Smart Health 30 (2023) 100432

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

A smartphone accelerometer data-driven approach to recognize activities of daily life: A comparative study

Faisal Hussain [a], Norberto Jorge Goncalves [b], Daniel Alexandre [b], Paulo Jorge Coelho [c,d], Carlos Albuquerque [e,f,g], Valderi Reis Quietinho Leithardt [h], Ivan Miguel Pires [i,j,*]

[a] Al-Khwarizmi Institute of Computer Science (KICS), University of Engineering & Technology (UET), 54890, Lahore, Pakistan
[b] Escola de Ciências e Tecnologia, University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801, Vila Real, Portugal
[c] Institute for Systems Engineering and Computers at Coimbra (INESC Coimbra), DEEC, Pólo II, 3030-290 Coimbra, Portugal
[d] Polytechnic of Leiria, 2411-901 Leiria, Portugal
[e] Health Sciences Research Unit: Nursing (UICISA: E), Nursing School of Coimbra (ESEnfC), 3004-011 Coimbra, Portugal
[f] Higher School of Health, Polytechnic Institute of Viseu, 3500-843 Viseu, Portugal
[g] Child Studies Research Center (CIEC), University of Minho, 4710-057 Braga, Portugal
[h] Instituto Superior de Engenharia de Lisboa (ISEL), Instituto Politécnico de Lisboa, 1959-007, Lisbon, Portugal
[i] Instituto de Telecomunicações, 6201-001 Covilhã, Portugal
[j] Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, 3750-127 Águeda, Portugal

ARTICLE INFO

Original content: Raw dataset with accelerometer, gyroscope and magnetometer data for activities with motion (Original data)

Keywords:
Human activities recognition
Activities recognition
Daily life activities
Human activities detection
Machine learning
Wearable sensors

ABSTRACT

Smartphones have become an indispensable part of our everyday life, influencing various aspects of our routines, from wake-up alarms to managing daily life activities. Nowadays, almost every smartphone has a built-in accelerometer sensor. Motivated by the notable increase in smartphone usage in our everyday life, in this research, we focus on harnessing the potential of smartphone accelerometers to recognize human daily life activities, aiming to leverage the usability and convenience of smartphones. We used smartphone accelerometer data from data collection to daily life activity recognition. To accomplish this, we first collected the smartphone's accelerometer data while performing five activities of daily living (ADLs) namely: moving downstairs, upstairs, running, standing, and walking, from 25 volunteers through a mobile application. After this, we extracted 15 statistical features from the smartphone's accelerometer data to efficiently classify the five referred ADLs. We then applied data pre-processing techniques, i.e., data cleaning and feature extraction. Afterward, we trained nine commonly used machine learning models to recognize five ADLs. Finally, we evaluated and compared the performance of all nine ML models to recognize each activity and analyzed the performance of these trained ML models to identify all five ADLs. The evaluated results revealed that the Adaboost (AB) classifier outperformed all other ML models with 100% area under the curve (AUC), precision, recall, accuracy, and F1-score for recognizing the five ADLs.

* Corresponding author.
E-mail addresses: faisal.hussain.engr@gmail.com (F. Hussain), njg@utad.pt (N.J. Goncalves), daniel@utad.pt (D. Alexandre), paulo.coelho@ipleiria.pt (P.J. Coelho), calbuquerque@essv.ipv.pt (C. Albuquerque), valderireis@deetc.isel.ipl.pt (V.R. Quietinho Leithardt), impires@ua.pt (I.M. Pires).

https://doi.org/10.1016/j.smhl.2023.100432
Received 17 December 2022; Received in revised form 17 July 2023; Accepted 21 October 2023
Available online 31 October 2023
2352-6483/

1. Introduction

Recognizing human daily living activities has become a significant field of research due to its growing applications in real life. Human activities recognition is essential in human health monitoring systems for efficiently measuring human physical activities to manage the pathologies like gait disorder, muscle movement disorder, depression, obesity, and others (Anikwe et al., 2022)– (Gjoreski et al., 2016). The correct measurement of human physical activities helps physicians to easily infer helpful information and prescribe intervention strategies (Ahmed & Ahmed, 2022). Therefore, the automatic recognition of human physical activities is significant for health monitoring systems. In addition, it has numerous applications in many other areas like virtual reality applications, surveillance, security, entertainment games, sports, and others (Pires et al., 2021)– (Pires, Hussain, et al., 2020).

With the tremendous advancement in artificial intelligence (AI) and sensor technology, human activities are now recognized by applying AI techniques over sensor data collected using various sources, including wearable and non-wearable sensors (Liu et al., 2021). Wearable sensors are embedded in a wearable device, including inertial sensors like accelerometers, gyroscopes, magnetometers, and force (pressure) sensors. They can be worn over different body locations to collect human activity data like wrist, neck, waist, foot, and others (Chen et al., 2022), (Kristoffersson et al., 2021). For example, devices like smartphones are also equipped with several inertial sensors to collect or monitor human activity data that can be used while keeping them in the pocket, hand, waist belt, and others (Chen et al., 2022), (Pires et al., 2018), (Dang et al., 2020), (Usmani et al., 2021). On the other hand, non-wearable sensors that can be (or are already) installed in the surrounding environment, like audio sensors, pressure sensors, infrared sensors, floor vibration sensors, and others, enable to monitor or collect human activity data (Ranieri et al., 2021), (Hussain et al., 2019). Altogether this can be used as human activities recognition (HAR) system, which consists of sets of sensors used to collect or monitor human activities.

Based on the sensors used for human activities monitoring, sensor based HAR systems are classified into two types, i.e., wearable HAR systems and ambient HAR systems. Non-wearable sensors include audio, pressure, infrared, floor vibration, and others installed in the surrounding environment to monitor or collect human activity data (Hussain et al., 2019).

HAR systems are broadly classified into two types: vision-based HAR systems and sensor-based HAR systems. The vision-based HAR systems consist of camera sets that collect or monitor human activities by capturing images or videos of a user (Kulsoom et al., 2022), (Moshiri et al., 2021). The data acquired from these vision-based or sensor-based HAR systems can be processed to classify the activities. Typically, the next step is removing or diminishing the raw data's noise, followed by a segmentation process. Afterward, the resultant matrix is used to classify the activities. With the advancement in deep learning models, there is no need to extract the features as the deep learning models can manually extract features and patterns from a given image or video themselves (Kulsoom et al., 2022), (Moshiri et al., 2021). Therefore, the vision-based HAR systems are easy to use because a user does not need to wear or carry the vision-based HAR system. However, the vision based HAR systems can only recognize human activities within a limited area range depending on the underlying cameras used in these systems. Furthermore, they can also affect the privacy of a user. Although with the usage of RGB depth cameras, the privacy issue has been resolved to a large extent, the disadvantage is that the vision-based HAR systems are usually expensive (Poulose et al., 2022).

The algorithms used for activity recognition are categorized into two types, i.e., analytical methods and machine learning methods. In analytical methods, statistical techniques, including thresholding (Poulose et al., 2022), fuzzy logic (Miranda et al., 2022), and hidden Markov models (Cheng et al., 2021), are used to classify human activities. On the other hand, machine learning methods are artificial intelligence-based methods that make HAR systems learn human activity patterns from given data to recognize human activities intelligently. Naïve Bayes, decision trees, support vector machines, k-nearest neighbors, random forest, k-means clustering, and other methods are the most common machine learning methods widely used for recognizing daily human activities. Fig. 1 shows the generic steps to develop machine learning based HAR systems. At first, the data is acquired from wearable or non-wearable sensors like

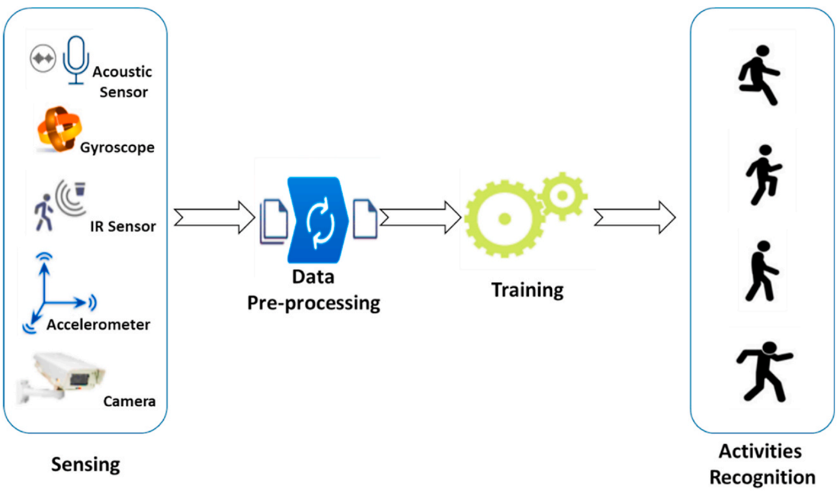

Fig. 1. Generic steps for developing machine learning-based smartphones HAR systems.

accelerometers, gyroscopes, and cameras. Afterward, the acquired data is pre-processed to remove noise, after which features are extracted. Finally, the machine learning algorithms are trained over the resultant feature vector to recognize human activities.

Motivated by the recent trends of sensor based HAR, in this work, we analyzed the performance of nine classical machine learning algorithms for recognizing five daily living human activities using accelerometer data. We intend to establish a baseline to implement a framework for identifying ADL based on wearable data sensors acquisition, in particular accelerometers. Furthermore, the authors seek to verify the algorithms and evaluate their performance according to accuracy, the F1-score, precision, recall, and area under the curve (AUC). The main objective of this work is to improve the current methods of the identification of ADL for the further development of a Personal Digital Life Coach personalized for the user and to perform the empowerment of daily life. The key contributions of this work are as follows.

- We proposed 15 statistical features extracted only from a single smartphone's sensor (accelerometer) data to classify five ADLs efficiently.
- A comparative analysis has been provided among the performances of nine machine learning classifiers for recognizing five ADLs.
- Finally, after a thorough comparison, we proposed the AB classifier to accurately recognize five ADLs as it outperformed all other classifiers and existing state-of-art.

2. Methods and materials

This section presents an overview of the methods and materials used in this comparative study. The methodology used in this paper consists of using a dataset with 15 features extracted from an accelerometer available in common smartphones. Firstly, the private dataset used in this work is analyzed considering the statistic range, minimum, maximum, mean, standard deviation, and variance. Furthermore, the features and target variables are represented. Secondly, nine machine learning methods are presented and applied to this dataset. The results are evaluated by considering the AUC, the classification accuracy (CA), the F1-Score, the precision, and recall metrics. The performance metrics have been analyzed accordingly to the separated classes. After that, a discussion of the overall results is presented to recommend the most appropriate method for activities of daily living (ADL) recognition. Fig. 2 represents the methodology of this comparative study.

2.1. Study design

The data (Pires & Garcia, 2020), (Pires, Marques, et al., 2020) has been collected using a custom-developed mobile application for Android Operating systems. The mobile application was installed in a BQ Aquaris 5.7 smartphone (DeviceSpecifications, 2022). It was placed in the front pocket of the users' pants in a non-intrusive manner, allowing the execution of the different ADLs. The mobile application captures the sensors' data related to ADLs' performance, and the user must select the ADL performed for labeling the dataset. Each activity was performed and captured for 5 s at a frequency rate of 40 Hz. In total, 25 volunteers (15 men and 10 women) have the mobile application and collected data in different environments in the city of Covilha, Portugal. The dataset was collected for approximately 180 h, consisting of 36 h for each ADL. These individuals have different lifestyles, aged between 16 and 60 years old. Regarding the users' lifestyles, 40% of the individuals practice physical exercise, and 60% have sedentary lifestyles. The ADLs analyzed are moving downstairs, upstairs, running, standing, and walking. The Anaconda software has been used as an environment to conduct machine learning experiments on a MacBook Pro. The device has a 2.6 GHz 6-core Intel Core i7, a Random Access Memory (RAM) of 16 GB running at 2400 MHz DDR4, and a Radeon Pro 560 × 4 GB graphics card. Table 1 presents the statistical analysis conducted on IBM SPSS v26. The University of Beira Interior Ethics Committee validated the study with the reference CE-UBI-Pj-2020-035:ID1965.

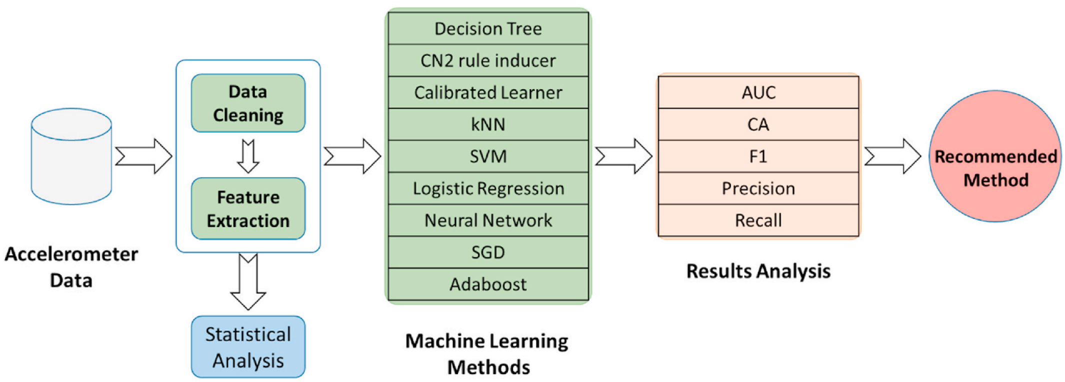

Fig. 2. Proposed methodology pipeline for recognizing five ADLs.

Table 1
Dataset statistical analysis.

Features		Range	Minimum	Maximum	Mean		Std. Deviation	Variance
Id	Names	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
F1	Distance between peaks 1	4677.00	4.00	4681.00	598.10	3.80	379.93	144349.54
F2	Distance between peaks 2	1686.00	2.00	1688.00	447.13	2.20	219.65	48246.16
F3	Distance between peaks 3	1132.00	2.00	1134.00	379.25	1.69	169.37	28684.50
F4	Distance between peaks 4	934.00	1.00	935.00	331.53	1.46	145.69	21224.35
F5	Distance between peaks 5	675.00	1.00	676.00	291.80	1.33	133.25	17756.04
F6	Mean of maximum peaks	17.65	7.31	24.96	13.06	0.04	3.59	12.88
F7	SD of maximum peaks	8.19	0.00	8.19	1.95	0.01	1.45	2.11
F8	Variance of maximum peaks	67.10	$6.314*10^{-6}$	67.10	5.91	0.07	7.49	56.06
F9	Median of maximum peaks	18.65	6.78	25.43	13.06	0.03	4.00	15.97
F10	SD of raw data	8.15	$6.227 * 10^{-3}$	8.16	2.54	0.02	2.03	4.13
F11	Mean of raw data	7.53	7.31	14.84	10.19	0.01	0.86	0.73
F12	Maximum of raw data	20.49	8.62	29.10	16.25	0.04	4.42	19.57
F13	Minimum of raw data	10.93	0.09	11.03	5.27	0.03	3.17	10.06
F14	Variance of raw data	66.56	$3.877 * 10^{-5}$	66.56	10.57	0.14	13.62	185.39
F15	Median of raw data	18.54	6.68	25.22	11.32	0.03	3.22	10.35

2.2. Data pre-processing

We performed two significant steps in the pre-processing data stage: data cleaning and feature extraction.

2.2.1. Data cleaning

During the data acquisition, two significant issues need to be added values and noise in the data. The raw data was processed in the data cleaning stage to handle these issues. While capturing the data, some signal values may be missed due to battery, sensor mis-detection, software issues, etc. Therefore, the samples with missing values were removed to provide refined signals for better training machine learning algorithms.

Table 2
Stratified 10-fold Cross-validation for target class methods specification.

Classifier	Model Parameters
DT	**Pruning:** at least two instances in leaves, at least five instances in internal nodes, maximum depth 100 **Splitting:** Stop splitting when the majority reaches 95% (classification only) **Binary trees:** Yes
RF	**Number of trees:** 10 **Replicable training:** No **Maximal tree depth:** default **Stop splitting nodes with maximum instances:** 5
NB	**Priors:** None **Variance smoothing:** 1e-9
kNN	**Number of neighbors:** 5 **Metric:** Euclidean **Weight:** Uniform
SVM	**SVM type:** SVM, C = 1.0, ε = 0.1 **Kernel:** RBF, exp(-auto$\|x-y\|^2$) **Numerical tolerance:** 0.001 **Iteration limit:** 100
LR	Ridge (L2), C = 1
NN	**Hidden layers:** 100 **Activation:** Logistic **Solver:** SGD **Alpha:** 0.0001 **Max iterations:** 200 **Replicable training:** True
SGD	**Classification loss function:** Hinge **Regularization:** Ridge (L2) **Regularization strength (α):** 1e-05 **Learning rate:** Constant **Initial learning rate (η_0):** 0.01 **Shuffle data after each iteration:** Yes
AB	**Base estimator:** tree **Number of estimators:** 50 **Algorithm (classification):** Samme.r **Loss (regression):** Linear

Besides missing values, the sensors can also capture the environmental noise while recording the activity data. Therefore, a low-pass filter was applied to process a refined signal and reduce noise from sensor values. This low-pass filter on each activity signal reduces the noise induced in the accelerometer signals while performing the ADLs. Moreover, the low-pass filter also minimizes the effect of different constraints caused by the signal.

2.2.2. Feature extraction

After the primary pre-processing stage, the cleaned data yields a refined and noise-free signal for further processing. However, this data is still raw for different stages, providing limited information to machine learning algorithms. The machine learning algorithms exhibit trivial performance for recognizing the activities (Maharana et al., 2022).

There was a need to extract more features based on the existing raw signal values to provide more helpful information for better activity recognition. Based on the current literature, 15 statistical features were extracted from each accelerometer data sample along three axes (x-axis, y-axis, and z-axis) to better train the machine learning algorithms to recognize human activities efficiently. Table 1 displays the features extracted from each sample for identifying the five human daily living activities. We calculated the distance among the six most enormous peaks in the first five features, i.e., F1 to F5. The following four features, i.e., F6 to F9, are extracted by calculating the mean, standard deviation (SD), variance, and median value of the maximum peaks detected previously.Similarly, the last six features, i.e., F10 to F15, are extracted by calculating the SD, mean, maximum, minimum, variance, and median value of the whole raw accelerometer data detected, respectively. The statistical analysis was also performed to observe the overall data distribution of all 15 features on the entire dataset. The results are enlisted in Table 1.

2.3. Models configuration and training

After the data pre-processing, the next step is to train the machine learning algorithms for recognizing five daily human activities. However, before training started, the pre-processed dataset was split into train and test sets. The train set was passed to the machine learning algorithms to train them for the five human daily living activities recognition. On the other hand, the test set was kept hidden during the training stage to evaluate the performance of machine learning algorithms after the training.

The 10-fold cross-validation was applied to split the pre-processed data into train and test sets. Once the train and test sets were generated, the next step was to train the machine learning models. However, before starting the training, there is a need to configure the machine learning models. Table 2 shows the model's parameters used while training these models in a 10-fold cross-validation manner. We used nine commonly used machine learning algorithms, including AdaBoost (AB), Random Forest (RF), k-Nearest Neighbors (kNN), Decision Tree (DT), Logistic Regression (LR), Stochastic Gradient Decent (SGD), Naïve Bayes (NB), Neural Network (NN) and Support Vector Machine (SVM) algorithms to recognize the chosen five human daily living activities.

2.4. Model performance assessment and validation

After training the machine learning models with different configurations, the evaluation of their performances for recognizing human daily living activities was performed. First, five commonly used parameters were calculated to analyze the performance of the nine machine learning models. The five performance parameters include area under the curve (AUC), classification accuracy (CA), F1-score, precision, and recall.

Some values are derived from the basic cardinalities of the confusion matrix to obtain these measures, namely the true positives (TP), the false positives (FP), the true negatives (TN), and the false negatives (FN) (Taha & Hanbury, 2015). TP represents the true positive, where the model correctly predicts the positive class. The TN represents the true negative, where the model correctly predicts the negative class. The FP stands for false positive, meaning the model incorrectly predicts the positive class. Finally, the FN stands for false negative, an outcome where the model incorrectly predicts the negative class. The positive class is the class that contains all the target classes for this subject (prediction of ADL), and the negative class is all the other possibilities.

2.4.1. Classification accuracy (CA)

It is defined as the proportion of accurate predictions (i.e., TP and TN) concerning all predictions. Mathematically, it is calculated as follows:

$$CA = \frac{TP + TN}{TP + FN + TN + FP} \times 100 \tag{1}$$

2.4.2. Precision

It tells about the percentage of actual true positives concerning all positive predictions, i.e., TP and FP. The precision also provides information about Type-I errors, i.e., FP. A Type-I error occurs when a classifier wrongly predicts an ADL and rejects a null Hypothesis (Ho). Mathematically, it is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{2}$$

2.4.3. Recall

It presents the proportion of the truly predicted ADL concerning actual true positive data in the ground truth. It is also called sensitivity. Mathematically, it is given as:

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{3}$$

2.4.4. F1-score

It is the harmonic mean of recall and precision. It refers to the percentage of correctly classified events. Mathematically, it is expressed as:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

The parameters defined above are used for the binary classification problem. While performing the multi-class classification, the one-vs-all (OVA) method was followed, i.e., it was calculated the performance parameters for each class concerning all other classes by considering one class at a time. Finally, the overall results by averaging the values of each parameter for all classes were calculated.

3. Results

As discussed earlier, based on the performance parameters defined in Section 2, it was evaluated the performance of all trained classifiers. Therefore, this section presents the performance of all classifiers for each ADL detection. Afterward, the average performance, calculated using the earlier OVA method, is represented.

Table 3 shows the performance of nine machine learning classifiers for recognizing a person's activity while moving downstairs over all the test set samples. It can be observed that AB and RF have 100% AUC, CA, F1-Score, Precision, and Recall. Thus, AB and RF outperformed all other classifiers for recognizing the Class-1 activity, i.e., moving downstairs. On the other hand, despite having 99.06% AUC, SVM performed worst for identifying the Class-1 activity since it has the lowest F1-Score, Precision (i.e., highest FP or Type-I error), and Recall (i.e., highest FN or Type-II error) results, i.e., 78.66%, 82.14%, 75.45% respectively.

Similarly, Table 4 presents the performance of the machine learning classifiers for recognizing a person's activity while moving upstairs over all the test set samples. Again, the AB classifier has the highest scores, i.e., 100% AUC, 100% CA, 100% F1-Score, 100% Precision, and 100% Recall. So, AB outperformed all other classifiers for recognizing the Class-2 activity, i.e., moving upstairs. On the other hand, despite having 91.28% AUC, SVM performed worst for identifying the Class-2 activity since it has the lowest F1-Score, Precision, and Recall results, i.e., 56.36%, 50.93%, and 63.10%, respectively.

Table 5 also exhibits the performance of previously mentioned machine learning classifiers for recognizing a person's activity while running. It can be observed that AB and RF classifiers have the highest scores, i.e., 100% AUC, 100% CA, 100% F1-Score, 100% Precision, and 100% Recall. Thus, both AB and RF classifiers outperformed all other classifiers for recognizing the Class-3 activity, i.e., running. On the other hand, SVM has the lowest scores for identifying the Class-3 activity, i.e., 97.20% for F1-Score, 96.45% for Precision, and 97.95% for Recall results. Although SVM scores are the lowest for recognizing Class-3 activity, it still shows a better performance score than for identifying the Class-1 and Class-2 activities.

Table 6 displays the classifiers' performance in recognizing a person's activity while standing. It can be observed that AB, RF, and kNN classifiers have the highest scores, i.e., 100% AUC, 100% CA, 100% F1-Score, 100% Precision, and 100% Recall. Thus, AB, RF, and kNN classifiers outperformed all other classifiers for recognizing the Class-4 activity, i.e., standing. On the other hand, SVM has the lowest scores for identifying the Class-4 activity, i.e., 97.16% F1-Score, 96.92% Precision, and 97.40% Recall results. Although SVM scores are the weakest for recognizing Class-4 activity, it still shows a better performance score than for identifying the Class-1 and Class-2 activities.

Finally, Table 7 depicts the performance of nine machine learning classifiers for recognizing a person's walking activity. The AB classifier has the highest scores, i.e., 100% AUC, 100% CA, 100% F1-Score, 100% Precision, and 100% Recall, outperforming all other classifiers for recognizing the Class-5 activity, i.e., walking. On the opposite side, despite having 94.86% AUC, SVM performed worst for identifying the Class-5 activity since it has the lowest F1-Score, Precision, and Recall results, i.e., 57.58%, 64.26%, and 52.15%, respectively.

Table 3
Stratified 10-fold Cross-validation for target Class-1 (i.e., Moving Downstairs).

Model	AUC	CA	F1	Precision	Recall
AB	100.00%	100.00%	100.00%	100.00%	100.00%
RF	100.00%	100.00%	100.00%	100.00%	100.00%
kNN	100.00%	99.56%	98.91%	97.85%	100.00%
DT	99.43%	98.87%	97.18%	97.10%	97.25%
LR	98.89%	95.86%	89.36%	91.91%	86.95%
SGD	91.93%	95.55%	88.53%	91.33%	85.90%
NB	96.43%	94.17%	84.42%	90.70%	78.95%
NN	96.52%	93.97%	84.53%	86.82%	82.35%
SVM	99.06%	91.81%	78.66%	82.14%	75.45%

Table 4
Stratified 10-fold Cross-validation for target Class-2 (i.e., Moving Upstairs).

Model	AUC	CA	F1	Precision	Recall
AB	100.00%	100.00%	100.00%	100.00%	100.00%
RF	100.00%	99.97%	99.93%	99.85%	100.00%
kNN	99.47%	98.94%	97.37%	96.56%	98.20%
DT	99.95%	98.78%	97.04%	94.25%	100.00%
LR	96.56%	93.24%	83.66%	80.96%	86.55%
SGD	90.53%	92.95%	83.07%	79.91%	86.50%
NB	95.84%	92.36%	81.35%	79.48%	83.30%
NN	94.85%	90.87%	78.04%	75.20%	81.10%
SVM	91.28%	80.46%	56.36%	50.93%	63.10%

Table 5
Stratified 10-fold Cross-validation for target Class-3 (i.e., Running).

Model	AUC	CA	F1	Precision	Recall
AB	100.00%	100.00%	100.00%	100.00%	100.00%
RF	100.00%	100.00%	100.00%	100.00%	100.00%
kNN	98.76%	99.45%	98.61%	99.64%	97.60%
DT	99.46%	99.35%	98.37%	98.89%	97.85%
LR	99.97%	99.26%	98.15%	98.05%	98.25%
SGD	99.00%	99.17%	97.92%	98.09%	97.75%
NB	99.47%	99.07%	97.71%	96.09%	99.40%
NN	99.73%	98.98%	97.47%	96.61%	98.35%
SVM	99.37%	98.87%	97.20%	96.45%	97.95%

Table 6
Stratified 10-fold Cross-validation for target Class-4 (i.e., Standing).

Model	AUC	CA	F1	Precision	Recall
AB	100.00%	100.00%	100.00%	100.00%	100.00%
RF	100.00%	100.00%	100.00%	100.00%	100.00%
kNN	100.00%	100.00%	100.00%	100.00%	100.00%
DT	99.76%	99.67%	99.18%	98.47%	99.90%
LR	100.00%	99.64%	99.11%	98.47%	99.75%
SGD	99.95%	99.64%	99.11%	98.47%	99.75%
NB	99.98%	99.49%	98.74%	97.51%	100.00%
NN	99.82%	99.35%	98.40%	96.85%	100.00%
SVM	99.90%	98.86%	97.16%	96.92%	97.40%

Table 7
Stratified 10-fold Cross-validation for target Class-5 (i.e., Walking).

Model	AUC	CA	F1	Precision	Recall
AB	100.00%	100.00%	100.00%	100.00%	100.00%
RF	100.00%	99.97%	99.92%	100.00%	99.85%
kNN	98.28%	98.16%	95.27%	97.94%	92.75%
DT	99.75%	97.60%	93.72%	98.35%	89.50%
LR	96.64%	95.71%	89.18%	89.97%	88.40%
SGD	92.53%	95.42%	88.45%	89.22%	87.70%
NB	96.18%	94.37%	85.79%	86.60%	85.00%
NN	95.83%	92.32%	80.83%	80.71%	80.95%
SVM	94.86%	84.63%	57.58%	64.26%	52.15%

4. Discussion

The average AUC scores of all trained ML classifiers over the test data for recognizing the five ADLs are shown in Fig. 3. The AB and RF classifiers outperformed the other, resulting in a 100% average AUC score for recognizing five ADLs. At the same time, the SGD classifier showed poor performance, i.e., 94.79% average AUC score for recognizing the five ADLs.

Similarly, the average CA scores of all trained ML classifiers over the test data are displayed in Fig. 4. Again, the AB classifier outperformed all other classifiers by resulting 100% average CA score for recognizing the five ADLs. On the other hand, the SVM classifier showed poor performance, i.e., a 90.93% average CA score for recognizing the five ADLs.

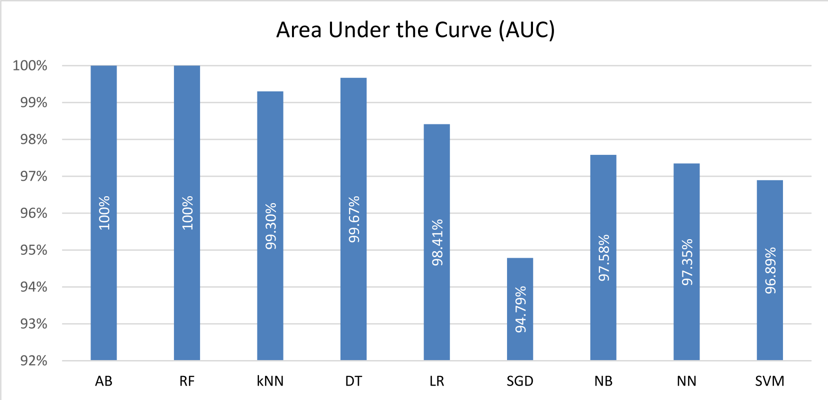

Fig. 3. Average AUC scores of all classifiers for recognizing five ADLs.

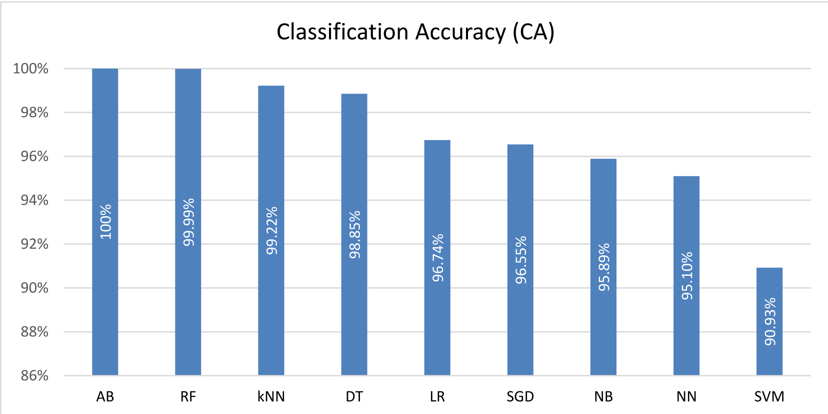

Fig. 4. Average CA scores of all classifiers for recognizing five ADLs.

Likewise, the average F1 scores of all trained ML classifiers over the test data are shown in Fig. 5. Again, the AB classifier surpassed all other classifiers by resulting 100% average F1- score for recognizing the five ADLs. The SVM classifier performs less, with a 77.39% average F1-score for recognizing the five ADLs.

The average precision scores of all trained ML classifiers over the test data are displayed in Fig. 6. The AB classifier outperformed all other classifiers by resulting 100% average precision score for recognizing the five ADLs. However, the SVM classifier also showed poor performance, i.e., a 78.14% average precision score for recognizing the five ADLs.

Similarly, the average recall scores of all trained ML classifiers over the test data are shown in Fig. 7. Again, the AB classifier outperformed all other classifiers by resulting 100% average recall score for recognizing the five ADLs, and the performance of the SVM classifier is lower than others, i.e., 77.21% average recall for recognizing the five ADLs.

In summary, the analysis of the performance of all nine trained ML classifiers for recognizing the five ADLs over the test data shows that the AB and RF classifiers resulted in the highest scores compared to all other classifiers' performance. Further, the performance of AB and RF classifiers is quite similar. However, if a deep comparison is made, the AB classifier outperformed all other classifiers, including RF classifiers, for recognizing the five ADLs. On the other hand, the general performance of the SVM classifier is the lowest when compared to all different classifiers. Henceforth, from the analysis of experimental results, it is concluded that the SVM classifier underperforms when compared to its performance in the AB and RF classifiers.

In contrast, the AB classifier outperformed all other classifiers for recognizing the five ADLs, i.e., moving downstairs, moving upstairs, running, standing, and walking. The key reason behind the outstanding results of the AB classifier is that it leverages ensemble learning in which multiple weak classifiers are combined to form a robust classifier. This approach allows the AB classifier to effectively handle the complex patterns and outliers in the accelerometer data, leading to efficient results in activity recognition. Additionally, the iterative nature of the AB classifier enables it to assign higher weights to misclassified instances, improving its overall performance.

The proposed AB classifier results have also been compared with an existing state-of-the-art in Table 8. Furthermore, the key difference between the performance of both methods can be significantly visualized as you look at Fig. 8. The proposed method outperformed the existing state-of-the-art by yielding 14.78% higher average CA score, 14.87% higher average F1 score, 14.92%

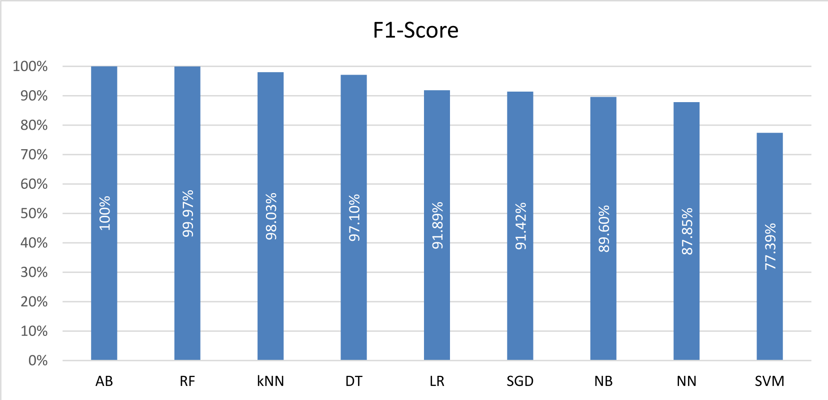

Fig. 5. Average F1-scores of all classifiers for recognizing five ADLs.

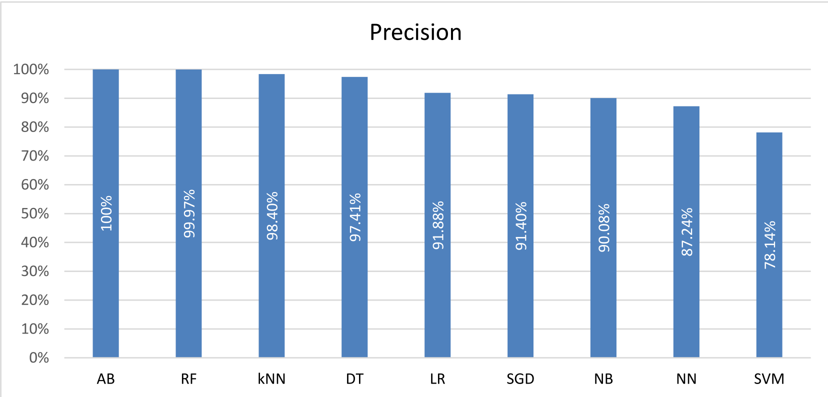

Fig. 6. Average Precision scores of all classifiers for recognizing five ADLs.

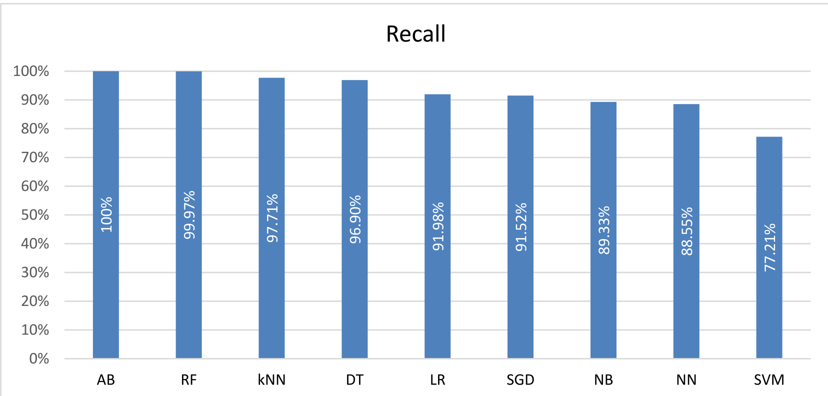

Fig. 7. Average Recall scores of all classifiers for recognizing five ADLs.

Table 8
Performance comparison of the proposed methodology with the existing state-of-the-art for recognizing five ADLs.

Model	AUC	CA	F1	Precision	Recall
AB (Proposed)	100.00%	100.00%	100.00%	100.00%	100.00%
DT (Pires, Marques, et al., 2020)	N/A	85.22%	85.13%	85.08%	85.22%

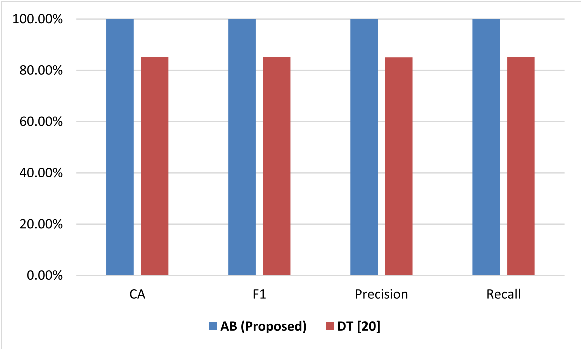

Fig. 8. Average scores comparison the proposed methodology with the existing state-of-the-art for recognizing five ADLs.

higher average precision score, and 14.78% higher average recall score compared to the existing state-of-the-art. Overall, the proposed methodology achieved 14.78% higher average results as compared to the state-of-the-art. Hence, it can be concluded that the proposed methodology is significantly efficient for recognizing five ADLs.

5. Conclusion

Motivated by the notable increase in the usage of smartphones in our daily life, in this work, we used smartphone accelerometer data to recognize five everyday human activities. We first applied data-cleaning techniques to remove missing values and noise removal. We then extracted fifteen features over the smartphone accelerometer data to better train nine commonly used machine learning (ML) classifiers. Afterward, we trained nine commonly used ML models to recognize five activities of daily living (ADLs): moving downstairs, moving upstairs, running, standing, and walking. Lastly, we evaluated the performance of nine trained ML models over the test dataset and compared their performance for recognizing daily life activity.

Furthermore, we also analyzed the overall performance of these trained ML models to identify all five ADLs. The experimental results showed that the AdaBoost (AB) classifier outperformed all other ML models with 100% area under the curve (AUC), precision, recall, accuracy, and F1-score for recognizing the five ADLs. On the other hand, the support vector machine (SVM) classifier underperformed all different classifiers.

In the current work, five ADLs are included and only statistical features are extracted. In future work, we aim to have more ADLs to further explore the smartphone accelerometer data applicability in recognizing other day-to-day activities for human well-being. Similarly, we can incorporate some non-statistical features and other sensor modalities to improve activity recognition accuracy while broadening the range of activities. Moreover, we can investigate real-time activity recognition and system optimization for energy efficiency to enhance the practicality of smartphone-based activity recognition systems.

Declaration of competing interest

There are no conflicts of interest.

Data availability

I have shared the link to my data in the Attach Files step

Raw dataset with accelerometer, gyroscope and magnetometer data for activities with motion (Original data) (Mendeley Data)

Acknowledgements

This work is funded by FCT/MEC through national funds and co-funded by FEDER—PT2020 partnership agreement under the project UIDB/50008/2020.This work is also funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project UIDB/00742/2020.This work is also funded by FCT/MEC through national funds and, when applicable, co-funded by the FEDER-PT2020 partnership agreement under the project UIDB/00308/2020.This article is based upon work from COST Action IC1303–AAPELE–Architectures, Algorithms and Protocols for Enhanced Living Environments and COST Action CA16226–SHELD-ON–Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology). More information is available at www.cost.eu.

References

Ahmed, J., & Ahmed, M. (2022). Machine learning and artificial intelligence in healthcare. In *Empowering artificial intelligence through machine learning* (pp. 43–56). Apple Academic Press.

Anikwe, C. V., et al. (2022). *Mobile and wearable sensors for data-driven health monitoring system: State-of-the-Art and future prospect*. Expert Systems with Applications, Article 117362.

Chen, M, Wang, H, Yu, L, Yeung, EHK, Luo, J, Tsui, K-L, & Zhao, Y (2022). A systematic review of wearable sensor-based technologies for fall risk assessment in older adults. *Sensors, 22*(18), 6752.

Cheng, X., Huang, B., & Zong, J. (2021). Device-free human activity recognition based on GMM-HMM using channel state information. *IEEE Access, 9*, 76592–76601.

Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition, 108*, Article 107561.

Gjoreski, H., Bizjak, J., Gjoreski, M., & Gams, M. (2016). 'Comparing deep and classical machine learning methods for human activity recognition using wrist accelerometer'. In *Presented at the proceedings of the IJCAI 2016 workshop on deep learning for artificial intelligence* (p. 970). New York, NY, USA.

Hussain, F., Hussain, F., Ehatisham-ul-Haq, M., & Azam, M. A. (2019). Activity-aware fall detection and recognition based on wearable sensors. *IEEE Sensors Journal, 19*(12), 4528–4536.

Kristoffersson, A., Du, J., & Ehn, M. (2021). Performance and characteristics of wearable sensor systems discriminating and classifying older adults according to fall risk: A systematic review. *Sensors, 21*(17), 5863.

Kulsoom, F., Narejo, S., Mehmood, Z., Chaudhry, H. N., & Bashir, A. K. (2022). 'A review of machine learning-based human activity recognition for diverse applications'. *Neural Computing & Applications*, 1–36.

Liu, R., Ramli, A. A., Zhang, H., Henricson, E., & Liu, X. (2021). An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence. In *International conference on internet of things* (pp. 1–14). Springer.

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings, 3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020. Jun.

Miranda, L., Viterbo, J., & Bernardini, F. (2022). A survey on the use of machine learning methods in context-aware middlewares for human activity recognition. *Artificial Intelligence Review, 55*(4), 3369–3400.

Moshiri, P. F., Shahbazian, R., Nabati, M., & Ghorashi, S. A. (2021). A CSI-based Human Activity Recognition using deep learning. *Sensors, 21*(21), 7225.

Pires, I., & Garcia, N. M. (2020). Raw dataset with accelerometer, gyroscope and magnetometer data for activities with motion. *Mendeley, V2*. https://doi.org/10.17632/XKNHPZ5T96.2. Aug.

Pires, I. M., Garcia, N. M., Pombo, N., Flórez-Revuelta, F., Spinsante, S., & Teixeira, M. C. (2018). Identification of activities of daily living through data fusion on motion and magnetic sensors embedded on mobile devices. *Pervasive and Mobile Computing, 47*, 78–93.

Pires, I. M., Hussain, F., Garcia, N. M., & Zdravevski, E. (2020). *Improving Human Activity Monitoring by Imputation of Missing Sensory Data: Experimental Study. Future Internet, 12(9), 155*.

Pires, I. M., Hussain, F., Marques, G., & Garcia, N. M. (2021). Comparison of machine learning techniques for the identification of human activities from inertial sensors available in a mobile device after the application of data imputation techniques. *Computers in Biology and Medicine, 135*, Article 104638.

Pires, I. M., Marques, G., Garcia, N. M., & Zdravevski, E. (2020). Identification of activities of daily living through artificial intelligence: An accelerometry-based approach. *Procedia Computer Science, 175*, 308–314.

Poulose, A., Kim, J. H., & Han, D. S. (2022). HIT HAR: Human image threshing machine for human activity recognition using deep learning models. *Computational Intelligence and Neuroscience, 2022*.

Ranieri, C. M., MacLeod, S., Dragone, M., Vargas, P. A., & Romero, R. A. F. (2021). Activity recognition for ambient assisted living with videos, inertial units and ambient sensors. *Sensors, 21*(3), 768.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging, 15*(1), 29. https://doi.org/10.1186/s12880-015-0068-x. Dec.

Usmani, S., Saboor, A., Haris, M., Khan, M. A., & Park, H. (2021). Latest research trends in fall detection and prevention using machine learning: A systematic review. *Sensors, 21*(15), 5134.

DeviceSpecifications, BQ Aquaris 5.7 - Specifications (2022). URL: https://www.devicespecifications.com/en/model/59bb30eb (accessed July. 17, 2022).