Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 2416 – 2436
© TÜBİTAK
doi:10.3906/elk-2010-75

# Ordered physical human activity recognition based on ordinal classification

**Duygu BAĞCI DAŞ**[1,*], **Derya BİRANT**[2]
[1]Graduate School of Natural and Applied Sciences, Dokuz Eylül University, İzmir, Turkey
[2]Department of Computer Engineering, Faculty of Engineering, Dokuz Eylül University, İzmir, Turkey

**Received:** 17.10.2020 • **Accepted/Published Online:** 21.05.2021 • **Final Version:** 23.09.2021

**Abstract:** Human activity recognition (HAR) is a critical process for applications that focus on the classification of human physical activities such as jogging, walking, downstairs, and upstairs. Ordinal classification (OC) is a special type of supervised multi-class classification in which an inherent ordering among the classes exists, such as low, medium, and high. This study combines these two concepts and introduces an approach to "human activity recognition based on ordinal classification" (HAROC). In the proposed approach, ordinal classification is applied to human activity recognition where the physical activities can be ordered by using their signals' band power values. This is the first study that investigates the performance of the HAROC approach by combining the ordinal classification with eight different base learners. Besides, this study is also original in that it examines the effects of the demographic characteristics of the participants (i.e., sex, age, weight, and height) on the classification performance. The experiments carried out on a real-world dataset show that the proposed HAROC approach is an effective method for human activity recognition tasks.

**Key words:** Human activity recognition, ordinal classification, machine learning

## 1. Introduction

Human activity recognition (HAR) is defined as the automatic identification of human physical activities, such as eating, walking, sitting, running, and so forth. Recently, HAR has been employed in various fields such as rehabilitation [1], health care [2], military [3], and security [4]. HAR process is performed via machine learning algorithms that use data obtained from cameras, radars, or wearable sensors. While cameras and radars have some drawbacks (i.e. light, limited area, operation range), wearable sensors become useful thanks to technological advancements. The wearable sensors can be placed directly or indirectly on a human body or on an accessory such as a watch, belt, smartphone, goggles, and shoes. The sensors that are currently being used are accelerometer, gyroscope, microphone, biosensor, and magnetometer. The sensors can be placed directly or indirectly on a human body or on an accessory such as a watch, belt, smartphone, goggles, and shoes. It is possible to obtain significant data such as pulse, speech, posture, motion, or body temperature to be used in the HAR process. In our study, we used the sensor data that was collected by the accelerometer sensor of the smartphone placed in the front pockets of the participant's trousers.

Ordinal classification (OC) solves the classification problems where the class labels in the dataset follow an order, such as hot $\succ$ warm $\succ$ cold class labels in the weather prediction problem. The ordinal classification has recently received much attention in the machine learning community with a consequence of the growing amount of real-world applications, such as sentiment analysis (i.e. sad, natural, and happy), human age estimation (i.e.

*Correspondence: duygu.bagci@deu.edu.tr

young, adult, and old), customer segmentation (i.e. gold, silver, and bronze), credit scoring (i.e. low, medium, and high-risk levels), survey analysis (i.e. agree, neutral, and disagree), and medical diagnosis (i.e. health status, cancer stage, or tumor grade). The importance of considering inter-classes relations has already been proven in [5].

The standard HAR applications disregard the ordering information in class labels (human activities); however, it usually leads to a significant loss of performance. As proven in many studies [5–7], the ordinal classification approach usually outperforms the nominal classification approach for the datasets that include ordered class targets. If we use a nominal classification technique for an ordinal dataset, ordering information is lost since the relationship among the class labels is discarded. Therefore, the classification performance may decrease as demonstrated in [5–7]. In this study, we put forward that OC can be effectively applied to HAR since it is possible to order human activities in terms of energy consumption. For instance, "walking" requires more energy than "sitting"; whereas, it requires less energy than "jogging".

The novelty and main contributions of this study can be summarized in three-folds. (i) This is the first study that combines two concepts HAR and OC, and it proposes an approach of *"human activity recognition based on ordinal classification* (HAROC)". (ii) It is the first time that this study investigates which base learner is effective when used with ordinal classification for HAR. (iii) This study is also original in that it examines the effects of the demographic characteristics of the participants (i.e. sex, age, weight, and height) on the ordinal classification performance.

In the experimental studies, we demonstrated the effectiveness of the ordinal classification on the HAR for the first time. The proposed HAROC method was tested on a real-world dataset collected from 24 participants with different ages and sexes. Each human physical activity in the dataset requires a different amount of power. Such actions were ordered by considering the band powers of their unique signals. Afterward, an OC algorithm was performed by using those ordered activities to build a HAR model. The experimental results showed that the proposed HAROC approach accurately recognized six human physical activities, including walking, jogging, standing, sitting, downstairs, and upstairs.

The paper is structured as follows. Section 2 gives the related studies on HAR and OC topics. Section 3 presents the background information about HAR and OC. This section also explains the proposed HAROC approach in detail. Section 4 gives a brief description of the dataset and presents the experimental results. Finally, a brief conclusion and possible future works are given in Section 5.

## 2. Related work

*Human activity recognition* (HAR) has attracted many researchers due to its importance and wide usage in many fields. There are numerous studies on HAR reported in the literature. Chen and Shen [8] presented a detailed performance analysis of phone sensors for HAR. They measured the classification accuracy by considering the phone placement settings, the combination of sensors, user-space sensibility, and data imbalance. Chen et al. [9] used the support vector machine technique to perform HAR by considering the orientation of the smartphone (portrait or landscape), the placement of the phone (i.e. leg, arm), and participant variations. Since the feature extraction and feature selection phases play an important role in the final system performance, some previous studies [10, 11] concentrated especially on these issues.

In the literature, most of HAR studies [12] focus on detecting daily living human activities such as walking, standing, downstairs, and upstairs. However, nowadays, some HAR studies try to identify different types of activities such as transportation-based activities [13] (i.e. riding a car, bike, bus, or train), military-based activities [3] (i.e. run, jump, or jump-rope), and sports activities [14] (i.e. playing basketball, football, or volleyball). In addition to these high-level activities, some HAR studies have been focused on the transitions between the activities such as sit-to-stand or stand-to-sit [2].

The sensors can be easily placed on many different body parts (i.e. arm, leg, waist, or wrist) [8, 9]. Janidarmian et al. [15] investigated the ideal sensor location depending on the physical activities being studied. Likewise, Shoaib et al. [10] explored the possible smartphone positions to understand whether changing phone locations improves activity recognition accuracy or not. The most widely-used sensors for HAR are accelerometer (A) and gyroscope (G), whereas different sensors such as magnetometer (M) and linear acceleration (LA) may also be utilized. Some studies [8, 10–12] evaluated the different combinations of sensor types (accelerometer, gyroscope, magnetometer, and linear acceleration) to find an optimal sensor combination and the most relevant sensor types for distinguishing different activities. Rather than using a single device, some studies [16] used multiple devices (i.e. smartphones and smartwatches) to perform HAR.

The HAR studies aforementioned disregard the demographic characteristics of the participants. However, human activities have a strong subjective characteristic, which is associated with many different factors, such as age, sex, height, and weight. For example, the youth's stride frequency is usually faster compared to the elder; men jump higher than women in general; a tall person's step length is longer than a short person's step. For this reason, unlike these previous studies, we took into account the demographic characteristics of the participants.

Ordinal classification (OC) has been studied by machine learning researchers in many different fields such as environment [17], finance [18], history [19], health, and transportation. However, there is no study reported in the literature combining HAR and ordinal classification. The OC approaches can be categorized into three groups: threshold approaches, naive approaches, and binary decomposition-based (BDB) approaches [20]. The threshold approaches obtain a set of thresholds by dividing the target variables into successive intervals, where each class label belongs to an interval limited by these thresholds. The naive approaches use an appropriate simplifying assumption on the class labels to treat OC problems as if they were standard classification problems. For instance, one possible solution is to use different weights for different class labels, or another alternative solution is to map the class labels into numeric values and then implement a standard regression algorithm such as the support vector regression. The ordinal binary decomposition (OBD) approaches [5] transform the original ordinal classification task into a set of binary classification tasks. Each sub-problem is separately solved by a binary classification algorithm, and then the final class labels are determined by ultimately combining the outputs into one label. In this study, a new method based on the BDB approach is proposed.

In our study, we did not employ the traditional (nominal) classification technique like the previous studies; rather, we implement a more complex yet effective human activity recognizer based on ordinal classification. To the best of our knowledge, OC-based HAR has not been studied until now. Here, we present a detailed analysis of the impacts of two main factors (base learners and the demographic characteristics of the participants) together on OC-based HAR for the first time.

## 3. Materials and methods

### 3.1. Human activity recognition

*Human activity recognition* (HAR) is defined as the process of detection the human-based actions from observations obtained by sensors or cameras [2]. There are two main categories in sensor-based HAR, namely, the *ambient sensor-based* HAR (ASHAR) and the *wearable sensor-based* HAR (WSHAR). ASHAR methods use fixed sensor data attached in a specific location such as a wall, fridge, door, or placed in an environment like factories, parking lots, and smart homes. ASHAR applications are suitable for security or elderly care purposes at home. Light, temperature, flow, vibration, pressure, reed switch, infrared sensors, microphones, video cameras, or RFID tags can be given as examples to ASHAR sensors. Such sensors work in a limited area since they are fixed to a particular position where they are placed. WSHAR methods perform the HAR process by considering the raw data obtained from a wearable smart apparatus such as smartwatches, smartphones, and smart glasses. In this study, we focused on the wearable sensor-based HAR due to its practicality and popularity.

The existing HAR algorithms are capable of processing nominal data; however, they do not capture and reflect the orderings among human activities. HAR based on ordinal classification is yet to be explored. Our study bridges this gap by proposing a new approach for the first time.

### 3.2. Ordinal classification

Ordinal classification (OC) is a special type of multi-class classification that deals with class attributes whose labels exhibit some form of natural ordering. For instance, an ordinal class attribute can represent the grades of students that there exists an order among the class labels such that *excellent* $\succ$ *good* $\succ$ *average* $\succ$ *bad*, where $\succ$ represents that the former class label is better than the latter class label. In the case of ordinal classification, a sample from the highest class is significantly different from a sample from the lowest class, whereas two samples, one from the highest and the other from the middle class, are relatively close to each other. Therefore, considering such ordering information aims to improve the performances of classification models. Unlike the previous OC studies [21], our study has the ability to consider the natural order of the physical activities according to their signals' band power values such as *jogging* $\succ$ *walking* $\succ$ *downstairs* $\succ$ *upstairs* $\succ$ *standing* $\succ$ *sitting*.

### 3.3. Proposed approach (HAROC)

The traditional HAR task is limited to using only nominal data to construct a classification model. Hence, it does not capture and reflect the orderings among human activities, which may lead to building inefficient models. It is possible to state that human activities are in order according to their characteristics. For example, a jogging activity requires a higher level of energy demand than a standing activity, since jogging involves more intense muscle contractions. Based on this motivation, we proposed a novel approach, called human activity recognition based on ordinal classification (HAROC).

Figure 1 shows the general workflow of the proposed HAROC approach. The method consists of nine main phases: data acquisition, data preprocessing, segmentation, feature extraction, OC representation, model training, performance evaluation, activity classification, and decision support. The data acquisition phase comprises obtaining raw signal data from an accelerometer sensor embedded in a wearable device (i.e. a smartphone, smartwatch, or smart glass) during human movement. Afterward, the obtained raw data is transferred to a data center via a specific communication technology such as WiFi or Bluetooth. In the data preprocessing phase, the raw signal data is filtered to remove the noise using a filtering algorithm. In the segmentation phase, the filtered large signal data is split into smaller fixed-size blocks, called windows, by utilizing a sliding window method. Here, each window corresponds to a single human activity such as walking, standing, or sitting. The feature extraction phase includes the extraction of informative features from each window by further processing such as mean, standard deviation, peak-to-peak value, kurtosis, root mean squared (RMS) value, the number of zero crossings, crest factor, skewness, signal entropy, and RMS velocity. The OC representation phase transforms the original dataset into a set of binary datasets to encode the ordering information among the class labels. Each binary dataset represents a class and it is created by taking the ranking of this class into account. The binary dataset includes a target value, which is evaluated by considering whether the class value of the original dataset is equal/below or above the rank of the corresponding class. In the model training phase, each binary dataset is trained by a base learning algorithm such as decision tree (DT), neural network (NN), support vector machine (SVM), naive Bayes (NB), partial decision tree (PART), $k$-nearest neighbor (KNN), random forest (RF), and AdaBoost (AB). The evaluation phase assesses the accuracy of the human activity recognition model by employing the $k$-fold cross-validation method. The model that is the best in classification is selected to recognize human activities. The activity classification phase uses the selected model to recognize the physical activities of persons such as standing, sitting, walking, jogging, downstairs, and upstairs. The final prediction is utilized in the decision support phase to provide guidance to the decision-maker.

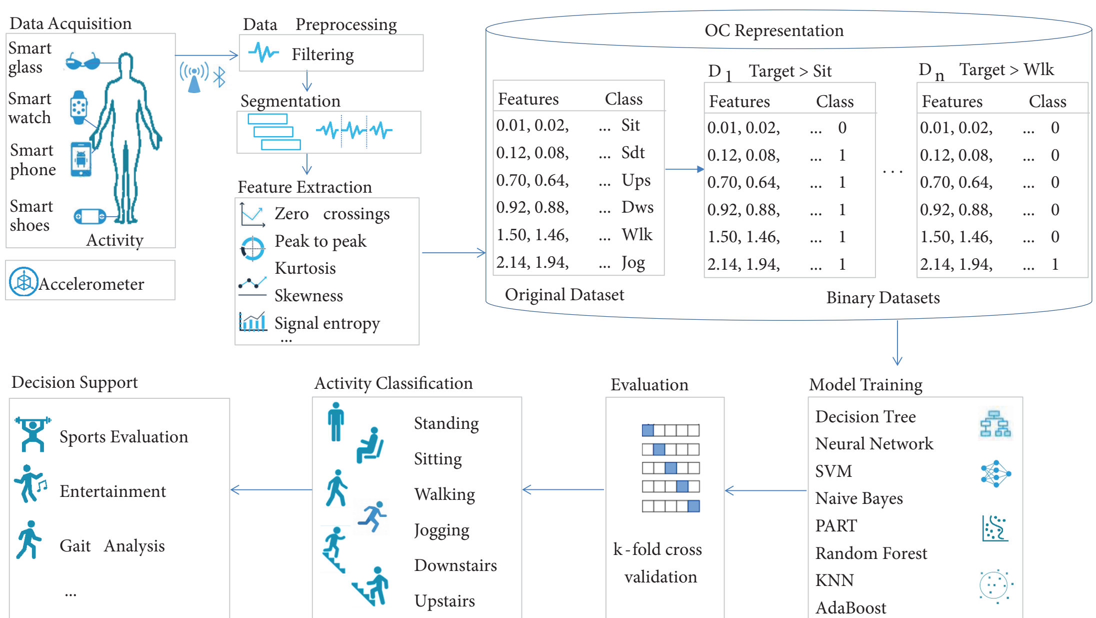


**Figure 1**. The main phases of the proposed human activity recognition approach based on ordinal classification (HAROC).

In this study, we propose an approach that has the ability to consider the natural order of the physical activities. To determine the ranking among the activities, in this study, we calculated the band powers of the signals by using MATLAB. Here, it is possible to say that some of the human activities can close to each other in terms of energy expenditure such as ”reading a book” and ”watching TV”. However, when we mathematically compute their band powers of the sensor signals, close but different values are obtained, so the similar activities can also be ordered according to the obtained numerical values. Another alternative is to assume that similar activities have the same rank. The proposed approach in this study is fundamentally designed for physical human activities which generally require different energy consumption. Some of the activities can involve exercises that include regular and repetitive body movements, some of the activities can use the muscles and joints more heavily, some of them can increase heart and respiratory rate and result in fatigue of different intensities.

It can be noted here that downstairs and upstairs activities are relatively close to each other. The difference between these activities can be changed according to some parameters such as duration of collected data, rate of oxygen consumption (as an indirect measure of the rate of energy expenditure) [22], the demographic characteristics of pedestrian (i.e. age, weight), the movement characteristics of pedestrians, and the staircase geometry including step height, tread depth, step width, the slope of stairs, total moving height on a stair, and presence and location of handrails [23]. For instance, related to the step height, it is reported that the energy expended to climb a stairway with a one-step climbing strategy is greater than with a two-step climbing strategy [22].

Figure 2 shows sample accelerometer sensor signals in the $x$, $y$, and $z$ axes ($Ax$, $Ay$, and $Az$) over time for each activity. Based on the calculated band powers using MATLAB and signal graphs drawn via R-Studio, the descending order among the human activities were obtained as follows: *jogging* $\succ$ *walking* $\succ$ *downstairs* $\succ$ *upstairs* $\succ$ *standing* $\succ$ *sitting*. For example, going upstairs requires higher power and effort than standing, since it involves more intense muscle contractions. It is noted that pedestrians walk downstairs faster than upstairs [23], and due to increased speed, a higher z-direction of the accelerometer (Az) for going downstairs vs upstairs can be achieved [24]. Furthermore, when moving downstairs, they have to notice the steps and should also do extra work to prevent themselves from falling from the stairs [23]. Moreover, a pedestrian may require different relaxation times when walking upstairs and downstairs [23]. A pedestrian has walked upstairs for a long time, moves slowly, sometimes stops to have a rest and spends more time since he/she feels tired. Climbing invokes a high rate of energy expenditure; however, downstairs can be energetically more expensive per time unit due to a shorter period. Since step rate is higher during going downstairs this may result in faster rates of muscle shortening, which increases energy turnover.

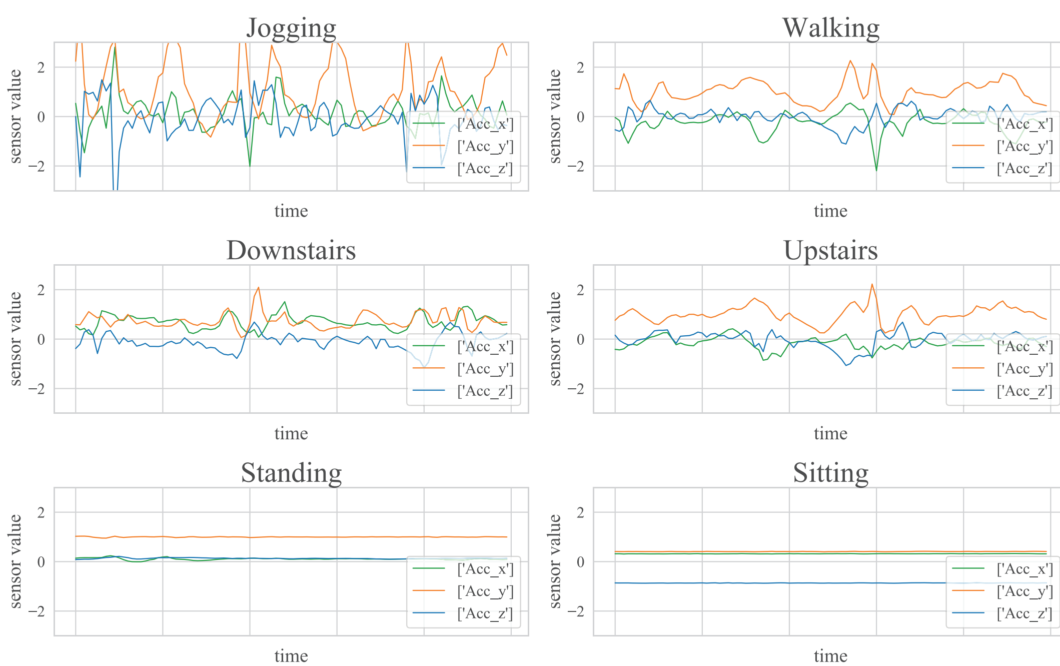


**Figure 2**. Sample accelerometer sensor signals in the x, y, z-axis over time for each activity.

#### 3.3.1. The formal definition of the proposed method

Consider a dataset $D$ having $n$ instances such that $D = \{(x_i, y_i)|i = 1, 2, ..., n\}$. Each data point $(x_i, y_i)$ in the dataset $D$ consists of $x_i$ and $y_i$, which are the input and the relevant human activity class label, respectively. The $d$-dimensional input vector $x_i$ belongs to the feature space $X \subseteq \mathbb{R}^d$, while the class label $y_i$ is related to the activity set $Y = \{c_1, c_2, ..., c_k\}$ such as jogging, walking, and sitting, where $k$ is the number of activities (number of classes). The activity classes are oredered in a consistent manner such that $c_1 \prec c_2 \prec ... \prec c_k$, where $\prec$ denotes this order information. The main goal in this kind of problem is to find a decision function $f : X \rightarrow Y$ that can be able to predict the activity class for any wearable sensor data $x$ according to the best fit possible. It can be noted here that this classification task consists of estimating porbability $P(y|x)$ for any $y \in Y$.

**Definition 1** *Human activity recognition based on ordinal classification (HAROC) refers to the problem that uses signal data obtained from sensors embedded in a wearable device during human movement, builds a prediction model by considering the orders among human activities, and aims to correctly predict the activity class label y of some unseen input x.*

The proposed HAROC method consists of three main steps. In the first step, called activity decomposition, the ordinal HAR problem involving $k$ activities is converted into $k - 1$ binary classification problems to represent the ordering among the activities such that *jogging* $\succ$ *walking* $\succ$ *downstairs* $\succ$ *upstairs* $\succ$ *standing* $\succ$ *sitting*. In the second step, a base learning algorithm is applied to build a model for each binary dataset separately. In the last step, for prediction, given a search sample $x$ is processed by each of the $k - 1$ model, the probability of each activity is calculated such as P(class $\succ$ walking), P(class $\succ$ downstairs), P(class $\succ$ upstairs), P(class $\succ$ standing), P(class $\succ$ sitting), and the class having the highest probability is assigned to the given sample.

**Definition 2** *Activity decomposition is to decompose the ordinal HAR problem involving k activities into k - 1 binary classification problems to encode the ordering information among the activities. In this process, for the* $i^{th}$ *binary problem, the label* $y_j$ *associated with the instance* $x_j$ *is replaced with* $y_j = 0, \forall y_j \preceq c_i$, *and,* $y_j = 1, \forall y_j \succ c_i$.

**Definition 3** *Assume that the human activities are denoted as* $Y = \{c_1, c_2, ..., c_k\}$, *and the activities are ordered according to the ranking structure* $c_1 \prec c_2 \prec ... \prec c_k$. *Let* $M_i$ *for i = 1, 2, ..., k - 1, denotes the model constructed for the* $i^{th}$ *binary classification problem. Given a search instance x, a prediction* $M_i(x)$ *is considered as an estimation of the probability* $P(A_x \succ c_i)$, *where* $A_x$ *denotes the activity class label of x. Activity probabilities on* $Y$ *are formulated as follows:*

$$\begin{aligned} P(c_1) &= 1 - P(A_x \succ c_1) \\ P(c_i) &= P(A_x \succ c_{i-1}) \times (1 - P(A_x \succ c_i)), 1 < i < k \\ P(c_k) &= P(A_x \succ c_{k-1}) \end{aligned} \tag{1}$$

At the end of the calculation of the activity probabilities, the class that has the highest probability is predicted as the final label for a given instance. The outputs of the classifiers are obtained by considering the evaluation of class probabilities, as presented in many studies [25–27]. For example, the class probability estimation of the decision tree algorithm is explained in [25], the probability calculation of KNN is given in

[25], and the probabilistic output of the Naive Bayes algorithm is presented in [25]. For instance, the class probability of KNN is calculated as $P(c_i|d) = k_m/k$, where $k_m$ denotes the number of samples belonging to $c_i$ in $k$ neighbor samples.

#### 3.3.2. The algorithm of the proposed method

Algorithm 1 shows the pseudocode of the proposed HAROC method. The algorithm consists of three steps. In the first step, called *activity decomposition*, the method converts the original ordinal dataset $D$ into a set of binary datasets $\{D_i\}_{i=1}^{k}$ to represent the ranking among activities. In this process, the label $y_j$ associated with the instance $x_j$ is replaced with $y_j = 0, \forall y_j \preceq c_i$, and, $y_j = 1, \forall y_j \succ c_i$. In the second step, a separate model $M_i$ is built for each binary dataset $D_i$ by using a base learning algorithm. In the last step (classification step), a new unseen sample $x$ is processed by all models $M^*$, and their predictions $M_i(x)$ are compared, and the activity class having the highest probability is assigned to $x$ as a label.

**Algorithm 1:** Human activity recognition based on ordinal classification (HAROC).

```
Inputs:  D: the ordinal HAR dataset D = {(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)} with n intances
Y: ordinal activity class labels y ∈ {c_1, c_2, ..., c_k} with an order c_1 ≺ c_2 ≺ ... ≺ c_k
k: the number of activities (classes)
Outputs: M*: ordinal prediction model
M*(x): activity of an unseen sample x
// Step 1 - Construction of binary datasets (activity decomposition)
for i ← 1 to k − 1 do
    foreach (x_j, y_j) in D do
        if (y_j ⪯ c_i) then
            D_i.Add(x_j, 0);
        else
            D_i.Add(x_j, 1);
        end
    end
end
// Step 2 - Construction of models
for i ← 1 to k − 1 do
    M_i = Train(D_i)
    M* = M* ∪ M_i
end
// Step 3 - Classification of a new sample x
y = M*(x) = MAX (
P(c_1) = 1 − P(A_x ≻ c_1)
for j ← 2 to k − 1 do
    P(c_j) = P(A_x ≻ c_{j−1}) × (1 − P(A_x ≻ c_j))
end
P(c_k) = P(A_x ≻ c_{k−1}))
```

The time complexity of the proposed HAROC algorithm is given by $O(n.(k-1) + T(n).(k-1))$, where $n$ is the number of instances in the dataset, $k$ is the number of activities (classes), and $T(n)$ refers to the time required for the execution of a base learning algorithm on $n$ instances. The time complexity depends on the number of activities $(k)$ since the method builds $k-1$ models.

#### 3.3.3. The advantages of the proposed method

The proposed method (HAROC) has a number of advantages that can be summarized as follows:

- The traditional HAR is limited to using only nominal data to build a model and disregards the ordering information among the activities. The main advantage of the HAROC method is that it overcomes this limitation and designs a classification model by considering the ranking among the activities.
- An important advantage of the HAROC method is that it can be used with the combination of any base learner such as DT, KNN, SVM, and NN. The base learner is entirely unaware of the HAROC method, in fact, it simply learns from the ordinal instances as if they were regular instances.
- Another advantage is that the HAROC method can be applied to any HAR data without any prior information about the given dataset. It doesn't make any specific knowledge and specific assumptions for the given data. The band powers of signals for each activity can be automatically calculated to determine the ordering among the activities.
- HAR based on ordinal classification is yet to be explored. Hence, our method expands the application field of the ordinal classification.
- The HAROC method enables enormous applications since a large amount of ordinal HAR data can be easily generated in real-life for different purposes such as gait analysis, sports evaluation, and entertainment. Therefore, it covers a relatively wide domain.

## 4. Experimental studies

The effectiveness of the proposed method (HAROC) was demonstrated on a real-world dataset by combining the method with eight different base learning algorithms (SVM, NN, RF, AB, DT, KNN, PART, and NB) by using the WEKA machine learning tool [28]. In the experiments, the different parameter values of the algorithms were tested and the optimal values were determined since default parameters may mislead the result. The best parameter settings among the tested ones were selected as shown in Table 1. During the parameter tuning for KNN, the maximum value of $k$ (the number of neighbors) was selected as $log_2(n)$ as suggested in many previous studies [29–31], where $n$ is the number of instances in the dataset. The maximum $k = log_2(n)$ is a reasonable choice because if $k$ is determined as larger than this, the neighborhood may cover many samples from other classes and increases the calculation time [29]; on the other hand, if $k$ is selected as smaller than this, then the classification result can be highly sensitive to noisy samples. Therefore, the logarithm is a reasonable choice because the probability of overfitting dramatically increases if $k$ is determined as too large or too small. For instance, with a sample size $n$ of 80 to 800, this corresponds to a selection of $k$ ranging from 6 to 10, which are at acceptable levels. For AB, the RF algorithm was selected as the base classifier due to the good generalization ability. In the experiments, the performances of the classification models were evaluated according to the three metrics: accuracy, F-measure, and training time. The experiments were performed on a personal computer with a 3.2 GHz quad-core processor and 8 GB of RAM.

The results were obtained using the 10-fold cross-validation technique. In this technique, the data is randomly divided into ten equal-sized and non-overlapped partitions. One of the partitions is kept for testing, while the remaining partitions are used for training. Each test and training partitions differ for every fold. The average error calculated at the end stands for the validity of the model. Since the MotionSense data contains 24 sub-datasets, we applied the 10-fold cross-validation technique to each sub-dataset, each corresponds to a different participant. In other words, we run 8 different algorithms on 24 sub-datasets with 10 different train-test sets, which corresponds to 1920 runs. This process was also repeated for parameter tuning.

Since people differ in age, sex, and body size, they usually show different movement patterns even for the same type of activity. Therefore, only a person's own motion data can reflect his/her own movement completely

**Table 1**. The optimal parameter settings determined by hyperparameter tuning for each algorithm.

| Algorithm | Parameter | Description | Values tested | Optimal value |
|---|---|---|---|---|
| DT | C | Confidence Factor | 0.1, 0.2, 0.25 (d), 0.3 | 0.1 |
| SVM | K | Kernel type | PolyKernel, RBFKernel | PolyKernel |
| RF | I | Number of Iterations | 100 (d), 200, 300 | 300 |
| AB | BE | Base Estimator | RF, DT | RF |
| | I1 | Number of Iterations (RF) | 100 (d), 200, 300 | 300 |
| | I2 | Number of Iterations (AB) | 10 (d), 50, 100 | 10 |
| KNN | K | Number of Neighbors | 5, 7, 9, 11 | 5 |
| PART | C | Confidence Factor | 0.1, 0.2, 0.25 (d), 0.3 | 0.25 |
| | M | Minimum Number of Objects | 1, 2 (d), 3 | 1 |

and accurately. For this reason, in this study, personalization models were studied to take into account these physical and demographic factors.

### 4.1. Data description and preparation

In this study, we used the MotionSense dataset [32], where six different activities (jogging, walking, downstairs, upstairs, standing, and sitting) were performed 15 times by 24 participants with different ages, sexes, heights, and weights as given in Table 2. We used the observations collected by the accelerometer sensor embedded in a smartphone placed in the front pocket of the trousers of the participant. It contains a multivariate time-series data for each long trial (2 to 3 min duration). All raw data were collected at 50 Hz sample rate in the three-dimensional space ($x$, $y$, and $z$ axes) in a university campus environment. The dataset consists of 1,412,865 records in total.

**Table 2**. Summarized demographic characteristics of the participants in the dataset.

| | Weight (kg.) | Height (cm.) | Age |
|---|---|---|---|
| Min. | 48 | 161 | 18 |
| Max. | 102 | 190 | 46 |
| Mean | 72.13 | 174.21 | 28.79 |
| Std. | 16.21 | 8.90 | 5.44 |
| Participant count | | | |
| Male | | Female | |
| 14 | | 10 | |

Directly using raw signal data in machine learning is generally not practical because it does not carry sufficient information to identify a human physical activity and to distinguish different activities. Therefore, in this study, three data preparation techniques were performed to extract meaningful information from the raw data. (i) Data preprocessing – the signal was filtered by a low-pass filter method with 0.6 Hz cutoff frequency. (ii) Segmentation – to split the large signal data into smaller fixed-size blocks, we used a sliding window method with an overlap of 50% between consecutive windows and with the window length of 1 sec. (iii) Feature extraction – both time-domain features and frequency-domain features were extracted from accelerometer sensor data, including min, max, mean, median, standard deviation, the number of zero crossings, peak-to-peak value, root mean squared (RMS) value, kurtosis, skewness, crest factor, RMS velocity, and signal entropy. These features were extracted from each window using both single-axial and multi-axial methods, giving a total of 182 attributes.

### 4.2. Experimental results

In this study, the HAROC method with each different base learning algorithm is used in combination to build an independent classifier. From here on, the abbreviation of the method followed by the base learning algorithm is utilized to indicate to the related approach. For example, HAROC.NN refers to the HAROC method with the NN base learning algorithm.

Table 3 presents the classification accuracy results. Accuracy is the ratio of correct predictions to total predictions. Accuracy is a useful measure for the degree of correctness of the model and how it may perform generally. In other words, it gives information about the predictive power of the model. As can be seen in Table 3 the proposed method successfully classifies all the human physical activities with high accuracies (>97.5% in general). From the results, it is clearly seen that the best accuracy was achieved by the AdaBoost (AB) algorithm (97.96%) on average. HAROC.AB performed better than the rest for 14 out of 24 participants. It is followed by the Random Forest (RF) algorithm (97.94%). This is probably because of the fact that both RF and AB algorithms are ensemble learning methods that construct many prediction models to improve classification performance. It seems that the Naive Bayes algorithm is not a good choice as a base learner for HAROC since their combinations showed the worst performance (83.51%) for all the participants. In particular, the biggest accuracy difference of HAROC.AB and HAROC.NB was observed on the $4^{th}$ participant, where HAROC.AB increased accuracy by over 20%.

In addition to accuracy, we compared the performances of the methods in terms of F-measure. Although accuracy is the most common measurement in classification performance, global accuracy alone is not sufficient to determine the quality of the prediction; because it makes no distinction between the classes. *F-measure* is a useful metric because it considers the three quantities: true-positive (TP), false-positive (FP), and false-negative (FN). F-measure is a summary performance measurement, which is the harmonic mean of precision and recall. Using this metric is useful since it represents both precision and recall by a single score. The values of these measures are ranged between 0 and 1, where 1 is the best value. As seen from Table 4 all the F-measure values obtained by the HAROC-based methods are very close to 1, except the HAROC.NB method. For example, HAROC.AB and HAROC.RF methods achieved the F-measure values of 0.980 and 0.979 on average, respectively. According to the obtained results, it can be concluded that the models constructed by HAROC have good generalization capability to distinguish all the human activities, so they can be effectively used to recognize them well.

Figure 3 compares the average ranks of the methods for all participants. This process was started by assigning rank 1 to the most accurate method, rank 2 to the second-best one, and was continued to increase the rank parameter until giving rank 8 to the worst one. In the case of a tie, the average rank value was assigned to each method. According to the results, HAROC.AB exhibited the best performance among the others because it had the lowest rank value (1.79). The HAROC.RF method followed it with a rank value of 2.0. The HAROC.SVM and HAROC.KNN methods achieved similar performance, hence their ranks were close to each other, 3.79 and 3.83, respectively.

Table 5 presents the confusion matrix of the HAROC.AB model for the last participant to show the prediction performance of the proposed method on each human activity separately. Since the best accuracy

**Table 3**. Comparison of HAROC methods with different base learners in terms of accuracy (%) values.

| | HAROC. DT | HAROC. KNN | HAROC. PART | HAROC. NB | HAROC. SVM | HAROC. NN | HAROC. RF | HAROC. AB |
|---|---|---|---|---|---|---|---|---|
| P1 | 95.50 | **98.44** | 96.39 | 88.28 | 97.78 | 98.28 | 97.89 | 97.83 |
| P2 | 95.97 | 97.85 | 96.02 | 84.73 | 98.01 | 97.69 | 98.28 | **98.49** |
| P3 | 97.39 | 98.90 | 97.14 | 86.02 | 98.45 | 98.65 | 99.00 | **99.10** |
| P4 | 90.85 | 94.88 | 91.93 | 76.58 | 95.22 | 96.70 | **96.99** | 96.76 |
| P5 | 95.16 | **97.46** | 94.98 | 84.31 | 96.78 | 96.28 | 96.96 | 97.02 |
| P6 | 94.36 | 97.00 | 94.59 | 82.77 | 96.41 | 96.88 | **97.47** | 97.35 |
| P7 | 95.69 | 97.32 | 96.00 | 80.54 | 97.63 | 97.53 | 97.79 | **97.95** |
| P8 | 97.36 | 98.60 | 96.71 | 84.41 | 98.22 | 98.44 | **98.65** | **98.65** |
| P9 | 93.80 | 94.97 | 94.58 | 83.47 | 95.42 | 95.59 | 96.93 | **96.98** |
| P10 | 94.82 | 97.44 | 94.10 | 82.69 | 97.05 | 97.83 | 97.94 | **98.22** |
| P11 | 96.73 | 98.36 | 97.12 | 85.67 | 98.14 | 97.63 | 98.42 | **98.59** |
| P12 | 95.99 | 97.11 | 95.40 | 86.26 | **98.35** | 98.05 | 97.46 | 97.46 |
| P13 | 96.86 | 98.21 | 97.04 | 89.96 | 97.91 | 98.46 | **98.89** | **98.89** |
| P14 | 95.56 | 97.62 | 95.99 | 87.15 | 97.29 | 97.45 | **97.78** | **97.78** |
| P15 | 94.78 | 97.27 | 95.53 | 85.37 | 97.39 | 97.33 | **98.14** | **98.14** |
| P16 | 94.35 | 95.93 | 94.25 | 77.38 | 96.53 | 96.63 | 97.27 | **97.52** |
| P17 | 95.64 | 97.22 | 95.41 | 83.68 | 97.11 | 97.34 | 97.73 | **97.85** |
| P18 | 93.45 | 97.08 | 94.94 | 79.13 | 97.24 | 97.19 | **97.90** | 97.70 |
| P19 | 97.40 | 98.64 | 98.02 | 85.48 | **99.08** | 98.90 | 98.90 | 98.90 |
| P20 | 94.73 | 97.48 | 96.02 | 80.94 | **97.98** | 97.87 | 97.65 | 97.70 |
| P21 | 94.38 | **98.14** | 94.10 | 81.99 | 96.01 | 97.96 | 98.10 | 98.10 |
| P22 | 94.76 | 97.83 | 94.94 | 81.16 | 97.23 | 96.93 | **98.19** | 98.07 |
| P23 | 93.34 | 96.08 | 93.57 | 80.42 | 97.14 | 96.90 | **97.49** | **97.49** |
| P24 | 95.43 | 97.68 | 96.18 | 85.72 | 98.12 | 98.43 | **98.62** | 98.56 |
| AVG | 95.18 | 97.40 | 95.46 | 83.51 | 97.35 | 97.54 | 97.94 | **97.96** |

values were usually obtained by the AdaBoost algorithm, the confusion matrix was evaluated for this algorithm only. It can be seen from the matrix that the constructed model usually had no difficulty in identifying all human activities. For example, 362 out of 367 walking activities were predicted correctly; however, only 5 of them were misclassified by the constructed model.

Although each human physical activity was distinguished and classified with high accuracy, downstairs and upstairs activities were slightly confused by the algorithm during classification. It can be concluded from the matrix that the best performance was achieved on the sitting and standing activities.

Figure 4 shows the training time comparison of the methods in seconds. HAROC.KNN seems like the fastest one; however, it does not hold a separate training phase since it is a lazy learning method. The HAROC.NB method is an efficient method in terms of execution time (0.17 sec.); however, it has the worst performance (83.51%) in terms of classification accuracy. It is followed by the HAROC.SVM method (0.29 sec.).The HAROC.NN method takes a very long time of learning (1026 sec.) since it requires many iterations when converging to the optimum solution.

**Table 4**. Comparison of HAROC methods with different base learners in terms of F-measure values.

| | HAROC. DT | HAROC. KNN | HAROC. PART | HAROC. NB | HAROC. SVM | HAROC. MLP | HAROC. RF | HAROC. AB.RF |
|---|---|---|---|---|---|---|---|---|
| P1 | 0.955 | **0.984** | 0.964 | 0.880 | 0.978 | 0.983 | 0.979 | 0.978 |
| P2 | 0.960 | 0.978 | 0.961 | 0.841 | 0.980 | 0.977 | 0.983 | **0.985** |
| P3 | 0.974 | 0.989 | 0.972 | 0.851 | 0.984 | 0.987 | 0.990 | **0.991** |
| P4 | 0.909 | 0.947 | 0.921 | 0.744 | 0.953 | 0.967 | **0.970** | 0.967 |
| P5 | 0.952 | **0.975** | 0.950 | 0.842 | 0.968 | 0.963 | 0.970 | 0.970 |
| P6 | 0.944 | 0.970 | 0.947 | 0.820 | 0.964 | 0.969 | **0.975** | 0.974 |
| P7 | 0.957 | 0.973 | 0.960 | 0.799 | 0.976 | 0.975 | 0.978 | **0.979** |
| P8 | 0.974 | 0.986 | 0.967 | 0.837 | 0.982 | 0.984 | **0.987** | **0.987** |
| P9 | 0.939 | 0.949 | 0.947 | 0.821 | 0.955 | 0.956 | 0.969 | **0.970** |
| P10 | 0.949 | 0.974 | 0.942 | 0.817 | 0.971 | 0.978 | 0.979 | **0.982** |
| P11 | 0.967 | 0.984 | 0.971 | 0.843 | 0.981 | 0.976 | 0.984 | **0.986** |
| P12 | 0.960 | 0.970 | 0.954 | 0.851 | **0.984** | 0.981 | 0.974 | 0.974 |
| P13 | 0.969 | 0.982 | 0.971 | 0.887 | 0.979 | 0.985 | **0.989** | **0.989** |
| P14 | 0.956 | 0.976 | 0.960 | 0.859 | 0.973 | 0.975 | **0.978** | **0.978** |
| P15 | 0.948 | 0.972 | 0.956 | 0.849 | 0.974 | 0.973 | **0.981** | **0.981** |
| P16 | 0.944 | 0.959 | 0.943 | 0.754 | 0.965 | 0.966 | 0.973 | **0.975** |
| P17 | 0.957 | 0.972 | 0.955 | 0.827 | 0.971 | 0.974 | 0.977 | **0.979** |
| P18 | 0.935 | 0.971 | 0.950 | 0.791 | 0.972 | 0.972 | **0.979** | 0.977 |
| P19 | 0.974 | 0.986 | 0.980 | 0.856 | **0.991** | 0.989 | 0.989 | 0.989 |
| P20 | 0.947 | 0.975 | 0.960 | 0.790 | **0.980** | 0.979 | 0.976 | 0.977 |
| P21 | 0.944 | **0.981** | 0.942 | 0.811 | 0.960 | 0.980 | **0.981** | **0.981** |
| P22 | 0.948 | 0.978 | 0.950 | 0.783 | 0.972 | 0.970 | **0.982** | 0.981 |
| P23 | 0.933 | 0.960 | 0.936 | 0.809 | 0.971 | 0.969 | **0.975** | **0.975** |
| P24 | 0.955 | 0.977 | 0.962 | 0.832 | 0.981 | 0.984 | **0.986** | **0.986** |
| AVG | 0.952 | 0.974 | 0.955 | 0.825 | 0.974 | 0.976 | 0.979 | **0.980** |

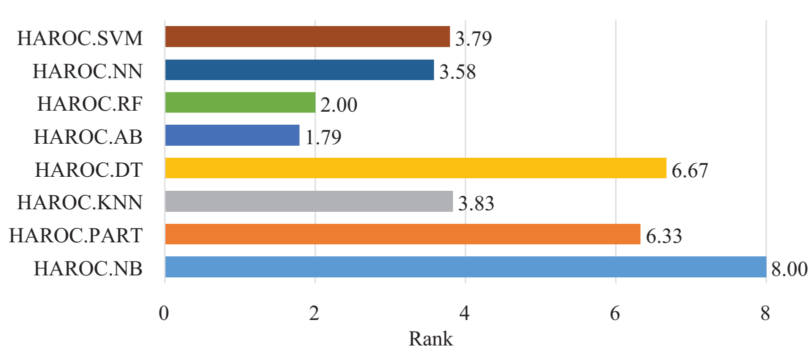


**Figure 3**. Comparison of HAROC methods with different base learners in terms of average rank.

According to kinesiology, the speed and range of a person's motion have corresponded with his/her sex, age, weight, and height. For example, the youth's stride frequency is usually faster compared to the elder; men

**Table 5**. Confusion matrix of the human activity recognition based on ordinal classification with AdaBoost algorithm (HAROC.AB).

| | Sitting | Standing | Upstairs | Downstairs | Walking | Jogging |
|---|---|---|---|---|---|---|
| Sitting | **431** | 0 | 0 | 0 | 0 | 0 |
| Standing | 0 | **443** | 0 | 0 | 0 | 0 |
| Upstairs | 0 | 0 | **107** | 10 | 0 | 0 |
| Downstairs | 0 | 0 | 5 | **118** | 2 | 0 |
| Walking | 0 | 1 | 1 | 3 | **362** | 0 |
| Jogging | 0 | 0 | 0 | 1 | 0 | **113** |

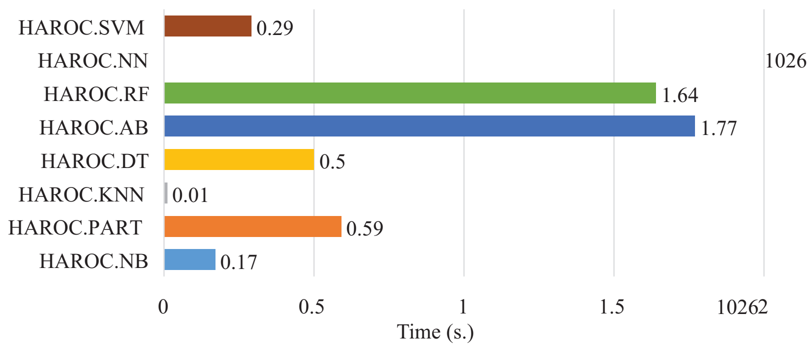


**Figure 4**. Comparison of HAROC methods with different base learners in terms of training time.

jump higher than women in general; a tall person's step length is longer than a short person's. Certainly, the aforementioned expressions are not always correct; however, in general, people having different demographic characteristics tend to behave differently for the same kind of movement. Since the demographic characteristics of the participants will affect the performance of the classifier, in this study, we investigate the effects of them on HAR.

In this study, we analyzed how the activity recognition performance differ in sex, age, weight, and height. Figure 5 shows the average accuracy values of the best models built by the HAROC approach in accordance with the sex, age, weight, and height of the participants. According to the results given in Figure 5a, the proposed HAROC method predicted human physical activities slightly better for females (98.20%) than males (98.02%). In terms of the ages of participants (Figure 5b), the HAROC approach achieved slightly more accurate results for the young persons, especially participants whose ages differ between 26-30. It is seen that the most influential factors that affect HAROC performance are weight and height. The classifiers that were built for the participants whose weight is smaller than 65 kg. (Figure 5c) or height smaller than 169 cm. (Figure 5d) achieved better performance than all the rest. Besides, the accuracy values decreased as the weight and the height of the participant increased. It can be noted here that the demographic characteristics of the person can affect his/her movement and hence they can also affect the sensor signals.

In this study, we also investigated the effects of different activity orderings on the performance of the proposed method. We considered 12 different human activity orders as given in Table 6, which are numbered from O1 to O12. While O1 and O12 refer to ordered (ascending and descending) class labels, the others (from O2 and O11) correspond to non-ordered class labels. Each number represents an activity as follows; 1: *sitting*,

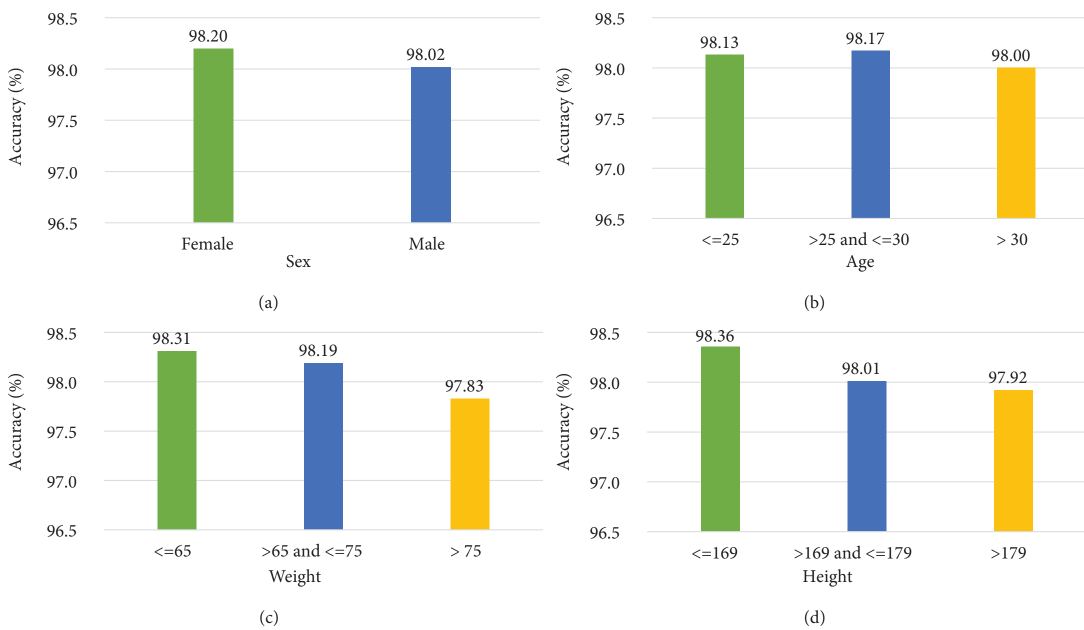


**Figure 5**. Evaluation of the HAROC method according to the demographic (age and sex) and physical (weight and height) characteristics of the participants.

2: *standing*, 3: *upstairs*, 4: *downstairs*, 5: *walking*, 6: *jogging*.

**Table 6**. Ordering of different human activities (1: sitting, 2: standing, 3: upstairs, 4: downstairs, 5: walking, 6: jogging).

| ID | Ordering | ID | Ordering |
|---|---|---|---|
| O1 | {1,2,3,4,5,6} | O7 | {4,1,6,2,5,3} |
| O2 | {2,6,4,1,3,5} | O8 | {4,2,5,1,6,3} |
| O3 | {3,1,5,2,6,4} | O9 | {5,2,4,1,6,3} |
| O4 | {3,5,1,6,2,4} | O10 | {5,2,6,3,1,4} |
| O5 | {3,6,2,5,1,4} | O11 | {5,3,1,6,2,4} |
| O6 | {4,1,5,2,6,3} | O12 | {6,5,4,3,2,1} |

Table 7 shows the accuracy values obtained when different activity orderings were considered. All orders were tested for each participant using the HAROC.DT method. The best accuracy value for each participant is highlighted in bold. It is seen from Table 7 that the performances of ordered activities (ascending or descending) are higher than that of randomly ordered activities. Therefore, it can be concluded that the order of class labels has a considerable impact on classification performance.

Figure 6 shows the average accuracy results obtained when different activity orderings were considered. It can be clearly seen that the best accuracy results were obtained when the activities were sorted in ascending order (O1) and descending order (O12). On the other hand, the accuracy of the classification model decreases when the activities (class labels) are unordered (from O2 to O11).

Table 8 gives the classification accuracy and F-measure values of ordinal classification (HAROC) and nominal classification methods presented in the previous studies [33–43]. Since the researchers used the same

**Table 7**. The effects of different human activity orderings on the performance of the HAROC.DT method.

| | O1 | O2 | O3 | O4 | O5 | O6 | O7 | O8 | O9 | O10 | O11 | O12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | **95.50** | 94.83 | 94.83 | 94.39 | 94.78 | 94.56 | 94.22 | 94.11 | 94.67 | 95.06 | 94.11 | **95.50** |
| P2 | 95.97 | 95.11 | 93.49 | 93.60 | 93.98 | 93.87 | 93.60 | 93.06 | 94.14 | 94.73 | 94.14 | **96.08** |
| P3 | 97.39 | 96.34 | 95.84 | 95.99 | 96.29 | 96.24 | 95.69 | 95.89 | 94.84 | 94.99 | 95.84 | **97.44** |
| P4 | **90.85** | 90.05 | 89.03 | 88.46 | 89.60 | 89.54 | 89.48 | 88.00 | 87.89 | 87.38 | 87.95 | **90.85** |
| P5 | 95.16 | 92.25 | 93.49 | 93.86 | 93.30 | 93.18 | 92.93 | 93.74 | 93.92 | 92.75 | 92.13 | **95.23** |
| P6 | **94.36** | 92.30 | 91.71 | 91.24 | 91.42 | 91.12 | 90.95 | 92.00 | 92.30 | 91.83 | 91.59 | 94.24 |
| P7 | 95.69 | 94.84 | 93.95 | 93.32 | 93.42 | 93.27 | 92.90 | 93.11 | 93.53 | 93.27 | 94.11 | **95.74** |
| P8 | **97.36** | 95.58 | 94.07 | 94.88 | 94.82 | 94.82 | 95.36 | 94.66 | 94.01 | 94.17 | 95.42 | **97.36** |
| P9 | **93.80** | 93.02 | 91.29 | 91.18 | 91.40 | 91.46 | 90.28 | 91.18 | 91.46 | 90.73 | 92.29 | **93.80** |
| P10 | **94.82** | 94.49 | 93.21 | 94.05 | 93.16 | 93.21 | 93.38 | 93.38 | 93.38 | 92.49 | 93.60 | **94.82** |
| P11 | **96.73** | 94.75 | 92.89 | 93.12 | 92.61 | 92.67 | 92.33 | 93.06 | 93.63 | 94.19 | 94.13 | 96.62 |
| P12 | **95.99** | 93.10 | 92.87 | 92.04 | 93.40 | 93.40 | 92.39 | 91.80 | 92.45 | 92.10 | 91.80 | **95.99** |
| P13 | 96.86 | **97.60** | 95.32 | 95.07 | 94.33 | 94.03 | 93.84 | 93.72 | 95.26 | 93.66 | 94.95 | 96.67 |
| P14 | **95.56** | 94.53 | 94.31 | 94.53 | 94.53 | 94.31 | 94.09 | 94.91 | 94.09 | 93.77 | 93.66 | **95.56** |
| P15 | **94.78** | 94.31 | 93.15 | 93.09 | 92.86 | 92.69 | 92.86 | 92.80 | 94.54 | 92.40 | 92.80 | 94.66 |
| P16 | **94.35** | 92.86 | 90.92 | 90.92 | 90.63 | 90.72 | 90.87 | 90.58 | 90.87 | 90.53 | 91.42 | **94.35** |
| P17 | **95.64** | 93.37 | 92.01 | 91.27 | 92.24 | 92.29 | 91.67 | 91.84 | 92.69 | 93.48 | 92.01 | **95.64** |
| P18 | 93.45 | 93.40 | 91.76 | 91.82 | 92.02 | 91.97 | 91.87 | 91.25 | 90.84 | 91.92 | 92.33 | **93.55** |
| P19 | 97.40 | **97.80** | 97.32 | 96.88 | 96.88 | 96.92 | 96.66 | 96.79 | 97.18 | 96.92 | 96.70 | 97.40 |
| P20 | 94.73 | 94.06 | 92.54 | 92.94 | 92.49 | 92.26 | 92.77 | 92.88 | 93.33 | 93.16 | 92.94 | **94.84** |
| P21 | 94.38 | 93.41 | 93.45 | 92.90 | 92.53 | 92.71 | 93.22 | 93.22 | 92.25 | 92.39 | 93.41 | **94.43** |
| P22 | 94.76 | 94.10 | 92.72 | 92.66 | 93.14 | 93.26 | 92.23 | 93.14 | 93.56 | 92.78 | 92.84 | **94.82** |
| P23 | 93.34 | 91.99 | 89.60 | 89.25 | 89.71 | 89.95 | 89.54 | 90.01 | 88.90 | 90.36 | 91.29 | **93.40** |
| P24 | 95.43 | **96.12** | 94.87 | 95.30 | 95.43 | 95.37 | 94.87 | 95.43 | 95.55 | 94.80 | 95.55 | 95.49 |

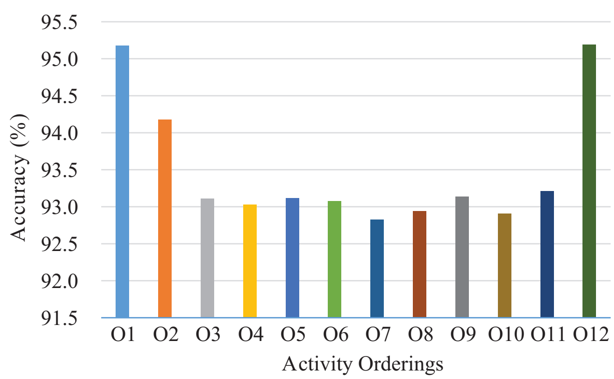


**Figure 6**. Average accuracy (%) values obtained for different human activity orderings.

dataset [32] as our study, the results were taken directly from the referenced studies. It can be clearly seen from Table 8 that all the ordinal classification methods (HAROC.KNN, HAROC.DT, HAROC.NN, HAROC.SVM, HAROC.RF, and HAROC.AB) outperformed their nominal counterparts (KNN, DT, NN, SVM, RF, and AB). For example, the HAROC.SVM algorithm achieved significantly better accuracy (97.35%) than the nominal SVM algorithm (78.61% [34], 86.33% [35], and 87.50% [36]). Similarly, the ordinal KNN method has higher accuracy (97.40%) than the nominal KNN method (68.47% [33], and 93.88% [34]).

**Table 8**. Comparison of ordinal and nominal classification performances in terms of accuracy (%) and F-measure values on the MotionSense [32] dataset.

<table>
<tr><th rowspan="2">Ref.</th><th rowspan="2">Nominal Classification</th><th colspan="2">Metrics</th><th rowspan="2">Ordinal Classification</th><th colspan="2">Metrics</th></tr>
<tr><th>Accuracy</th><th>F-measure</th><th>Accuracy</th><th>F-measure</th></tr>
<tr><td>[33]</td><td>KNN</td><td>68.47</td><td></td><td rowspan="2">HAROC.KNN</td><td rowspan="2">97.40</td><td rowspan="2">0.974</td></tr>
<tr><td rowspan="5">[34]</td><td>KNN</td><td>93.88</td><td></td></tr>
<tr><td>ELM</td><td>96.87</td><td></td><td>-</td><td></td><td></td></tr>
<tr><td>NN</td><td>77.77</td><td></td><td>HAROC.NN</td><td>97.54</td><td>0.976</td></tr>
<tr><td>RF</td><td>95.55</td><td></td><td>HAROC.RF</td><td>97.94</td><td>0.979</td></tr>
<tr><td>SVM</td><td>78.61</td><td></td><td rowspan="4">HAROC.SVM</td><td rowspan="4">97.35</td><td rowspan="4">0.974</td></tr>
<tr><td>[35]</td><td>SVM</td><td>86.33</td><td></td></tr>
<tr><td>[36]</td><td>SVM+PSO</td><td>87.50</td><td></td></tr>
<tr><td rowspan="2">[37]</td><td>SVM</td><td></td><td>0.867</td></tr>
<tr><td>DT</td><td></td><td>0.883</td><td rowspan="2">HAROC.DT</td><td rowspan="2">95.18</td><td rowspan="2">0.952</td></tr>
<tr><td rowspan="4">[38]</td><td>DT</td><td><95.00</td><td></td></tr>
<tr><td>LR, LSTM, DySan</td><td><=95.00</td><td></td><td>-</td><td></td><td></td></tr>
<tr><td>NN</td><td><95.00</td><td></td><td>HAROC.NN</td><td>97.54</td><td>0.976</td></tr>
<tr><td>RF</td><td>96.00</td><td></td><td rowspan="2">HAROC.RF</td><td rowspan="2">97.94</td><td rowspan="2">0.979</td></tr>
<tr><td rowspan="3">[39]</td><td>RF</td><td>90.00</td><td></td></tr>
<tr><td>NN</td><td>~87.50</td><td></td><td>HAROC.NN</td><td>97.54</td><td>0.976</td></tr>
<tr><td>XGBoost</td><td>~96.00</td><td></td><td>-</td><td></td><td></td></tr>
<tr><td>[40]</td><td>AB</td><td>78.06</td><td></td><td>HAROC.AB</td><td>97.96</td><td>0.980</td></tr>
<tr><td>[41]</td><td>CNN</td><td></td><td>0.903</td><td>-</td><td></td><td></td></tr>
<tr><td>[42]</td><td>CNN</td><td>95.08</td><td></td><td>-</td><td></td><td></td></tr>
<tr><td>[43]</td><td>CNN</td><td>96.77</td><td></td><td>-</td><td></td><td></td></tr>
</table>

## 5. Conclusions and future work

The standard human activity recognition is limited to using only nominal data to build a model and disregards the ordering information among the activities. However, human physical activities are in order according to their characteristics. For example, walking requires more energy than sitting, whereas it requires less energy than jogging since jogging involves more intense muscle contractions. Based on this motivation, we proposed a novel approach, called Human Activity Recognition based on Ordinal Classification (HAROC), which considers the ranking among the activities during the classification task. This is the first study that combines two concepts: HAR and OC.

It is the first time that this study investigates which base learner (SVM, NN, RF, AB, DT, KNN, PART, and NB) is effective when used with ordinal classification for HAR. This study is also original in that it examines the effects of the demographic characteristics of the participants (i.e. sex, age, weight, and height) on the ordinal classification performance. The effectiveness of the proposed HAROC method was demonstrated on a real-world dataset collected from 24 participants to recognize six human physical activities, including walking, jogging, standing, sitting, downstairs, and upstairs. Based on the analysis results, the following conclusions were drawn.

- The proposed HAROC approach is an effective approach for HAR since it designs a classification model by considering the ranking among the activities.
- HAROC combination with the AdaBoost (HAROC.AB) achieved the best performance (97.96%) on average among the other alternative base learners.
- HAROC.AB performed better than the rest since it had the lowest rank value (1.79) on average.
- It can be concluded from the confusion matrix that each human physical activity was distinguished and classified with high accuracies, however, downstairs and upstairs activities were slightly confused by the algorithm during classification. On the test set, the best performance was achieved on the sitting and standing activities.
- According to the training time comparisons, HAROC.KNN, HAROC.NB, and HAROC.SVM methods learned in much less time, while the HAROC.NN method took a very long time of learning.
- The classification performance of the HAROC method was slightly better for females than males.
- The HAROC approach achieved slightly more accurate results for the young persons, especially participants whose ages change between 26 and 30.
- According to the experimental results, the weight and height values of the participants are inversely proportional with the accuracy. The accuracy values decreased as the weight and the height of the participant increased. The highest accuracy values were obtained for the participants whose weight is lower than 65 kg or height smaller than 169 cm.
- It is seen that the best performance was obtained for ordered (ascending or descending) class labels, rather than unordered ones. Therefore, it can be concluded that considering ordered class labels increases the performance when compared with that of unordered class labels.
- Comparison of ordinal and nominal classification approaches showed that all the ordinal classification methods (HAROC.KNN, HAROC.DT, HAROC.NN, HAROC.SVM, HAROC.RF, and HAROC.AB) outperformed their nominal counterparts (KNN, DT, NN, SVM, RF, and AB).
- Finally, it can be concluded that the proposed HAROC method enables enormous applications in real-life and can be effectively used for different purposes such as gait analysis, sports evaluation, and entertainment.

As future work, a feature selection method may be used to identify and eliminate irrelevant and redundant attributes from data with the aim of reducing the training time. Furthermore, in future, the performance of OC-based HAR may be investigated on more specific activities, such as sports activities or military activities.

## References

[1] Tolba A, Al-Makhadmeh Z. Wearable sensor-based fuzzy decision-making model for improving the prediction of human activities in rehabilitation. Measurement 2020; 166: 1-11. doi: 10.1016/j.measurement.2020.108254

[2] Wang Y, Cang S, Yu H. A survey on wearable sensor modality centred human activity recognition in health care. Expert Systems with Applications 2019; 137: 167-190. doi: 10.1016/j.eswa.2019.04.057

[3] Mukherjee A, Misra S, Mangrulkar P, Rajarajan M, Rahulamathavan Y. SmartARM: a smartphone-based group activity recognition and monitoring scheme for military applications. In: IEEE 2017 ANTS International

Conference on Advanced Networks and Telecommunications Systems; Bhubaneswar, India; 2017. pp. 1-6. doi: 10.1109/ants.2017.8384149

[4] Supriyatna TB, Nasution SM, Nugraheni RA. Human activity recognition using support vector machine for automatic security system. Journal of Physics 2019; 1192: pp. 1-10. doi: 10.1088/1742-6596/1192/1/012017

[5] Frank E, Mark H. A simple approach to ordinal classification. In: European Conference on Machine Learning; Freiburg, Germany; 2001. pp. 145-156. doi: 10.1007/3-540-44795-4_13

[6] Nguyen B, Morell C, Baets BD. Distance metric learning for ordinal classification based on triplet constraints. Knowledge-Based Systems 2018; 142: 17-28. doi: 10.1016/j.knosys.2017.11.022

[7] Campoy-Muñoz P, Gutiérrez PA, Hervás-Martínez C. Addressing remitting behavior using an ordinal classification approach. Expert Systems with Applications 2014; 41 (10): 4752-4761. doi: 10.1016/j.eswa.2014.01.036

[8] Chen Y, Shen C. Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access 2017; 5: 3095–3110. doi: 10.1109/access.2017.2676168

[9] Chen Z, Zhu Q, Soh YC, Zhang L. Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. IEEE Transactions on Industrial Informatics 2017; 13 (6): 3070-3080. doi: 10.1109/tii.2017.2712746

[10] Shoaib M, Bosch S, Durmaz Incel O, Scholten H, Havinga PJM. Fusion of smartphone motion sensors for physical activity recognition. Sensors 2014; 14 (6): 10146-10176. doi: 10.3390/s140610146

[11] Saylam B, Shoaib M, Durmaz Incel O. Exploring the parameter space of human activity recognition with mobile devices. Turkish Journal of Electrical Engineering & Computer Sciences 2020; 28 (6). doi: 10.3906/elk-1910-139

[12] Kerdjidj O, Ramzan N, Ghanem K, Amira A, Chouireb F. Fall detection and human activity classification using wearable sensors and compressed sensing. Journal of Ambient Intelligence and Humanized Computing 2020; 11: 349-361. doi: 10.1007/s12652-019-01214-4

[13] Gjoreski M, Janko V, Slapnicar G, Mlakar M, Rescic N et al. Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors. Information Fusion 2020; 62: 47–62. doi: 10.1016/j.inffus.2020.04.004

[14] Zalluhoglu C, Ikizler-Cinbis N. Collective sports: a multi-task dataset for collective activity recognition. Image and Vision Computing 2020; 94: 1-11. doi: 10.1016/j.imavis.2020.103870

[15] Janidarmian M, Fekr AR, Radecka K, Zilic Z. A comprehensive analysis on wearable acceleration sensors in human activity recognition. Sensors 2017; 17 (3): 1-26. doi: 10.3390/s17030529

[16] San-Segundo R, Blunck H, Moreno-Pimentel J, Stisen A, Gil-Martín M. Robust human activity recognition using smartwatches and smartphones. Engineering Applications of Artificial Intelligence 2018; 72: 190-202. doi: 10.1016/j.engappai.2018.04.002

[17] Guijo-Rubio D, Gutiérrez PA, Casanova-Mateo C, Sanz-Justo J, Salcedo-Sanz S et al. Prediction of low-visibility events due to fog using ordinal classification. Atmospheric Research 2018; 214: 64-73. doi: 10.1016/j.atmosres.2018.07.017

[18] Manthoulis G, Doumpos M, Zopounidis C, Galariotis E. An ordinal classification framework for bank failure prediction: methodology and empirical evidence for US banks. European Journal of Operational Research 2020; 282 (2): 786-801. doi: 10.1016/j.ejor.2019.09.040

[19] Rahiche A, Hedjam R, Al-maadeed S, Cheriet M. Historical documents dating using multispectral imaging and ordinal classification. Journal of Cultural Heritage, in press. doi: 10.1016/j.culher.2020.01.012

[20] Gutiérrez PA, Pérez-Ortiz M, Sánchez-Monedero J, Fernández-Navarro F, Hervás-Martínez C. Ordinal regression methods: survey and experimental study. IEEE Transactions on Knowledge and Data Engineering 2016; 28 (1): 127-146. doi: 10.1109/TKDE.2015.2457911

[21] Sousa R, Yevseyeva I, Costa JFP, Cardoso JS. Multicriteria models for learning ordinal data: a literature review. In: Yang XS (editor). Artificial Intelligence, Evolutionary Computing and Metaheuristics. Berlin, Germany: Springer, 2013, pp.109-138.

[22] Halsey LG, Watkins DAR, Duggan BM. The energy expenditure of stair climbing one step and two steps at a time: estimations from measures of heart rate. Plos One 2012; 7 (12): 1-6. doi:10.1371/journal.pone.0051213

[23] Qu Y, Gao Z, Xiao Y, Li X. Modeling the pedestrian's movement and simulating evacuation dynamics on stairs. Safety Science 2014; 70: 189-201. doi: 10.1016/j.ssci.2014.05.016

[24] Hackett S, Cai Y, Siegel M. Activity recognition from sensor fusion on fireman's helmet. In: 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); Huaqiao, China; 2019. pp. 1-6. doi: 10.1109/CISP-BMEI48845.2019.8966005

[25] Wang L, Qian Q, Zhang Q, Wang J, Cheng W et al. Classification model on big data in medical diagnosis based on semi-supervised learning. The Computer Journal 2020; bxaa006: 1-15. doi:10.1093/comjnl/bxaa006

[26] Muñoz-Almaraz FJ, Zamora-Martínez F, Botella-Rocamora P, Pardo J. Supervised filters for EEG signal in naturally occurring epilepsy forecasting. Plos One 2017; 12: 1-17. doi:10.1371/journal.pone.0178808

[27] Xiao P, Qu W, Qi H, Li Z. Detecting DDoS attacks against data center with correlation analysis. Computer Communications 2015; 67: 66-74. doi:10.1016/j.comcom.2015.06.012

[28] Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[29] Jiang D, Ma P, Su X, Tiantian W. Distance metric based divergent change bad smell detection and refactoring scheme analysis. International Journal of Innovative Computing, Information and Control 2014; 10 (4): 1519-1531.

[30] Shmulevich I, Dougherty D. Genomic Signal Processing. Princeton, NJ, USA: Princeton University Press, 2007.

[31] Mercado P, Lukashevich H. Applying constrained clustering for active exploration of music collections. In: Proceedings of the 1st Workshop on Music Recommendation and Discovery (WOMRAD); Barcelona, Spain; 2010. pp. 39-46.

[32] Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Protecting sensory data against sensitive inferences. In: Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems; Porto, Portugal; 2018. pp. 1-6. doi: 10.1145/3195258.3195260

[33] Ferrari A, Mobilio M, Micucci D, Napoletano P. On the homogenization of heterogeneous inertial-based databases for human activity recognition. In: 2019 IEEE World Congress on Services (SERVICES); Milan, Italy; 2019. pp. 295-300. doi: 10.1109/SERVICES.2019.00084.

[34] Kuncan F, Kaya Y, Kuncan M. A novel approach for activity recognition with down-sampling 1D local binary pattern features. Advances in Electrical and Computer Engineering 2019; 19: 35-44. doi:10.4316/AECE.2019.01005

[35] Gil-Martín M, Sánchez-Hernández M, San-Segundo R. Human activity recognition based on deep learning techniques. Proceedings 2020; 42 (1): 1-6. doi: 10.3390/ecsa-6-06539

[36] Batool M, Jalal A, Kim K. Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm. In: 2019 International Conference on Applied and Engineering Mathematics (ICAEM); Taxila, Pakistan; 2019. pp. 145-150. doi: 10.1109/ICAEM.2019.8853770

[37] Jalal A, Batool M, Kim K. Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors. Applied Sciences 2020; 10 (20): 1-20. doi: 10.3390/app10207122

[38] Boutet A, Frindel C, Gambs S, Jourdan T, Ngueveu RC. DYSAN: dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks. In: 34th Conference on Neural Information Processing Systems (NeurIPS 2020); Vancouver, Canada; 2020. pp. 1-10.

[39] Fermanian A. Embedding and learning with signatures. Computational Statistics and Data Analysis 2021; 157: 1-23. doi: 10.1016/j.csda.2020.107148

[40] Ferrari A, Micucci D, Mobilio M, Napoletano P. On the personalization of classification models for human activity recognition. IEEE Access 2020; 8: 32066-32079. doi: 10.1109/ACCESS.2020.2973425

[41] Saeed A, Ozcelebi T, Lukkien J. Multi-task self-supervised learning for human activity detection. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2019; 3 (2): 1-30. doi: 10.1145/3328932

[42] Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Protecting sensory data against sensitive inferences. In: Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems; Porto, Portugal; 2018. pp. 1-6. doi: 10.1145/3195258.3195260

[43] Gamble JA, Huang J. Convolutional neural network for human activity recognition and identification. In: IEEE International Systems Conference (SysCon); Dublin, Ireland; 2018. pp. 1-7. doi: 10.1109/SysCon47679.2020.9275924