

Article

Recognition of Activities of Daily Living and Environments Using Acoustic Sensors Embedded on Mobile Devices

Ivan Miguel Pires ^{1,2,*}, Gonçalo Marques ^{1,†}, Nuno M. Garcia ^{1,†}, Nuno Pombo ^{1,†}, Francisco Flórez-Revueña ^{3,†}, Susanna Spinsante ^{4,†}, Maria Canavarro Teixeira ^{5,6,†} and Eftim Zdravevski ^{7,†}

- ¹ Instituto de Telecomunicações, Universidade da Beira Interior, 6200-001 Covilhã, Portugal; goncalosantosmarques@gmail.com (G.M.); ngarcia@di.ubi.pt (N.M.G.); ngpombo@di.ubi.pt (N.P.)
 - ² Computer Science Department, Polytechnic Institute of Viseu, 3504-510 Viseu, Portugal
 - ³ Department of Computing Technology, University of Alicante, P.O. Box 99, E-03080 Alicante, Spain; francisco.florez@ua.es
 - ⁴ Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona, Italy; s.spinsante@staff.univpm.it
 - ⁵ UTC de Recursos Naturais e Desenvolvimento Sustentável, Polytechnique Institute of Castelo Branco, 6001-909 Castelo Branco, Portugal; ccanavarro@ipcb.pt
 - ⁶ CERNAS-Research Centre for Natural Resources, Environment and Society, Polytechnique Institute of Castelo Branco, 6001-909 Castelo Branco, Portugal
 - ⁷ Faculty of Computer Science and Engineering, University Ss Cyril and Methodius, 1000 Skopje, Macedonia; eftim.zdravevski@finki.ukim.mk
- * Correspondence: impres@it.ubi.pt; Tel.: +351-966-379-785
† These authors contributed equally to this work.

Received: 23 November 2019; Accepted: 3 December 2019; Published: 7 December 2019



Abstract: The identification of Activities of Daily Living (ADL) is intrinsic with the user's environment recognition. This detection can be executed through standard sensors present in every-day mobile devices. On the one hand, the main proposal is to recognize users' environment and standing activities. On the other hand, these features are included in a framework for the ADL and environment identification. Therefore, this paper is divided into two parts—firstly, acoustic sensors are used for the collection of data towards the recognition of the environment and, secondly, the information of the environment recognized is fused with the information gathered by motion and magnetic sensors. The environment and ADL recognition are performed by pattern recognition techniques that aim for the development of a system, including data collection, processing, fusion and classification procedures. These classification techniques include distinctive types of Artificial Neural Networks (ANN), analyzing various implementations of ANN and choosing the most suitable for further inclusion in the following different stages of the developed system. The results present 85.89% accuracy using Deep Neural Networks (DNN) with normalized data for the ADL recognition and 86.50% accuracy using Feedforward Neural Networks (FNN) with non-normalized data for environment recognition. Furthermore, the tests conducted present 100% accuracy for standing activities recognition using DNN with normalized data, which is the most suited for the intended purpose.

Keywords: Activities of Daily Living (ADL); data fusion; environments; feature extraction; pattern recognition; sensors

1. Introduction

Data collection [1] can be conducted using different sensors existing on mobile devices, such as the microphone, the accelerometer, the magnetometer and the gyroscope. The acquired data from mobile sensors are related to the movement and environment where the activities are performed [2]. These data can also be used to develop a method for automatic Activities of Daily Living (ADL) and environment recognition [3].

In continuation of a previous study, available in Reference [4], this paper proposes the use of the microphone for environment identification, that is, bar, classroom, gym, street, kitchen, hall, living room, library and bedroom, which is fused with the data collected using the accelerometer, gyroscope and magnetometer sensors for the recognition of the standing activities, that is, sleeping and watching TV. These methods are included in the design of an ADL and environment recognition framework, proposed in References [5–7]. The advantages of environment recognition are not limited to the increasing number of ADL recognized. Furthermore, this allows the framework to combine the environments with ADL recognition, which returns different results, such as the user walking on the street.

The topic related to the recognition of the ADL has some studies available in the literature [8–13] but there are no studies that use all sensors incorporated in the mobile devices. However, the Artificial Neural Network (ANN) is one of the most used methods in this topic [14,15]. Based on our previous studies using motion and magnetic sensors for the development of an environment and ADL recognition framework [4,16], this paper proposes the creation of several methods to adapt the framework to all sensors incorporated in mobile devices. Some methods using different combinations of sensors are presented in previous studies [4,16], such as the accelerometer, using the accelerometer and magnetometer and using all of the previously described, along with the gyroscope. Thus, this study presents an approach using acoustic data for environment identification, as well as different methods, fusing the environment recognized with other data sources. The proposed method can use the accelerometer and the environment, the accelerometer, the magnetometer and environment but also can be performed using all the mobile sensors and the environment (accelerometer, magnetometer and gyroscope). For the implementation and testing of these methods, we propose the use of ANN [17–19] using three different implementations of ANN [4]. This research also includes the definition of the correct set of features needed and the best implementation of ANN for ADL and environment recognition. The best results are achieved with Feedforward Neural Network (FNN) with Backpropagation for environment recognition and with Deep Learning techniques for standing activities identification.

The main goal of this study is the design of an ADL and environment recognition framework. We discovered that the recognition of the environment increases the number of activities recognized, differentiating the standing activities, where the proposed standing activities are sleeping and watching TV. At this point, the framework will be able to recognize six activities and nine environments, utilizing the accelerometer, gyroscope, magnetometer and mobile microphone sensors.

The Introduction section is concluded in this paragraph and the remaining sections are structured as follows—Section 2 introduces a literature review focused on the use of acoustic sensors for ADL and environment recognition. The methods used for the development of the ADL and environment recognition framework are presented in Section 3. Section 4 presents the results of the implementation of different methods. Finally, the discussion about the results and implementation in the framework is presented in Section 5, the conclusions are presented in Section 6.

2. Related Work

There are no studies related to the use of the fusion of the data collected using all sensors incorporated in off-the-shelf portable devices, including accelerometer, gyroscope, magnetometer and microphone, for ADL and environment recognition [1]. However, numerous methods which incorporate subsets of these mobile sensors are presented in the literature.

The authors of Reference [20] used the Global Positioning System (GPS), accelerometer and microphone sensors for sleeping, walking, standing, running, and social interaction activities recognition using linear and logistic regression methods reporting an accuracy around 90%.

In Reference [21], the authors extracted the minimum, difference between axis, mean, standard deviation, variance, correlation between axis, sum of coefficients, spectral energy and spectral entropy from the accelerometer sensor. Moreover, they study the total spectrum power, zero-crossing rate, spectral centroid, sub-band powers, spectral spread, spectral roll-off, spectral flux and Mel-Frequency Cepstral Coefficients (MFCC) using the microphone. The proposed study applied Gradient Boosting Decision Tree methods and Support Vector Machine (SVM) to recognize several activities such as sitting on a chair, standing, lying, walking, going upstairs and downstairs, running, jogging and drinking. The results report 89.12% and 91.5% accuracy.

The authors of Reference [22] recognized several activities, including cycling, cleaning table, shopping, travelling by car, going to the toilet, cooking, watching television, eating, driving, working at a computer, reading and sleeping, using data acquired from the microphone and accelerometer sensors and applying the Gaussian mixture model (GMM) with log power and MFCC as features, reporting an accuracy of 77.9%.

In Reference [23], the accelerometer and microphone sensors were also used for the recognition of shopping, driving, travelling by car, cooking, washing dishes, cleaning with a vacuum cleaner, waiting in a queue, sleeping, working at a computer, watching television, sitting, being a bar, walking, lying and standing activities, using a J48 decision tree, logistic model tree (LMT) and functional tree (FT), and Instance-based k-Nearest Neighbour (IBk) lazy algorithm with mean, standard deviation, angular degree, range and MFCC as features. The reported accuracies are around 90%, where the LMT decision tree reports 90.4%, the J48 decision tree reports 90.7%, the IBk lazy algorithm reports 90.8% and the FT decision tree reports 90.7% [23].

The remaining studies available in the literature using acoustic sensors do not use data fusion techniques, because they only use the microphone signal. Based on the acoustic signal acquired from the microphone, the authors of Reference [24] used the SVM method with spectral roll-off, slope, minimum, median, coefficient of variation, inverse coefficient of variation, trimmed mean, skewness, kurtosis and 1st, 57th, 95th and 99th percentiles as features. This method presents an accuracy higher than 90% for the recognition of some environments such as restaurant, casino, playground, train, street with ambulance, street traffic, nature at day, nature at night, river and ocean.

In Reference [25], the Linear Discriminant Classifier (LDC) was used with microphone data to recognize several ADLs, including eating, drinking, clearing the throat, relaxing, laughing, coughing, sniffing and talking. This method uses several features including log power, total Root-Mean-Square (RMS) energy, spectral kurtosis, spectral centroid, spectral roll-off, spectral flux, spectral skewness, spectral slope, spectral variance, MFCC, zero crossing rate, minimum, mean, median, maximum, RMS, 1st and 3rd quartiles, interquartile range, standard deviation, skewness, kurtosis, quantity of peaks, mean peaks distance, mean peaks amplitude, mean crossing rate and linear regression slope. The best reported accuracy was achieved using the total RMS energy, spectral flux, spectral centroid, spectral skewness, spectral variance, spectral roll-off, spectral kurtosis, spectral slope and MFCC as features. The average of the reported accuracy was 66.5%.

Artificial Neural Networks (ANN) is one of the most used methods for ADL and environment identification using acoustic signals. In Reference [26], the authors implemented an ANN method, *i.e.*, (Multilayer Perceptron) MLP, with MFCC as features for the identification of acoustic warning signals of emergency units (police, fire department and ambulance), reporting a highest accuracy of 96.7%.

Another study [27] uses ANN for the recognition of several materials collisions such as boll, metal, wood and plastic. Moreover, this research also focuses on the identification of other activities such as door opening/closing, typewriting, knocking, a phone ringing, grains falling, spray and whistle, using time-variance and frequency-variance patterns as features, reporting an average accuracy of 98%.

In Reference [28], the ANN was used for the recognition of sneezing, dog barking, clock ticking, baby crying, crowing rooster, raining, sound of sea waves, fire crackling, sound of helicopter and sound of chainsaw with some features, such as zero crossing rate, MFCC, spectral flatness and spectral centroid, reporting an accuracy around 94.5%.

The authors of Reference [29] used the FNN for the recognition of the sound of sirens from emergency vehicles, automobile horns and normal street sounds with MFCC and zero crossing rate as features, reporting an accuracy between 80% and 100%.

Deep Neural Network (DNN) is another type of ANN used for laughing, singing, crying, arguing and sighing recognition with MFCC as features [30]. The authors of Reference [31] also used DNN for the ambient scene analysis (i.e., voice, music, water and traffic), stress, emotion and speaker recognition with MFCC as features, presenting an accuracy between 60% and 90%.

The SVM is another method used for ADL and environment recognition using acoustic signals. In Reference [32], the authors achieved an accuracy of 78.4% by using the SVM method for keystrokes identification with MFCC as features. Furthermore, the SVM method has been used by the authors of Reference [33] for the identification of several sounds, including beach, forest, street, shaver, crowd football, birds, dog, sink, dishwasher, washing machine, brushing teeth, speech, bus, car, restaurant, phone ringing, train station, chair, vacuum cleaner, coffee machine, raining and computer keyboard, using MFCC as features and reporting an accuracy around 80%. The SVM method is also used for the recognition of sleeping using MFCC and sound pressure level (SPL) as features, reporting accuracies between 75% and 81% [34,35].

The Hidden Markov model (HMM) is another method used for ADL and environment recognition using acoustic signals. In Reference [36], the authors used HMM for the recognition of several sounds such as automobile, aircraft, moped, train and truck. The proposed study has used calculation and storage of sound levels, statistical indices, one-third-octave spectra and noise events detection based on thresholds as features, presenting more than 95% accuracy. In Reference [37], the authors recognized the idle state and the cicada singing sounds with HMM, based on the frequency bands and ratio.

The Gaussian Mixture Model (GMM) is another method used for ADL and environment recognition using acoustic signals. In Reference [38], the authors used GMM with MFCC as features for the recognition of calls during driving, reporting an accuracy around 86%. On the other hand, the authors of Reference [39] used GMM with zero crossing rate, Root Mean Square (RMS), MFCC and low energy frame rate as features for the recognition of emotional states, reporting an accuracy between 65% and 100%.

The authors of Reference [40] used Random Forests and SVM methods for the recognition of street music, siren, gun shot, idling, drilling, dog bark, children playing, car horn and air conditioner sounds. This study used MFCC and motif features, reporting an accuracy between 26.45% and 55.68% with SVM, and between 70.55% and 85% with Random Forests.

In Reference [41], the authors used the decision tree and HMM approach for several ADL and environment identification including reading, meeting, chatting, assisting conference talks, lectures, music, driving, elevator, walking, airplane, fan, vacuuming, shower, clapping, raining, climbing stairs, and wind. The proposed method used a zero crossing rate, low energy frame rate, spectral roll-off, spectral flux, bandwidth, normalized weighted phase deviation, and Relative Spectral Entropy (RSE). The reported accuracy is higher than 78%.

The authors of Reference [42] implemented the GMM, Feed-Forward DNN, Recurrent Neural Networks (RNN), and SVM for the recognition of baby crying and smoking alarm, using MFCC, spectral centroid, spectral flatness, spectral roll-off, spectral kurtosis and zero crossing rate, reporting accuracies between 2% and 24%.

The SVM, diverse density (DD) and expected maximization (EM) methods were implemented in Reference [43] for the recognition of several sounds, including cutlery, water, voice, ambient and music. The proposed method uses MFCC, spectral flux, spectral centroid, bandwidth, Normalized

Mel-Frequency Bands, zero crossing rate and low energy frame rate as features, presenting 87% accuracy (average).

In Reference [44], several sounds were identified, including coffee machine brewing, hand washing, walking, elevator, door opening/closing and silence, using k-Nearest Neighbour (k-NN), SVM and GMM methods. This study use several features, such as zero crossing rate, short-time energy, temporal centroid, energy entropy, autocorrelation, RMS, spectral centroid, spectral roll-off point, spectral spread, spectral entropy, spectral flux, and MFCC methods. The highest accuracies achieved with the different methods are 97.9%, with k-NN, 90%, with GMM, and 100% with SVM [44].

The authors of Reference [45] implemented the Random Forests, HMM, GMM, SVM, ANN, k-NN, and deep belief network methods to recognize babble, driving, machinery, crowded restaurant, street, air conditioner, washer, dryer, and vacuum cleaner, with MFCC, band periodicity and band entropy.

In Reference [46], the authors implemented Naive Bayes, k-NN, Random Forests and Bayesian Networks methods for the recognition of several nursing activities, including the measurement of height, patient sitting, assisting doctor, attaching/measuring/removing electrocardiography (ECG), changing bandage, cleaning body, examining edema and washing hands. This method uses several features, including mean of intensity, mean, variance of intensity, variance, mean of Fast Fourier Transform (FFT)-domain energy, and covariance between intensities. The results reported are 56.10%, with k-NN and Naive Bayes, 73.18%, with k-NN and Bayesian Networks, 55.15%, with Naive Bayes only, 80.96%, with Naive Bayes and Bayesian Networks, 59.03%, with Random Forests and Naive Bayes, and 67.83%, with Random Forests and Bayesian Networks [46].

The identification of various sounds including alarms, birds, clapping, dogs, footsteps, motorcycles, raining, rivers, sea waves, and wind, using k-NN, Naive Bayes, SVM, C4.5 decision tree, logistic regression and ANN, imputing several features is proposed in Reference [47]. These features include skewness, zero crossing rate, kurtosis, spectral spread, spectral roll-off, spectral centroid, spectral flatness measure, spectral slope, spectral flux, spectral skewness, spectral kurtosis, spectral sharpness, spectral crest factor, spectral smoothness, spectral variability, Chroma vectors and MFCC. The highest reported accuracies are 45%, with k-NN, 45%, with Naive Bayes, 54%, with SVM, 45%, with a C4.5 decision tree, 44%, with logistic regression and 54%, with ANN [47].

In Reference [48], a fall detection method was developed with k-NN, SVM, least squares method (LSM), and ANN methods with spectrogram, MFCC, linear predictive coding (LPC) and matching pursuit (MP) as features, reporting 98% accuracy.

The Random Forests classifier was also implemented for the recognition of babble, driving, go to the supermarket, outdoor walking, multiple speakers and kitchen hood. This method use band-periodicity, bandentropy, spectrum flux (SF), subband short-time energy deviation (STED) and subband power spectral deviation (SPSD) as features extracted from the microphone, and present more than 70% accuracy [49]. In Reference [50], the Random Forest was also used to recognize several activities, including using an escalator, an elevator, a drink vending machine and a ticket vending machine, crossing a gate, climbing straight stairs, waiting, entering, queuing, and getting off a train. This study implemented several features extracted from the microphone, such as the step interval, the average step interval variances, the trajectory stretchiness, the peak and trough strength and the amplitude.

The cough sound was recently recognized with a microphone, implementing the k-NN with Hu moment as features [51], which reports accuracies over 93%. Moreover, the the k-NN and the SVM methods are implemented with MFCC, Spectral Centroid, Spectral Bandwidth, Spectral Crest Factor, Spectral Turbulence, Spectral Flux, Ratio f50 versus f90, Spectral Roll-off, Spectral Standard Deviation, Spectral Skewness, Spectral Kurtosis, Spectral Peak Entropy and Tsallis Entropy as features [52], which has accuracies around 99%.

The HMM was also used with the microphone and accelerometer incorporated in mobile and wearable devices for the recognition of different scenes, including meal, arm gestures of eating, conversations, participants, TV viewing, clattering sound, and voice. This study used MFCC,

the average X-axis acceleration and the changing rate were used as features, reporting a minimum accuracy of 88.7% [53].

In Reference [54], the authors used the SVM method for the classification of the different types of vehicles with the Zero Crossing Rate (ZCR), MFCC, Spectral centroid and Spectral flux as features extracted from the microphone, reporting a minimum accuracy equal to 78.95%.

The Adaboost method was proposed in Reference [55] with the maximum, minimum, mean, standard deviation, Root Mean Square (RMS), ZCR, bandwidth, normalized phase deviation and MFCC as features collected using the microphone, gyroscope and magnetometer to identify meals, cooking, TV viewing and conversations, reporting a minimum accuracy of 65%.

The authors of Reference [56] used the J48 decision tree for the recognition of chatting, coding, writing documents, and playing games, reporting 95% accuracy with the maximum, minimum and mean as features.

In Reference [57], the cycling activity was recognized with Weka (REPTree), reporting an accuracy of 97.4% with frequency spectrum as a feature.

Other studies have been done but they used big data and distributed systems and our proposal consists of the use of local processing for the recognition of ADL and its environments [58–60].

Table 1 present the ADL and environments identified using the microphone, verifying that the standing activities are well differentiated with acoustic data.

Table 1. Activities of Daily Living (ADL) and environments identified in the literature review.

ADL:	Number of Studies:
Street with emergency vehicles (police, fire department and ambulance)	6
Sleeping; walking; standing; street traffic; ocean	5
Driving; river	4
Sitting; cleaning with a vacuum cleaner; train; nature; typing; dog barking; baby crying; raining; music	3
Running; lying; going upstairs; going downstairs; drinking; shopping; travelling by car; cooking; watching television; eating; working on a computer; reading; washing dishes; restaurant; laughing; door opening/closing; telephone ringing; helicopter; speech; coffee machine; elevator	2
social interaction activities; jogging; cycling; cleaning table; going to toilet; waiting in a queue; being a bar; casino; playground; clearing the throat; relaxing; coughing; sniffing; talking; grains falling; whistle; sneezing; clock ticking; arguing; football; shaver; bird; dishwasher; brushing teeth; bus; calling; air conditioner; car horn; children playing; drilling; meeting; chatting; shower; clapping; smoking alarm; hand washing	1

Based on the previous studies, the features used for the recognition of ADL and environments with acoustic data are presented in Table 2, showing that the MFCC, zero crossing rate, spectral roll-off, spectral centroid, spectral flux, total RMS energy, mean, standard deviation, minimum, median and low energy frame rate are used in more than 3 studies, with more relevance for MFCC.

At the end, the ADL and environment identification can be executed using several methods shown in Table 3. We found that the approaches with the highest accuracy are ANN, k-NN, Gradient Boosting Decision Tree, IBk lazy algorithm, logistic regression, linear regression and FNN. Following the methods for ADL and environment identification using the acoustic signal, an average accuracy higher than 90% is reported. Moreover, the method that presents better accuracy for ADL and environment the recognition is the MLP, presenting 96% accuracy (average).

Table 2. Features identified in the literature review.

Features:	Number of Studies:
Mel-Frequency Cepstral Coefficients (MFCC)	21
zero-crossing rate	8
spectral roll-off	6
spectral centroid; spectral flux	5
total Root-Mean-Square (RMS) energy	4
Mean; standard deviation; minimum; median; low energy frame rate	3
spectral spread; log power; skewness; kurtosis; sound pressure level (SPL); bandwidth; Relative Spectral Entropy (RSE)	2
total spectrum power; sub-band powers; range; angular degree; slope; coefficient of variation; inverse coefficient of variation; trimmed mean; percentiles (1st, 57th, 95th and 99th); spectral variance; spectral skewness; spectral kurtosis; spectral slope; maximum; quartiles (1st and 3rd); interquartile range; number of peaks; mean distance of peaks; mean amplitude of peaks; mean crossing rate; linear regression slope; spectral flatness; threshold; noise level; one-third-octave spectra; statistical indices; motif; normalized weighted phase deviation; Normalized Mel-Frequency Bands; short-time energy; temporal centroid; energy entropy; autocorrelation; spectral entropy	1

Table 3. Classification methods identified in the literature review.

Methods:	Number of Studies:	Average of Reported Accuracy:
Multi-Layer Perceptron (MLP)	3	96%
k-Nearest Neighbour (k-NN)	3	95%
Gradient Boosting Decision Tree	1	92%
IBk lazy algorithm	1	91%
logistic regression	1	90%
linear regression	1	90%
Feedforward Neural Networks (FNN)	1	90%
Hidden Markov Models (HMM)	2	87%
diverse density (DD)	1	87%
expected maximization (EM)	1	87%
J48 decision tree	2	84%
FT decision tree	2	84%
LMT decision tree	2	84%
Support Vector Machine (SVM)	10	77%
Gaussian mixture model (GMM)	5	76%
Deep Neural Networks (DNN)	3	68%
Linear Discriminant Classifier (LDC)	1	67%
Random Forests	3	66%
Adaboost	1	65%
Recurrent Neural Networks (RNN)	1	24%

3. Methods

In this work, we propose a model for the detection and recognition of the environment detection. This model is based on acoustic sensors and a model for the recognition of standing activities based on motion and magnetic sensors as an enhancement of a previous developed framework for the recognition of ADL and their environments [4–7,16]. The framework was designed to recognize the following ADL—running, walking, going upstairs, sleeping, going downstairs, sleeping, watching TV and standing. In addition, the following scenarios are also recognized by the framework—bar, classroom, gym, kitchen, library, street, hall, living room and bedroom.

3.1. Data Acquisition

The data acquisition module aims to capture all the sensors' data, including accelerometer, magnetometer, gyroscope and microphone. Unlike the microphone, the data from which are saved in a raw forma, this data was acquired at the same time as the study available in Reference [4] and with the same individuals.

3.2. Data Processing

On the one hand, environment recognition comprehends the use of the microphone with the application of the Fast Fourier Transform (FFT) [61] to extract the relevant features. After the application of the FFT, several features were extracted, including 26 MFCC coefficients and standard deviation, average, maximum value, minimum value, variance and median of the raw signal.

On the other hand, the recognition of the standing activities makes use of the environment recognized and accelerometer, magnetometer and/or gyroscope sensors' data with the application of a low pass filter [62], extracting the same features presented in Reference [4].

3.3. Data Fusion

This module encompasses several databases obtained from the combination of different sensors, and features, which are depicted in Figure 1. The different combinations of sensors are:

- Microphone for the Environment Detection
- Accelerometer data plus Environment Recognized
- Accelerometer and Magnetometer data plus Environment Recognized
- Accelerometer, Magnetometer and Gyroscope data plus Environment Recognized

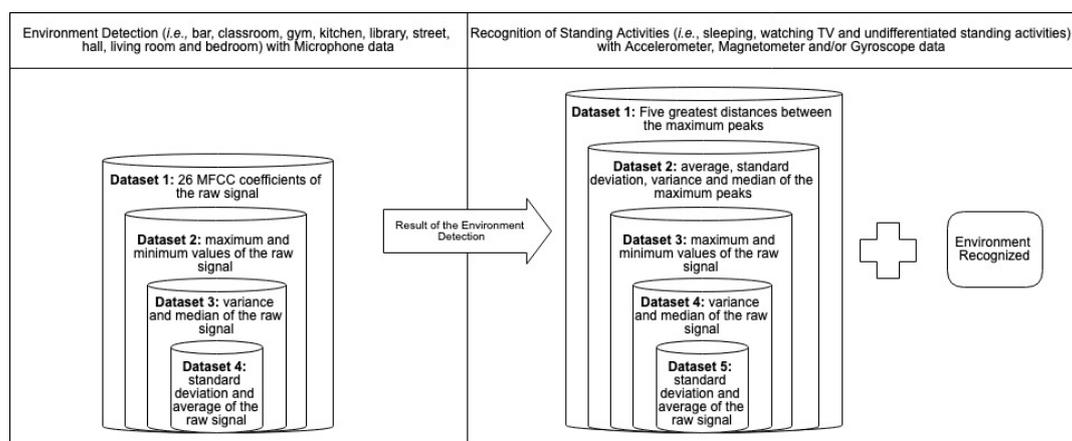


Figure 1. Different combinations of features for the recognition of environment and standing activities.

3.4. Classification

This study aims to recognize nine environments, including bar, classroom, gym, kitchen, library, street, hall, living room and bedroom using the same methods and implementations, which are

implemented and tested in Reference [4]. The different implementations were performed with non-normalized and normalized data, implementing a stop criterion related to the maximum number of training interactions tested with three limits, namely: 10^6 , 2×10^6 and 4×10^6 .

4. Results

4.1. Identification of the Environment of the Activities of Daily Living with Microphone

The implementation of MLP with Backpropagation reported the results presented in Figure 2, verifying that the accuracy reported is very low with all datasets. With non-normalized data (Figure 2a, the results achieved are between 10% and 15%. With normalized data (Figure 2b, the results obtained are between 10% and 20%, where the best results are achieved with dataset 1.

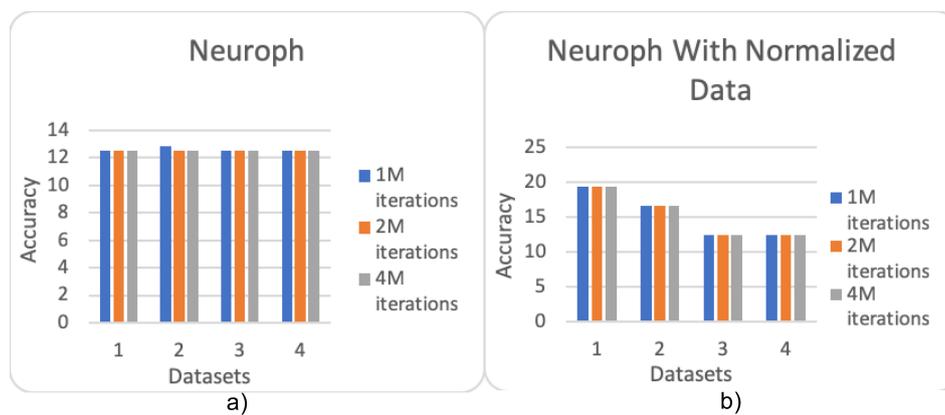


Figure 2. Results obtained with Multilayer Perceptron (MLP) with Backpropagation for the different datasets of microphone data. (a) shows the results with non-normalized data. (b) shows the results with normalized data.

Moreover, the results reported by the implementation of the FNN with Backpropagation are presented in Figure 3. In general, this implementation reports better results with non-normalized data. With non-normalized data (Figure 3a), the FNN reports results higher than 70% with dataset 1 with a maximum number of training iterations, dataset 2 with 10^6 of training iterations, and dataset 4 with 4×10^6 of training iterations. With normalized data (Figure 3b), the FNN reports results below than 60% but the results achieved are higher than 60% with the dataset 4 trained over 10^6 and 2×10^6 of iterations.

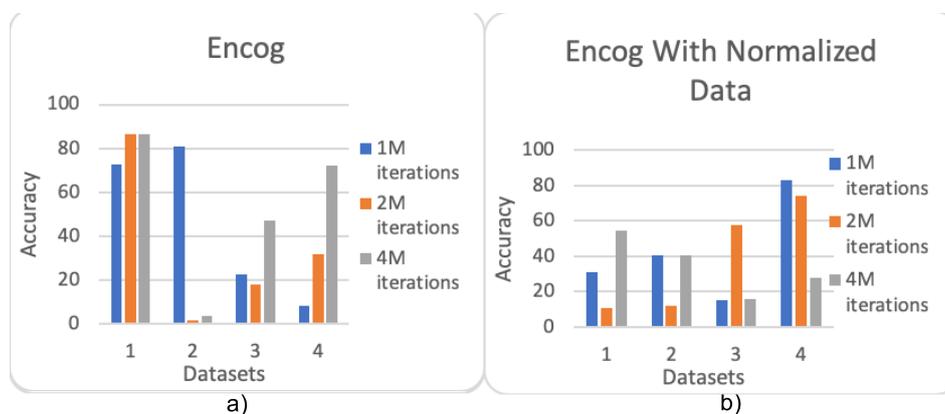


Figure 3. Results obtained with Feedforward Neural Network (FNN) with Backpropagation for the different datasets of microphone data. (a) shows the results with non-normalized data. (b) shows the results with normalized data.

The results of the implementation of DNN are presented in Figure 4, where, with non-normalized data (Figure 4a), the results obtained are below 20% with datasets 1 and 2, and the results obtained are higher than 40% with datasets 3 and 4. In addition, with normalized data (Figure 4b), the results reported are round 50% with all datasets.

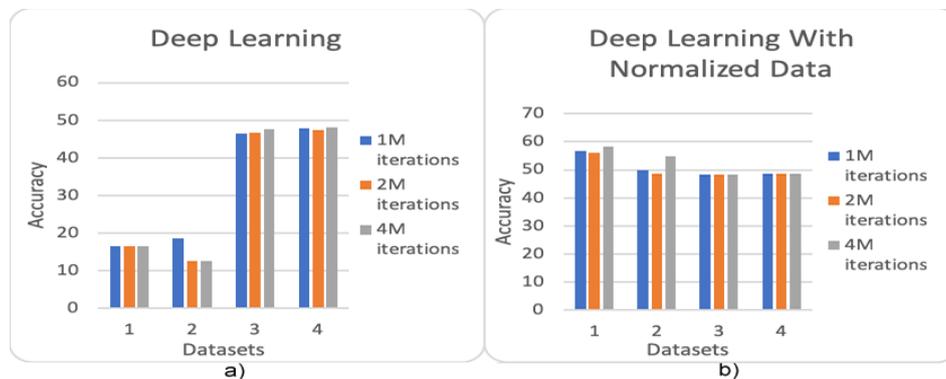


Figure 4. Results obtained with Deep Neural Network (DNN) for the different datasets of microphone data. (a) shows the results with non-normalized data. (b) shows the results with normalized data.

In Table 4, the maximum accuracies achieved with the different implementations of ANN are related to the different datasets used for the microphone data and the maximum number of training iterations, verifying that the best results are achieved with the FNN with Backpropagation with non-normalized data.

Table 4. Best accuracies obtained with the different frameworks, datasets and number of iterations for the recognition of environments using microphone data.

	Framework	Datasets	Iterations Needed for Training	Best Accuracy Achieved (%)
Non-normalized data	MLP with Backpropagation	2	10^6	12.86
	FNN with Backpropagation	1	2×10^6	86.50
	DNN	4	4×10^6	48.11
Normalized data	MLP with Backpropagation	1	10^6	19.43
	FNN with Backpropagation	4	10^6	82.75
	DNN	4	4×10^6	48.74

In conclusion, the method for the recognition of the environment that should be implemented in the framework for the recognition of ADL and their environments is the FNN with Backpropagation using non-normalized data, because it achieves results around 86.50% with the dataset 1.

4.2. Identification of the Standing Activities with the Environment Recognized and the Accelerometer Sensor

The use of normalized data resulted in the achievement of an accuracy of 100% with MLP with Backpropagation, FNN with Backpropagation and DNN methods, because the use of the correct recognition of environments with acoustic data provides a correct discretization of the accelerometer data.

Following the use of non-normalized data, Figure 5 shows the results obtained with MLP with Backpropagation, FNN with Backpropagation and DNN methods. MLP with Backpropagation (Figure 5a) reported results between 50% and 100%, where the better accuracy was achieved with the datasets 1 and 4. FNN with Backpropagation (Figure 5b) reported results around 100%, except with

dataset 1 that achieves an accuracy around 50%. DNN method (Figure 5c) reported results around 100% with datasets 2, 4 and 5 with all training iterations, and with dataset 3 with 4×10^6 iterations, but the results obtained with other combinations are below expectations.

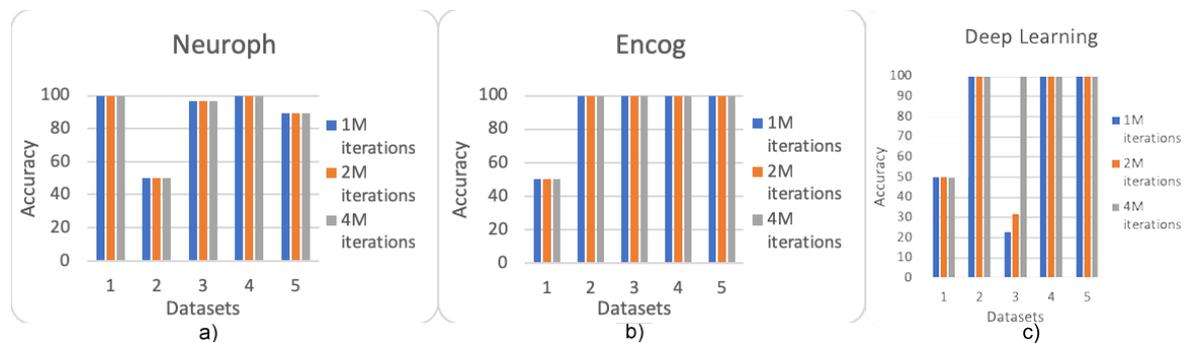


Figure 5. Results obtained with MLP with Backpropagation (a), FNN with Backpropagation (b) and DNN (c) methods for the different datasets of environment and accelerometer data.

In Table 5, the maximum accuracies achieved with the different types of ANN are presented with the relation of the different datasets used for the environment recognized and the accelerometer data and the maximum number of iterations.

Table 5. Best accuracies obtained with the different frameworks, datasets and number of iterations for the recognition of standing activities with the accelerometer data and the environments recognized.

	Framework	Datasets	Iterations Needed for Training	Best Accuracy Achieved (%)
Non-normalized data	MLP with Backpropagation	1	10^6	100.00
	FNN with Backpropagation	2	10^6	100.00
	DNN	2	10^6	100.00
Normalized data	MLP with Backpropagation	1	10^6	100.00
	FNN with Backpropagation	1	10^6	100.00
	DNN	1	10^6	100.00

Regarding the results obtained, in the case of the use of the environment recognized and the accelerometer data in the module for the recognition of standing activities in the framework for the identification ADL and their environments, the implementation that should be used is a DNN with normalized data because the results obtained are always 100%.

4.3. Identification of the Standing Activities with the Environment Recognized and the Accelerometer and Magnetometer Sensors

The use of normalized data resulted in the achievement of an accuracy of 100% with MLP with Backpropagation, FNN with Backpropagation and DNN methods, because the use of the correct recognition of environments with acoustic data provides a correct discretization of the accelerometer and magnetometer data.

Following the use of non-normalized data, Figure 6 shows the results obtained with MLP with Backpropagation, FNN with Backpropagation and DNN methods. MLP with Backpropagation (Figure 6a) reported results around 100%, except with the datasets 1 and 5 which achieved an accuracy around 50%. FNN with Backpropagation (Figure 6b) reported results around 100%. DNN method (Figure 6c) reported results around 100% with dataset 5 with all training iterations, and with dataset 4 with 10^6 of training iterations, but the results obtained with other combinations are below expectations.

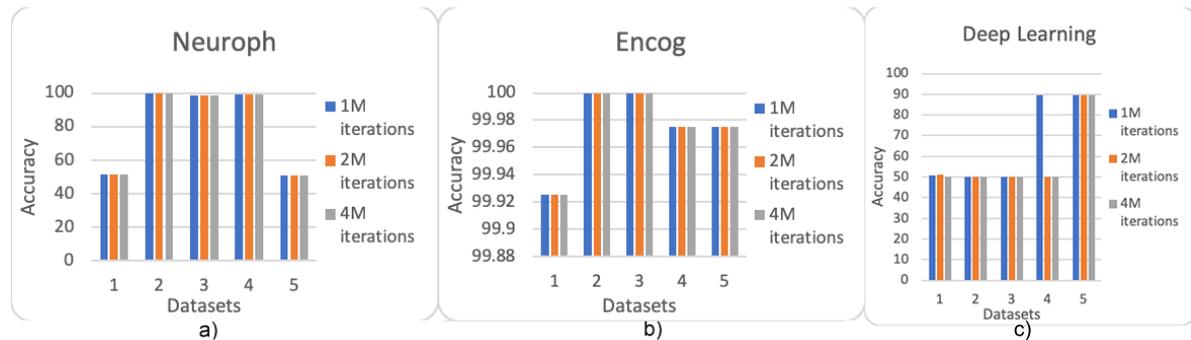


Figure 6. Results obtained with MLP with Backpropagation (a), FNN with Backpropagation (b) and DNN (c) methods for the different datasets of environment and accelerometer and magnetometer sensors' data.

In Table 6, the maximum accuracies achieved with the different implementations of ANN are presented with the relationship between the different datasets used for the environment recognized, and the accelerometer and magnetometer sensors' data, and the maximum number of iterations.

Table 6. Best accuracies obtained with the different frameworks, datasets and number of iterations for the recognition of standing activities with the accelerometer and magnetometer data, and the environments recognized.

	Framework	Datasets	Iterations Needed for Training	Best Accuracy Achieved (%)
Non-normalized data	MLP with Backpropagation	4	10^6	99.05
	FNN with Backpropagation	2	10^6	100.00
	DNN	3	10^6	89.55
Normalized data	MLP with Backpropagation	1	10^6	100.00
	FNN with Backpropagation	1	10^6	100.00
	DNN	1	10^6	100.00

DNN with normalized data always reported results equal to 100% with the use of the accelerometer and magnetometer sensors' data combined with the environment recognized. Thus, the framework for the identification ADL and their environments should implement the DNN with normalized data.

4.4. Identification of the Standing Activities with the Environment Recognized and the Accelerometer, Magnetometer and Gyroscope Sensors

On the one hand, the results reported by the implementation of the MLP with Backpropagation using the MLP with Backpropagation are presented in Figure 7. With non-normalized data (Figure 7a), the results achieved are around 100%, except with the datasets 1 that achieves an accuracy around 50%. With normalized data (Figure 7b), the results obtained are always around 100% with all datasets.

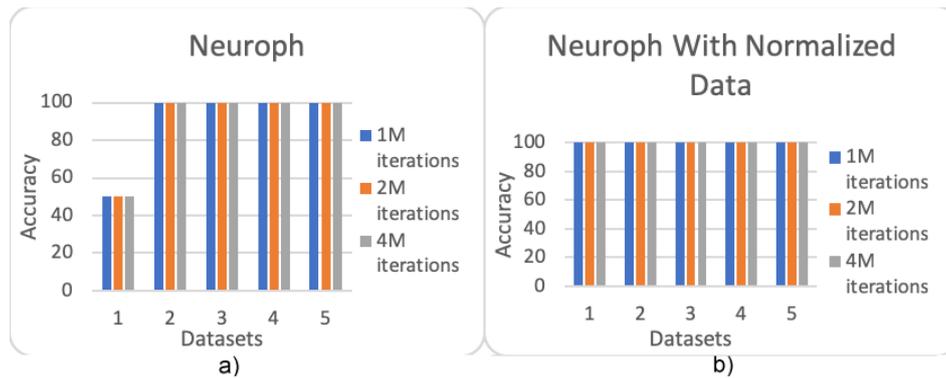


Figure 7. Results obtained with MLP with Backpropagation for the different datasets of environment, and accelerometer, magnetometer and gyroscope sensors' data. (a) shows the results with non-normalized data. (b) shows the results with normalized data.

On the other hand, the results reported by the implementation of the FNN with Backpropagation are presented in Figure 8. With non-normalized data (Figure 8a), the results achieved are always around 100%. With normalized data (Figure 8b), the results obtained are always around 100% with all datasets.

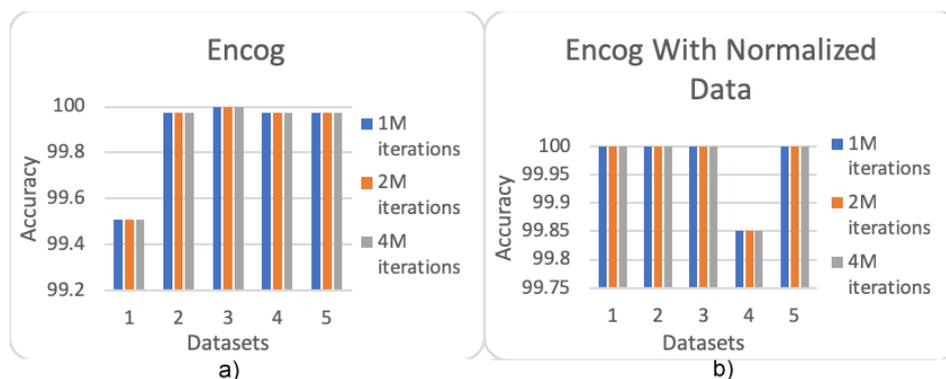


Figure 8. Results obtained with FNN with Backpropagation for the different datasets of environment and accelerometer, magnetometer and gyroscope sensors' data. (a) shows the results with non-normalized data. (b) shows the results with normalized data.

Additionally, the results reported by the implementation of DNN are presented in Figure 9. On the one hand, with non-normalized data (Figure 9a), the results obtained are around 90% with dataset 5 with all training iterations. However, the results obtained with other datasets are below the expectations. On the other hand, with normalized data (Figure 9b), the results obtained are always around 100% with all datasets.

The datasets acquired from the accelerometer, magnetometer and gyroscope combined with the environment recognized, the maximum number of iterations and the maximum accuracies reported by the different implementations of ANN are presented in Table 7.

Using the environment recognized and the accelerometer, magnetometer and gyroscope sensors' data in the module for the recognition of standing activities in the framework for the identification ADL and their environments, the reported results are always 100% with implementation of DNN with normalized data.

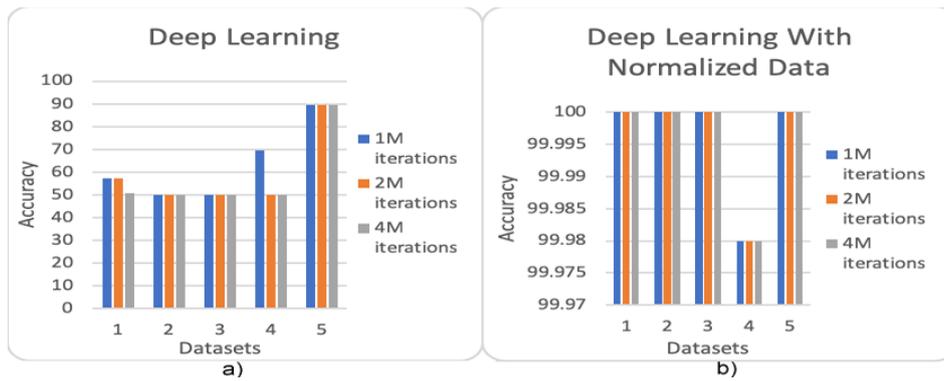


Figure 9. Results obtained with DNN for the different datasets of environment, and accelerometer, magnetometer and gyroscope sensors’ data. (a) shows the results with non-normalized data. (b) shows the results with normalized data.

Table 7. Best accuracies obtained with the different frameworks, datasets and number of iterations for the recognition of standing activities with the accelerometer, gyroscope and magnetometer data, and the environments recognized.

	Framework	Datasets	Iterations Needed for Training	Best Accuracy Achieved (%)
Non-normalized data	MLP with Backpropagation	2	10 ⁶	100.00
	FNN with Backpropagation	3	10 ⁶	100.00
	DNN	5	10 ⁶	89.55
Normalized data	MLP with Backpropagation	1	10 ⁶	100.00
	FNN with Backpropagation	1	10 ⁶	100.00
	DNN	1	10 ⁶	100.00

5. Discussion

This research is included in the development of the framework for the recognition of ADL and their environments, presented in References [5–7]. Furthermore, this study is composed by several modules such as data acquisition, data processing, data fusion, and classification methods. The definition of the method for the identification started in the previous studies [4,16]. These studies have used accelerometer, gyroscope and magnetometer sensors to identify several activities such as going downstairs, going upstairs, running, walking and standing with the DNN, data normalization and L₂ regularization. In Section 4.1, the results of the recognition of the environments using the microphone data, where the environments recognized are bar, classroom, gym, kitchen, library, street, hall, living room and bedroom with the FNN with non-normalized data are presented. Fusing the environment recognized with the accelerometer, gyroscope and magnetometer sensors’ data, the recognition of more standing activities (i.e., watching TV and sleeping) was allowed, increasing the number of ADL recognized at this stage of the development of the framework for the recognition of ADL and environments, as presented in Figure 10.

The characteristics of the mobile devices, that is, the number of sensors available, influences the methods for data fusion and artificial intelligence chosen. Ideally, all sensors available in the mobile device should be used to increase the accuracy of the method. In Figure 10, a simplified schema for the development of a framework for the identification of ADL is presented.

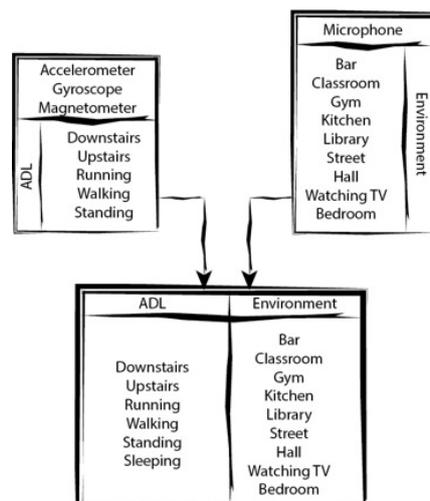


Figure 10. ADL and environments recognized by the framework for the recognition of ADL and environments.

Based on the results reported, the use of acoustic data revealed results with low accuracy because, due to the amount of data used, it reports that the ANN are overfitted. In order to avoid the overfitting problem, we used the early-stop technique, stopping the training of the ANN, when the reducing of the training error stopped. The recognition of standing activities includes only the results obtained with the recognition of the environment. The results obtained for the recognition of standing activities are around 100%, because we considered that the environment is correctly recognized. The results of the final framework will be different because of the recognition of environments that reported lower accuracy. This study only took into account the recognition of environments and standing activities separately. The use of the environment recognized correctly distinguish the activity performed.

The implementation of the framework for the recognition of ADL and their environments is composed by data acquisition, data processing, data cleaning, feature extraction, data fusion and data classification methods. Firstly, based on the results obtained in Section 4.1, the best results achieved for each implementation are presented in Table 4. The best method for the recognition of the environments is the FNN with non-normalized data, reporting an accuracy of 86.50%. Secondly, based on results obtained with the use of the environment recognized and the accelerometer data, presented in Section 4.2, the recognition of standing activities is allowed and the best results achieved for each implementation are presented in Table 4. The best method for the recognition of the standing activities is the DNN with normalization of the data and the application of L_2 regularization, reporting an accuracy of 100%. Thirdly, based on results obtained with the use of the environment recognized and the accelerometer and magnetometer sensors' data, presented in Section 4.3, the recognition of standing activities is allowed and the best results achieved for each implementation are presented in Table 5. The best method for the recognition of the standing activities is the DNN with normalization of the data and the application of L_2 regularization, reporting an accuracy of 100%. Finally, based on results obtained with the use of the environment recognized and the accelerometer, magnetometer and gyroscope sensors' data, presented in Section 4.4, the recognition of standing activities is allowed and the best results achieved for each implementation are presented in Table 6. The best method for the recognition of standing activities is the DNN with normalization of the data and the application of L_2 regularization, reporting an accuracy of 100%.

Our results and implementations cannot be directly compared with other studies because the datasets and implementation code used by other authors are not share. We asked other authors about the details of the implementation but they did not answer at the moment.

In conclusion, when the activity was recognized as standing and the environment is correctly identified, the accuracy for the recognition of standing activities is 100%. At this stage of the framework

for the recognition of ADL and their environments, two different classification methods are defined, these are:

- DNN with normalized data for the general identification of ADL;
- FNN with non-normalized data for the general identification of the environments;
- DNN with normalized data for the identification of standing activities.

6. Conclusions

The development of a framework for ADL [1] and environment recognition using mobile sensors, including accelerometer, gyroscope, magnetometer and microphone, with the architecture presented in References [5–7], has several steps including data acquisition, data processing, data fusion and classification methods. At this stage of the development, the proposed identified ADL are running, walking, standing, going downstairs and upstairs, and sleeping, and the proposed identified environments are bar, classroom, gym, kitchen, library, street, hall, watching TV and bedroom.

Depending on the types of sensors, several features were extracted from the sensors' data for further processing. The features extracted from the microphone are 26 MFCC coefficients and standard deviation, average, maximum value, minimum value, variance and median of the raw signal. Following the motion and magnetic sensors, we extracted the same features of the previous study [4]. The method developed should be adapted to the number of sensors available in the off-the-shelf mobile devices and adapted to the limited resources of these devices.

In coherence with the previous studies [4,16], this research includes the comparison of three different implementations of ANN, such as MLP and FNN with Backpropagation, and the DNN. The DNN is the best method for the recognition of general ADL and standing activities, but the FNN with Backpropagation is the best method for the recognition of environments. In Reference [4], the different parameters of the ANN implemented are detailed.

The accuracies of the recognition ADL and their environments are different depending on the different stages of the framework for the recognition of ADL and environments. Firstly, the best accuracy for the recognition of the general ADL, presented in previous studies [4,16], is 85.89%, implementing the DNN using L_2 regularization and normalized data. Secondly, the best accuracy for the recognition of the environments is 86.50%, implementing the FNN with Backpropagation using non-normalized data. Finally, the recognition of standing activities are always around 100% with all implementations studied, but, due to the performance, the best method for the implementation in the framework is the DNN using L_2 regularization and normalized data.

As future work, we intend to develop a framework for the identification of ADL and their environments, adapting the method to the number of sensors available on the mobile device. The recognition of the environments allows the framework for identifying the location in the indoor/outdoor environments, where the ADL were performed. The environment recognition can also improve the recognition of ADL, increasing the number of ADL recognized. The data related to this research are available in a free repository [63].

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, writing—original draft preparation, writing—review and editing: I.M.P., G.M., N.M.G., N.P., F.F.R., S.S., M.C.T. and E.Z.

Funding: This work is funded by FCT/MEC through national funds and co-funded by FEDER-PT2020 partnership agreement under the project UID/EEA/50008/2019.

Acknowledgments: This work is funded by FCT/MEC through national funds and when applicable co-funded by FEDER-PT2020 partnership agreement under the project UID/EEA/50008/2019. This work is funded by FCT/MEC through national funds and co-funded by FEDER-PT2020 partnership agreement under the project (*Este trabalho é financiado pela FCT/MEC através de fundos nacionais e cofinanciado pelo FEDER, no âmbito do Acordo de Parceria PT2020 no âmbito do projeto UID/EEA/50008/2019*). 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 in www.cost.eu.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Foti, D.; Koketsu, J.S. Activities of daily living. In *Pedretti's Occupational Therapy: Practical Skills for Physical Dysfunction*; Elsevier: Amsterdam, The Netherlands, 2013; Volume 7, pp. 157–232.
2. Salazar, L.H.A.; Lacerda, T.; Nunes, J.V.; von Wangenheim, C.G. A Systematic Literature Review on Usability Heuristics for Mobile Phones. *Int. J. Mob. Hum. Comput. Interact.* **2013**, *5*, 50–61. [[CrossRef](#)]
3. Garcia, N.M. *A Roadmap to the Design of A Personal Digital Life Coach*; Springer: Berlin, Germany, 2016.
4. Pires, I.M.; Garcia, N.M.; Pombo, N.; Flórez-Revuelta, F.; Spinsante, S.; Teixeira, M.C. Identification of Activities of Daily Living through Data Fusion on Motion and Magnetic Sensors embedded on Mobile Devices. *Pervasive Mob. Comput.* **2018**, *47*, 78–93. [[CrossRef](#)]
5. Pires, I.; Garcia, N.; Pombo, N.; Flórez-Revuelta, F. From Data Acquisition to Data Fusion: A Comprehensive Review and a Roadmap for the Identification of Activities of Daily Living Using Mobile Devices. *Sensors* **2016**, *16*, 184. [[CrossRef](#)] [[PubMed](#)]
6. Pires, I.M.; Garcia, N.M.; Flórez-Revuelta, F. Multi-sensor data fusion techniques for the identification of activities of daily living using mobile devices. In Proceedings of the ECMLPKDD 2015 Doctoral Consortium, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Porto, Portugal, 7–11 September 2015.
7. Pires, I.M.; Garcia, N.M.; Pombo, N.; Flórez-Revuelta, F. Identification of Activities of Daily Living Using Sensors Available in off-the-shelf Mobile Devices: Research and Hypothesis. In Proceedings of the Ambient Intelligence-Software and Applications-7th International Symposium on Ambient Intelligence (ISAmI 2016), Seville, Spain, 1–3 June 2016; pp. 121–130.
8. Banos, O.; Damas, M.; Pomares, H.; Rojas, I. On the use of sensor fusion to reduce the impact of rotational and additive noise in human activity recognition. *Sensors* **2012**, *12*, 8039–8054. [[CrossRef](#)] [[PubMed](#)]
9. Akhoundi, M.A.A.; Valavi, E. Multi-Sensor Fuzzy Data Fusion Using Sensors with Different Characteristics. *arXiv* **2010**, arXiv:1010.6096.
10. Paul, P.; George, T. An Effective Approach for Human Activity Recognition on Smartphone. In Proceedings of the 2015 IEEE International Conference on Engineering and Technology (Icotech), Coimbatore, India, 20 March 2015; pp. 45–47. [[CrossRef](#)]
11. Hsu, Y.-W.; Chen, K.-H.; Yang, J.-J.; Jaw, F.-S. Smartphone-based fall detection algorithm using feature extraction. In Proceedings of the 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, China, 15–17 October 2016; pp. 1535–1540.
12. Dernbach, S.; Das, B.; Krishnan, N.C.; Thomas, B.L.; Cook, D.J. Simple and Complex Activity Recognition through Smart Phones. In Proceedings of the 8th International Conference on Intelligent Environments (IE), Guanajuato, Mexico, 26–29 June 2012; pp. 214–221.
13. Shen, C.; Chen, Y.F.; Yang, G.S. On Motion-Sensor Behavior Analysis for Human-Activity Recognition via Smartphones. In Proceedings of the IEEE International Conference on Identity, Security and Behavior Analysis (Isba), Sendai, Japan, 29 February–2 March 2016; pp. 1–6.
14. Wang, D. Pattern recognition: Neural networks in perspective. *IEEE Expert* **1993**, *8*, 52–60. [[CrossRef](#)]
15. Doya, K.; Wang, D. Exciting Time for Neural Networks. *Neural Netw.* **2015**, *61*. [[CrossRef](#)]
16. Pires, I.M.; Garcia, N.M.; Pombo, N.; Pires, F.F.L.; Spinsante, S.; Teixeira, M.C.; Zdravevski, E. Pattern Recognition Techniques for the Identification of Activities of Daily Living using Mobile Device Accelerometer. *PeerJ Prepr.* **2019**. [[CrossRef](#)]
17. Gripenberg, G. Approximation by neural networks with a bounded number of nodes at each level. *J. Approx. Theory* **2003**, *122*, 260–266. [[CrossRef](#)]
18. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)]
19. Costarelli, D.; Vinti, G. Pointwise and uniform approximation by multivariate neural network operators of the max-product type. *Neural Netw.* **2016**, *81*, 81–90. [[CrossRef](#)] [[PubMed](#)]
20. Lane, N.D.; Mohammad, M.; Lin, M.; Yang, X.; Lu, H.; Ali, S.; Doryab, A.; Berke, E.; Choudhury, T.; Campbell, A. Bewell: A smartphone application to monitor, model and promote wellbeing. In Proceedings of the 5th international ICST conference on pervasive computing technologies for healthcare, Dublin, Ireland, 23–26 May 2011.

21. Mengistu, Y.; Pham, M.; Do, H.M.; Sheng, W. AutoHydrate: A Wearable Hydration Monitoring System. In Proceedings of the IEEE/Rsj International Conference on Intelligent Robots and Systems (Iros 2016), Daejeon, Korea, 9–14 October 2016; pp. 1857–1862. [[CrossRef](#)]
22. Nishida, M.; Kitaoka, N.; Takeda, K. Daily activity recognition based on acoustic signals and acceleration signals estimated with Gaussian process. In Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, China, 16–19 December 2015; pp. 279–282.
23. Filios, G.; Nikolettseas, S.; Pavlopoulou, C.; Rapti, M.; Ziegler, S. Hierarchical Algorithm for Daily Activity Recognition via Smartphone Sensors. In Proceedings of the IEEE 2nd World Forum on Internet of Things (Wf-Iot), Milan, Italy, 14–16 December 2015; pp. 381–386. [[CrossRef](#)]
24. Delgado-Contreras, J.R.; Garaeia-Vázquez, J.P.; Brena, R.F.; Galván-Tejada, C.E.; Galván-Tejada, J.I. Feature Selection for Place Classification through Environmental Sounds. *Procedia Comput. Sci.* **2014**, *37*, 40–47. [[CrossRef](#)]
25. Rahman, T.; Adams, A.T.; Zhang, M.; Cherry, E.; Zhou, B.; Peng, H.; Choudhury, T. BodyBeat: A mobile system for sensing non-speech body sounds. In Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, Bretton Woods, NH, USA, 16–19 June 2014.
26. Mielke, M.; Brück, R. Smartphone application for automatic classification of environmental sound. In Proceedings of the 20th International Conference Mixed Design of Integrated Circuits and Systems-MIXDES, Gdynia, Poland, 20–22 June 2013; pp. 512–515.
27. Guo, X.; Toyoda, Y.; Li, H.; Huang, J.; Ding, S.; Liu, Y. Environmental sound recognition using time-frequency intersection patterns. In Proceedings of the 3rd International Conference on Awareness Science and Technology (iCAST), Ypsilanti, MI, USA, 3–5 October 2011; pp. 243–246.
28. Pillos, A.; Alghamidi, K.; Alzamel, N.; Pavlov, V.; Machanavajhala, S. A real-time environmental sound recognition system for the Android OS. In Proceedings of the Detection and Classification of Acoustic Scenes and Events, Budapest, Hungary, 3 September 2016.
29. Mielke, M.; Brueck, R. Design and evaluation of a smartphone application for non-speech sound awareness for people with hearing loss. In Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 5008–5011.
30. Dubey, H.; Mehl, M.R.; Mankodiya, K. BigEAR: Inferring the Ambient and Emotional Correlates from Smartphone-Based Acoustic Big Data. In Proceedings of the IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Washington, DC, USA, 27–29 June 2016; pp. 78–83.
31. Lane, N.D.; Georgiev, P.; Qendro, L. DeepEar: Robust smartphone audio sensing in unconstrained acoustic environments using DNN. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015.
32. Wang, J.; Ruby, R.; Wang, L.; Wu, K. Accurate Combined Keystrokes Detection Using Acoustic Signals. In Proceedings of the 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), Shenyang, China, 9–14 December 2016.
33. Rossi, M.; Feese, S.; Amft, O.; Braune, N.; Martis, S.; Tröster, G. AmbientSense: A real-time ambient sound recognition system for smartphones. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), San Diego, CA, USA, 18–22 March 2013; pp. 230–235.
34. Nishijima, K.; Uenohara, S.; Furuya, K. A Study on the Optimum Number of Training Data in Snore Activity Detection Using SVM. In Proceedings of the 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), Fukuoka, Japan, 6–8 July 2016; pp. 582–584.
35. Nishijima, K.; Uenohara, S.; Furuya, K. Snore activity detection using smartphone sensors. In Proceedings of the IEEE International Conference on Consumer Electronics-Taiwan, Taipei, Taiwan, 6–8 June 2015; pp. 128–129.
36. Gaunard, P.; Mubikangiey, C.G.; Couvreur, C.; Fontaine, V. Automatic classification of environmental noise events by hidden Markov models. *IEEE Int. Conf. Acoust. Speech Signal Process.* **1998**, *3*, 3609–3612.
37. Zilli, D.; Parson, O.; Merrett, G.V.; Rogers, A. A Hidden Markov Model-Based Acoust. Cicada Detect. Crowdsourced Smartphone Biodivers. *Monit. J. Artif. Int. Res.* **2014**, *51*, 805–827.

38. Song, T.; Cheng, X.; Li, H.; Yu, J.; Wang, S.; Bie, R. Detecting driver phone calls in a moving vehicle based on voice features. In Proceedings of the IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, 10–14 April 2016.
39. Chen, Y.A.; Chen, J.; Tseng, Y.C. Inference of Conversation Partners by Cooperative Acoustic Sensing in Smartphone Networks. *IEEE Trans. Mob. Comput.* **2016**, *15*, 1387–1400. [[CrossRef](#)]
40. Gomes, E.F.; Batista, B.; Jorge, P.M. Using Smartphones to Classify Urban Sounds. In Proceedings of the Ninth International Conference on Computer Science & Software Engineering, Porto, Portugal, 20–22 July 2016.
41. Lu, H.; Pan, W.; Lane, N.D.; Choudhury, T.; Campbell, A.T. SoundSense: Scalable sound sensing for people-centric applications on mobile phones. In Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, Kraków, Poland, 22–25 June 2009.
42. Sigtia, S.; Stark, A.M.; Krstulovic, S.; Plumbley, M.D. Automatic Environmental Sound Recognition: Performance Versus Computational Cost. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 2096–2107. [[CrossRef](#)]
43. Kelly, D.; Caulfield, B. Pervasive Sound Sensing: A Weakly Supervised Training Approach. *IEEE Trans. Cybern.* **2016**, *46*, 123–135. [[CrossRef](#)]
44. Abreha, G.T. An Environmental Audio-Based Contextrecognition System Using Smartphones. Master's Thesis, University of Twente, Enschede, The Netherlands, August 2014.
45. Saki, F.; Sehgal, A.; Panahi, I.; Kehtarnavaz, N. Smartphone-based real-time classification of noise signals using subband features and random forest classifier. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 2204–2208.
46. Inoue, S.; Ueda, N.; Nohara, Y.; Nakashima, N. Mobile activity recognition for a whole day: Recognizing real nursing activities with big dataset. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015.
47. Bountourakis, V.; Vrysis, L.; Papanikolaou, G. Machine Learning Algorithms for Environmental Sound Recognition: Towards Soundscape Semantics. In Proceedings of the Audio Mostly 2015 on Interaction with Sound, Thessaloniki, Greece, 7–9 October 2015.
48. Cheffena, M. Fall Detection Using Smartphone Audio Features. *IEEE J. Biomed. Health Inf.* **2016**, *20*, 1073–1080. [[CrossRef](#)]
49. Sehgal, A.; Saki, F.; Kehtarnavaz, N. Real-time implementation of voice activity detector on ARM embedded processor of smartphones. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017. [[CrossRef](#)]
50. Elhamshary, M.; Youssef, M.; Uchiyama, A.; Yamaguchi, H.; Higashino, T. CrowdMeter: Congestion Level Estimation in Railway Stations Using Smartphones. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–12. [[CrossRef](#)]
51. Hoyos-Barceló, C.; Monge-Álvarez, J.; Shakir, M.Z.; Alcaraz-Calero, J.-M.; Casaseca-de-la-Higuera, P. Efficient k-NN Implementation for Real-Time Detection of Cough Events in Smartphones. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1662–1671. [[CrossRef](#)]
52. Monge-Alvarez, J.; Hoyos-Barcelo, C.; Lesso, P.; Casaseca-de-la-Higuera, P. Robust Detection of Audio-Cough Events using local Hu moments. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 184–196. [[CrossRef](#)]
53. Bi, C.; Xing, G.; Hao, T.; Huh, J.; Peng, W.; Ma, M. FamilyLog: A mobile system for monitoring family mealtime activities. In Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom), Seattle, WA, USA, 21–25 March 2017; pp. 21–30. [[CrossRef](#)]
54. Soni, S.; Aggarwal, N.; Vij, D.; Doegar, A. Acoustic Scene Classification for Personal Commuting Mode: Detecting Polluting vs. Non Polluting Vehicles. In Proceedings of the 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 11–12 January 2018; pp. 274–279. [[CrossRef](#)]
55. Gu, F.; Niu, J.; He, Z.; Jin, X.; Rodrigues, J.J.P.C. SmartBuddy: An Integrated Mobile Sensing and Detecting System for Family Activities. In Proceedings of the 2017 IEEE Global Communications Conference (GLOBECOM 2017), Singapore, 4–8 December 2017; pp. 1–7. [[CrossRef](#)]
56. Yu, Z.; Du, H.; Xiao, D.; Wang, Z.; Han, Q.; Guo, B. SmartBuddy: An Integrated Mobile Sensing and Detecting System for Family Activities. *IEEE Internet Things J.* **2018**, *5*, 1156–1168. [[CrossRef](#)]

57. Kawanaka, S.; Kashimoto, Y.; Firouzian, A.; Arakawa, Y.; Pulli, P.; Yasumoto, K. Approaching vehicle detection method with acoustic analysis using smartphone for elderly bicycle driver. In Proceedings of the 2017 Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), Toyama, Japan, 3–5 October 2017; pp. 1–6. [CrossRef]
58. Su, X.; Sperli, G.; Moscato, V.; Picariello, A.; Esposito, C.; Choi, C. An Edge Intelligence Empowered Recommender System Enabling Cultural Heritage Applications. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4266–4275. [CrossRef]
59. Chen, L.; Nugent, C.D. Sensor-Based Activity Recognition Review. In *Human Activity Recognition and Behaviour Analysis*; Springer: Cham, Switzerland, 2019; pp. 23–47.
60. Amato, F.; Moscato, V.; Picariello, A.; Sperli, G. Extreme events management using multimedia social networks. *Future Gener. Comput. Syst.* **2019**, *94*, 444–452. [CrossRef]
61. Rader, C.; Brenner, N. A new principle for fast Fourier transformation. *IEEE Trans. Acoust. Speech Signal Process.* **1976**, *24*, 264–266. [CrossRef]
62. Graizer, V. Effect of low-pass filtering and re-sampling on spectral and peak ground acceleration in strong-motion records. In Proceedings of the 15th World Conference of Earthquake Engineering, Lisbon, Portugal, 24–28 September 2012; pp. 24–28.
63. ALLab. August 2017-Multi-Sensor Data Fusion in Mobile Devices for the Identification of Activities of Daily Living-ALLab Signals. Available online: https://allab.di.ubi.pt/mediawiki/index.php/August_2017-_Multi-sensor_data_fusion_in_mobile_devices_for_the_identification_of_activities_of_daily_living (accessed on 2 September 2017).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).