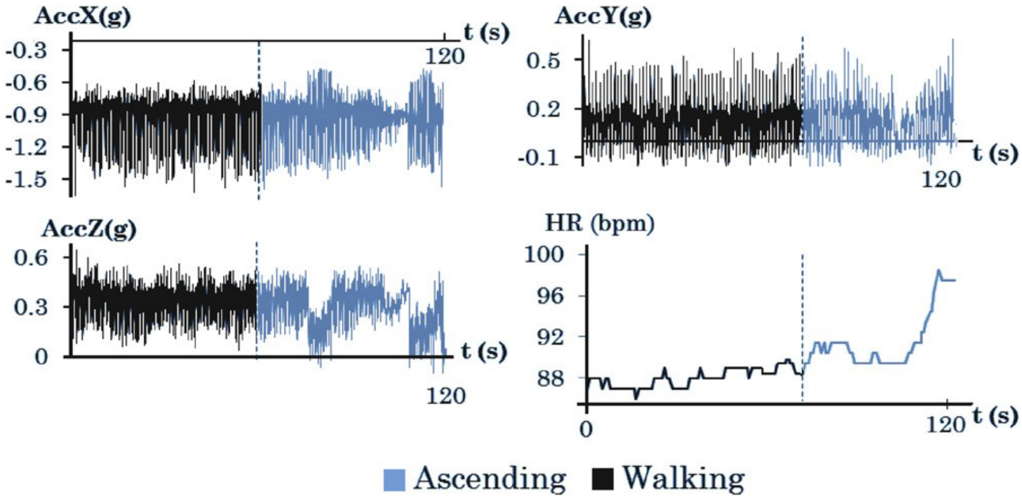


Fig. 1. Acceleration signals and heart rate for the activities *walking* and *ascending*.

This paper presents Centinela, a human physical activity recognition system for five different activities: *sitting, walking, running, ascending*, and *descending*. The proposed methodology encompasses (1) collecting vital sign and acceleration data from human subjects; (2) extracting features from the measured attributes; (3) building supervised machine learning models for activity classification; and (4) evaluating the accuracy of the models under different parameter configurations. The main contributions of this work are listed below:

- Centinela combines acceleration data with vital signs to achieve highly accurate activity recognition. In fact, it provides higher accuracy than other approaches under the same conditions.
- Since vital signs are not expected to change abruptly, Centinela applies structure detectors [5], i.e., linear and non-linear functions, to extract

features.

- Two new features are proposed for vital signs: *magnitude of change* and *trend*, intended to discriminate among activities during periods of vital sign stabilization.
- Centinela features a portable and unobtrusive real-time data collection platform, which allows not only for activity recognition but also for monitoring health conditions of target individuals.
- Several classifiers are analyzed in the study, allowing other researchers and application developers to use the most appropriate classifiers for specific activities.

The rest of the paper is organized as follows: Section 2 analyzes the state of the art in human activity recognition. Later, Section 3 introduces the global structure of Centinela. Section 3.1 describes the data acquisition architecture, as well as the data collection protocol. Section 3.2 covers the methods applied for feature extraction, i.e., statistical, structural, and transient features. Then, Section 4 presents the methodology of the experiments and the main results. Finally, Section 5 summarizes the most important conclusions and findings.

2. Related work

Although the first works in human activity recognition (HAR) date back to the late '90s [6], there are still many issues that motivate the development of new techniques to improve accuracy under more realistic conditions. These issues concern to four different phases: *data collection*, *feature extraction*, *classification*, and *evaluation*.

2.1. *Data collection*

With respect to data collection, it is crucial to make an appropriate selection of the attributes to be measured and the sensors to be used. Many previously proposed schemes use triaxial accelerometers on different parts of the body (e.g., wrist, thigh, leg, pocket, etc.) to recognize ambulation activities (e.g., walking, running, lying, etc.) [7–12]. Other methods are based upon environmental variables and utilize microphones, light sensors, and humidity sensors, among others [13,14]. Nevertheless, little work has been done using vital sign data. Tapia et al. [15] proposed an activity recognition system that combines data from five triaxial accelerometers and a heart rate monitor. However, they concluded that the heart rate is not useful to discriminate between activities. Their argument, which is valid, is that after performing physically demanding activities (e.g., running) the heart rate remains high for a while, even if the individual is lying or sitting. To deal with this issue, Centinela utilizes new feature extraction techniques that allow for activity recognition during periods of vital sign stabilization.

It is also important to build an effective data collection system (i.e., hardware and software) in terms of portability, reliability, energy consumption, comfort, and cost. Some methods require four or five accelerometers in different parts of the body [12,15,16], or need the user to carry a heavy rucksack with a computer and other recording devices [14]. This might be invasive, uncomfortable, expensive, and hence not suitable for online activity recognition. Centinela requires one single sensing device, which is comfortable and unobtrusive (see Section 3.1), and a Java-enabled cellphone with Bluetooth connectivity.

2.2. *Feature extraction*

Existing HAR systems based on accelerometer data employ statistical feature extraction. Most of them apply either time- domain features such as mean, variance, energy, correlation between axes, etc. [11–17], or frequency-domain features, such as entropy and the coefficients of the Fourier transform. Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) have also been applied with promising results [10], as well as autoregressive model coefficients [7]. All these techniques are conceived to handle the high variability of acceleration signals. In contrast, vital signs fluctuate smoothly and are not expected to suddenly change in short periods of time. Therefore, structure detectors [5] are utilized in this work to approximate vital sign time series by means of linear and non-linear functions. Moreover, two new features are proposed: the magnitude of change and trend of vital signs, intended to discriminate among activities during periods of vital sign stabilization.

2.3. *Classification*

Many classification algorithms have been applied for activity recognition: decision trees, such as C4.5 and ID3 [9,12– 14,16,18]; Bayesian methods, such as Naïve Bayes (NB) and Bayesian Networks (BN) [12,15,18]; Nearest Neighbor [13,18], Fuzzy Logic [11,17], Neural Networks [19], and Support Vector Machines [7,8,10], among others. In this work we not only evaluate traditional classifiers, such as Naïve Bayes, Bayesian Networks, C4.5, and Multilayer Perceptron, but also classifier ensembles with methods such as Bagging and Boosting. The main idea behind these techniques is to make decisions based upon the output of a set of classifiers rather than considering one single learning method. Section 4 shows the methodology and results of the classifier evaluation.

2.4. *Evaluation*

Two types of analyses have been proposed to evaluate activity recognition systems: *subject-dependent* and *subject- independent* evaluations [15]. In the first one, a classifier is trained and tested for each individual with his/her own data and the average accuracy for all subjects is computed. In the second one, only one classifier is built splitting the data of all individuals into a training set and a testing set.

It is important to emphasize that each person may perform activities in a different manner, which makes subject- independent analysis more challenging. In practice, a real-time activity recognition system should be able to fit any individual. It would not be convenient to train the system for each new user, especially when (1) there are too many activities; (2) some activities are not desirable for the subject to carry out (e.g., falling downstairs); or (3) the subject would not cooperate with the data collection process (e.g., patients with dementia and other mental pathologies). Thus, a subject- independent analysis, as the one presented in this paper, is preferred.

A comparison of the classification accuracy given by Centinela and other state-of-the-art approaches is included in Section 4.2.5.

3. Description of the system

Fig. 2 illustrates the process for activity recognition. First, data are collected from accelerometer and vital sign sensors, as described in Section 3.1. Then, time-domain and frequency-domain statistical feature extraction is applied to the acceleration signals (Section 3.2.1), as well as structural and transient features are extracted from vital signs (Sections 3.2.2 and 3.2.3). Next, the dataset with the extracted features is given as input to various classification algorithms and the classification accuracy of each one is calculated by means of cross validation (Section 4). Finally, the best classifier is selected as the result of a non-parametric statistical test.

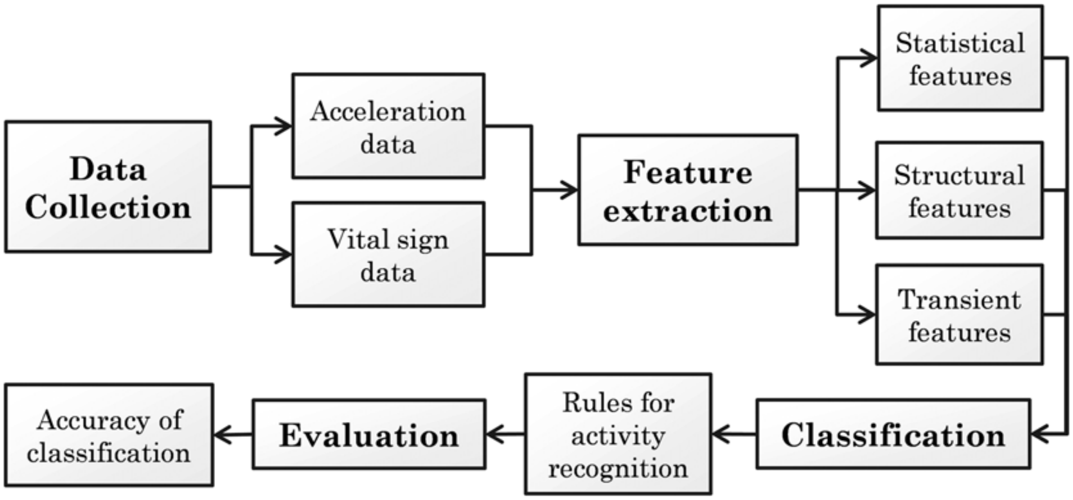


Fig. 2. Centinela’s system overview.

3.1. *Data collection*

Fig. 3 shows the system architecture for the data collection phase. The sensing device (see Section 3.1.1 for more details) communicates via Bluetooth with an Internet-enabled cellphone. There is a mobile application which decodes the packets and sends labeled data to the application server via the Internet. The server then receives these data and stores them into a relational database

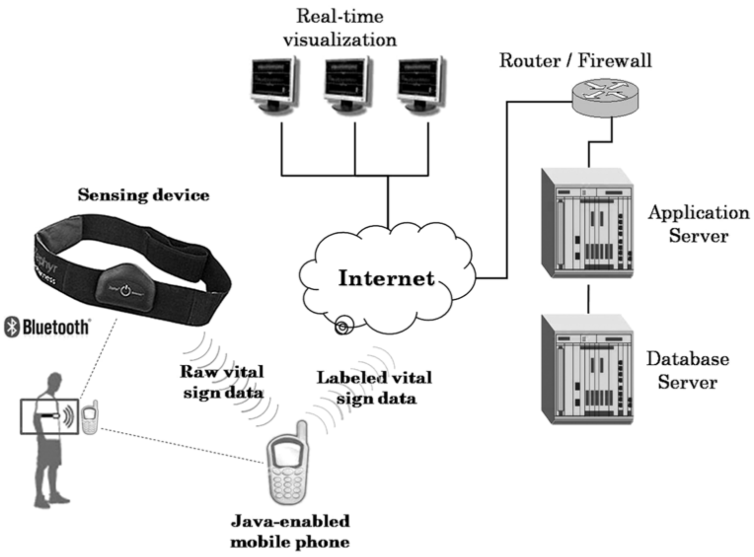
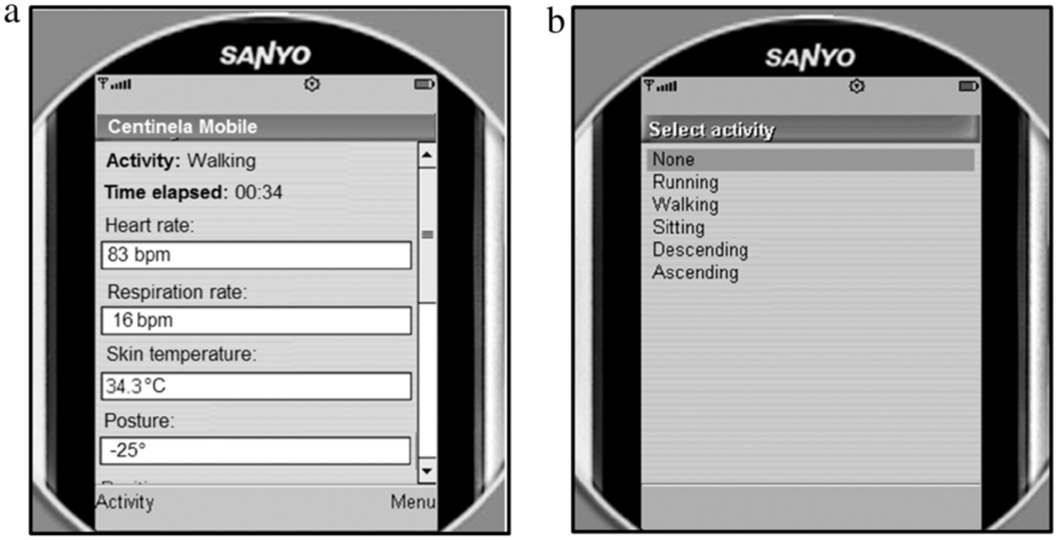


Fig. 3. Data collection system architecture.

3.1.1. *Sensing device*

We are using the BioHarness^{TM} BT chest sensor strap [20] manufactured by Zephyr shown in Fig. 3. This device features a triaxial accelerometer and allows for measuring vital signs as well. The strap is unobtrusive, lightweight, and can be easily worn by any person. The measured attributes are: heart rate, respiration rate, breath amplitude, skin temperature, posture (i.e., inclination of the sensor), electrocardiogram amplitude, and 3D acceleration. The accelerometer records measurements at 50 Hz, each one between $-3g$ to $3g$, where g stands for the acceleration due to gravity. Acceleration samples are aggregated in packets sent every 400 ms, so every packet contains twenty acceleration measurements in all three dimensions. On the other hand, the vital signs are sampled at 1 Hz, since they are not expected to change considerably in short periods of time.

In the literature, accelerometers are commonly placed either on the wrist [12,14,16], ankle [12,16], or in the trouser's pocket [7,8,10], yet a person might be moving his/her arms or legs while been seated. This fact may introduce noise to the data, thereby causing misclassification. We believe that placing the accelerometer on the chest makes our system more noise tolerant, and our results support this hypothesis.



Centinela Mobile

Activity: Walking

Time elapsed: 00:34

Heart rate: 83 bpm

Respiration rate: 16 bpm

Skin temperature: 34.3°C

Posture: -25°

Fig. 4. Mobile application user interface [21].

Table 1

Physical characteristics of the participants.

	Avg	Min	Max
Age (years)	24	9	34
Weight (kg)	76.5	27	95
Height (m)	1.74	1.35	1.88
BMI	24.23	20.96	29

3.1.2. *Mobile application*

A mobile software application was built to collect training data under the Java ME platform. This allows Centinela to run on any mobile phone that supports Java, thereby avoiding the inconvenience of requiring the user to carry additional devices. The mobile application receives and decodes the raw data sent from the sensor via Bluetooth, visualizes the measurements (see Fig. 4(a)), and labels each measurement according to the option selected by the user, either: *running*, *walking*, *sitting*, *ascending,* or *descending* (see Fig. 4(b)). The samples are sent in real time, via UDP, to the application server, which stores the labeled data into a relational PostgreSQL database.

3.1.3. *Data collection protocol*

The data were collected in a naturalistic fashion, thus, no specific instructions were given to the participants. The speed, intensity, gait, and other environmental conditions were arbitrarily chosen by the subjects. Eight individuals, 7 males and 1 female, participated in this study. Their physical characteristics, namely age, weight, height, and body mass index are shown in Table 1.

Unlike accelerometer signals, vital signs do not abruptly vary after the person changes activities. On the contrary, the values of vital signs during time interval I_j depend of the activity during I_j-1. Therefore, the data should be collected so that the recognition of each single activity can be independent of the previous state. If we required, for instance, the individuals to be at rest before recording each session, the system would not be trained to recognize interleaving activities! Consequently, we have collected data from subjects while performing successive pairs of activities, e.g., running before sitting, walking before descending, and so on. This was carried out for all twenty possible combinations of pairs of consecutive activities.

3.2. *Feature extraction*

In general, two approaches have been proposed to extract features in time series data: *statistical detectors* and *structure detectors* [5]. Statistical detectors, such as the Fourier transform and the Wavelet transform, use quantitative characteristics of the data to extract features. On the other hand, structure detectors take into account the interrelationship among data. Hence, they have been widely used for image processing and time series analysis. Due to both acceleration and physiological signals being distinct in nature, we have employed methods from statistical and structural feature extraction.

Now, to overcome the problem of detecting transitions between activities, all measured signals were divided into fixed size 50% overlap time windows [11,12]. Three different window sizes were tried: 5s, 12s, and 20s. For every time window, 90 features were extracted as follows: eight statistical features for each of the acceleration signals (i.e., 24 features), nine structural features for each of the physiological signals (i.e., 54 features), and two transient features for each of the physiological signals (i.e., 12 features). Table 2 summarizes the feature set computed from raw signals in this work. The definitions of these features are presented in the following subsections.

Table 2

Complete set of features extracted in this work.

Measured signals	Extracted features			
	Statistical	Structural	Transient	Total
AccX (g)	X			8
AccY (g)	X			8
AccZ (g)	X			8
Heart rate		X	X	11
Respiration rate		X	X	11
Breath amplitude		X	X	11
Skin temperature		X	X	11
Posture		X	X	11
ECG amplitude		X	X	11
Total	24	54	12	90

3.2.1. *Statistical features*

Time-domain and frequency-domain features [11–17] have been extensively used to filter the relevant information of acceleration signals. In this work, eight features were calculated for all three acceleration signals (a total of 24 features). These are: *mean*, *variance*, *standard deviation*, *correlation between axes, interquartile range, mean absolute deviation*, and *root mean square,* from the time domain; and, *energy* from the frequency domain. The interested reader might refer to [11] for the definition of all these features.

Table 3

Structure detectors evaluated in this work.

Function	Equation	Parameters
Linear	$F(t) = mt + b$	$\{m, b\}$
Polynomial	$F(t) = a_0 + a_1t + \cdots + a_n{-}1\ t^n{-}1$	$\{a_0, \ldots, a_n{-}1\}$
Exponential	$F(t) = a\|b\|^t + c$	$\{a, b, c\}$
Sinusoidal	$F(t) = a \ast sin(t + b) + c$	$\{a, b, c\}$

3.2.2. *Structural features*

Since vital signs have much lower variability than acceleration signals, structure detectors turn out to be a suitable approach to extract features from vital sign time series. Structure detectors use an arbitrary function f with a set of free parameters $\{a_0, \ldots, a_n\}$ to fit the points of a given time series S [5]; these parameters are, in fact, the extracted features. In order to evaluate the goodness of fit of f to S, the sum of squared error (SSE) is computed. For each measured attribute, the goal was to find the function f^* with the smallest SSE, so that the feature extraction process relies on the calculation of the free parameters of f^*. Table 3 summarizes the different types of functions that have been evaluated in this work.

The median of the SSE was calculated for all time windows from all six physiological signals and all four structure detectors. The median was preferred over the mean to prevent noisy instances to bias the goodness of fit of the feature detectors. From the evaluation, polynomial functions of third degree had the lowest SSE for all six vital signs. Polynomials of degree higher than three were not considered to avoid overfitting due to Runge's phenomenon [22]. A total of nine structural features were extracted from each vital sign time window, i.e., the coefficients of the polynomials of degree one, two, and three that best fit the points in the time window.

3.2.3. *Transient features*

Consider, for instance, that someone is running for one minute and then sits for two minutes. Even though the individual is seated, the vital signs (e.g., heart rate, respiration rate, etc.) remain as if he/she was running for an interval of time that we have called the *transient period*. To overcome this issue, two new features are proposed in this work, the *trend τ* , and the *magnitude of change κ*, intended to describe the behavior of the vital signs during transient periods. The trend indicates whether the time series is increasing, decreasing, or constant. Notice that, due to the nature of the human activities considered in this work, it is expected that vital signs are either strictly increasing, strictly decreasing, or remain constant while an individual is performing one single activity.

Definition 1.

Trend.

Let *m* be the slope of the line that best fits the series *S*. Then, the trend *τ(S, r)* of *S* is defined as follows:

$$\tau(S, r) = \begin{cases} 1 \text{ (increasing)} & \text{if } (m \geq r) \\ -1 \text{ (decreasing)} & \text{if } (m \leq -r) \\ 0 \text{ (constant)} & \text{if } (|m| < |r|) \end{cases} \tag{1}$$

where *r* is a positive real number that stands for the slope threshold. This value was set to tan*(*15°*)* after doing an experimental analysis over the entire the dataset. The *trend* can be computed in *O(*1*)*, given that the slope of the line that best fits the data points was calculated beforehand as one of the structural features.

Now, it is important not only to detect whether the vital signs increased or decreased, but also to measure how much they varied. For this purpose the magnitude of change feature is presented as follows:

Definition 2.

Magnitude of change.

Let *S* be a given time series define from $t_{\min}$ to $t_{\max}$. Let S_p^- be a subset of *S* which contains all measurements between $t_{\min}$ and $t_{\min} + (t_{\max} - t_{\min})p$, where $0 < p < 1$ is a percentage of the series. Let S_p^+ be a subset of S which contains all samples between $t_{\min} + (t_{\max} - t_{\min})(1 - p)$ and $t_{\max}$. Then, the magnitude of change $\kappa(S, p)$ is defined as

$$\kappa(S,p) = \max\left\{\left|\max\left(\overset{+}{\underset{p}{S}}\right) - \min\left(\overset{-}{\underset{p}{S}}\right)\right|, \left|\max\left(\overset{-}{\underset{p}{S}}\right) - \min\left(\overset{+}{\underset{p}{S}}\right)\right|\right\}. \qquad (2)$$

The value of *p* was set to 0.2 after doing an experimental analysis over the entire dataset. This implies that S^- is the first 20% of the series and S^+ would be the last 20% of the series. The purpose of the magnitude of change is to estimate the maximum deviation between the beginning and the end of the series, and it can be calculated in linear time. Fig. 5 illustrates the process of calculating this feature.

Using transient features together, our hypothesis is that a classifier could generate rules such as: if $\kappa(S_{HR}, p)$ is *large* and $\tau(S_{HR}, r)$ is *increasing*, then *activity* is *running*, for a given heart rate time series S_{HR}.

Even though both *magnitude of change* and *trend* are strongly related to the slope of the line that best fits a time series, they are different measures of the data shape. Section 4.2.4 analyzes the effectiveness of the two new features proposed in this work.

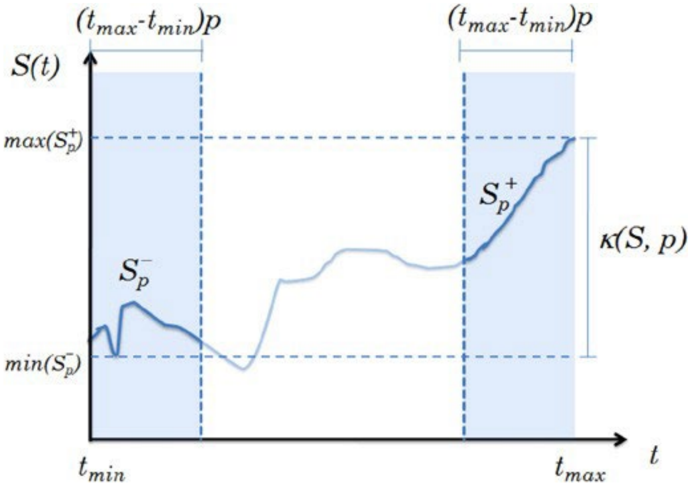


Fig. 5. Calculation of the magnitude of change feature.

4. Evaluation

This section describes the methodology to evaluate the system and provides further analysis and discussion of our results and main findings.

4.1. *Design of the experiments*

Activity recognition was fulfilled by assessing three different datasets: the first one, D_{acc} , solely contains the features extracted from acceleration data; the second, D_{nt} , has statistical and structural features; and, the last one, D_{VS}, includes all features (i.e., statistical, structural and transient). This is with the purpose of measuring the impact of vital signs features in the classification accuracy. Eight classification algorithms were also evaluated for each dataset:

1. Naïve Bayes (NB) [23].
2. Bayesian Network (BN) using K2 search algorithm [24].
3. J48 decision tree, which is an implementation of the C4.5 algorithm [25].
4. Multilayer Perceptron (MLP), which relies on a Backpropagation Neural Network [25].
5. Additive Logistic Regression (ALR) [26], performing Boosting with an ensemble of ten Decision Stump classifiers.
6. Bagging using an ensemble of ten Naïve Bayes classifiers (BNB) and each bag having the same size than the training set.
7. Bagging using an ensemble of ten Bayesian Network classifiers (BBN) and each bag having the same size than the training set.
8. Bagging using an ensemble of ten J48 classifiers (BJ48) and each bag having the same size than the training set.

We do not elaborate on the classification algorithms since they are not part of the contributions of this work. The interested reader may refer to [23–26] for a complete description of them.

The evaluation encompasses two parts: first, the selection of the best classifier(s), and later the calculation of their accuracy. This process was completed for all three datasets.

In order to determine whether a classifier is better than another, a 5 × 2 fold cross validation [25] was performed. In general, a 5 × 2 fold cross validation is preferred over other approaches (e.g., 10 fold cross validation or percentage split) because it has a smaller probability of concluding that one classifier is better than another when it is not the case. As we do not have information regarding the probability distribution of the accuracy of each classifier, the non-parametric *Sign Test* [27] was utilized. This test allows us to determine whether there is any statistical difference between the probability distribution functions of two independent random variables. The test is defined as follows:

Definition 3. Sign test.

- Let X, Y be two independent random variables.
- Let $H_0 : P(X > Y) = P(X < Y)$ be the null hypothesis.
- Let (x_i, y_i) be a set of n matched pairs from X and Y respectively.
- Let n' be the number of observations such that $x_i \neq y_i$.
- Let T be the number of observations such that $x_i > y_i$. T follows the binomial distribution under H_0 with parameters n' (i.e., the number of trials) and $p = 0.5$.

The p-values for the test are as follows:

- Lower: $p_{low} = P(T \leq t_{obs})$. Alternative hypothesis: $H_1 : P(X > Y) < P(X < Y)$.
- Upper: $p_{upper} = P(T \geq t_{obs})$. Alternative hypothesis: $H_1 : P(X > Y) > P(X < Y)$.
- Two-sided: $p = 2 \min(p_{low}, p_{upper})$. Alternative hypothesis: $H_1 : P(X > Y) \neq P(X < Y)$.

In our case, the random variables X and Y are the values from the accuracy of two classifiers, and there are five matched pairs for all different random seeds. A way to extend this test to deal with n random variables (since we are considering eight algorithms) is to make a k-rounds binary tournament among pairs of classifiers. If the null hypothesis is rejected, the best classifier (i.e., the winner) goes to the next round, and the other one (i.e., the loser) is discarded. If the null hypothesis is not rejected, both classifiers pass to the next round. The process repeats until no more classifiers can be discarded. At the very end, the classifiers which did not lose any match are selected as the best classifiers.

All the classification algorithms were tested in the Waikato Environment for

Knowledge Analysis (WEKA) [28]. This is a well-known software tool developed by the University of Waikato, New Zealand, which allows for easy evaluation of machine learning algorithms. Three window sizes were evaluated, namely 5s, 12s, and 20s. The significance level for all sign tests was fixed to $\alpha = 0.05$.

4.2. *Results*

An interesting fact in machine learning is that the performance of a classification algorithm depends on which dataset it is applied to. As we are to decide on whether vital signs allow for more accurate classification, it is required to determine the best classifier for each dataset.

4.2.1. *Dataset with features from vital signs and acceleration*

The best classifiers on D_{VS}, according to the 5 × 2 cross validation tournament, are shown in Table 4, for all three different window sizes, and five random seeds s_i: 1, 128, 255, 1023, and 4095. As a notation, the name of the classifiers will be henceforth accompanied with the window size written as superscript and the dataset as subscript. For example, ALR^{12s} stands for ALR over the D_{VS} dataset using 12s time windows.

In order to select the most appropriate window size, the same concept of *k*-rounds binary tournament of sign tests was applied among the best eight classifiers. As a result, four of them were selected: ALR_{vs}^{5s}, BBN_{vs}^{5s}, $BJ48_{vs}^{5s}$, and ALR_{vs}^{12s}. Notice that having time windows as long as 20s considerably affected the accuracy of all the classifiers. This was expected since, in such a case, more than one activity might be performed within one single time window [29].

Let us emphasize that the 5×2 fold cross validation is not intended to measure classification accuracy but to discriminate whether or not there is a statistically significant difference among them! This is because a 5 × 2 fold cross validation only uses half of the data as training set. Now, the actual accuracy of the four classifiers is measured by a 5 × 10 fold cross validation for each activity. The average accuracies are shown in Table 5. Classifiers using time windows of five seconds had the lowest overall accuracy. This is reasonable since vital signs cannot properly describe activity patterns in time intervals as short as five seconds. Instead, Table 5 suggests that features extracted from vital signs within a window of 12s allow the system to discern among activities that might be ambiguous for accelerometers, such as *walking* and *ascending* (see Fig. 1).

Although there is no sufficient statistical evidence to assure that any of the classifiers is better than the others, we have chosen 12s as the best window size, and ALR as the best classifier for the *Dvs* dataset. This is not only because it reaches the highest accuracy in the 5 × 10 fold cross validation, but also because of how each activity was classified. Observe that ALR_{vs}^{12s}(i.e., ALR with 12s time windows over the *Dvs* dataset) classified activities such as *running*, *ascending*, and *descending* with the highest accuracy. For real applications in health care or tactic scenarios, it may be more important to detect these types of activities.

4.2.2. *Dataset with features from acceleration only*

The same procedure was carried out over the dataset that only contains features from the accelerometer data. That is, a tournament of 5 × 2 fold cross validations and sign tests for all eight algorithms. Table 6 contains the results from the best classifiers given by the 5 × 2 fold cross validation over the Dacc dataset, namely 〖ALR〗_acc^5s and 〖MLP〗_acc^5s. Their accuracy was computed by means of a 5 × 10 fold cross validation (see Fig. 6).

Table 4

Percentage classification accuracy for the best classifiers given by the 5 × 2 fold cross validation on *Dvs* . Five different random seeds s_i were utilized.

	s_1	s_2	s_3	s_4	s_5	Average
ALR_{vs}^{5s}	89.22	89.49	91.91	94.34	92.18	91.98
BBN_{vs}^{5s}	90.84	90.57	91.64	90.03	89.76	90.57
$BJ48_{vs}^{5s}$	88.41	90.03	93.53	87.6	88.14	89.54
ALR_{vs}^{12s}	92.857	92.86	87.857	95	88.57	91.43
BBN_{vs}^{12s}	84.286	87.86	89.286	91.43	90.71	89.82
BBN_{vs}^{20s}	80.822	84.931	84.931	80.82	82.19	82.74

Table 5

Per-class mean percentage accuracy of the 5 × 10 fold cross validation among the best classifiers for the *Dvs* . *s*s_i dataset.

Activity	ALR_{vs}^{5s}	BBN_{vs}^{5s}	$BJ48_{vs}^{5s}$	ALR_{vs}^{12s}
Walking	92.98	97.62	94.72	94.3
Running	98.56	98.8	98.32	100
Ascending	83	69.24	86.56	92.84
Descending	91.06	80.4	89.52	91.36
Sitting	100	95.58	96.36	100
Total	93.12	88.328	94.24	95.7

Table 6

Percentage classification accuracy for the best classifiers given by the 5×2 fold cross validation using accelerometer data only. Five different random seeds s_i were ultilized

	s_1	s_2	s_3	s_4	s_5	Average
MLP_{acc}^{5s}	92.45	91.11	94.61	88.14	92.99	91.86
ALR_{acc}^{5s}	88.95	91.37	90.84	92.45	91.91	91.105

4.2.3. *Analyzing the impact of vital signs*

To quantify the improvement achieved by incorporating vital sign data, the best classifiers for each dataset are now compared. Fig. 6 summarizes the classification accuracy per activity for ALR_{vs}^{12s}, ALR_{acc}^{5s}, andMLP_{acc}^{5s}. Note that ALR_{vs}^{12s} reached the highest overall accuracy (i.e., 95.7%), perfectly classifying activities such as *sitting* and *running*. The activity labeled as *ascending* reported the most significant improvement (between 10% and 13%). This was expected as acceleration signals are similar for *ascending* and *walking* whereas vital signs provide more clear patterns to distinguish between these activities (see Fig. 1).

Notice that MLP_{acc}^{5s} yields higher mean accuracy than ALR_{vs}^{12s} for descending (roughly 6%). This brings a new point to the discussion: depending on the application and the activities that are to be recognized, it might (or might not) be useful to consider vital sign data to recognize human activities. According to our results, if the target activities are *descending* or *walking*, the data from accelerometers would be sufficient to discover activity patterns. On the other hand, if activities such as *running*, *sitting*, or *ascending* need to be recognized, vital signs would definitely provide the system with a more reliable output.

4.2.4. *Analyzing the impact of transient features*

To evaluate the effectiveness of transient features, an additional 5 × 2 fold cross validation tournament was applied to a new dataset D_{nt} (where the subscript *nt* stands for no transient) which only includes statistical features from acceleration data and structural features from vital signs. After evaluating all eight classification algorithms and all three window sizes, two classifiers were chosen, ALR_{nt}^{5s} and ALR_{nt}^{12s}. These are now compared to ALR_{vs}^{12s} (which does include transient features), and the results are in Fig. 7. Despite the overall accuracy improvement was between 3% and 4%, transient features enhanced between 4% and 10% for *ascending*. We believe this improvement is worthwhile as transient features are inexpensive, computationally speaking. Finally, Table 7 shows the average of the confusion matrices from the 5 × 10 fold cross validation using the ALR_{vs}^{12s} classifier. Confusions are, on average, less than 5% and only among three activities: *walking*, *ascending*, and *descending*. This is reasonable since these three activities might have similar patterns depending on the intensity at which they are performed by the individual.

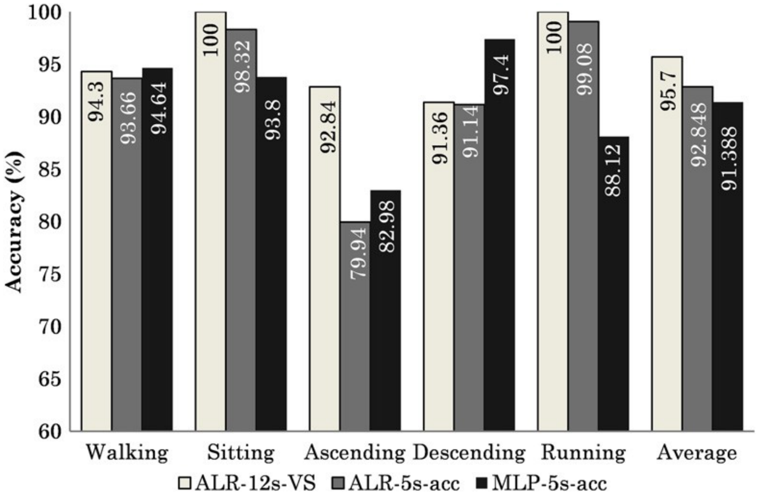


Fig. 6. Average accuracy of the 5 × 10 fold cross validation for the best classifiers in each dataset: *Dvs* and D_{acc} .

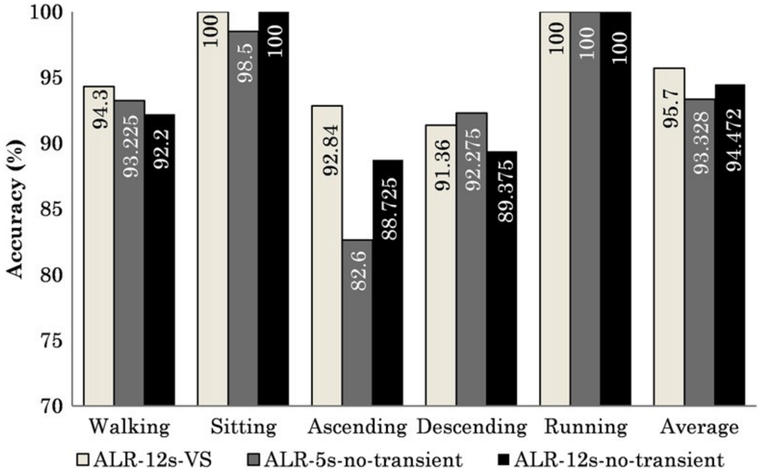


Fig. 7. Average accuracy of the 5 × 10 fold cross validation for the best classifiers in each dataset: *Dvs* and D_{nt} .

4.2.5. *Centinela vs. other state-of-the-art approaches*

It is worth mentioning that we cannot directly compare Centinela to all other HAR systems. This is mainly because each approach carries out a different experimental setup in terms of (1) the physical characteristics of the individuals the data were collected from, (2) the activities to be recognized, and (3) the evaluation methodology. However, to have a general idea of the benefits provided by Centinela, Table 8[1] is presented to compare our system to three others working with the same set of activities.

Table 7

Average of the confusion matrices from the 5 × 10 fold cross validation using *ALR*^{12s}.

Walking		Sitting		Descending	Running
		Ascending			
Walking	94.28	0	7.16	3.64	0
Sitting	0	100	0	0	0
Ascending	2.71	0	92.84	5	0
Descending	3.01	0	0	91.36	0
Running	0	0	0	0	100

Firstly, eWatch is introduced in [13] as an online activity recognition system which embeds sensors and a microcontroller within a device that can be worn as a sport watch. Four sensors are included, namely accelerometer, light sensor, thermometer, and microphone. Although eWatch features up to 92.5% overall accuracy, it achieves less than 70% of accuracy for activities such as *descending* and *ascending*. Centinela reduces the misclassification of these activities by considering vital signs and it reaches 91.36% and 92.84% respectively. Also, in eWatch, data were collected under controlled conditions, thus, a lead experimenter supervised and gave specific guidelines to the subjects on how to perform the activities [13]. In 1999, Foerster et al. [6] demonstrated 95.6% of accuracy for ambulation activities in a controlled data collection experiment, but in a natural environment, the accuracy dropped to 66%!

Table 8

Per-activity percentage accuracy of Centinela and other state-of-the-art approaches for human activity recognition.

	Centinela	eWatch [13]	HAAR [9]	HMM [30]
Walking	94.28	92*	**94.43**	90.3
Running	**100**	93*	99.89	87.01
Ascending	**92.84**	68*	84.12	89.15
Descending	91.36	67*	86.42	**92.5**
Sitting	**100**	99*	99.15	N/A
Total	**95.7**	92.8*	93.91	89.74

Secondly, the system proposed in [9] uses HAAR filters to extract features and the C4.5 algorithm for classification purposes. In activities such as *running, walking* and *sitting*, their results are fairly close to Centinela's; yet for *ascending* and *descending*, Centinela is slightly better. Furthermore, only four individuals with unknown physical characteristics participated in the study presented in [9]. Collecting data from such number of people might be insufficient to provide flexible recognition of activities on new users.

Finally, the system presented in [30] applies Hidden Markov Models and Neural Networks resulting in almost the same accuracy than Centinela's for *ascending* and *descending*. But, Centinela recognizes the activities *running* and *walking* more accurately. In addition, data collected in [30] are from one single individual, which implies that a subject-dependent analysis was performed (refer to Section 2.4 for the definition and disadvantages of a subject-dependent analysis).

[1] Values marked with an asterisk (*) are approximated. They were obtained from a chart included in [13].

In the present work, all the data were collected under naturalistic conditions and a subject-independent analysis was applied for the evaluation. Next, Centinela is qualitatively compared to other approaches that recognize a different set of activities:

In 2002, Randel et al. [19] introduced a system to recognize ambulation activities which makes use of Root Mean Square for feature extraction and Backpropagation Neural Networks for classification. The authors claim to have reached up to 95% of accuracy but also emphasize that results were analyzed after further person specific training. This implies a subject- dependent analysis which, again, might not be convenient for real applications. Additionally, the paper does not include information regarding the characteristics of the subjects, the data collection protocol, and the confusion matrix.

In [15], the authors claim that the average classification accuracy for ambulation activities is 94.6%, but this is only for subject-dependent analysis. They hardly reach 56% of accuracy in the subject-independent evaluation. The same situation occurs in the system proposed in [18]. They compare different classification algorithms to recognize ambulation activities with a 95% subject-dependent accuracy, but only reach 86% of accuracy for the subject-independent analysis.

Ermes et al. [16] developed an online system that reaches 94% overall average accuracy but they only applied a subject- dependent evaluation. Besides, their data were collected from only three subjects. Kao et al. [17] also present an online system with 94.71% overall accuracy, but they include other activities such as *hitting*, *knocking*, *working at PC*, and *brushing teeth*. The activities *descending* and *ascending*, included in this work, are not considered there. He et al. [7,8,10] achieved up to 97% of accuracy but only considered four activities: *running*, *being still*, *jumping*, and *walking*. These activities are quite different in nature, which considerably reduces the level of uncertainty thereby enabling higher accuracy. In this work, we consider other activities such as *ascending* and *descending* stairs which open new possibilities for real applications and require a higher level of discrimination.

In 2010, Berchtold et al. introduced *ActiServ* as an activity recognition service for mobile phones [31]. They make use of a fuzzy inference system to classify daily activities, achieving up to 97% of accuracy. However, this requires a runtime duration in the order of days! When their algorithms are executed to meet a feasible response time, the accuracy drops to 71%. ActiServ can also reach up to 90% after personalization, in other words, a subject-dependent analysis.

The approach proposed in [11] exhibits high recognition accuracy (about 93%). Nonetheless, all the data were collected inside the laboratory, under controlled conditions.

4.3. *Beyond the recognition of human activities*

Centinela can be visualized as a general tool for pattern recognition in time series data. As a matter of fact, it could be extended to provide inference not only on the individual's activity, but also their gender and other personal information. While Centinela can accurately make statements such as *the individual is running*, it would be even more useful to deliver additional information, such as *a female individual between 130 and 150 pounds is running*. However, in order to achieve gender, weight, and activity recognition, these new attributes have to be included as goal attributes in the classification context as well. And, in this case, instead of a single attribute classification problem, we would be dealing with a multiattribute classification problem (activity, gender, and weight). As most classification algorithms only support a single class attribute, this problem becomes quite challenging. One possible solution might be to create a composite class attribute whose domain is the Cartesian product of the atomic class attributes (i.e., all possible combinations of activities, weights, and genders). Of course, many more confusions are expected due to additional uncertainty being introduced to the system. Moreover, it would be required to acquire greater amount of data from individuals throughout the entire spectrum of possible values, i.e., females and males, from all possible weights, heights, ages, and so forth. Such a multiattribute classification problem is beyond the goals of this paper, but is part of our current research work.

5. Conclusions

This paper presents Centinela, a human activity recognition system based upon acceleration and vital sign data. An extensive evaluation was performed for three feature sets (i.e., statistical, structural, and transient), eight classification algorithms, and three different window sizes. Overall, the highest mean accuracy achieved was 95.7% for the Additive Logistic Regression algorithm with a window size of 12s and considering both vital signs and acceleration data. This fact supports the hypothesis that vital signs together with acceleration data can be useful for recognizing certain human activities more accurately than by considering acceleration data only. There are some activities, however, for which acceleration data are enough to perform accurate classification.

Ensembles of classifiers turned out to have the highest accuracy, yet they require more training and testing time. This introduces new challenges to achieve online activity recognition. Another important point of discussion is the placement of the sensor. We believe that placing the accelerometer on the chest of the person avoids confusions that may arise if it is placed on the wrist [17]. As a matter of fact, Centinela reaches 92.84% of accuracy with acceleration data only, which is better than most of the previously proposed approaches. Centinela also features a portable and unobtrusive real-time data collection platform, which allows not only for activity recognition but also for monitoring health conditions of target individuals.

References

[1] A.J. Perez, M.A. Labrador, S.J. Barbeau, G-sense: a scalable architecture for global sensing and monitoring, IEEE Network 24 (4) (2010) 57–64.

[2] J. Yin, Q. Yang, J. Pan, Sensor-based abnormal human-activity detection, IEEE Transactions on Knowledge and Data Engineering 20 (8) (2008) 1082–1090.

[3] E. Kim, S. Helal, D. Cook, Human activity recognition and pattern discovery, IEEE Pervasive Computing 9 (1) (2010) 48–53.

[4] P. Turaga, R. Chellappa, V. Subrahmanian, O. Udrea, Machine recognition of

human activities: a survey, IEEE Transactions on Circuits and Systems for Video Technology 18 (11) (2008) 1473–1488.

[5] R.T. Olszewski, C. Faloutsos, D.B. Dot, Generalized feature extraction for structural pattern recognition in time-series data, Tech. rep., In Time-Series Data, Ph.D. dissertation, Carnegie Mellon University, 2001.

[6] F. Foerster, M. Smeja, J. Fahrenberg, Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring, Computers in Human Behavior 15 (5) (1999) 571–583.

[7] Z.-Y. He, L.-W. Jin, Activity recognition from acceleration data using ar model representation and svm, in: Proceedings of International Conference on Machine Learning and Cybernetics, vol. 4, 2008, pp. 2245–2250.

[8] Z. He, Z. Liu, L. Jin, L.-X. Zhen, J.-C. Huang, Weightlessness feature; a novel feature for single tri-axial accelerometer based activity recognition, in: Proceedings of the 19th International Conference on Pattern Recognition, 2008, pp. 1–4.

[9] Y. Hanai, J. Nishimura, T. Kuroda, Haar-like filtering for human activity recognition using 3d accelerometer, in: Proceedings of the 13th IEEE Digital Signal Processing Workshop, 2009, pp. 675 –678.

[10] Z. He, L. Jin, Activity recognition from acceleration data based on discrete consine transform and svm, in: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 5041–5044.

[11] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee, J.-S. Wang, Online classifier construction algorithm for human activity detection using a tri-axial accelerometer, Applied Mathematics and Computation 205 (2) (2008) 849–860.

[12] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, in: IEEE Pervasive, 2004, pp. 1–17.

[13] U. Maurer, A. Smailagic, D.P. Siewiorek, M. Deisher, Activity recognition and monitoring using multiple sensors on different body positions, in: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, 2006, pp. 113–116.

[14] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, I. Korhonen, Activity classification using realistic data from wearable sensors, IEEE Transactions

on Information Technology in Biomedicine 10 (1) (2006) 119–128.

[15] E.M. Tapia, S.S. Intille, W. Haskell, K. Larson, J. Wright, A. King, R. Friedman, Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor, in: Proceedings of the International Symposium on Wearable Computing, 2007.

[16] M. Ermes, J. Parkka, L. Cluitmans, Advancing from offline to online activity recognition with wearable sensors, in: Proceedings of the 30th IEEE Annual International Conference of Engineering in Medicine and Biology Society, 2008, pp. 4451–4454.

[17] T.-P. Kao, C.-W. Lin, J.-S. Wang, Development of a portable activity detector for daily activity recognition, in: Proceedings of IEEE International Symposium on Industrial Electronics, 2009, pp. 115–120.

[18] L.C. Jatoba, U. Grossmann, C. Kunze, J. Ottenbacher, W. Stork, Context-aware mobile health monitoring: evaluation of different pattern recognition methods for classification of physical activity, in: Proceedings of the 30th Annual IEEE International Conference in Engineering in Medicine and Biology Society, 2008, pp. 5250–5253.

[19] C. Randell, H. Muller, Context awareness by analysing accelerometer data, in: Proceedings of the Fourth International Symposium on Wearable Computers, 2000, pp. 175–176.

[20] Zephyr Bioharness BT Website, http://www.zephyr-technology.com/bioharness-bt.html.

[21] Sprint Java ADP, http://developer.sprint.com/site/global/home/p_home.jsp.

[22] G. Dahlquist, A. Bjorck, Numerical methods in scientific computing, vol. 1, SIAM, 2008.

[23] H. Zhang, The optimality of naive Bayes, in: Proceedings of FLAIRS Conference, 2004.

[24] P. Antal, Construction of a classifier with prior domain knowledge formalised as Bayesian network, in: Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, vol. 4, 1998, pp. 2527–2531.

[25] I.H. Witten, E. Frank, Data Mining, Practical Machine Learning Tools and

Techniques, 2nd ed., Elsevier, 2005.

[26] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, Annals of Statistics 28 (1998) 2000.

[27] W.J. Conover, Practical Nonparametric Statistics, 3rd ed., Wiley, 1999.

[28] Waikato Environment for Knowledge Analysis (WEKA), http://www.cs.waikato.ac.nz/ml/weka/.

[29] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, M. Srivastava, Using mobile phones to determine transportation modes, ACM Transactions on Sensor Networks 6 (2) (2010) 1–27.

[30] C. Zhu, W. Sheng, Human daily activity recognition in robot-assisted living using multi-sensor fusion, in: Proceedings of the IEEE International Conference on Robotics and Automation, 2009, pp. 2154–2159.

[31] M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, M. Beigl, Actiserv: Activity recognition service for mobile phones, in: Proceedings of the 2010 International Symposium on Wearable Computers, 2010, pp. 1–8.