

Article

Classification of Transition Human Activities in IoT Environments via Memory-Based Neural Networks

Giovanni Acampora *, Gianluca Minopoli, Francesco Musella and Mariacarla Staffa 

Department of Physics “Ettore Pancini”, University of Naples Federico II, Complesso di Monte Sant’Angelo, Via Cintia 21, 80126 Napoli, Italy; lucaminopoli1@gmail.com (G.M.); fran.musella@gmail.com (F.M.); mariacarla.staffa@unina.it (M.S.)

* Correspondence: giovanni.acampora@unina.it; Tel.: +39-081-676150

Received: 31 December 2019; Accepted: 14 February 2020; Published: 28 February 2020



Abstract: Human activity recognition is a crucial task in several modern applications based on the Internet of Things (IoT) paradigm, from the design of intelligent video surveillance systems to the development of elderly robot assistants. Recently, machine learning algorithms have been strongly investigated to improve the recognition task of human activities. Though, in spite of these research activities, there are not so many studies focusing on the efficient recognition of complex human activities, namely transitional activities, and there is no research aimed at evaluating the effects of noise in data used to train algorithms. In this paper, we bridge this gap by introducing an innovative activity recognition system based on a neural classifier endowed with memory, able to optimize the performance of the classification of both transitional and non-transitional human activities. The system recognizes human activities from unobtrusive IoT devices (such as the accelerometer and gyroscope) integrated in commonly used smartphones. The main peculiarity provided by the proposed system is related to the exploitation of a neural network extended with short-term memory information about the previous activities’ features. The experimental study proves the reliability of the proposed system in terms of accuracy with respect to state-of-the-art classifiers and the robustness of the proposed framework with respect to noise in data.

Keywords: internet of things; human activity recognition; machine learning; neural networks; sensors

1. Introduction

Currently, the wide adoption of Internet of Things (IoT) technologies is enabling a transparent way to gather useful data about human beings’ status in uncontrolled environments and paves the way towards a plethora of new application domains where embedded sensors, such as accelerometers and gyroscopes, can be used as indirect and unobtrusive sources of data able to infer new knowledge about the presence and behaviors of human beings who act in a given living space. In this scenario, Human Activity Recognition (HAR) is one of the research areas that has profited from this possibility. Indeed, it aims to develop innovative applications able to identify humans’ activities and behaviors by analyzing data gathered from sensors and deliver a set of smart services capable of improving humans’ comfort and wellness. Thus, HAR techniques have been applied in many different fields such as healthcare [1], smart environments, and robotics [2–4]. As an example, by using IoT technologies, it is possible for a robot assistant to understand the human’s behavior and use this inferred information to enable appropriate user authentication tasks [5,6]. Moreover, HAR can be useful to support a remote monitoring between doctors and elderly patients [7].

However, in spite of their wide adoption in different application domains, some of the HAR approaches tend to discard data associated with so-called transient actions, which correspond to the transition between two significant activities, or to consider these actions as belonging to a single action

class labeled as “other”, meaning everything that is a basic and well-defined human activity. Namely, the “transitional” activities are defined as those transitory movements performed by humans between a sequence of two specific and well-defined activities (e.g., from standing to sitting or from laying down to standing up). Different from the above-mentioned approaches, our research strongly focuses on the identification and classification of human transitional actions that, in our opinion, should not be considered as noise between two specific human actions, but they must be considered as actions in themselves, although transitional, able to provide additional and useful knowledge to IoT-based environments. To this extent, there is a strong emergence of innovative classification algorithms devoted to enable an efficient recognition of human transitional actions. This paper bridges this gap by introducing a novel framework for HAR where conventional neural networks are extended with a memory to enable a successful recognition of transitional activities. The proposed approach can be considered as fully alternative to previous studies where transitional activities are considered as not semantically useful in classifying human behavior, and their recognition is performed just to enable a better classification of non-transitional actions. Vice versa, the main goal of this paper is to provide a semantic interpretation of transitional actions and to design an efficient algorithm to recognize them so as to provide additional knowledge to IoT environments where the proposed approach will be embedded. Precisely, the proposed architecture uses a set of data collected by an accelerometer and a gyroscope, integrated in a conventional smartphone and input in an Artificial Neural Network (ANN) together with a set of information related to previously classified human actions (the memory) in order to allow this classification algorithm to better identify dynamical behaviors as the aforementioned transitional activities. Moreover, the proposed method, in addition to its capabilities in classifying human transitional activities, shows a high tolerance to the noise in data. This feature is particularly useful in IoT environments, where a set of sensors collects data to be transferred to computing devices by means of not fully reliable wireless connections. As will be shown in the experimental section, the proposed method yields better performance in the recognition of non-transitional activities than state-of-the-art methodologies for HAR, such as hierarchical continuous HMM [8], convolutional neural networks [9,10], SVM [11], and so on. Unfortunately, it is not possible to perform a full comparison with other classification approaches specifically designed for the recognition of temporal events (recurrent neural networks, time-delay neural networks, and so on) because, so far, there are no studies where these techniques have been applied to the recognition of transitional activities.

The remainder of this work is organized as follows: Section 2 provides related works on the recognition of transitional activities and related approaches. Section 3 overviews the design of the proposed ANN architecture and reports details about the two step memory approach. Section 4 presents the validation and evaluation activity performed by comparing the proposed solution with state-of-the-art classifiers. Finally, Section 5 concludes the document with a final discussion about the achieved results.

2. Related Works

The design of enhanced classifiers for HAR has emerged as a popular research topic in recent years, mainly thanks to the rise of IoT frameworks capable of collecting a huge amount of data related to human actions to be analyzed and inferring additional knowledge about the human status so as to deliver an appropriate set of services aimed at improving human wellness. In this scenario, the state-of-the-art research strongly focuses on the design of classifiers using temporal information as input so as to improve the accuracy in the recognition of dynamic human activities. In general, these approaches are based on a hierarchical/two stage architecture often integrating an ensemble of classifiers, each one aimed at recognizing a different component of a specific human activity. As an example, in [12], a two stage classifier, combining a tracking algorithm, time-delay neural networks, and fuzzy inference systems, was proposed to classify human behaviors from video sequences; in [13], inherent sequential characteristics of the activities were used within a two stage classification structure deploying continuous hidden Markov models to discriminate between stationary and moving activities;

finally, the authors in [14] presented a three layer architecture incorporating different methods such as binary classification, KNN, and hidden Markov models to recognize simple human activities. Within the above-mentioned approaches, the transition activities were not taken into consideration or they were considered as a single cluster of actions useful to identify everything not belonging to a proper class of human actions.

Other HAR classification systems exploit the dynamic nature of human actions to design algorithms capable of identifying the temporal dependencies in data collected by sensors and properly detect human activities. To this extent, several approaches have been presented in recent years, such as those based on Deep Recurrent Neural Networks (DRNN) [15], and on Long Short-Term Memory cells (LSTMs) [16]; moreover, in [17], a novel approach based on a combination of feature extraction using time-delay embedding and supervised learning was applied to extract significant features from time series related to human actions so as to yield good results in terms of activity recognition. However, in spite of their evaluable performance, these approaches provided significant results with respect to periodic human activities, resulting in being less suitable for the classification of non-periodic activities.

A problem with RNN is related to the choice of the window length for sliding window segmentation since the HAR context is strictly affected by the bias among individual samples. Namely, an individual sample of movement data is not well defined, at least beyond immediate correlations between neighboring samples, and also depends on subjective behavior (e.g., different subjects performing the same action, but within different time windows) and on the specific movement performed (complex movements can be generated with higher biases among subjects) [18]. For this reason, RNN are usually combined with a CNN, where they are deployed just to model temporal dependencies on a higher level, meaning that they do not work on individual samples recorded by the sensor(s), making their sole application to complete movements (i.e., individual sample sequences recorded by sensors) inadequate.

With respect to transitional activities, a few works exist that exploit their identification, mainly in order to improve the classification performance of the non-transitional/static activities. Namely, in [19], Salarian et al. proposed the use of multi-stage classification for human activities with postural transitions characterized by the logistic regression method used to predict discrete outcomes (transition/non-transition activities). In [20], the authors proposed a two-layer and multi-strategy framework for HAR, where the first layer was used to recognize classes of activities, while the second one to discriminate within the classes. The two stage approach was used to mitigate the effect of orientation and position variation introduced by the use of the smartphone as the acquisition device. In particular, each layer was implemented through the Random Forest (RF) method, which has been shown to be effective in HAR [21]. However, focusing on postural transitional activities can be useful for several reasons. As highlighted above, some studies showed that, by classifying postural transition, basic activity classification could be improved: in [22] Capela et al. proposed a hierarchical classifier that included the transition phases into and out of a sitting state to improve sitting and standing classification. In [20], Quian Guo et al. used a two layer classifier: in the first layer, the activities were classified into different groups, and in the second layer, for each group, the appropriate strategy was designed according to the characteristics of the group to improve the recognition performance. For static activities, the transitional activities were introduced to help classify the activities indirectly. Besides, as a main advantage, another reason to study the classification of postural transition is that, as some studies showed [23], these are associated with a high risk of falling, and thus, correctly classifying them can improve healthcare, which is a very promising area, where AI models and techniques are finding widespread applicability.

For these reasons, conversely to the presented works, this paper is focused on the classification of transitional activities that until now have only be used as additional information to improve the classification of static activities. We aim to prove that is also possible to improve the classification of transitional activities by using the previous non-transitional/static activities. In particular, the proposed classifier is endowed with a short-term memory able to trace the previous activities at different times

in the past based on a feature augmentation procedure. Namely, the previous classified state is not considered, but the features characterizing the previous state.

Khan et al. proposed an accelerometer-based approach for human activities, which used a hierarchical scheme: at the lower level, ANNs discriminated between static, transition, and dynamic activities, and at the upper level, other ANNs discriminated inside each category, using an augmented feature space obtained by autoregressive modeling of acceleration signals. However, in this work, the authors presented the results actuated on a home-made dataset, which was not available for a possible comparison. Additionally, they adopted a testing procedure where the items used for the validation were within the training set, thus affecting the performance evaluation.

In [24], Reyes-Ortiz et al. proposed a system architecture for activity recognition with a smartphone using a real-time classification, addressing the issue of transition activities by applying a heuristic filtering approach. This was achieved by appending consecutive probability vectors given as SVM predictions and using temporal filtering techniques. In this work, also a temporal activity filtering technique was used that considered the statistical evaluation of the duration of transitions against static and dynamic activities. The authors evaluated the proposed method on publicly available HAR datasets. In particular, they used the SBHAR (Smartphone-based HAR Dataset with Postural Transitions) dataset [11]. The approach of this work, based on an artificial neural network with two step memory, was compared with state-of-the-art methodologies such as hierarchical continuous HMM [8], convolutional neural networks [9,10], and SVM [11], including that of Reyes-Ortiz et al. [24], applied on the SBHAR Dataset. However, the comparison was presented only on the non-transitional activities, since the other approach collapsed the transitional activities into only one class, meaning that they did not present the classification results of single transitional activities for a comparison. With regard to the transitional activities' classification, the results of our approach are presented before and after the optimization. The approach proposed in this paper is very different from the state-of-the-art methods because it does not use a hierarchical architecture to detect human activities, but it is based on a single computational layer able to perform a correct recognition of human activities. Moreover, different from recent research on HAR where innovative and complex methodologies such as CNN and RNN are used to take into strong consideration the dynamic nature of human actions, the proposed system uses a conventional neural network opportunely extended with a memory scheme; the proposed approach results in being very simple to design and deploy in a real IoT environment. Moreover, the MANN-based approach is the first HAR architecture specifically designed to improve the classification of both transitional and non-transitional human activities, as shown in the experimental section.

3. The Proposed Architecture

This work aims at developing a classification algorithm able to identify both "Basic" Activities (BAs) (or "Non-Transition Activities" (NTAs)) and "Transition Activities" (TAs), which are characterized respectively by the dynamic or static activities of a long duration (e.g., walking, standing, etc.) and by activities of a brief duration (i.e., seconds) such as Postural Transitions (PTs) (e.g., sit-to-stand, lie-to-sit, etc.) [24]. In particular, this paper proposes a so-called Memory-based Artificial Neural Network (MANN), a machine learning framework able to improve the classification accuracy of TAs with respect to state-of-the-art classifiers, thanks to the exploitation of a short-term memory in the learning process.

Figure 1 shows the proposed architecture at work in an IoT environment populated by devices aimed at reading raw information about human status and a set of smart services opportunely activated when a human activity is recognized. The figure highlights both the training step, where a database is used to train a MANN classifier, and the classification step, where a trained MANN is used to classify a user's activities.

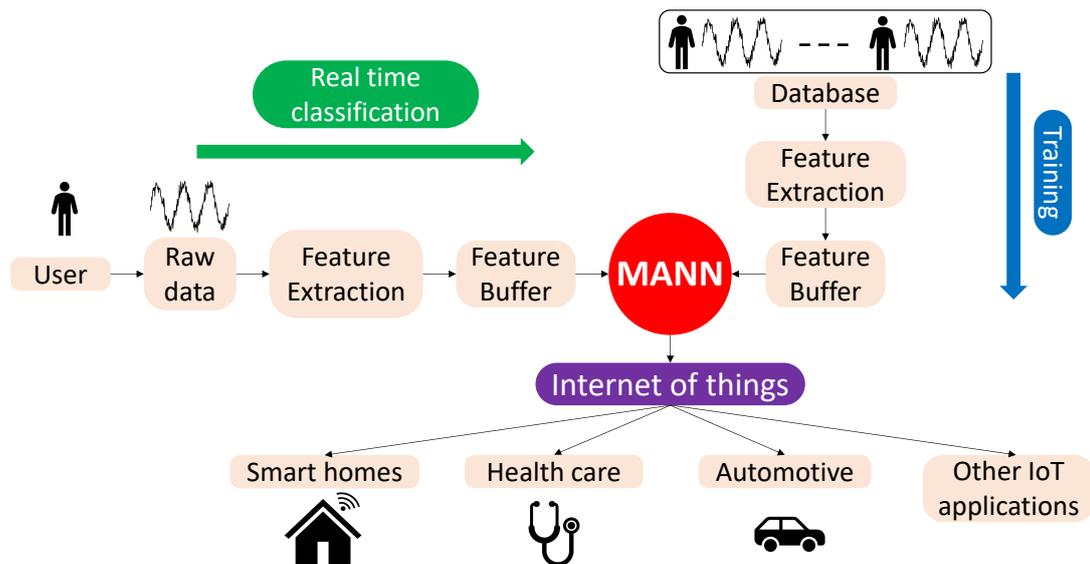


Figure 1. The HAR based on MANN in IoT applications.

With respect the training step, a set of raw data collected by an accelerometer and gyroscope of a smartphone is input in a feature extraction process able to infer relevant knowledge useful to improve the quality of the learning algorithm. With respect to the classification step, the extracted feature space is augmented with features coming from previous classification steps so as to implement a sort of short-memory mechanism storing past classified behaviors to be eventually used to improve the performance of the proposed method in predicting current human activities. Thus, the central hypothesis of this work is that short-term memory empowers the classic architecture for HAR by improving its classification performance mainly with respect to the recognition of TAs, and it improves the robustness of the proposed approach in being tolerant to data noise. In the following subsections, a detailed description of the proposed architecture for HAR is introduced.

3.1. MANN

Figure 2 shows the machine learning algorithm for HAR proposed in this work. It represents an extended version of a classical artificial neural network with back-propagation, where a memory buffer is used to store information of features related to previous states. This memory usage is able to improve the classification accuracy of time-dependent classification problems, especially when data are highly correlated in time, as for the case of HAR tasks.

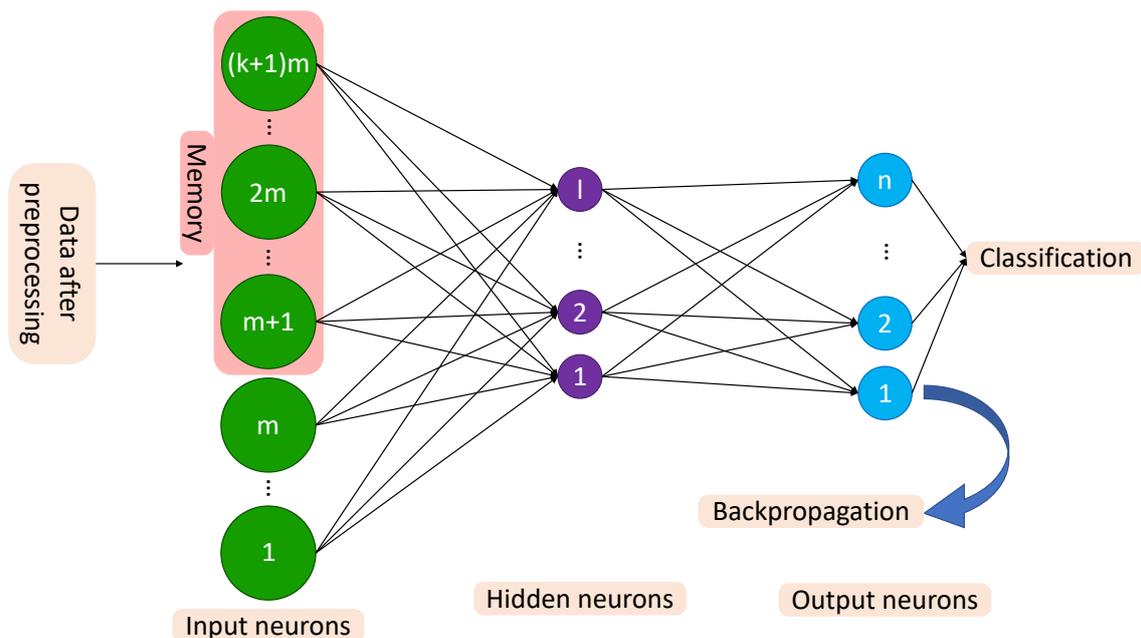


Figure 2. A MANN classifier based on k memory steps. The number of hidden neurons is set as $l = \sqrt{(k+1)mn}$.

There are other networks that have been proven to be able to deal with time-dependent classification problems successfully, as for instance recurrent neural networks [25–27] and time-delay neural networks [28–30], which are specifically designed for processing sequential data. However, MANN employs a much simpler topology, implementing a slight change with respect to standard ANNs, and yet, it is able to treat time-dependent data series successfully and, as will be shown, to improve the classification performances significantly with respect to state-of-the-art classifiers for TAs' classification in HAR applications. Additionally, this intrinsic simplicity allows MANN-based algorithms to be paired with more sophisticated approaches to increase classification performance.

The MANN-based classification algorithm is particularly able to deal with discrete time-dependent problems, i.e., when data are sampled at various time steps. When data are stored continuously with time, a procedure of sampling and feature extraction is required. This procedure is also helpful when the time steps are too close to each other: by averaging over several time steps, a bigger time window can be used.

As can be seen in Figure 2, in order to classify a human activity related to the time step t_n , the MANN network requires features read at the instant t_n and some features from the previous ones. In particular, features from k previous time steps are required, i.e., features from $t_{n-1}, t_{n-2}, \dots, t_{n-k}$. Consequently, the total number of input neurons is $(k+1)m$, where m is the number of features per item and k the number of time steps used as memory in the MANN. By using this network topology, the proposed framework is able to use information about the human status from previous times to classify the current human activity successfully. According to [31,32], the number of hidden neurons l is set by using a heuristic rule taking into consideration the square root of the product of input and output neurons ($l = \sqrt{(k+1)mn}$), while the number of output neurons is set equal to the number of classification labels.

Finally, the activation function is set as the hyperbolic tangent function $f(h) = \tanh h$, and the solver for weight optimization is the Adam algorithm [33].

3.2. Data Acquisition and Feature Extraction for HAR

In this paper, MANN networks are applied to a HAR in IoT application. In particular, the main goal of the proposed approach is the classification of human activities starting from raw data collected

by an accelerometer and a gyroscope on mobile devices, such as a smartphone or a smartwatch. Since the data acquisition process is usually continuous (or discrete with an extremely small time window for every time step, making it almost continuous), as discussed in the previous section, a process of data collection and a subsequent feature extraction algorithm are required to enable MANN to be utilized in the given IoT scenario. In this section, a feature extraction algorithm, which was directly taken from [24], is described.

In the proposed application, data are collected from a smartphone mounted on the waist, but it is possible to consider other positions, such as the pocket [34] or grasped by hand [35]. It is important to use the gyroscope in addition to the accelerometer since it has been shown that the use of both sensors is more beneficial than solely relying on the accelerometer readings [36]. The process of extracting meaningful information from raw data is shown in Figure 3 [24].

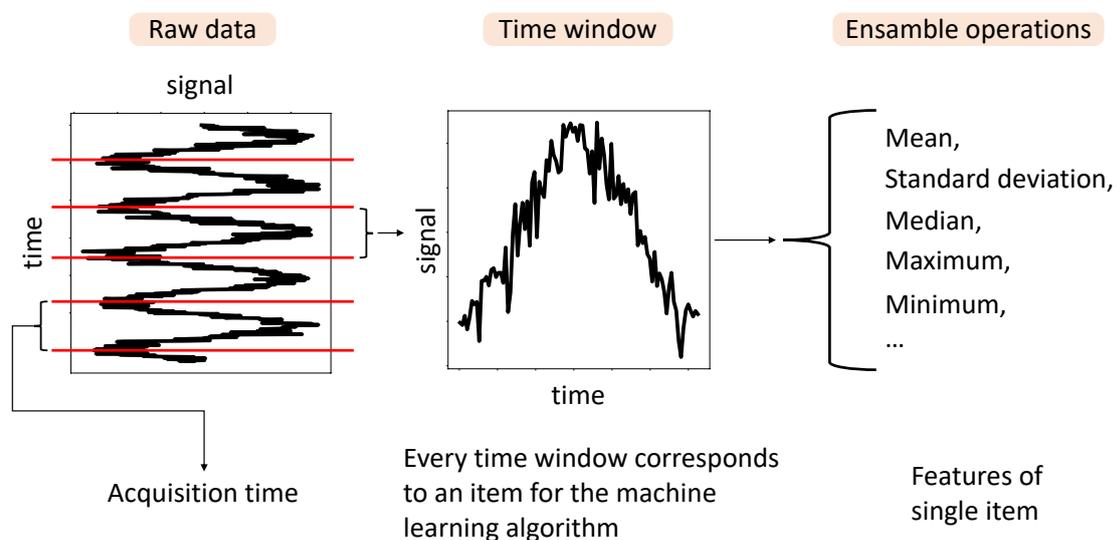


Figure 3. Data collection process for HAR dataset composition [24].

Raw data are collected in a series of single time windows with defined duration given by the acquisition time. Every time window is then treated separately as an instance for learning algorithms, whereas features are extracted through ensemble operations (e.g., mean, standard deviation, median, maximum, minimum). The acquisition time is a very important parameter because it should be sufficiently large to include one and only one human activity. By keeping this acquisition process, data are collected as a set of discrete time-dependent items with a fixed set of features.

As explained before, there are two categories of activities to be classified: TAs (Transition Activities) and NTAs (Non-Transition Activities, or basic activities). The core idea of this work is that by using MANN instead of standard neural networks, it is possible to improve the classification of these activities, especially TAs. Indeed, these activities are usually extremely short and consequently highly correlated with data from previous movements. For instance, the TA “stand-to-sit” performed at the time step t_n is a transition from the NTA “stand” at t_{n-1} to the NTA “sit” at t_{n+1} : using information related only to t_n ignores the time correlation, which could however help classification accuracy. Thus, providing the network the information at t_{n-1} can greatly improve performances. Indeed, in the following section, this will be proven to be the case: MANN can improve HAR classification, mainly of TAs, but also of NTAs.

4. Experiments

This section shows and discusses the results obtained by using a MANN in a HAR application for IoT environments. In the first subsection, the dataset called the “SBHAR dataset” is described. In the next subsection, results yielded by using MANN on the SBHAR dataset are compared to those obtained from other studies on the same dataset, so as to prove the proposed approach is able to compete with state-of-the-art classifiers. However, it should be pointed out that no research studies related to TAs have been performed on the SBHAR dataset, and consequently, only results related to NTAs could be compared. Finally, in the last subsection, MANN is compared to standard methodologies used in HAR applications, showing that it outperformed other architectures in performing TAs classification. The comparison was performed by considering the classification accuracy, but also the robustness with respect to the noise, as well as the execution times, where MANN showed good performances as well. The code was written in Python by using the scikit-learn library [37].

4.1. SBHAR Dataset

In this section, the SBHAR dataset [24], used to test MANN on HAR tasks, is briefly described. A group of 30 volunteers aged between 19 and 48 was selected for the collection of data for the SBHAR dataset. A smartphone (Samsung Galaxy S II) was mounted on each volunteer’s waist. The volunteers performed a protocol of activities composed of six NTAs: three static postures (i.e., standing, sitting, lying) and three dynamic activities (i.e., walking, walking downstairs, walking upstairs). The SBHAR dataset also included TAs occurring between static postures (i.e., stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, lie-to-stand). By using the embedded accelerometer and gyroscope of the smartphone, 3-axis linear acceleration and 3-axis angular velocity were captured at a constant rate of 50 Hz. Features were extracted from raw data as explained in Section 3.2. A detailed description of this process can be found in the SBHAR dataset (<http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human\+Activities+and+Postural+Transitions>). The final result was a dataset composed of 10,929 items with 561 features each, as shown in Table 1.

Table 1. Distribution of dataset items for each activity of the Smartphone-based HAR Dataset with Postural Transitions (SBHAR) dataset [24].

	Activities	Number of Terms	
NTAs	Walking	1722	10,411
	Walking upstairs	1544	
	Walking downstairs	1407	
	Sitting	1801	
	Standing	1979	
	Laying	1958	
			10,929
TAs	Stand to sit	70	518
	Sit-to-stand	33	
	Sit-to-lie	107	
	Lie-to-sit	85	
	Stand-to-lie	139	
	Lie-to-stand	84	

In this scenario, the training of a MANN classifier has required $m = 561$ features per item, corresponding to the features extracted in the SBHAR dataset, and $n = 12$ output neurons, corresponding to the number of activities to be classified. As for the number of memory steps k , some tests were performed to find the value maximizing the performance of the network: the result was $k = 2$, and consequently, the number of hidden neurons was $l = \sqrt{3mn} = 142$.

4.2. Comparison of MANN Results to Other Studies on the SBHAR Dataset

In this section, the performance of MANN is evaluated with respect to classifiers proposed by other studies on the SBHAR dataset. It should be stressed that no research studies related to TAs

have been performed on the SBHAR dataset, and thus, only classification performances on NTAs could be compared. Of course, since MANN was mainly built to improve the classification of highly time-correlated items in time-dependent classification tasks, a particular improvement on NTAs was not expected.

In order to compare the MANN results to those in the literature, a standard user-independent train-test split, proposed by the authors of the SBHAR dataset, was employed: out of the 30 volunteers, 21 were used for training and the remaining 9 for testing. This user-independent train-test split allowed a good generalization of the classification accuracy, since no data from the same user were used either in training or testing, and consequently, there was no risk of the classifier learning patterns specific to a certain user (thus the name “user-independent”).

The performance of the architecture was evaluated by the accuracy, which is defined as:

$$\text{accuracy} = \frac{\text{number of correct classifications}}{\text{total number of test items}} \quad (1)$$

The result for MANN’s accuracy was 96.24%, which was comparable to the highest accuracy values obtained by other studies. The complete comparison between the MANN results and those of other works is shown in Table 2: MANN was third in terms of overall accuracy, over-performing, for instance, several convolutional neural network implementations (e.g., 95.18% for [10], 94.79% for [9], 90.89% for [38]). The architectures which proved to have a superior accuracy to MANN were an SVM classifier [11], with a 96.37% accuracy, and a convolutional neural network [39], with a 97.63% accuracy. I should be stressed again that these results were only related to NTAs, which were the activities less likely to be improved by the MANN architecture, but nevertheless, MANN obtained a solid result compared to the other studies.

Table 2. Comparison between the MANN results and the literature results on the SBHAR dataset [39].

Paper	Method	Accuracy
[40]	Hidden Markov Models	83.51
[41]	Dynamic Time Warping	89.00
[42]	Handcrafted Features + SVM	89.00
[38]	Convolutional Neural Network	90.89
[13]	Hidden Markov Models	91.76
[43]	PCA + SVM	91.82
[43]	Stacked Autoencoders + SVM	92.16
[8]	Hierarchical Continuous HMM	93.18
[9]	Convolutional Neural Network	94.79
[10]	Convolutional Neural Network	95.18
[9]	FFT + CNN Features	95.75
	MANN	96.24
[11]	Handcrafted Features + SVM	96.37
[39]	Convolutional Neural Networks	97.63

In order to benchmark MANNs on the SBHAR dataset including TAs, the confusion matrix was computed on the entire dataset (including both TAs and NTAs) employing the same 21 + 9 user-independent train-test split: results are shown in Table 3. The overall accuracy was 95.48% (in the bottom-right of the matrix), which was lower than that measured when considering only NTAs, since TAs are usually tougher to classify. The matrix showed also two additional performance parameters specific to each activity: recall and precision. The former is defined as:

$$r_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

where r_i is the recall for the activity i , TP_i is the number of “true positives” (i.e., the number of test items belonging to class i and correctly identified), and FN_i is the number of “false negatives”

(i.e., the number of test items belonging to class i and incorrectly classified), whereas the latter is defined as:

$$p_i = \frac{TP_i}{TP_i + FP_i}, \quad (3)$$

where p_i is the precision for the activity i and FP_i is the number of “false positives” (i.e., the number of test items not belonging to class i and incorrectly classified to class i). It was clear that NTAs were usually better classified by MANN than TAs: the average recall for NTAs was 96.38%, while that for TAs was 77.74%. It is interesting to note that, although some TAs were extremely difficult to recognize (e.g., “lie-to-sit” had a recall of 64.00% and “lie-to-stand” one of 59.26%), others were successfully classified, as for instance “sit-to-lie” and “stand-to-lie”, with recalls of respectively 96.88% and 83.67%. Indeed, as will be shown in the next section, these two activities were those that were improved the most by the use of MANN’s memory buffer.

Table 3. Confusion matrix related to the MANN classification with respect to all activities belonging to the SBHAR dataset; S1: walking, S2: walking upstairs, S3: walking downstairs, S4: sitting, S5: standing, S6: laying, T1: stand-to-sit, T2: sit-to-stand, T3: sit-to-lie, T4: lie-to-sit, T5: stand-to-lie, T6: lie-to-stand.

	S1	S2	S3	S4	S5	S6	T1	T2	T3	T4	T5	T6	Recall
S1	491	0	5	0	0	0	0	0	0	0	0	0	98.99%
S2	17	450	4	0	0	0	0	0	0	0	0	0	95.54%
S3	5	7	408	0	0	0	0	0	0	0	0	0	97.14%
S4	0	1	0	455	50	1	0	0	0	0	1	0	89.57%
S5	0	0	0	11	527	0	0	0	0	0	0	0	97.96%
S6	0	0	0	0	4	540	1	0	0	0	0	0	99.08%
T1	0	0	0	2	2	0	19	0	0	0	0	0	82.61%
T2	0	0	0	0	0	0	2	8	0	0	0	0	80.00%
T3	0	0	0	0	0	0	0	0	31	0	1	0	96.88%
T4	0	0	0	1	0	0	0	0	0	16	0	8	64.00%
T5	1	0	0	1	1	1	0	0	4	0	41	0	83.67%
T6	0	0	0	1	0	0	0	0	0	9	1	16	59.26%
Precision	95.53%	98.25%	97.84%	96.60%	90.24%	99.63%	86.36%	100%	88.57%	64.00%	93.18%	66.67%	95.48%

4.3. Comparison of MANN to Standard Methodologies in HAR

In the previous section, MANN was compared to architectures proposed in other studies. However, since TAs in the SBHAR dataset have not been treated by anyone else, it was not clear how much MANN improved the classification accuracy for these activities. To test this, here, MANN was compared to other architectures commonly used in HAR applications [19,31] considering both TAs and NTAs. In particular, the following were considered: a Logistic Regressor (LR) with the L2 regularization strength set as 1, a Support Vector Machine (SVM) with the penalty parameter set as 1, a Random Forest (RF) with 300 trees, a K-Nearest Neighbors (KNN) with $k = 5$, and an Artificial Neural Network (ANN) with one hidden layer containing 82 neurons.

To make the analysis statistically more robust, in this section, a personalized train-test split is employed, since using the standard 21+9 train-test split of the previous section only provided a single confusion matrix. In particular, a stratified 3-fold cross validation with 10 random sub-samples was used: data were divided into 3 folds, making sure that labels (i.e., activities to be recognized) were equally distributed in these 3 portions of the dataset, and in turn, two of the three folds were used as training sets and the other one as test set, where a confusion matrix was computed. By doing so, the performance of the architecture was evaluated without any bias, since no data from the training set were included in the test set. This procedure was repeated 10 times, randomly shuffling data before partitioning, and thus, 30 independent confusion matrices, i.e., performance measurements, were obtained. Finally, these 30 matrices were averaged to produce the mean and standard deviation for each value of interest.

It is important to point out that here, the train-test split was not user-independent: when randomly shuffling the dataset, data for each user were included in both the training and the test

sets. Consequently, the performance scores presented here should not directly be compared to those discussed in Section 4.2.

Since the analyses of this section required the knowledge of the classification accuracy for each activity, the performance parameter taken into consideration was the recall r_i (Equation (2)), which was the ratio of correctly classified test items for the activity i . For any of the 30 pairs of training and testing, the recall of each activity was measured and then averaged to produce the mean and standard deviation of the mean as the final evaluation:

$$\bar{r}_i = \frac{\sum_{\alpha=1}^{30} r_i^\alpha}{30}, \quad (4)$$

$$\sigma_i = \sqrt{\frac{\sum_{\alpha=1}^{30} (r_i^\alpha - \bar{r}_i)^2}{29}}, \quad (5)$$

$$\bar{\sigma}_i = \frac{1}{\sqrt{30}} \sigma_i, \quad (6)$$

where r_i^α is the recall for the activity i measured in the train-test split with index α (α runs from 1 to 30, which is the number of independent train-test splits), \bar{r}_i is the mean recall for the activity i , and $\bar{\sigma}_i$ is the standard deviation of the mean for the activity i , which is the statistical error of \bar{r}_i . As an overall performance score, we then computed the mean recall for all activities:

$$\bar{r}_m = \frac{\sum_{i=1}^{12} \bar{r}_i}{12}, \quad (7)$$

$$\bar{\sigma}_m = \frac{\sqrt{\sum_{i=1}^{12} \bar{\sigma}_i^2}}{12}. \quad (8)$$

Figure 4 shows the comparison between the performances of all classifiers, plotting the mean recall \bar{r}_i of all activities and the average mean recall \bar{r}_m (where the errors $\bar{\sigma}_i$ and $\bar{\sigma}_m$ are indicated as error bars).

MANN soundly outperformed the other classifiers: its average recall was 90.18%, which was considerably higher than both ANN (87.93%) and LR (87.85%), which were the most efficient out of the other classifiers. It was clear that most of the improvement by MANN could be tracked down to TAs, and in particular to the activities “sit-to-lie” and “stand-to-lie”, where MANN reached a recall of respectively 89.78% and 90.57%, whereas ANN only had 77.10% and 76.83%, and LR obtained 74.28% and 77.91%. These two TAs were difficult to classify because they had a similar pattern, since in both cases, the user lied down, either from a sitting position or from a standing one. MANN was able to improve the classification performance for these TAs thanks to the memory buffer: accessing the information of previous times, it was able to discriminate between sitting and standing and, consequently, to correctly identify from which position the user was laying down. On the other hand, no particular improvement was observed for NTAs, although it is interesting to observe a slight improvement in classifying “sitting” and “standing” activities, which have been proven to be tough to distinguish by machine learning algorithms [22]; MANN obtained a recall of 96.34% for the former and 97.09% for the latter, while ANN obtained 95.93% and 96.62% and LR 94.90% and 95.78%.

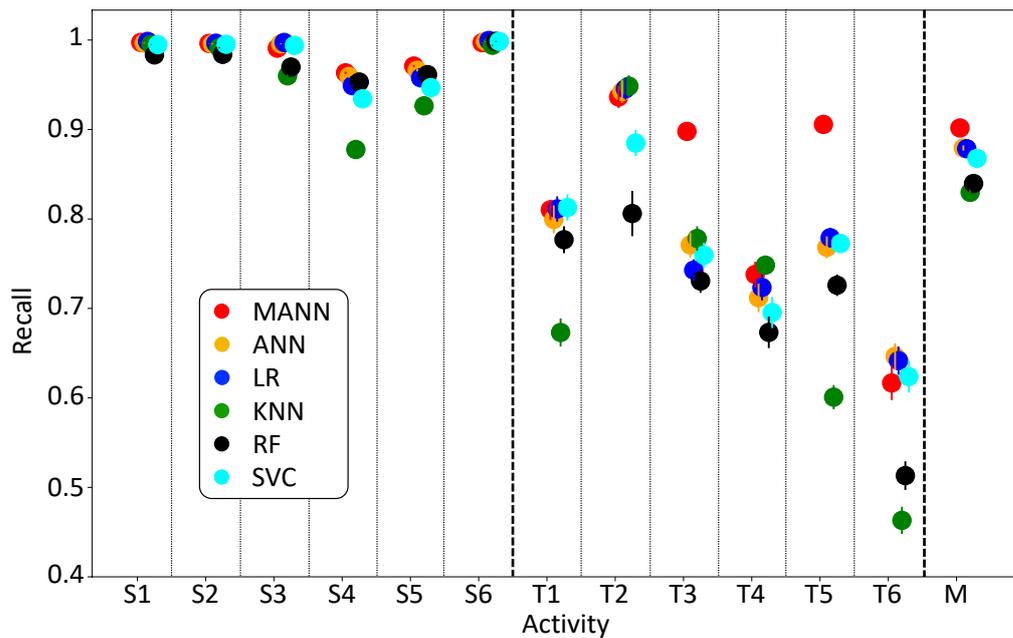


Figure 4. Comparison between classifiers' classification performances. For each classifier, we compare the average recall related to each activity (S1: walking, S2: walking upstairs, S3: walking downstairs, S4: sitting, S5: standing, S6: laying, T1: stand-to-sit, T2: sit-to-stand, T3: sit-to-lie, T4: lie-to-sit, T5: stand-to-lie, T6: lie-to-stand, M: mean recall). MANN outperforms other classifiers.

Successively, it was investigated how MANN was able to treat noise compared to the other architectures. Indeed, a classification algorithm in an HAR application should be resilient to noise contamination, since in real life, applications data are not as clean as in laboratory conditions (e.g., data from the accelerometer and the gyroscope are affected by an error due to an uneven surface on which the user is walking). To perform this analysis in a controlled way, data were contaminated with noise after the feature extraction procedure explained in Section 3.2. In particular, after data were processed and separated into training and test sets, the test data were contaminated with a noise source, simulated as a Gaussian distribution with zero mean and the standard deviation set as a parameter α (the bigger α was, the stronger the noise source was). Since data were previously standardized (i.e., values for each feature were linearly transformed so that items had 0 mean and a standard deviation 1), this approach allowed the noise to be equally weighted for every feature extracted. The parameter used to quantify noise resilience was the loss L , defined as:

$$L(\alpha) = \bar{r}_m(\alpha = 0) - \bar{r}_m(\alpha) \quad (9)$$

where $\bar{r}_m(\alpha = 0)$ was the average of the 12 recalls for each activity (measured, as described previously, using the 30 independent train-test pairs) with no noise source (i.e., $\alpha = 0$) and $\bar{r}_m(\alpha)$ is the average recall when the noise source has strength α . Thus, the loss quantified how much the noise reduced the overall recall, and consequently, if a classifier had a strong noise resilience, its loss should be small (ideally zero) even for high values of α . The results are shown in Figure 5, where α ranges from 0.1 to 1.5.

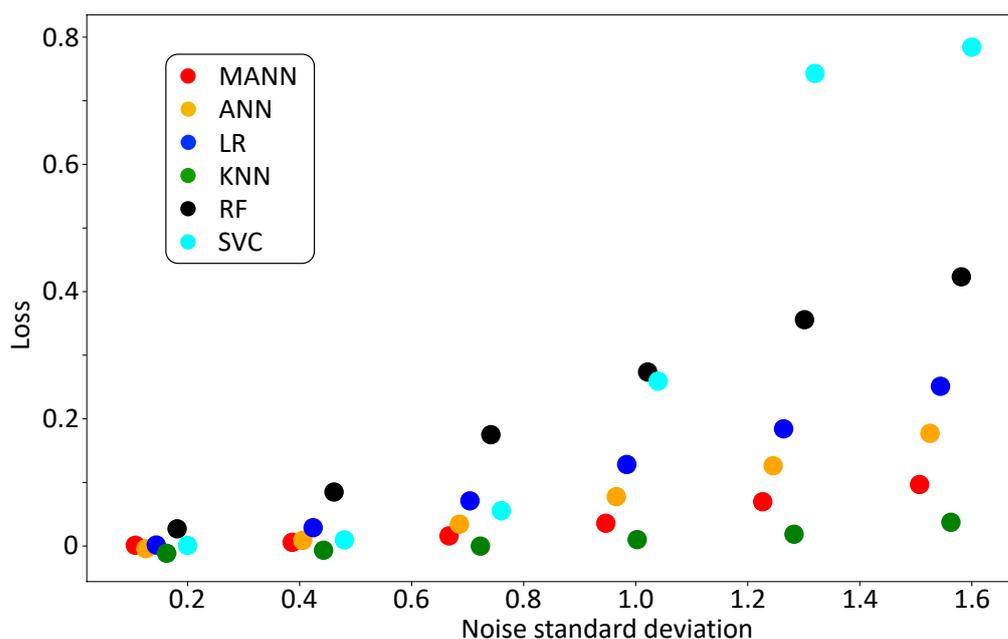


Figure 5. Noise robustness comparison. It shows how, as noise grows, MANN has a good noise resilience.

Overall, MANN had a good noise resilience: for small α values, the loss was negligible (0.1% for $\alpha = 0.1$), and it increased up to 10% for $\alpha = 1.5$: this meant that the performance worsening due to random noise in the test set was always inferior to 10%. It is particularly interesting that MANN increased noise robustness with respect to standard ANN, where the loss was almost double for a strong noise source (20% for $\alpha = 1.5$). Thus, the introduction of the memory buffer to the neural network not only improved the classification accuracy, but also made it more robust to random noise sources. The classifier which showed the best behavior for noise contamination was KNN ($L = -1\%$ for $\alpha = 0.1$ and $L = 4\%$ for $\alpha = 1.5$), which was understandable given the nature of its classification approach relying on the neighborhood. On the other hand, both RF and SVC showed poor noise robustness performances, the former being the worst for low noise source and the latter being the worst for high ones: for $\alpha = 0.4$, $L = 8\%$ for RF, and $L = 0.1\%$ for SVC, while, for $\alpha = 1.5$, $L = 40\%$ for RF, and $L = 80\%$ for SVC, which meant that these classifiers lost almost all predictive capabilities if a strong noise source contaminated the data.

Lastly, learning times for each classifier were compared to test whether the memory buffer increased them dramatically for MANN. For each training set, the time required by each classifier to learn its weights was measured (each training set was composed of approximately 3000 items), and then, these times were averaged over all sets. The results are shown in Figure 6.

MANN was comparatively quick at learning the data, requiring about 7 seconds for each training set. ANN and LR were the quickest, requiring only 4 seconds; this was understandable, since the memory buffer in MANN increased the number of input and hidden neurons, thus having more weights to optimize. However, the increase in learning time observed for MANN was not too high, making it feasible to use it in real-time classification problems. Finally, the other classifiers had higher learning times (11 seconds for RF, 18 seconds for SVC and 20 seconds for KNN).

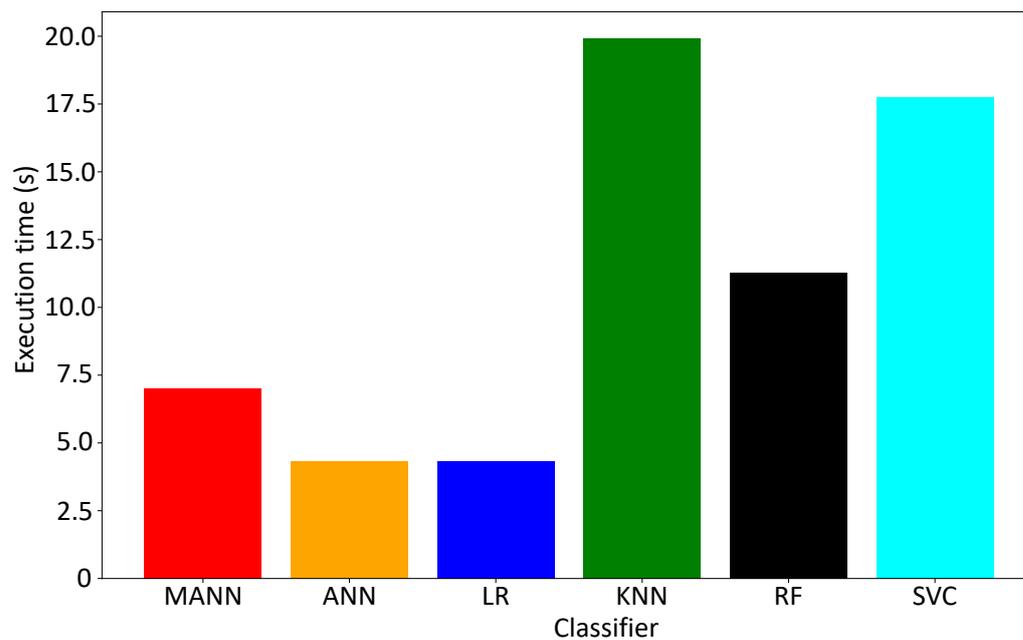


Figure 6. Comparison between classifiers' training times. The usage of a memory does not crucially slow down the proposed neural network.

5. Conclusions

This research introduced an innovative machine learning approach for human activity recognition in IoT environments. Specifically, a neural network extended with a short memory was designed to enable the implementation of algorithms aimed at efficiently classifying human activities. In particular, different from other approaches from the literature, the proposed method was proven to be particularly suitable for detecting transitional human activities, and moreover, it provided a high level of robustness with respect to noise in data. Indeed, the accuracy of the MANN method was comparable to existing approaches when applied to non-transitional activities (96.24%), and it provided a very high accuracy (95.48%), with respect to some of the most used machine learning algorithms, when transitional activities were included. Finally, the proposed method yielded good performance also in terms of training time, that was very low with respect to other machine learning algorithms such as random forests, k-nearest neighbors, and support vector machine. As a result of this, it can be stated that the MANN algorithm represents a suitable method to be embedded in IoT frameworks where the human activity recognition represents a critical task to enable the delivery of the right set of services to the users' systems. In the future, the proposed approach for HAR will be compared to other similar techniques incorporating memory to further prove its capabilities in classifying non-transitional and transitional human activities. Specifically, the proposed method will be compared with recurrent neural networks [25–27] and time-delay neural networks [28–30] on a set of datasets different from SBHAR.

Author Contributions: Supervision G.A.; conceptualization F.M. and G.M.; methodology F.M. and G.M.; investigation F.M., G.M. and M.S.; writing—original draft preparation F.M., G.M. and M.S.; writing—review and editing G.A. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Abbreviations

The following abbreviations are used in this manuscript:

HAR	Human Activity Recognition
TA	Transitional Activity
NTA	Non-Transitional Activity
NN	Neural Networks
ANN	Artificial Neural Networks
MANN	Memory Artificial Neural Networks
LR	Logistic Regressor
SVC	Support Vector Classifier
RF	Random Forest
KNN	K-Nearest Neighbor

References

1. Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Comparison of situation awareness algorithms for remote health monitoring with smartphones. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; pp. 2454–2459.
2. Rossi, S.; Capasso, R.; Acampora, G.; Staffa, M. A Multimodal Deep Learning Network for Group Activity Recognition. In Proceedings of the International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 8–13 July 2018; doi:10.1109/IJCNN.2018.8489309. [[CrossRef](#)]
3. Vitiello, A.; Acampora, G.; Staffa, M.; Siciliano, B.; Rossi, S. A neuro-fuzzy-Bayesian approach for the adaptive control of robot proxemics behavior. In Proceedings of the IEEE International Conference on Fuzzy Systems, Naples, Italy, 9–12 July 2017; doi:10.1109/FUZZ-IEEE.2017.8015647. [[CrossRef](#)]
4. Rossi, S.; Staffa, M.; Bove, L.; Capasso, R.; Ercolano, G. User's Personality and Activity Influence on HRI Comfortable Distances. In *Social Robotics. ICSR 2017*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2017; pp. 167–177, doi:10.1007/978-3-319-70022-9_17. [[CrossRef](#)]
5. Amroun, H.; Ouarti, N.; Ammi, M. Recognition of human activity using Internet of Things in a non-controlled environment. In Proceedings of the 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), Phuket, Thailand, 13–15 November, 2016; pp. 1–6.
6. Batool, S.; Saqib, N.A.; Khan, M.A. Internet of Things data analytics for user authentication and activity recognition. In Proceedings of the 2017 Second International Conference on Fog and Mobile Edge Computing (FMEC), Valencia, Spain, 8–11 May 2017; pp. 183–187.
7. Osmani, V.; Balasubramaniam, S.; Botvich, D. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *J. Netw. Comput. Appl.* **2008**, *31*, 628–655. [[CrossRef](#)]
8. Ronao, C.A.; Cho, S.B. Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models. *Int. J. Distrib. Sens. Netw.* **2017**, *13*, doi:10.1177/1550147716683687. [[CrossRef](#)]
9. Ronao, C.; Cho, S.B. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **2016**, *59*, 235–244. doi:10.1016/j.eswa.2016.04.032. [[CrossRef](#)]
10. Jiang, W.; Yin, Z. Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; ACM: New York, NY, USA, 2015; pp. 1307–1310, doi:10.1145/2733373.2806333. [[CrossRef](#)]
11. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition using Smartphones. In Proceedings of the ESANN, Bruges, Belgium, 24–26 April 2013.
12. Acampora, G.; Foggia, P.; Saggese, A.; Vento, M. A hierarchical neuro-fuzzy architecture for human behavior analysis. *Inf. Sci.* **2015**, *310*, 130–148, doi:10.1016/j.ins.2015.03.021. [[CrossRef](#)]
13. Ronao, C.A.; Cho, S.B. Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models. In Proceedings of the 2014 10th International Conference on Natural Computation (ICNC), Xiamen, China, 19–21 August 2014; pp. 681–686.
14. Kozina, S.; Gjoreski, H.; Gams, M.; Lustrek, M. Three-layer Activity Recognition Combining Domain Knowledge and Meta-classification. *J. Med. Biol. Eng.* **2013**, *33*, 406–414, [[CrossRef](#)]

15. Hammerla, N.Y.; Halloran, S.; Plötz, T. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 9–15 July 2016; pp. 1533–1540.
16. Zhao, Y.; Yang, R.; Chevalier, G.; Gong, M. Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors. *Math. Probl. Eng.* **2018**, *2018*, 7316954. [[CrossRef](#)]
17. Frank, J.; Mannor, S.; Precup, D. Activity and Gait Recognition with Time-Delay Embeddings. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*; Fox, M., Poole, D., Eds.; AAAI Press: Palo Alto, CA, USA, 2010.
18. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 33:1–33:33. [[CrossRef](#)]
19. Salarian, A.; Russmann, H.; Vingerhoets, F.J.G.; Burkhard, P.R.; Aminian, K. Ambulatory Monitoring of Physical Activities in Patients With Parkinson's Disease. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 2296–2299. doi:10.1109/TBME.2007.896591. [[CrossRef](#)]
20. Guo, Q.; Liu, B.; Chen, C.W. A two-layer and multi-strategy framework for human activity recognition using smartphone. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6.
21. Abdullah, M.F.A.; Negara, A.F.P.; Sayeed, S.; Choi, D.; Anbananthen, K. Classification algorithms in human activity recognition using smartphones. *Int. J. Comput. Inf. Eng.* **2012**, *6*, 415–432.
22. Capela, N.A.; Lemaire, E.D.; Baddour, N. Improving classification of sit, stand, and lie in a smartphone human activity recognition system. In Proceedings of the 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, Turin, Italy, 7–9 May 2015; pp. 473–478, doi:10.1109/MeMeA.2015.7145250. [[CrossRef](#)]
23. Najafi, B.; Aminian, K.; Loew, F.; Blanc, Y.; Robert, P.A. Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 843–851. doi:10.1109/TBME.2002.800763. [[CrossRef](#)]
24. Reyes-Ortiz, J.L.; Oneto, L.; Samà, A.; Parra, X.; Anguita, D. Transition-Aware Human Activity Recognition Using Smartphones. *Neurocomputing* **2016**, *171*, 754–767. doi:10.1016/j.neucom.2015.07.085. [[CrossRef](#)]
25. Hochreiter, S.; Schmidhuber, J. LSTM can solve hard long time lag problems. In *Advances in Neural Information Processing Systems*; NeurIPS Press: San Diego, CA, USA, 1996; pp. 473–479.
26. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* **2018**, *6*, 1662–1669. [[CrossRef](#)]
27. Martinez, J.; Black, M.J.; Romero, J. On human motion prediction using recurrent neural networks. *arXiv* **2017**, arXiv:1705.02445.
28. Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; Lang, K.J. Phoneme recognition using time-delay neural networks. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 328–339. doi:10.1109/29.21701. [[CrossRef](#)]
29. Peddinti, V.; Povey, D.; Khudanpur, S. A time delay neural network architecture for efficient modeling of long temporal contexts. In Proceedings of the INTERSPEECH, Dresden, Germany, 6–10 September 2015.
30. Huang, X.; Zhang, W.; Xu, X.; Yin, R.; Chen, D. Deeper Time Delay Neural Networks for Effective Acoustic Modelling. *J. Phys. Conf. Ser.* **2019**, *1229*, 012076. doi:10.1088/1742-6596/1229/1/012076. [[CrossRef](#)]
31. Micucci, D.; Mobilio, M.; Napolitano, P. UniMiB SHAR: a new dataset for human activity recognition using acceleration data from smartphones. *Appl. Sci.* **2017**, *7*, 1101. doi:10.3390/app7101101. [[CrossRef](#)]
32. Xu, H.; Liu, J.; Hu, H.; Zhang, Y. Wearable Sensor-Based Human Activity Recognition Method with Multi-Features Extracted from Hilbert-Huang Transform. *Sensors* **2016**, *16*, 2048. doi:10.3390/s16122048. [[CrossRef](#)]
33. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
34. Kwapisz, J.R.; Weiss, G.M.; Moore, S. Activity recognition using cell phone accelerometers. *SIGKDD Explor.* **2010**, *12*, 74–82. [[CrossRef](#)]
35. Lee, Y.; Cho, S.B. Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer. In *Hybrid Artificial Intelligent Systems. HAIS 2011*; Lecture Notes in Computer Science; Corchado, E.; Kurzynski, M.; Wozniak, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6678, pp. 460–467.
36. Wu, W.; Dasgupta, S.; Ramirez, E.E.; Peterson, C.; Norman, G.J. Classification accuracies of physical activities using smartphone motion sensors. *J. Med. Internet Res.* **2012**, *14*. doi:10.2196/jmir.2208. [[CrossRef](#)]

37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
38. Ronao, C.; Cho, S. Evaluation of Deep Convolutional Neural Network Architectures for Human Activity Recognition with Smartphone Sensors. In Proceedings of the KIISE Korea Computer Congress, Jeju Island, Korea, 24–26 June 2015; pp. 858–860.
39. Andrey, I. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Appl. Soft Comput.* **2017**, *62*. doi:10.1016/j.asoc.2017.09.027. [[CrossRef](#)]
40. Pires, I.M.; Garcia, N.M.; Pombo, N.; Flórez-Revuelta, F.; Spinsante, S.; Teixeira, M.C.C.; Zdravevski, E. Pattern recognition techniques for the identification of Activities of Daily Living using mobile device accelerometer. *PeerJ Prepr.* **2018**, *6*, e27225.
41. Seto, S.; Zhang, W.; Zhou, Y. Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 7–10 December 2015.
42. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. In *Ambient Assisted Living and Home Care*; Bravo, J., Hervás, R., Rodríguez, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 216–223.
43. Li, Y.; Shi, D.; Ding, B.; Liu, D. Unsupervised Feature Learning for Human Activity Recognition Using Smartphone Sensors. In *Mining Intelligence and Knowledge Exploration*; Prasath, R., O'Reilly, P., Kathirvalavakumar, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 99–107.



© 2020 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/>).