This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

IoT Wearable Sensor and Deep Learning: an Integrated Approach for Personalized Human Activity Recognition in a Smart Home Environment

Valentina Bianchi, Marco Bassoli, Gianfranco Lombardo, Paolo Fornacciari, Monica Mordonini, Ilaria De Munari, *member IEEE*

***Abstract* — Human Activity Recognition (HAR) is currently recognized as a key element of a more general framework designed to perform continuous monitoring of human behaviors in the area of Ambient Assisted Living (AAL), well-being management, medical diagnosis, elderly care, rehabilitation, entertainment and surveillance in smart home environments. In this paper, an innovative HAR system, exploiting the potential of wearable devices integrated with the skills of deep learning techniques, is presented with the aim of recognizing the most common daily activities of a person at home. The designed wearable sensor embeds an Inertial Measurement Unit (IMU) and a Wi-Fi section to send data on a cloud service and to allow direct connection to the Internet through a common home router so that the user themselves could manage the installation procedure. The sensor is coupled to a CNN network designed to make inferences with the minimum possible resources to keep open the way of its implementation on low-cost or embedded devices.**

The system is conceived for daily activity monitor and nine different activities can be highlighted with an accuracy of 97%.

***Index Terms*— activity recognition, Internet of Things, machine learning, wearable sensor.**

I. INTRODUCTION

IN recent years, wearable sensors have gained considerable importance both in research and application fields. The reason for such interest lies in their use in many applications, which has been made possible by the progressive reduction of their size and costs. Some examples are wearable sensors in sport and physical activities [1]–[3], surveillance [4], human computer interaction [5], [6], rehabilitation [7], [8], monitoring elderly people for Ambient Assisted Living (AAL) purposes [9], [10]. The latter has particular importance, since nowadays the progressive increment of the average population age is widely recognized as a big deal in both social and economic contexts [11]. AAL technologies may contribute to the construction of active aging scenarios [12]–[14] in order to preserve quality of life of aging population in an affordable way, reducing the need of social and health-care services. The importance of this topic is evidenced worldwide by the great number of research initiatives and programs (e.g. "active and assisted Living joint program" (AAL-JP)).

In this framework, wearable sensors can be exploited in different ways, from the simplest panic button function up to the continuous monitoring of the user's physiological parameters [10], [15]. Among these, Human Activity Recognition (HAR) plays an important role: it is indeed recognized that an active lifestyle is the basis for a healthy life [16]. Therefore, users' lifestyle can be assessed monitoring the amount of daily activity and, eventually, building-up a behavioral model which in turn can be useful for the early detection of anomalies possibly relevant to wellbeing [16]. Moreover, to construct a precise behavioral profile, it is of utmost importance to accurately assess the type of the user's activity (i.e. walking, climbing stairs up/down, etc.). For example, a user could continue to move regularly (e.g. walking) but begin to avoid more difficult movements (e.g. climbing stairs). This behavior could signal an increase in fatigue, which may indicate a possible deterioration in the health conditions worthy of further study.

A behavioral model can be constructed also with environmental sensors [17], but aging at home makes it possible to have the presence of two or more elderly people in the same environment, or in the same building, each with its own diseases that must be monitored in a personalized manner. Thanks to the use of wearable sensors, the single person's activities can be recognized. More information can be acquired integrating the sensor in a more complete IoT system [17]. Moreover, from the point of view of HAR algorithms, a lot of work has been done in automatic recognition of human activities through the analysis of data coming from a video camera [18], [19] or from integration of data obtained from different types of sensors [20], [21]. The potential of deep learning, however, allows to explore new developments in

Manuscript received 15 February 2019.
The authors are with the Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy (e-mail: valentina.bianchi@unipr.it; marco.bassoli@studenti.unipr.it;gianfranco.lombardo@unipr.it; paolo.fornacciari@unipr.it;monica.mordonini@unipr.it; ilaria.demunari@unipr.it).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

intelligent assisted living home environments also in the presence of chronic patients [13].

It is important to note that, in the context of AAL, ergonomics and ease of use is of utmost importance: the device must be worn 24 hours a day without being perceived as a burden to the user. Platform such as, for example, smartphones are not conceived for this use, even if they integrate all the sensors needed and they are used for many hours a day. It therefore becomes important to design ad-hoc sensors that meet these needs. Moreover, it is worth noting that the quality of the models coming from machine learning techniques greatly depends on the dataset used for the training. A dataset similar to the actual use case and trained on the specific monitored user could guarantee the highest activity recognition ability. In addition, the capability of update the trained model is important to reach the best system performance with an arbitrary users' number. In these context, taking advantage of cloud computing can help. Typically, wearable sensors base their connectivity on body area networks (BAN) [15], with the sensor connected through a low power protocol (e.g Bluetooth Low Energy) to a gateway (e.g. the smartphone). With the development of the Internet of Things (IoT) paradigm, it is now possible to exploit system architectures connecting the device directly with cloud services in order to minimize the number of devices involved in data acquisition/transmission and therefore the costs, users' burden, technological skills needed [8], [22]–[24] and allowing also the sharing of the device outcomes with other users (e.g. physicians, caregivers, etc.).

In this paper, we present an HAR system, based on a Wi-Fi wearable sensor and Deep Learning techniques, conceived to exploit the potential of smart and wearable devices in order to recognize daily activities of single users within an AAL environment. In a general view, the proposed architecture exploits Wi-Fi connectivity and a neural network designed to be used on cloud for the most computational demanding task (the training phase) and on an embedded unit or on a low cost local device for the daily activity recognition. In this view, the connection to the cloud service is necessary only when a new person begins to be monitored, providing a small re-training phase, to create a complete dataset perfectly fitting to the use case.

The purpose of this work is not a real time activity recognition but a long-term personalized monitoring of the activity performed during the day by the elderly to infer abnormal behaviors, often relevant to unhealthy states or emergent situations.

The paper is organized as follow: in section II some related works are reported, in section III the proposed system is described, detailing the Wi-Fi wearable device, the choice of the neural network architecture and its optimization. Then, in section IV, to validate the system some tests are described and results are discussed. In section V conclusion are drawn.

II. Related Works

Deep learning approach is based on computational models that are composed of multiple processing layers. This allows an automatic learning of the intrinsic structure in complex and large data [25].

In the field of healthcare, deep learning is widely used to carry tasks based on data that comes from mobile systems [26]. In [27] the authors describe how sensor-equipped smartphones and wearables are transforming the health monitoring. These devices could collect many people analytic data and deep learning is considered to be a key element in analyzing this new type of information, but using deep learning approach in healthcare sensing domain is also an open challenge due mainly to hardware limitations.

In [28] the authors argue about HAR using wearable sensors and deep learning: rather than analyzing handcrafted features from time-series sensor signals, they arrange signal sequences of accelerometers and gyroscopes into an activity image, which enables Deep Convolutional Neural Networks (DCNN) to learn the discriminative features automatically for activities classification. Their results demonstrated a better performance respect to the state-of-the-arts on three public datasets. In particular, an accuracy of 99.93% has been obtained. Despite this very high performance, the system requires a quite complex preprocessing phase: considering the application context, a light learning model with automatic feature extraction techniques is desirable.

Also in [29] the HAR task is carried out by Convolutional Neural Network (CNN) that automatically extract human activity features without any domain knowledge (such as activities in kitchen or jogging, walking, etc.), while prior works have shown that some heuristically-defined features can perform well in recognizing one activity, but badly for others. The authors focus on the fact that a CNN method can capture the local dependencies and scale-invariant features of activity signals and so, variations of the same activity can be effectively captured through the extracted features. Also in this case, the network is tested on three public dataset: the best accuracy obtained by the authors is 96.88%.

The CNN used in [30] performs a HAR task by using the data coming from a single accelerometer, making possible to construct an acceleration-based HAR on the mobile platform, with no extra hardware demand. The results show a quite good accuracy of 93,8% on a dataset acquired by volunteers which uses an Android application to record tri-axial accelerometer data. The experiments were repeated with the device placed in 3 different body parts with the purpose of maintaining diversity of data. The comparison on the same dataset with other popular classifiers, such as Support Vector Machine (SVM), shows that CNN must have extracted more effective features than Fast Fourier Transform (FFT) and Discrete Cosine Transform (DCT), which were calculated by hand as input features for the SVM.

Instead, in the paper [31] a deep learning technique based on Long Short-Term Memory (LSTM), or Recurrent Neural Network, is introduced to perform HAR task from data captured with wearable sensors in ubiquitous computing. This approach was chosen in order to exploit the temporal dependencies within the movement data allowing an immediate, real-time inference at the same rate as the data is collected by the sensors. In fact, in system that use a segmentation step, that is some kind of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

sliding window extraction procedure, a delay between the recording of movements using inertial sensors and the inference procedure is present. The authors illustrate how their results, verified on a large benchmark dataset, are the state-of-the-art.

The works described above, report HAR computation, by using deep learning, that are quite suitable for mobile application. In [32] HAR technique based on a deep learning methodology is designed to enable accurate and real-time classification for low-power wearable devices. The proposed approach has been deployed as an app for Android devices and also as an embedded algorithm for the Intel Edison Development Platform in order to demonstrate on-node human activity classification using the trained classification model. To avoid the problem of know the position of the inertial sensor and to reduce the complexity of the classification task, the raw inertial data is projected to the spectral domain. The output, the spectrogram representation, is the input for the deep learning activity recognition model, which shows results that are comparable to existing state-of-the-art approaches that utilize many more nodes and layers but with computation times that are consistent with the requirements for real-time on-node human activity recognition.

Finally in [33] the authors propose a system that is able to identify human's physical activities automatically through analysis of the signals acquired (in real time) from multiple body-worn (or body-embedded) inertial sensors. Multiple sensor is used to overcome the issues of intra-class variability (each subject may perform the same activity with somewhat different movements) and the inter-class similarity (for example, jogging and running). The method builds a deep architecture based on CNN to investigate the multichannel time series data with better results than other state-of-the-art methods.

All these results are obtained in general purpose experiments, where the type of activity and the related datasets are not intended for an AAL field. In [8] an activity recognition approach applied to the AAL context and based on support vector machines (SVMs), decision trees, and dynamic time warping is presented. The dataset is acquired by eight volunteers using different smartphones. For the activity recognition task, the results obtained reach an average accuracy of 82% on the three tested methods.

III. THE PROPOSED SYSTEM

The system architecture is sketched in Fig. 1. During the training phase the data, collected by the wearable sensor, are

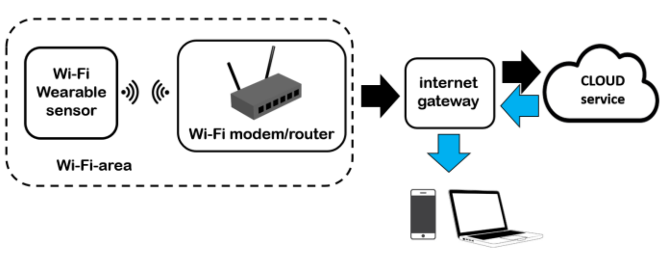

Fig. 1. The system architecture.

sent to the cloud, while in daily use they can be elaborated on board or sent to another device for further processing. For prototypal purposes, the data are elaborated offline and used to create the datasets exploited to design the neural network proposed in this work.

A. Hardware Platform

Data acquisition system is composed by a Wi-Fi wearable sensor sending collected data to a cloud service in IoT compliant vision. The prototype of the wearable sensor (Fig. 2) is based on the MPU9250 integrated inertial measurement unit (IMU, with a 3D accelerometer, gyroscope and magnetometer) connected, in the prototype version, to a development board (LaunchPadXL board) including the CC3200 system-on-chip (SoC) by Texas Instruments, which integrates the ARM Cortex-M4 MicroController Unit (MCU). It features a 32-bit

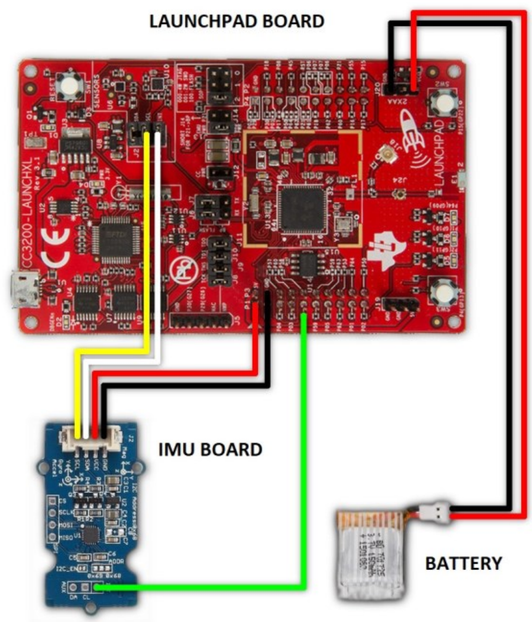

Fig. 2. The WiFi wearable sensor including the LaunchPadXL board and the inertial measurement unit (IMU) board

architecture at 80 MHz clock and a network processor compliant with the IEEE 802.11b/g/n network protocol radio. From the ergonomic point of view, a more compact board redesign is planned.

The sensors full-scales are programmable and a sensing acceleration up to ±8g and a magnetometer sensitivity of ±4800µT have been set, while the gyroscope was set on a full scale range of ±250°/sec. All the sensors have been configured with a sampling rate of 50Hz. The physical sensor output consists of an ordered list of values (3D output from accelerometer, gyroscope and magnetometer) corresponding to 16bit with sign for axis for a total of 18bytes. For testing purposes, two bytes have been added indicating respectively the identification number of the user, and a code related to the activity actually performed. The latter has been obtained asking the user to follow a specified protocol during the test (i.e. activities sequence) and to press a button available on the wearable sensor at every activity change. Pressing the button triggers a counter in the device firmware: every activity the subject was performing was then identified with the corresponding counter number. No noise filtering was carried out on the data.

A key issue in device usability is the network configuration procedure: the wearable sensor is directly connected to the Internet through the Wi-Fi home router. The commissioning procedure is very simple and relies on Wi-Fi Protected Setup (WPS) standard: the user is asked to push a button on the device, and a single LED signaling pattern has been implemented to provide feedback. No technical skill is required, so that the user themselves could manage the installation procedure. Data are sent to the online Watson IoT platform, inside the IBM Bluemix cloud services, via MQTT protocol with a Quality of Service (QoS) equal to 2: this ensures the necessary reliability to the transmission process, since the protocol itself guarantee that the message is received by the broker once and once only. The payload of the message is a string in a JSON format, reporting the sensors status. The device firmware, nevertheless, is suitable for connecting to other platforms as well (e.g. Amazon AWS, Microsoft Azure, Thingspeak, etc.). Each wearable device features its own unique ID, so that the association to the cloud environment can be managed by the service provider; this results in a truly "plug-and-play" approach.

The collected data are in general not significant in themselves, but they need interpretation to infer meaningful information (e.g., information about user actual activity such as walking, sitting, etc.). We may think to transmit every sensor sample to the cloud, for subsequent processing and interpretation. This implies that the radio section, the most energy-hungry part of the sensor, is always active reducing the battery lifetime. In order to prove the validity of the whole approach, sensor energy consumption has to be considered. In order to account for actual use scenarios, a low capacity battery (Li-Ion 4.2V, 500mAh) has been used (battery capacity is limited due to ergonomic constraints: size and weight).

The analysis of human motion features requires a sampling rates as high as 50 samples per second [34]: assuming a sampling frequency of 50 Hz, a full 9 degrees-of-freedom datum sent as soon as sampled and a battery of 500mAH, it can be demonstrated [22] that a lifetime shorter than 8 hours could be obtained. This prevents such approach to be actually usable in a real-life context. Improvement in the battery lifetime can be obtained implementing compression algorithms [35] or other on-board processing features. For instance, we may greatly reduce transmission overhead by packing data into larger bursts, exploiting internal memory for buffering. An expected battery lifetime (at 500 mAh capacity) of 2.07 days was measured in this case, as reported in [22]. This result turns out to be fully compatible with the purposes of this work.

B. Selection of the Neural Network Model

The advent of deep learning has widely modified the approaches in signal processing and features extraction fields. In the past years, in fact, the features extraction was performed by a manual analysis of the signals in its components, with the aim of creating domain-specific features [8]. Statistical and classical machine learning models were then trained on the processed version of the data. A limitation of these approaches is that signal processing and domain expertise are required to analyze the raw data and collect the features to fit a model and this expertise would be required for each new dataset or sensor. Another important limit of a classic machine learning approach in activities classification is the difficulties in being able to generalize the models against the variety of movements performed by different subjects.

Considering the application context and with the aim of not imposing any constraints on the computational resources made available by a cloud service, we focused on learning models that could combine performance with system requirements. In particular, we considered:

- **The need of automatic feature extraction**. In light of this, we focused on deep learning architectures that can ensure our requirement on feature extraction and optimal learning capacity. This need is also motivated by the recent HAR literature [36].
- **The need of a light learning model**. We designed and trained a light enough neural network capable of making inferences with the minimum possible resources and potentially easily portable on low-cost or embedded devices.

Thus, in this paper we considered a deep learning based approach to classify signal data, without the need to manually engineer features but with only a simple preprocessing phase described later.

The state-of-the-art of deep learning architectures are represented by Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) [37], [38]. RNNs, and in particular the version with Long-Short-Term Memory gates (LSTM) [39], are commonly adopted in the HAR field [31].

In order to choose which architecture could reach the best results in our context, a dataset has been collected with the proposed device and preliminary experiments have been carried out on a LSTM [40] and on a CNN [36] already presented in literature.

The data collection phase involved 15 subjects (12 male and 3 female aging between 25 and 50 years) repeating the same test 9 times in different days. During the test, the subject was asked to wear the sensor in a belt and to execute the most common activities performed in the daily life (Table I), in a predefined order.

TABLE I
ACTIVITY SET

Activity ID	Activity
1	Walking
2	Stand
3	Sitting-down
4	Stay seated
5	Standing-up
6	Running
7	Climbing stairs down
8	Climbing stairs up
9	Lie down

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

In Table II the protocol for data acquisition is reported.

TABLE II
ACQUISITION DATA PROTOCOL

counter	Activity ID	Activity
1	1	Walking
2	2	Stand
3	3	Sitting-down
4	4	Stay seated
5	5	Standing-up
6	2	Stand
7	6	Running
8	8	Climbing stairs up
9	7	Climbing stairs down
10	1	Walking
11	6	Running
12	8	Climbing stairs up
13	7	Climbing stairs down
14	1	Walking
15	3	Sitting-down
16	4	Stay seated
17	5	Standing-up
18	9	Lie down

Data have been organized as follows: userID, counter, 3D linear acceleration (x,y,z), 3D angular rate, and 3D magnetic data.

In Fig.3 an example of x-axis readings of the accelerometer for various activities has been reported.

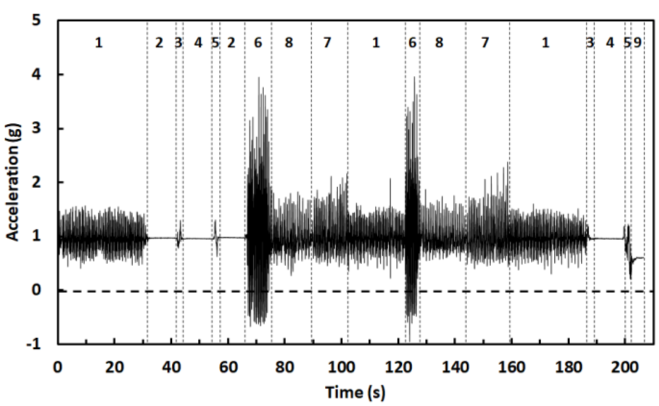

Fig. 3. Example of the accelerometer x-axis reading. The activities performed during data acquisition have been highlighted, reporting the corresponding activity ID.

Sensor signals (accelerometer, gyroscope and magnetometer) have been collected and pre-processed by sampling in fixed-width sliding windows of 2.56s and 50% overlap (128 readings/window). During this phase all those ranges of values that are not part of a single category are discarded as showed in Fig. 4.

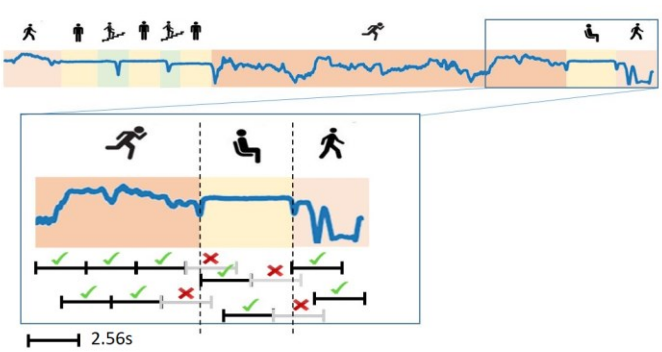

Fig. 4. Example of sampling in fixed-width sliding windows of 2.56s.

The window is rather short: this would allow to process data into real time mode, opening the way to new development such as the embedding of the inference phase into the sensor node.

The result was a dataset of 15616 instances, which have been subdivided as reported in the Table III.

TABLE III
COMPOSITION OF DATASET
(15616 INSTANCES)

% of Instances	Activity
26.1	Walking
12.0	Stand
3.6	Sitting-down
11.9	Stay seated
2.6	Standing-up
9.4	Running
14.8	Climbing stairs down
16.5	Climbing stairs up
2.7	Lie down

From this dataset both the training set and the test set were obtained. Three different partitions of the data in training set and test set have been evaluated to assess the performance of LSTM and CNN networks in different cases and to define the more appropriate architecture for our system.

In the first scenario, we considered a random split (about 60% of the instances have been included into the training set and the rest into test set): even if this subdivision is widely used, this is a quite unrealistic case, since an unbalanced distribution of the instances could occur between the training and test set.

In the second scenario, we considered the nine repetitions of ten users for the training set, while five users (different from those of the training set) for the test set. In this case, thus, the network is trained on a set of individuals not considered during the network working phase. In a real scenario, this consists of training the network once before the actual use with a set of pre-gathered data: the network does not have to be retrained if a new individual is to be monitored.

Finally, in the third scenario, we have distributed the nine repetitions of all the 15 users between training set (6 repetitions) and test set (3 repetitions). In this case the network has been trained on all users monitored in the running phase. With this scenario the best results are expected. Training the network on the same users that will be monitored is a demanding method in a real context, requiring high computational resources. Actually, taking advantage of the Wi-Fi transceiver on the sensor and cloud connection, a different approach can be conceived. A brief training phase can be planned every time a new user is considered: the user is asked to follow a defined activity protocol, then, the data acquired by the sensor are sent to the cloud, where, exploiting cloud analytics, a training phase of the network can be performed. The new parameters of the network tuned to detect the activities with the best performance can be then downloaded on the wearable sensor to perform HAR on board if a light enough neural network is available. Then, if necessary, the outcome of the HAR analysis can be transmitted to the cloud service for sharing with other users (physicians, caregivers, etc.).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

Experiments were conducted using the same number of training epochs (1000) and the same batch size (600).

The results have shown that CNN can reach better accuracies compared to LSTMs architectures (about 2% higher).

Accuracy obtained with CNN in the preliminary tests have been reported in Table IV.

TABLE IV
PRELIMINARY RESULTS

Architecture	scenario 1	scenario 2	scenario 3
CNN	91%	81%	92%

These results confirm that the third scenario shows the best performance.

C. Optimization of the CNN Model

In light of the preliminary results obtained, we selected the CNN architecture for the development of the system and for the optimization of the model using the best practices suggested in literature [41].

The optimization has been obtained performing several experiments with different network structures (changing number of hidden layers and typology of these layers), applying the Learning Rate Decay technique (LRD) [25], and a grid-search over the main hyper-parameters that can affect the model performance.

The network structure with the highest accuracy has been selected. For each structure a Dropout layer set to 0.5 before the output layer was introduced. A Softmax (eq. 1) has been used as activation function for the output layer.

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}} \quad (1)$$

Where a_i is an element of the input vector of N real numbers and p_i is the corresponding probability, obtained by normalizing with the sum of all exponentials. The Softmax function is used to map the non-normalized output of a network to a probability distribution.

Moreover, for each structure a grid-search over the following hyper-parameters has been performed:

- **Learning Rate**: search range (0.0001 – 0.04) with an incremental step of 0.005
- **Batch size**: evaluated values 64,128,256,600
- **Number of training epochs**: search range (500 – 2500) with an incremental step of 500

Results are presented in Table V.

TABLE V
COMPARISON OF DIFFERENT CNN STRUCTURES

N° Convolutional layers	N° Fully Connected layers	Accuracy
2	1	92.80%
3	1	93.80%
3	2	93.60%
4	**1**	**94.20%**
4	1	93.80%

The structure that achieves the best accuracy has 4 convolutional and 1 fully connected layers with Learning Rate equal to 0.0001, batch size equal to 256 and training epochs equal to 2500. This architecture, represented in Fig. 5, has been adopted for the proposed system.

The aim of each convolutional layer is to learn features from the network input. It is composed by two components, a convolutional operator (eq. 2) applied between I (training example i×j input matrix) and K (filter m×n matrix) and a max pooling operator [42].

$$g(i,j) = (K * I)(i,j) = \sum_{m}\sum_{n} I(i-m, j-n)K(m,n) \quad (2)$$

The K values are learned subsequently with the Back-Propagation Algorithm [43]. Once the resulting matrix g(i,j) is computed, the max pooling operator implements a down-sampling of it. This last operation is useful to reduce the computational cost by reducing the number of parameters to be learned. Finally, the fully connected layer is based on the classical neural network structures [44].

As suggested in [44], once the structure has been determined,

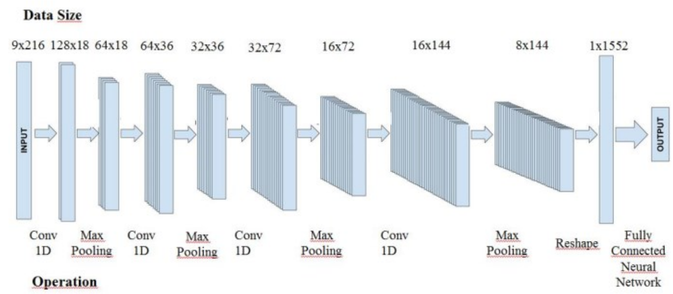

Fig. 5. The final architecture of the trained CNN.

the model has been optimized using the LRD with another grid search. The LRD requires a starting value for the learning rate parameter in order to start its optimization during the training epochs. This last grid search (Fig. 6) enabled the model to reach the accuracy of 94.75 % with the best configuration of the hyper-parameters shown in Table VI.

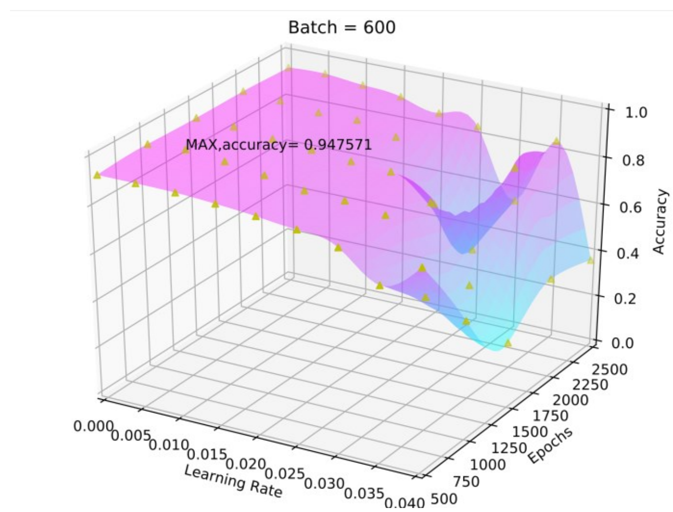

Fig. 6. The final grid-search with the best configuration

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

TABLE VI
THE OPTIMAL CONFIGURATION OF THE HYPER PARAMETERS
.

Hyperparameter	Best Value
Learning Rate	0.0050
Learning Rate Decay	0.96
Epochs	1000
Batch size	600

The network model so optimized can be trained, exported and used in the framework of our system to make new inferences.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the optimized CCN model presented above, some tests have been carried out. Two different datasets have been exploited: a standard dataset from UCI Machine Learning Repository [45] and a new dataset created collecting data with the proposed sensor. The parameters used to compare performance are Accuracy, Precision, Recall and F-Measure, defined as follows:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (3)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (4)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (5)$$

$$F\text{-}Measure = 2 \times \frac{Precision \times Recall}{\mathrm{Precision} + \mathrm{Recall}} \quad (6)$$

Where TP are the True Positives, TN the True Negatives, FP the False Positives and FN the False Negatives.

For each dataset accuracy has been evaluated to assess the global CNN performance (i.e. considering the whole set of classes), while F-Measure, Precision and Recall have been computed to give a more precise indication on how the CNN behave in recognizing a particular class.

The UCI dataset is show in Table VII.

TABLE VII
COMPOSITION OF UCI DATASET
(10299 INSTANCES)

% of Instances	Activity
16.72	Walking
14.99	Climbing stairs up
13.65	Climbing stairs down
17.25	Sitting-down
18.50	Standing up
18.87	Lie down

The confusion matrix related to this test is reported in Table VIII, together with Recall (RCL) Precisions (PRC) and F-Measure (FM) per class. The most significant FPs and FNs are highlighted in bold. The resulting global accuracy is 92.5%.

Table VIII
THE CONFUSION MATRIX RELATED TO THE TEST-SET OF THE UCI DATASET. IN BOLD THE MOST SIGNIFICANT FALSE POSITIVES AND FALSE NEGATIVES

	Activity	1	2	3	4	5	6	**RCL**	**FM**
1	Walking	491	1	4	0	0	0	0.99	0.98
2	Climbing stairs up	11	433	27	**0**	0	0	0.92	0.92
3	Climbing stairs down	0	1	419	0	0	0	0.99	0.96
4	Sitting-down	0	**6**	0	414	**70**	1	0.84	0.84
5	Standing-up	0	2	0	**86**	444	0	0.83	0.85
6	Lie down	0	**27**	0	0	0	510	0.95	0.97
	PRC	0.97	0.92	0.93	0.83	0.86	0.99		

The accuracy obtained has been compared (Fig. 7) with other results achieved with other machine learning models, presented in literature [28], [45]–[57], applied to the same dataset. In these works, when more than one machine learning architectures is presented, the one performing better is selected. As can be seen, our result is comparable with the state-of-the-art, although the network has been optimized for a different context. It can be concluded that the proposed network shows a good generalization capability.

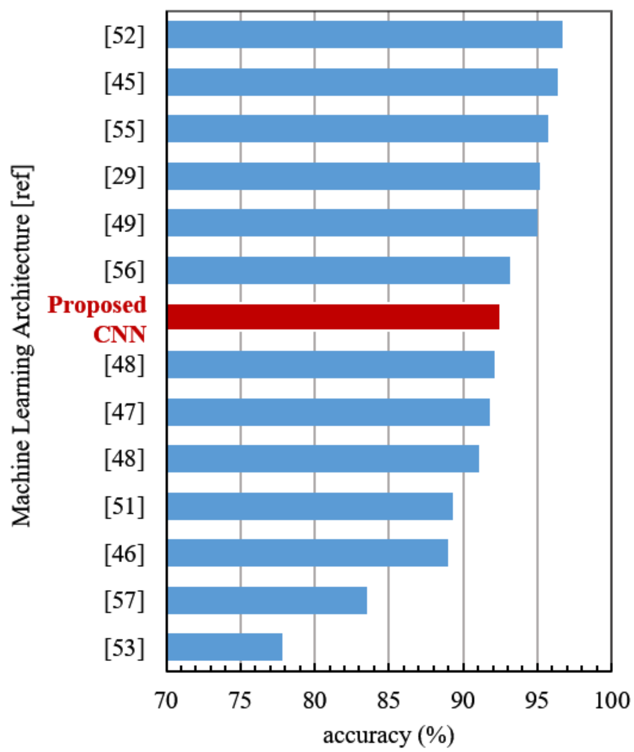

Fig. 7. Comparison between different architectures reported in literature using the UCI Dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

In Table IX the dataset obtained with a new data acquired with our device is shown.

TABLE IX
COMPOSITION OF THE NEW DATASET
(38764 INSTANCES)

% of Instances	Activity
20.6	Walking
11.6	Stand
7.9	Sitting-down
10.6	Stay seated
7.1	Standing-up
8.5	Running
10.3	Climbing stairs down
11.7	Climbing stairs up
11.7	Lie down

This dataset features more balanced instances compared to the one used in the optimization phase of the CNN architecture (Table III), and can be used to analyze in a more accurate way the performance of our network.

The training set and the test set have been created from this dataset following the third scenario described above (i.e. the network has been trained on all users monitored in the running phase). Each user was asked to do the test three times: the first and the second test have been used for the training phase, while the third for the test.

A global accuracy of 97% has been reached.

In Table X, the confusion matrix related to the test-set is shown. As before, the most significant false positives and false negatives are highlighted in bold. ID is the Activity ID reported in Table I.

Table X
THE CONFUSION MATRIX RELATED TO THE TEST-SET.
IN BOLD THE MOST SIGNIFICANT FALSE POSITIVES AND FALSE NEGATIVES

ID	1	2	3	4	5	6	7	8	9	**RCL**	**FM**
1	2584	**17**	0	0	1	0	**17**	**17**	1	0.98	0.98
2	4	1452	5	10	5	1	1	2	0	0.98	0.96
3	1	11	963	16	13	0	2	1	5	0.95	0.96
4	1	25	4	1328	2	0	0	0	0	0.97	0.97
5	1	**34**	20	11	834	0	0	0	2	0.92	0.95
6	11	3	0	0	0	1046	11	10	3	0.96	0.97
7	21	0	0	0	0	12	1270	15	1	0.96	0.96
8	13	2	0	0	0	8	8	1469	1	0.98	0.97
9	1	2	3	1	3	5	3	4	1476	0.98	0.99
PRC	0.98	0.94	0.97	0.97	0.97	0.97	0.97	0.97	0.99		

The result obtained is positive when compared to the other studies discussed in section II [8], [28]–[33] as presented in Fig. 8. In [28], [32] a better accuracy has been reached, but the classifying systems require a quite complex pre-processing phase. In our context a light model is desired with an automatic features extraction process to simplify the training procedure when a new user has to be monitored. Moreover, the number of activities recognized are smaller (7 and 6 respectively), influencing in positive the results. Finally, in [32] the hardware is a smartphone: this is a good solution when a short-term monitoring has to be done (e.g. rehabilitation session), but in the case of continuous monitoring, these device usually do not ensure the required ergonomics.

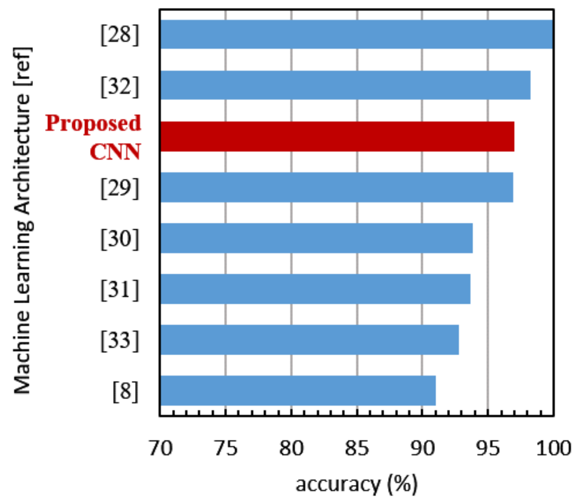

Fig. 8. Comparison between different architectures reported in literature using datasets acquired in an AAL Environment.

The architecture designed is then fully compliant with the requirements of our HAR system exploiting cloud scenario through a wearable sensor with Wi-Fi connection and a CNN model.

V. CONCLUSIONS AND FUTURE WORK

In this paper, an innovative IoT system for long-term personalized monitoring of the activities performed by a person at home is proposed. The system integrates a Wi-Fi wearable sensor and Deep Learning Techniques to give information on a number of activities with the aim to infer abnormal behaviors. The approach presented has been conceived to be extended to systems requiring multiple wearable sensors (such as in the case of a home habited by more people) giving information in a personalized manner.

The activity classification has been performed by a CNN-based architecture (with four convolutional layers, one fully connected layer and a sliding processing window of 2.56s conceived for real time elaboration), able to classify data coming from nine different activities with an accuracy of 97%, while having a relatively small training set. This result is interesting because it shows the possibility to implement, quite easily, different HAR systems calibrated on different classes of problems (e.g. age groups of people). Like all Machine Learning algorithms, Deep Learning requires a large amount of computer power to train the network, while the use of a pre-trained network to make inference is less expensive in terms of computing resources. Nowadays, to improve performance in computational terms, graphics processing units (GPUs) are widely used (both for the training and the inference phases). With the increasing complexity and flexibility of embedded devices, we can consider the use of them as an interesting alternative for the implementation of the pre-trained CNN-based model.

The presented system architecture exploits on-board Wi-Fi connectivity and cloud computing to ensure a constantly update of the network with new training sets when users are added. To this purposes every data sample acquired by the sensor is transferred to the cloud: this has been recognized as an energy consumption problem. Energy saving can be obtained elaborating data on the sensor itself buffering them in the internal memory and sending to the cloud compressed data. Further improving in battery consumption could be obtained, in the future, implementing on the sensor the inference phase based on the CNN model proposed in this paper. The training phase, instead, has been designed to remain on the cloud, in consideration of the fact that typically it requires the greatest computational power and that the network is updated only when a new user needs to be added.

The system architecture designed open the door to an alternative approach that could take advantage on the use of FPGA technologies for the implementation of complex signal processing systems to produce tiny, wearable and autonomous embedded HAR systems.

Acknowledgment

The authors would like to thank Stefano Volponi for his important contribution in the optimization phase of the architectures presented in this paper.

References

[1] H. Qiu, X. Wang, and F. Xie, "A Survey on Smart Wearables in the Application of Fitness," in *Proceedings - 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 2017 IEEE 15th International Conference on Pervasive Intelligence and Computing, 2017 IEEE 3rd International Conference on Big Data Intelligence and Compu*, 2018, vol. 2018-Janua, pp. 303–307.

[2] M. N. S. Zainudin, M. N. Sulaiman, N. Mustapha, and T. Perumal, "Monitoring daily fitness activity using accelerometer sensor fusion," in *2017 IEEE International Symposium on Consumer Electronics (ISCE)*, 2017, pp. 35–36.

[3] A. Ahmadi *et al.*, "Toward Automatic Activity Classification and Movement Assessment During a Sports Training Session," *IEEE Internet Things J.*, vol. 2, no. 1, pp. 23–32, Feb. 2015.

[4] V. Bianchi, P. Ciampolini, and I. De Munari, "RSSI-Based Indoor Localization and Identification for ZigBee Wireless Sensor Networks in Smart Homes," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 566–575, 2019.

[5] A. Ancans, A. Rozentals, K. Nesenbergs, and M. Greitans, "Inertial sensors and muscle electrical signals in human-computer interaction," in *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*, 2017, pp. 1–6.

[6] N. Mora, V. Bianchi, I. De Munari, and P. Ciampolini, "A low cost brain computer interface platform for AAL applications," *Assist. technol. Res. Ser.*, vol. 33, pp. 946–952, 2013.

[7] S. Qiu, Z. Wang, H. Zhao, L. Liu, and Y. Jiang, "Using Body-Worn Sensors for Preliminary Rehabilitation Assessment in Stroke Victims With Gait Impairment," *IEEE Access*, vol. 6, pp. 31249–31258, 2018.

[8] I. Bisio, A. Delfino, F. Lavagetto, and A. Sciarrone, "Enabling IoT for In-Home Rehabilitation: Accelerometer Signals Classification Methods for Activity and Movement Recognition," *IEEE Internet Things J.*, pp. 1–1, 2016.

[9] F. Montalto, C. Guerra, V. Bianchi, I. De Munari, and P. Ciampolini, "MuSA: Wearable multi sensor assistant for human activity recognition and indoor localization," *Biosyst. Biorobotics*, vol. 11, pp. 81–92, 2015.

[10] V. Bianchi, C. Guerra, M. Bassoli, I. De Munari, and P. Ciampolini, "The HELICOPTER project: Wireless sensor network for multi-user behavioral monitoring," in *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings*, 2018, vol. 2018-Janua, pp. 1445–1454.

[11] "World Population Prospects: The 2017 Revision | Multimedia Library - United Nations Department of Economic and Social Affairs." [Online]. Available: https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html. [Accessed: 22-Jan-2019].

[12] S. Malwade *et al.*, "Mobile and wearable technologies in healthcare for the ageing population," *Comput. Methods Programs Biomed.*, vol. 161, pp. 233–237, 2018.

[13] C. Fonseca, D. Mendes, M. Lopes, A. Romao, and P. Parreira, "Deep Learning and IoT to Assist Multimorbidity Home Based Healthcare," *J. Heal. Med. Informatics*, vol. 08, no. 03, pp. 1–4, Jun. 2017.

[14] C. Guerra, V. Bianchi, I. De Munari, and P. Ciampolini, "CARDEAGate: Low-cost, ZigBee-based localization and identification for AAL purposes," in *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2015, vol. 2015-July, pp. 245–249.

[15] S. C. Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review," *IEEE Sens. J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.

[16] D. S. Kehler *et al.*, "A systematic review of the association between sedentary behaviors with frailty ☆," *Exp. Gerontol.*, vol. 114, pp. 1–12, 2018.

[17] M. Bassoli, V. Bianchi, I. Munari, M. Bassoli, V. Bianchi, and I. De Munari, "A Plug and Play IoT Wi-Fi Smart Home System for Human Monitoring," *Electronics*, vol. 7, no. 9, p. 200, Sep. 2018.

[18] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1473–1488, Nov. 2008.

[19] S. Cagnoni, G. Matrella, M. Mordonini, F. Sassi, and L. Ascari, "Sensor fusion-oriented fall detection for assistive technologies applications," in *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications*, 2009, pp. 673–678.

[20] R. Ugolotti, F. Sassi, M. Mordonini, and S. Cagnoni, "Multi-sensor system for detection and classification of human activities," *J. Ambient Intell. Humaniz. Comput.*, vol. 4, no. 1, pp. 27–41, 2013.

[21] K. Davis *et al.*, "Activity recognition based on inertial sensors for Ambient Assisted Living," in *FUSION 2016 - 19th International Conference on Information Fusion, Proceedings*, 2016, pp. 371–378.

[22] M. Bassoli, V. Bianchi, I. De Munari, and P. Ciampolini, "An IoT Approach for an AAL Wi-Fi-Based Monitoring System," *IEEE Trans. Instrum. Meas.*, 2017.

[23] C. Tsirmpas, A. Anastasiou, P. Bountris, and D. Koutsouris, "A New Method for Profile Generation in an Internet of Things Environment: An Application in Ambient-Assisted Living," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 471–478, Dec. 2015.

[24] M. Bassoli, V. Bianchi, I. De Munari, and P. Ciampolini, "An unobtrusive Wi-Fi system for human monitoring," in *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin*, 2017.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[26] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *MADiMa 2016 - Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, co-located with ACM Multimedia 2016*, 2016, pp. 41–49.

[27] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, no. 6, pp. 1236–1246, 2018.

[28] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *MM 2015 - Proceedings of the 2015 ACM Multimedia Conference*, 2015, pp. 1307–1310.

[29] M. Zeng *et al.*, "Convolutional Neural Networks for human activity recognition using mobile sensors," in *Proceedings of the 2014 6th International Conference on Mobile Computing, Applications and Services, MobiCASE 2014*, 2015, pp. 197–205.

[30] Y. Chen and Y. Xue, "A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer," in *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, 2016, pp. 1488–1492.

[31] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *IJCAI International Joint Conference on Artificial Intelligence*, 2016, vol. 2016-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2920283, IEEE Internet of Things Journal

Janua, pp. 1533–1540.

[32] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *BSN 2016 - 13th Annual Body Sensor Networks Conference*, 2016, pp. 71–76.

[33] P. P. San, P. Kakar, X.-L. Li, S. Krishnaswamy, J.-B. Yang, and M. N. Nguyen, *Deep Learning for Human Activity Recognition*. 2017.

[34] A. Khan, N. Hammerla, S. Mellor, and T. Plötz, "Optimising sampling rates for accelerometer-based human activity recognition ☆," *Pattern Recognit. Lett.*, vol. 73, pp. 33–40, 2016.

[35] M. Hooshmand, D. Zordan, D. Del Testa, E. Grisan, and M. Rossi, "Boosting the Battery Life of Wearables for Health Monitoring Through the Compression of Biosignals," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1647–1662, Oct. 2017.

[36] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *IJCAI International Joint Conference on Artificial Intelligence*, 2015, vol. 2015-Janua, pp. 3995–4001.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[38] D. Yu and J. Li, "Recent progresses in deep learning based acoustic models," *IEEE/CAA J. Autom. Sin.*, vol. 4, no. 3, pp. 396–409, 2017.

[39] R. D. Connor, Jerome T. MARTIN and L. E. ATLAS, "Recurrent neural networks and robust time series prediction," *IEEE Trans. neural networks*, vol. 5, no. 2, pp. 240–254, 1994.

[40] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," Jan. 2016.

[41] B. K. James Bergstra, Remi Bardenet, Yoshua Bengio, "Algorithms for Hyper-Parameter Optimization," *Adv. Neural Inf. Process. Syst.*, pp. 2546–2554.

[42] Y.-T. Zhou and R. Chellappa, "Computation of optical flow using a neural network," in *IEEE International Conference on Neural Networks*, 1988, pp. 71–78.

[43] Y. LeCun, D. Touresky, G. Hinton, and T. Sejnowski, "A theoretical framework for back-propagation," in *Proceedings of the 1988 connectionist models summer school*, 1988, pp. 21–28.

[44] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[45] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *ESANN 2013 proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013, pp. 437–442.

[46] S. Seto, W. Zhang, and Y. Zhou, "Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition," in *EEE Symposium Series on Computational Intelligence*, 2015, pp. 1399–1406.

[47] C. A. Ronao and S. B. Cho, "Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models," in *International Conference on Natural Computation*, 2014, pp. 681–686.

[48] Y. Li, D. Shi, B. Ding, and D. Liu, "Unsupervised feature learning for human activity recognition using smartphone sensors," in *Mining Intelligence and Knowledge Exploration: Second International Conference*, 2014, pp. 99–107.

[49] I. Masaya, I. Sozo, and N. Takeshi, "Deep Recurrent Neural Network for Mobile Human Activity Recognition With High Throughput," *Artif. Life Robot.*, vol. 23, no. 2, pp. 173–185, 2018.

[50] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors," *Math. Probl. Eng.*, 2018.

[51] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, *Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine*, vol. 7657 LNCS. 2012.

[52] A. MURAD and J.-Y. PYUN, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, 2017.

[53] B. Kolosnjaji and C. Eckert, "Neural network-based user-independent physical activity recognition for mobile devices," in *International Conference on Intelligent Data Engineering and Automated Learning*, 2015, pp. 378–386.

[54] C.A. Ronao and S. B. Cho, "Evaluation of Deep Convolutional Neural Network Architectures for Human Activity Recognition with Smartphone Sensors," *Yonsei Univ.*, 2015.

[55] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, 2016.

[56] C. A. Ronao and S.-B. Cho, "Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models," *Int. J. Distrib. Sens. Networks*, 2017.

[57] D. K. Y.J. Kim, B.N. Kang, "Hidden Markov model ensemble for activity recognition using tri-axis accelerometer," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 3036–3041.