

# Human Activity Recognition System Using Artificial Neural Networks

Vinícius Ferreira De Almeida and Rodrigo Varejão Andreão

## Abstract

Population aging and the increasing costs of health care, especially for the elderly affected by chronic diseases, requires new medical assistance strategies that makes it possible to monitor these people remotely and provide reliable information on their routines. In this context, human activity recognition (HAR) systems are an important element to overcoming the problem. Therefore, this paper proposes a HAR system prototype containing a multilayer perceptron (MLP) as a classifier. The model hyperparameters were selected using a publicly available dataset. Then, data was collected from accelerometers and gyroscopes embedded in wearable devices of 15 subjects while performing six basic activities (walking, sitting, lying down, standing, walking upstairs and walking downstairs). The system reached an average accuracy of 90.74% and weighted F-measure of 90.03% based on leave-one-subject-out cross-validation.

## Keywords

Human activity recognition • Artificial neural networks • Wearable devices

## 1 Introduction

The population aging is a global scale phenomenon. As a result, virtually all countries around the world are experiencing an increase in the proportion of elderly people among their inhabitants. In Brazil, according to the 2018 review of the Population Projection, conducted by the Brazilian Institute of Geography and Statistics (IBGE), about 25.5% of Brazilians will be over 65 years old in the year 2060 [1]. In this sce-

nario, the health care of these people needs special attention, especially when they are affected by chronic diseases.

The main types of these diseases are cardiovascular diseases, cancer, diabetes and others that together are responsible for a high number of deaths and life quality reduction in several countries. In addition, when neglected or not adequately treated, these diseases are responsible for reducing family income, since treatment is usually a prolonged and expensive process [2]. Thus, in order to mitigate such consequences, it is necessary to employ complementary strategies to monitor people's habits and health, making it possible to recognize changes that may highlight more serious conditions, especially for individuals who are in advanced age.

In this context, human activity recognition (HAR) systems present a growing importance, since it is a strategy capable of assisting medical teams in the accompaniment of their patients, especially the chronically ill, who must follow a well structured routine of activities and exercises in their daily lives [3].

Several studies on HAR use publicly available datasets with samples generated by inertial sensors located at different parts of the human body. The data are used for training and testing machine learning models, where the most frequently employed are: k-Nearest Neighbors (kNN) [4–6], Support Vector Machines (SVM) [4–7], Artificial Neural Networks (ANN) [6–8] and even complex deep learning models [9, 10].

Considering the importance of building a prototype for a real application with chronic elderly patients, this work proposes and validates a prototype of an inertial sensor data acquisition and activity recognition system. The data acquisition system is based on a mobile application developed to capture the data generated by the inertial sensors through a *smartphone*. 15 healthy subjects participated in this study. The recorded data are used in the tests of the human activity recognition system, which employs an ANN as a classification algorithm. The activities performed belong to two categories: static activities (lying, sitting and standing) and dynamic activities (walking, walking upstairs and walking

V. F. De Almeida (✉) · R. V. Andreão  
Department of Electrical Engineering, Federal Institute of Espírito Santo, Vitória Avenue, 1729, Vitória, Brazil  
e-mail: [viniciusf92@gmail.com](mailto:viniciusf92@gmail.com)

downstairs). After the acquisition, the data are provided to the other stages of the process, which consist of: filtering, segmentation, feature extraction, classification and evaluation.

Another contribution of the work lies in the fact that the selection of parameters of the ANN is performed in a publicly available dataset and is trained and evaluated in a different dataset, built with the developed prototype.

## 2 Materials and Methods

### 2.1 Application

The application developed for data acquisition was programmed in Java through the *Android Studio* IDE. Its interface can be seen in Fig. 1. The software connects to the monitoring devices through bluetooth low energy (BLE). Then, some parameters can be defined, such as the sampling frequency, the label of the activity that will be executed and the identification code of the participant. The time period for data collection is pre-determined by the application.

Once these parameters are established, it is possible to start capturing the signals by pressing the “start acquisition” button. Thus, the application sends a command that enables the sensors and starts the storage of the data that are returned by them. The information received is organized in a table format, where each sample obtained receives a time label, the identification code of the subject and the label of the activity being developed.

### 2.2 Data Acquisition

In compliance with the National Health Council Resolution No. 466 of December 2012, which establishes rules and guidelines regulating research involving human beings, this project was submitted to and approved by the Ethics and Research Committee of the Federal Institute of Espírito Santo through the “Plataforma Brasil” (CAAE 89787518.5.0000.5072).

Data collection was carried out in a laboratory environment with individuals over 18 years of age. A total of 15 healthy volunteers participated. All of them wore a device tied to the right wrist and another at the waist. The devices adopted were the *SimpleLink™ SensorTag CC2650STK* from Texas Instruments, which contains a number of ten sensors, including an inertial measurement unit MPU-9250 from InvenSense, used in this work. The place of attachment on the body was defined based on the activities of interest. Figure 2 illustrates the positions at which the devices were attached.

The sensors were configured with a range of  $\pm 8G$  for the accelerometer and  $\pm 250^\circ/s$  for the gyroscope. According to [11], human activity recognition improves with higher sampling frequencies, but such gains are smaller at rates above

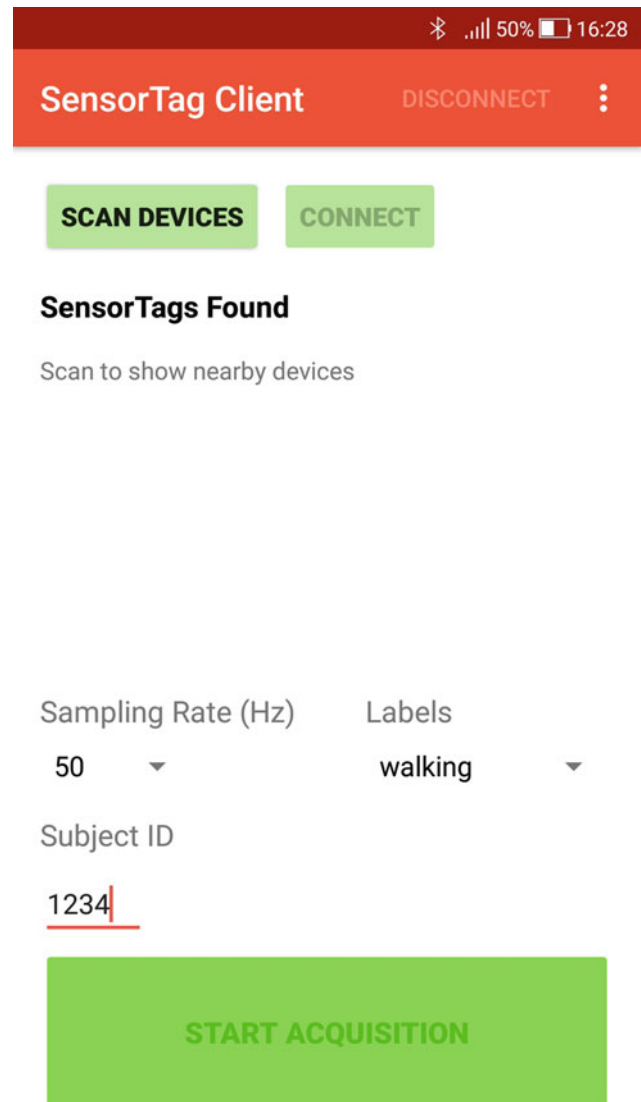


Fig. 1 App interface



Fig. 2 Body locations where *SensorTags* were attached: one at the waist and the other at the right wrist

20 Hz. Therefore, in order to ensure better results, a frequency of 50 Hz has been set for capturing data from the sensors used.

### 2.3 Signal Preprocessing

In its raw form, the recorded data may present noises and errors related to the acquisition process, interfering negatively on the performance of the HAR system. Therefore, preprocessing is necessary in order to prepare the data for the consecutive steps.

Thus, to recover the corrupted information, a linear interpolation was adopted at these locations. Furthermore, a 3rd order *Butterworth* low-pass filter (LPF) with a cutoff frequency of 20 Hz was used to reduce noise present in the signals. This choice is related to the characteristics of human body movements, which mainly have frequency components lower than 20 Hz [12].

### 2.4 Segmentation

The signal segmentation is intended to accommodate the data in reduced blocks, from which will be extracted the features that will allow the classifier to distinguish one activity from another. The determination of the type and size of such block must consider not only the characteristics of the activities of interest, but also the balance between amount of information and computational cost.

The activities of interest of this work have periodicity characteristics. In this case, based on previous proposals found in the literature, the sliding windows method presents promising results, with two second windows showing a good relationship between amount of information and computational cost [3, 13, 14].

Thus, segmentation in two seconds blocks was adopted, resulting in data windows with 100 samples. In addition, an overlap of 50% between adjacent windows was defined. This

strategy provides a greater amount of data, in addition to ensuring smooth transitions between neighboring windows, a desirable feature when handling continuous data [15].

### 2.5 Feature Extraction

The features, or attributes, can be classified according to which domain they belong, and frequently those of the time and frequency domains are adopted.

In this way, attributes from both domains were extracted in the proposed system. Additionally, new data were generated from the raw sensor readings, such as the root mean square (RMS) of the accelerometer and gyroscope readings, the extraction of the gravitational component of the accelerometer by applying an LPF with a cutoff frequency of 0.3 Hz, and the application of the first derivative in certain components.

Tables 1 and 2 show the features that were adopted and from which data were extracted (marked with “X”).

Based on the tables, the resulting feature vector has a total of 156 attributes. In order to minimize influences caused by the different orders of magnitude of the sensors to the classification, the attributes were scaled by the Z-score.

### 2.6 Model Selection

The model selection aims to optimally combine the internal parameters of a machine learning algorithm in order to improve its performance during the execution of a given task. Thus, the following parameters of MLP were evaluated for best performance: the number of hidden layers, the number of neurons in these layers, the initial learning rate, the activation and optimization functions, and the *momentum*.

As highlighted in [7, 16], adopting different datasets negatively influences the overall accuracy of the classifier due to variations in the data of one dataset in relation to the other. However, since the performance achieved by the HAR system

**Table 1** Features extracted from the accelerometer data

Features	Raw (X,Y,Z)	RMS	Grav. (X,Y,Z)
Mean	X	X	X
Median	–	X	X
Standard deviation	X	X	X
Variance	–	X	X
Minimum value	–	X	X
Maximum value	–	X	X
Kurtosis	–	X	X
Skewness	–	X	X
Spectral energy	X	–	–
Spectral entropy	X	–	–

**Table 2** Features extracted from the gyroscope data

Features	1st deriv. (X,Y,Z)	RMS	1st deriv. (RMS)
Mean	X	X	X
Median	X	X	–
Standard deviation	X	X	X
Variance	X	X	–
Minimum value	X	X	–
Maximum value	X	X	–
Kurtosis	X	X	–
Skewness	X	X	–
Spectral energy	–	–	–
Spectral entropy	–	–	–

becomes less dependent on a specific dataset, this approach can benefit your evaluation by making it more realistic. Thus, a publicly available dataset, distinct from the one developed in this work, was adopted for model selection.

The dataset selected was the *Opportunity* dataset [17, 18]. This set contains information from multiple sensors modalities, collected while four subjects performed daily activities in a laboratory similar to a residential kitchen. However, only data from accelerometers and gyroscopes tied to the user's body and in positions similar to those defined in this work were selected. Then, the same procedures described above were applied to this subset.

Therefore, the selection was made using the grid search tool, present in the *Scikit-learn* machine learning library [19]. This strategy performs a search within a range of predefined parameters and selects the configuration that obtained the best performance based on some evaluation metrics, in this work, the weighted F-measure. The data splitting and the classifier evaluation were implemented based on leave-one-subject-out cross-validation (LOSOVC). This technique ensures that data from the same individual does not appear in training and test sets at the same time.

## 2.7 Classification

Once the MLP network hyperparameters were defined, the feature vectors of the 15 subjects of this work were provided to the classification stage. The training and test sets were also created based on the LOSOCV. Thus, the performance of the classifier consists on the average of the results obtained in each partition created by the cross-validation.

## 3 Results and Discussion

### 3.1 Dataset Development

In data acquisition, each activity (lying down, sitting, standing, walking, walking upstairs and walking downstairs) was executed for 1 min. However, the activities “walking upstairs and walking downstairs”, were performed in sessions of 10 s and up to a total of 30 s, due to the physical limitation of the stairs used and in order to mitigate any discomfort to the subjects.

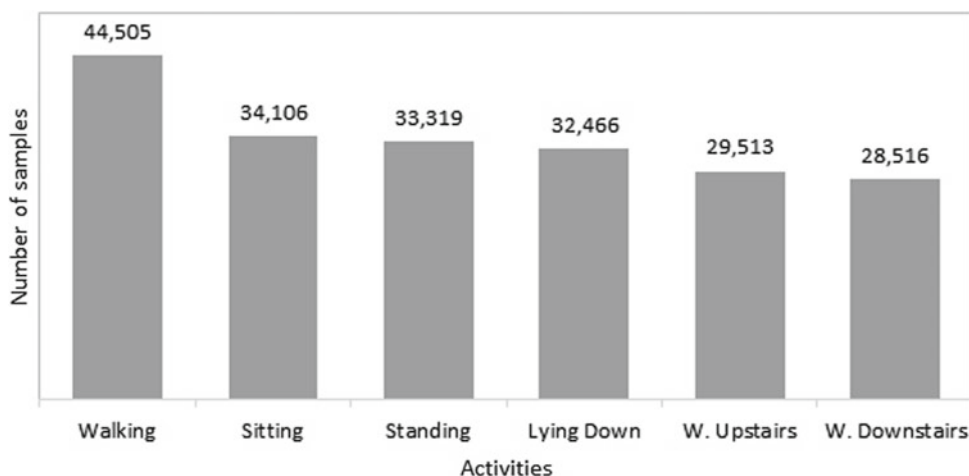
During the data acquisition of the static activities, samples of transitions between postures were also collected, such as “sitting-standing”, “standing-lying down” etc. In addition, fall simulations were performed (a mattress was used for reduce impact), such as: “forwards”, “backwards”, “lateral” and another one that scenes a fall after getting up quickly from a chair. However, the samples of both activity categories were not used in this work.

A total of 202,425 samples of the activities of interest were collected. Figure 3 shows the number of samples for each activity.

### 3.2 Model Selection

The following intervals from Table 3 were defined for the parameters tested in the model selection.

A total of 960 combinations were evaluated based on the weighted F-measure, which performs a harmonic mean between recall and precision, weighted by the number of sam-

**Fig. 3** Number of samples collected for each activity**Table 3** Hyperparameters evaluated in model selection

Hyperparameter	Values
Hidden layers (HL)	1–2
Neurons in each hidden layer (NHL)	80–225 (increments of 5)
Initial learning rate (ILR)	0.001; 0.002; 0.01; 0.02
Activation function (AF)	Tanh and ReLU
Optimization function (OF)	stochastic gradient descent (SGD) and Adam
<i>Momentum</i> <sup>a</sup>	0.9

<sup>a</sup>Only in combination with SGD

**Table 4** MLP networks that have achieved the best performance (opportunity dataset)

HL	NHL	ILR	AF	OF	F-measure
2	130	0.01	Tanh	Adam	85.48%±1.85%
2	205	0.002	Tanh	Adam	85.27%±2.82%
1	165	0.002	Tanh	Adam	85.02%±1.95%

**Table 5** Classifier performance based on LOSOCV

Accuracy	Recall	Precision	F-measure
90.74% ± 7.72%	92.26% ± 5.60%	90.74% ± 7.72%	90.03% ± 9.20%

ples present in each class. This metric was chosen because it reliably evaluates the performance of classifiers in an unbalanced dataset, that is, where there is the predominance of one class over another.

Thus, Table 4 presents the models that obtained the three best performances during the grid search. As it can be observed, an MLP network with 2 hidden layers and 130 neurons in these layers, initial learning rate of 0.01, hyperbolic tangent activation function and Adam optimization function presented the highest performance, being adopted in the data classification of the volunteers of this work.

It should be noted that in the tests of all configurations the network was initiated in an identical way, a factor that

eliminates a possible favorable condition to a certain parameter combination due, for example, to the initial values of the synaptic weights of the ANN.

### 3.3 Classification Results

The classification performance in the dataset developed in this study was assessed based on the LOSOCV. The MLP network that was elected in the model selection reached the results shown in Table 5 and in the confusion matrix of Table 6.

From the analysis of the confusion matrix, there was difficulty in recognizing the classes “walking upstairs”, “walking

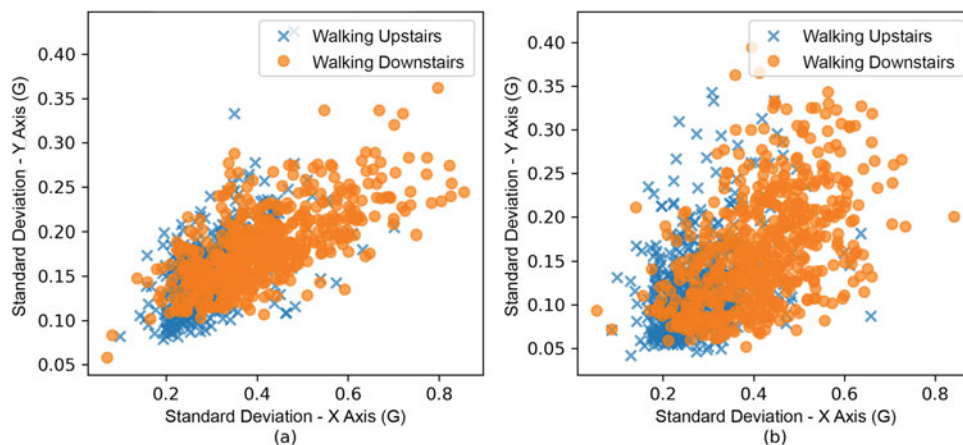


**Table 6** Confusion matrix for the MLP classifier based on LOSOCV

	Standing (%)	Lying down (%)	Sitting (%)	W. Downstairs (%)	W. Upstairs (%)	Walking (%)
Standing	<b>93.30</b>	0.61	4.72	0.00	0.61	0.76
Lying down	1.38	<b>97.08</b>	0.92	0.00	0.61	0.00
Sitting	12.52	1.60	<b>85.44</b>	0.00	0.15	0.29
W. Downstairs	0.00	0.18	0.00	<b>87.97</b>	10.23	1.62
W. Upstairs	0.85	0.17	0.00	12.90	<b>83.02</b>	3.06
Walking	0.23	0.00	0.23	1.24	2.60	<b>95.71</b>

The bold values related to the percentage of correctly classified samples for each activity

**Fig. 4** Comparison between walking upstairs and walking downstairs activities based on standard deviation of accelerometers data attached on **a** waist and **b** wrist



downstairs” and “sitting”. Such behavior can be explained because these activities present similar characteristics to each other (in the case of walking upstairs and walking downstairs) or to others activities present in the dataset.

As an illustration of this behavior, Fig. 4 makes a comparison based on the standard deviation of the acceleration observed by the X and Y axis of the waist and wrist accelerometers during the “walking upstairs” and “walking downstairs” activities. It can be seen that such activities keep a high similarity between them, so that the values appear overlapped in the figure.

However, it was easier for the classifier to distinguish static activities from dynamic activities, since such classes present a clearer separation, as shown in Fig. 5.

### 3.4 Comparison with Related Works

Table 7 presents a comparison of results based on the accuracy metric, commonly adopted in other studies. The listed works were chosen because they performed the activity recognition task with similar approaches to the one presented in this study.

However, some approaches differ in the way that data were collected and in the method in which the classification was performed or evaluated, as an example, by the application of k-fold cross-validation or the simple division of data between training and testing set, which gives overestimated results

when compared to the LOSOCV technique used in this study. Even so, the ANN chosen presented results comparable to the studies shown in Table 7.

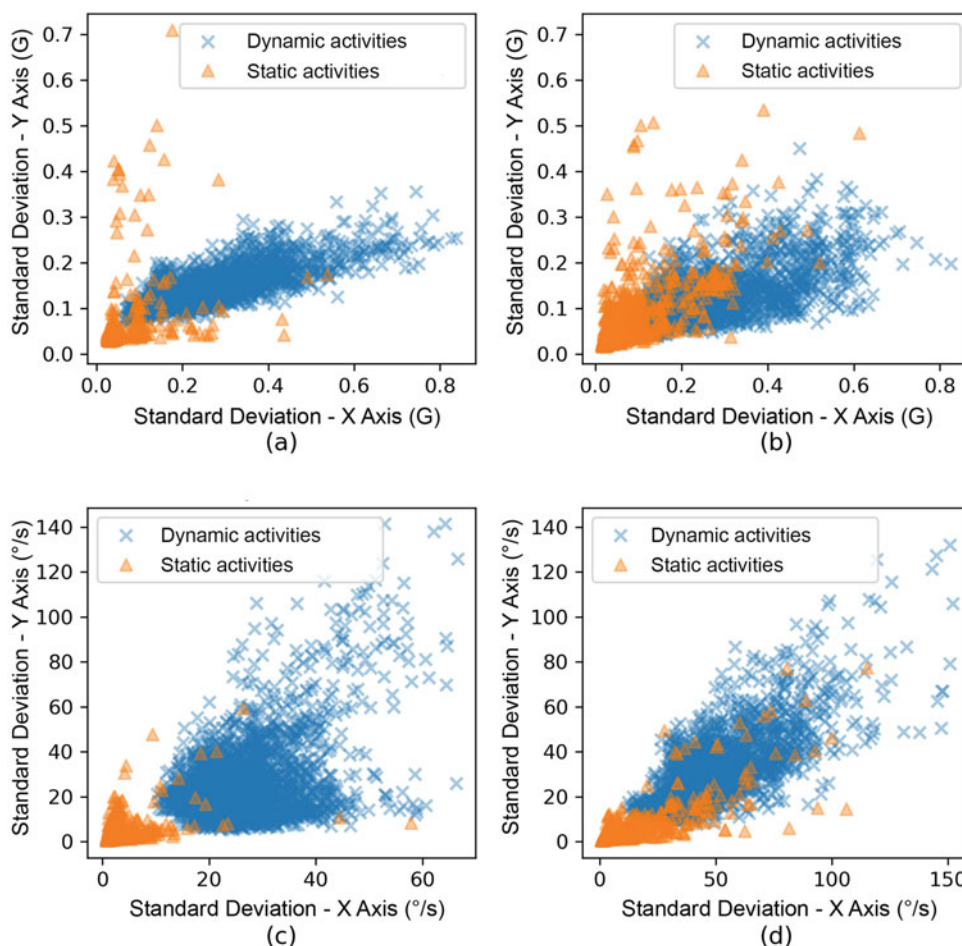
## 4 Conclusion

From the results presented, it can be seen that the system developed achieved satisfactory performance compared to others found in the literature. In this work, collection sessions with volunteers were conducted, acquiring accelerometers and gyroscopes data from simple activities. These samples were processed and later used in the training and classification of a multilayer perceptron neural network. The hyperparameters of this algorithm were defined using the grid search technique, using the content of the *Opportunity* dataset.

Although the classifier has a fundamental role in a HAR system, requiring a careful definition of its internal parameters, special attention should be given to the previous steps, as they are crucial factors for the best performance of the algorithm.

Future improvements can be made to the proposed system, making it suitable for a real context, such as the inclusion of sensors in the environment and the consequent acquisition of new samples of activities, tests with unhealthy and older subjects and real-time response, with the ability to generate emergency alerts, for example. Also, the evaluation of the

**Fig. 5** Comparison between static and dynamic activities based on standard deviation of accelerometer data from **a** waist and **b** wrist and gyroscope data from **c** waist and **d** wrist



**Table 7** Comparison with related studies

Classifier	Evaluation	Accuracy	Reference
MLP	LOSOVC	90.74%	This work
MLP	Train/Test split	86.20%	[8]
MLP	LOSOVC	83.00%	[20]
kNN	k-fold CV	94.13%	[21]

system performance on a short, medium and long term basis need to be assessed.

**Conflict of interest** The authors state no conflict of interest.

**References**

1. Agência IBGE Notícias (2018) Projeção da População 2018: número de habitantes do país deve parar de crescer em 2047
2. Goulart FAA (2011) Doenças Crônicas Não Transmissíveis: Estratégias de Controle e Desafios Para os Sistemas de Saúde. PAHO
3. Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor* 15:1192–1209
4. Gao L, Bourke AK, Nelson J (2014) Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Med Engi Phys* 36:779–785
5. Shoaib M, Bosch S, Incel O et al (2014) Fusion of smartphone motion sensors for physical activity. *Recogn Sens* 14:10146–10176
6. Yurtman A, Barshan B (2017) Activity recognition invariant to sensor orientation with wearable motion sensors. *Sensors* 17:1838
7. Janidarmian M, Fekr AR, Radecka K et al (2017) A comprehensive analysis on wearable acceleration sensors in human activity. *Recogn Sens* 17:529
8. Lubina P, Rudzki M (2015) Artificial neural networks in accelerometer-based human activity recognition. In: 2015 22nd international conference on mixed design of integrated circuits systems (MIXDES). IEEE
9. Ordóñez F, Roggen D (2016) Deep Convolutional and LSTM recurrent neural networks for multimodal wearable activity. *Recogn Sens* 16:115

10. Murad A, Pyun JY (2017) Deep recurrent neural networks for human activity recognition. *Sensors* 17:2556
11. Bersch S, Azzi D, Khusainov R et al (2014) Sensor data acquisition and processing parameters for human activity. *Classif Sens* 14:4239–4270
12. Karantonis DM, Narayanan MR, Mathie M et al (2006) Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Trans Inform Technol Biomed* 10:156–167
13. Banos O, Galvez JM, Damas M et al (2014) Window size impact in human activity recognition. *Sensors* 14:6474–6499
14. Bulling A, Blanke U, Schiele B (2014) A tutorial on human activity recognition using body-worn inertial sensors *ACM Comput Surv* 46:1–33
15. Wang A, Chen G, Yang J et al (2016) A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J* 16:4566–4578
16. Igual R, Medrano C, Plaza I (2015) A comparison of public datasets for acceleration-based fall detection. *Med Eng Phys* 37:870–878
17. Roggen D, Calatroni A, Rossi M et al. Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh international conference on networked sensing systems (INSS). IEEE
18. Chavarriaga R, Sagha H, Calatroni A et al (2013) The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. *Pattern Recogn Lett* 34:2033–2042
19. Pedregosa F, Varoquaux G., Gramfort A., et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
20. De Vries S, Garre FG, Engbers LH et al (2011) Evaluation of neural networks to identify types of activity using accelerometers medicine. *Sci Sports Exercise* 43:101–107
21. Chernbumroong S, Atkins AS, Yu H (2011) Activity classification using a single wrist-worn accelerometer. In: 2011 5th international conference on software, knowledge information, industrial management and applications (SKIMA). IEEE