Human Activity Recognition Based on Convolutional Neural Network

Yves Coelho, Luara Rangel, Francisco dos Santos, Anselmo Frizera-Neto, and Teodiano Bastos-Filho

Abstract

It is increasingly essential to monitor clinical signs and physical activities of elderly, looking for early warning signs or to recognize abnormal situations, such as a fall. In recent years, the usage of wearable sensors has increased significantly. Data from wearable devices can be used to recognize human movement patterns while performing various activities. Accelerometers have been widely used in human activity recognition systems, however, instead of traditional techniques used for feature extraction, the scientific community is currently developing classifiers based on deep learning techniques, seeking better performance and lower computational cost. Convolutional neural networks (CNN) are the main deep learning technique used in this context. These networks adjust filter coefficients that are applied to small regions of the data, extracting local patterns and their variations. This paper presents a human activity recognition system based on convolutional neural networks to classify six activities—walking, running, walking upstairs, walking downstairs, standing and sitting—from accelerometer data. Results demonstrate the ability of the proposed CNN-based model to obtain a state-of-art performance, with accuracy of 94.89% and precision of 95.78% for the best configuration.

Keywords

Human activity recognition • Convolutional neural networks • Wearables

Y. Coelho (✉) · A. Frizera-Neto · T. Bastos-Filho
Postgraduate Program in Electrical Engineering, Federal University of Espirito Santo, Vitoria, Brazil
e-mail: yvesluduvico@gmail.com

L. Rangel · A. Frizera-Neto · T. Bastos-Filho
Department of Electrical Engineering, Federal University of Espirito Santo, Vitoria, Brazil

F. dos Santos
Department of Computer Science and Electronics, Federal University of Espirito Santo, Sao Mateus, Brazil

1 Introduction

The deceleration of world population growth caused by the reduction of the birth rate, together with the higher life expectancy, results in the increase of aging population. Therefore, it is increasingly important to give the elderly the opportunity to have a more independent life, with safety and autonomy to carry out their daily tasks. In addition, it is essential to monitor clinical signs and physical activities of this population, to early detect serious medical conditions or to recognize an abnormal situation, such as a fall.

In recent years, the usage of wearable sensors has increased significantly. These systems monitor and store real-time information on the physiological condition and movements of a person through the use of different types of flexible sensors that can be integrated into clothing, elastic bands or directly attached to the human body [1]. The data from these lightweight and power efficient wearable sensors can be used to recognize human movement patterns while performing various activities [2].

Among these wearable sensors, accelerometers have been widely used in human activity recognition (HAR) systems [3–7]. According to [7], most of the works choose the waist to place the accelerometer, given its central location, which is more stable and better for recognition of global movements, such as walking and running. However, the accelerometer can be used on the body extremities to recognize a greater number and more complex activities.

For feature extraction, handcrafted features are widely used in HAR systems. Those features can be extracted from time domain or fast-Fourier transform coefficients in the frequency domain [8]. Time-domain features include mean, median, variance, skewness, kurtosis, range, root mean square, mean absolute deviation [3–5]. Those features are fed into different classification algorithms, such as Decision Tree, k-Nearest Neighbors (k-NN), Naïve Bayes, Support Vector

© Springer Nature Singapore Pte Ltd. 2019

R. Costa-Felix et al. (eds.), *XXVI Brazilian Congress on Biomedical Engineering*,
IFMBE Proceedings 70/2, https://doi.org/10.1007/978-981-13-2517-5_38

Machine (SMV), Hidden Markov Model (HMM) and Multilayer Perceptron (MLP) [3–5, 9]. Despite the great performance found in most of the previous work [3–5], the generalization of handcrafted features to new data sources is usually poor and a prior domain knowledge is needed to find the more efficient features.

Instead of using techniques that rely on handcrafted feature extraction or others traditional techniques, the scientific community is currently developing classifiers based on deep learning techniques for HAR systems, seeking better performance and lower computational cost. In this technique, the feature extraction and classification procedures are often performed simultaneously, and the features are learned automatically through the network training process instead of being manually designed [10].

This paper presents both a HAR system based on convolutional neural network (CNN) to classify six activities (walking, running, walking upstairs, walking downstairs, standing and sitting) and the development of a wearable device for data gathering. In next section, recent state-of-art performances in HAR systems are presented, followed, in Sect. 3, by the wearable device designed and constructed for gathering data. Section 4 describes the proposed CNN and its characteristics, and the results are discussed in Sect. 5. Finally, we conclude this paper in Sect. 6.

2 Related Work

Proposals using deep-learning approach, such CNN, for HAR systems using wearable sensors have multiplied recently, always achieving state-of-the-art performances [6, 11–13]. CNN is a neural network that uses convolution in place of general matrix multiplication in at least one of their layers [14]. CNN extracts features from signals, which has shown promising results in image classification, speech recognition, and text analysis. When dealing with time series classification like HAR, CNN has two main advantages over previously used models: local dependency and scale invariance [10].

In [6], the authors proposed a CNN to classify three activities (walking, running, and staying still) using accelerometer data from smartphone, used in various positions. The authors worked with the raw x, y, and z-axis acceleration data transformed into a magnitude data. The CNN was constructed with one convolutional layer, one max-pooling layer, one fully connected layer and one softmax layer. They reached 91.32% and 92.71% of accuracy, using 10 and 20 s data frame as input, respectively, and then concluded that the dimension of the input vector affects the activity recognition performance.

Pham et al. [11] developed a classifier for seven activities (running, kicking, jumping, cycling, standing, walking, and an arbitrary) using a CNN composed of two convolutional layers, two max pooling layers, two fully connected layers, a dropout layer and a softmax layer. The authors used 32 and 64 filters, respectively, for first and second convolutional layers. The fully connected layers have 500 and 100 neurons. Using approximately 780 sample frames for each activity from 10 subjects, and testing with 10-fold cross validation, they reached a precision of 93.41% and a recall of 93.16%.

In [12], accelerometer and gyroscope tri-axial data were collected from 30 subjects who performed six different activities carrying the smartphone in their pockets. Using a 6-channel 1-D CNN, the authors achieved accuracy of 94.79% with raw sensor data, and 95.75% with additional information of FFT from the HAR data set. They concluded high accuracy was mostly due to great classification of dynamic activities, especially those very similar, like walking upstairs and downstairs.

Zebin et al. [13] proposed a CNN to classify the following activities: walking, walking upstairs, downstairs, sitting, standing, lying down. The authors collected data from 3-axis accelerometer and 3-axis gyroscope from 12 healthy volunteers using the sensors on pelvis, thigh or shank. They performed a 6-channel 1D convolution with three convolutional layers. Results show an improvement in accuracy using CNN, reaching 97.01%, against 96.40% and 91.70%, using SVM and MLP, respectively. According to the authors, tuning the filter and pooling sizes showed that larger filter and lower pooling size improves the network performance.

3 Wearable Device

The hardware used on our wearable device was initially developed to monitor the user's physical activity level, calculating the absolute mean deviation from accelerometer data. Given the requirements of the previous research, in principle, it was not necessary to choose components of high-performance regarding processing and memory resources. For this reason, the electronic components chosen have limited resources for the current research, but they are low-energy and low-cost components. This section presents the wearable device, its main components, how to wear it, and the hardware's limited resources.

The wearable device is composed, basically, of a 3-axis digital accelerometer (ADXL362), an 8-bit microcontroller (ATtmega328P) and a Bluetooth Low Energy (BLE) module (HM-10). Figure 1a shows the developed prototype comprised of the electronic board, the battery (CR2032) and the encapsulation. The user wears the device on waist (see Fig. 1b). Figure 2a shows the axes of movement related to the human body, while Fig. 2b illustrates the orientation of accelerometer axes according to the worn position.

Fig. 1 **a** Developed wearable device. **b** Device worn on the user's waist

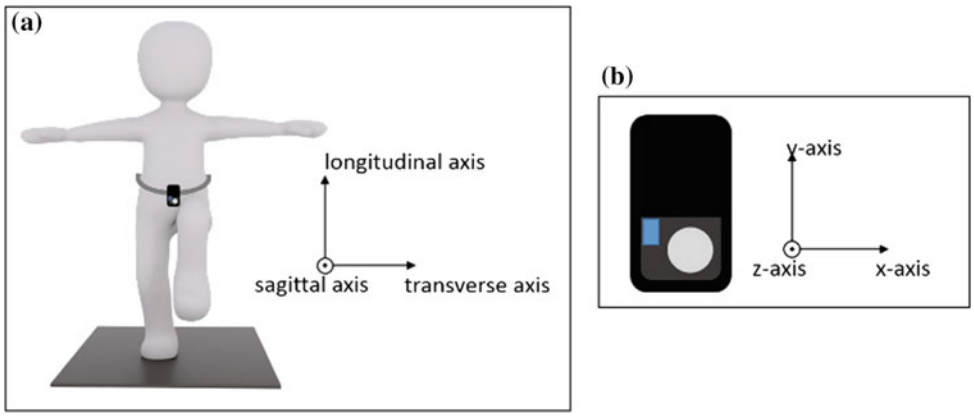

Fig. 2 **a** Axes of movement. **b** Orientation of device's accelerometer axes according to the worn position

At this stage of the research, the wearable device was used to collect acceleration data while subjects perform activities using the device. The data was then used to construct the proposed model. Currently, a device with higher processing capacity is being developed, in which the classifier algorithm will be embedded on it for real-time activity classification.

4 Human Activity Recognition (HAR) System

In this section, we describe the basic architecture of CNNs and present the proposed HAR system based on CNN.

4.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) are neural networks that use convolution in place of general matrix multiplication in at least one of their layers [14]. In training process, CNNs adjust the filter coefficients from convolutional layers. These filters are applied to the data, capturing local patterns. In CNN architectures, the feature extraction is performed inherently in the convolutional layers, and the produced features are inputted to the fully-connected layers where classification occurs [15].

The convolutional layer is composed of several filters that are applied to the data to form feature maps. The new feature map can be obtained by convolving the input data with a learned filter, and then applying an element-wise non-linear activation function on the convolved results. Commonly used activation functions are sigmoidal, hyperbolic tangent and linear rectified activation function [15].

Pooling is a very important tool of CNN, because it reduces the computational cost by reducing the number of connections between convolutional layers [15]. According to [14], pooling helps to make the representation approximately invariant to small translations of the input.

After convolutional and max-pooling layers, the processed output from these layers is then flattened to be used as input of the fully connected layer. Finally, the output of the fully connected layer is passed to the soft-max layer that computes probability distributions.

4.2 Proposed Architecture

In this paper, we propose the CNN model presented in Fig. 3 to classify six activities from accelerometer data. The network input data are extracted from accelerometer's y and z axes. The input data frame was evaluated in different sizes: 20, 50 and 100 elements, corresponding to a 2-s, 5-s and

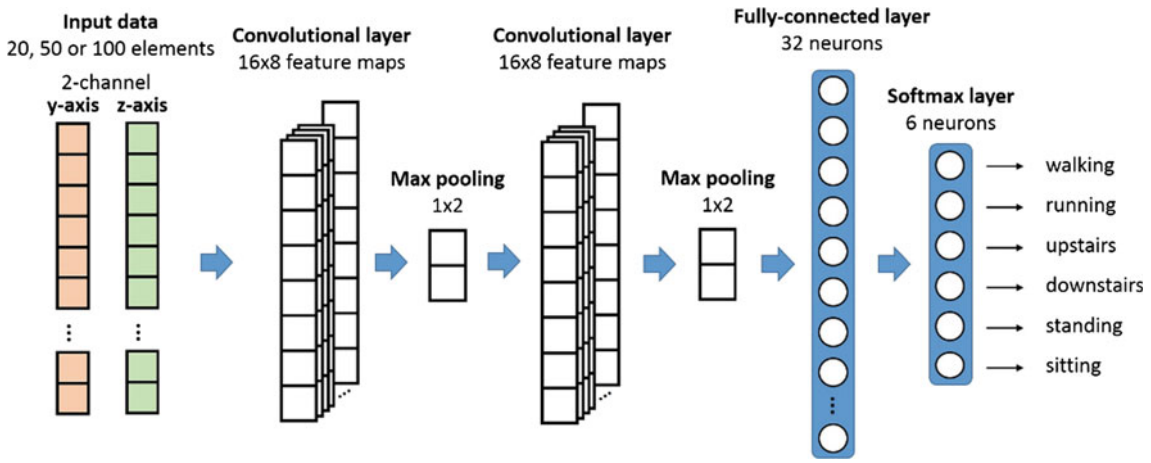

Fig. 3 Proposed CNN architecture

10-s window, respectively. Therefore, the input of the network has two channels, each channel being formed by a vector of dimensions 1×20, 1×50 or 1×100.

The input is then applied to the first convolutional layer, and its output, the filtered data, passes through the first max pooling layer. The following stage is another pair of convolutional layer and max pooling layer. Both convolutional layers have 16 filters of 1×8 size and rectified linear unit as activation function, while the max-pooling layers have filters of dimension 1×2.

In the next step, the processed data are flattened to a row-vector and applied to the fully connected layer, formed by 32 neurons, with sigmoid activation function. The output layer has six neurons, each one representing one class of activity, and softmax activation function.

The algorithm was developed in Python 3.6, using NumPy, Keras and scikit-learn libraries, in Spyder integrated development environment. The CNN model was trained using the stochastic gradient descent optimizer (SGD function from Keras) with learning rate of 0.01, learning rate decay of 10^{-5} and Nesterov momentum of 0.9.

5 Results and Discussion

Data were collected from five healthy individuals, three males and two females, with a mean age of 28.80 ± 3.42 years. Each subject was instructed to perform each one of the six activities during about 30 s to 1 min. While each subject performed the activities, the acceleration data was transferred to a smartphone via Bluetooth (see Fig. 4, step 1) and saved to a text file (see Fig. 4, step 2). Finally, when concluding tests, the text files, separated by subjects and activities, were manually transferred to a computer, where data were extracted and used to create and evaluate the proposed classifier model.

Due to the limited processing and memory resources of the microcontroller used, it was necessary to reduce the size of data packets. For this reason, data from only two axes of the accelerometer were collected (y and z—respectively longitudinal and sagittal axes of movement—as shown in Fig. 2). Data acquisition from accelerometer was performed using a sample rate of 100 Hz, applying an average

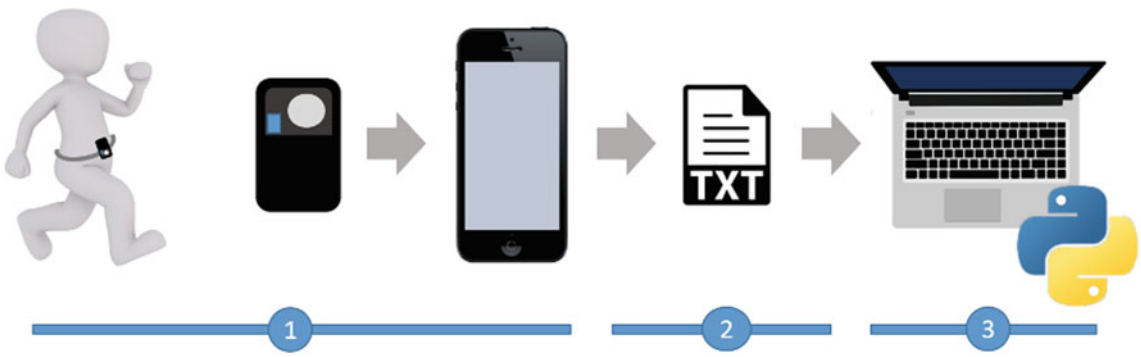

Fig. 4 Steps of data collection

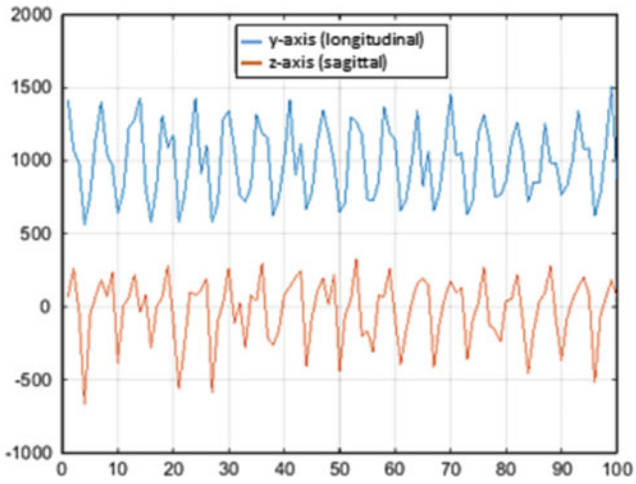

Fig. 5 Samples of acceleration data from walking activity

downsampling from 100 to 10 Hz, resulting in a smooth signal with an averaged sample acquired each 100 ms. Figure 5 shows a piece of data gathered from walking activity. In total, 28,640 samples were collected, distributed per activity, as shown in Table 1.

For training of the classifier, we used the k-fold cross validation technique with four partitions. The training was carried out in 300 epochs, an amount that was sufficient for network convergence. Five rounds of training and test were performed, and we used average accuracy, precision and recall as metrics. The model was evaluated with different dimensions of input vector, with size of 20, 50 and 100 elements, which represent a data frame of 2, 5 and 10 s. Table 2 presents the performance results.

In our preliminary tests, statistical domain features (handcrafted) were extracted from accelerometer data. In that case, we worked with 2-s window and 10 features (five from y-axis and five from z-axis): mean, median, standard deviation, signal magnitude area and amplitude. Different classifiers were tried for comparison, including MLP, k-NN (k = 5), Decision Tree and Naïve Bayes. Those performance results are also shown in Table 2 for comparison.

For comparison, we used macro-averaged precision (P) and recall (R) as metrics, calculated according the following equations, for each class i, for all c classes. True positives (TP) refer to the number of instances of a class k classified as k. False positives (FP) refer to the number of all non-k instances classified as k. While false negatives (FN) refer to the number of instances of class k not classified as k.

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{1}$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{2}$$

$$P = \frac{1}{c}\sum_{i=1}^{c} P_i \times 100\% \tag{3}$$

$$R = \frac{1}{c}\sum_{i=1}^{c} R_i \times 100\% \tag{4}$$

Table 1 Number of samples collected and number of frames according to the data window size

Activity	Samples	Frames of 2-s window	Frames of 5-s window	Frames of 10-s window
Walking	5740	287	114	57
Running	4180	209	83	41
Walking downstairs	5700	285	114	57
Walking upstairs	6500	325	130	65
Standing	3460	173	69	34
Sitting	3220	161	64	32

Table 2 Performance of proposed classifiers for the HAR system

Classifier	Accuracy (%)	Precision (%)	Recall (%)
MLP	85.47	88.86	88.15
Decision tree	82.05	84.31	81.54
k-NN	88.62	91.32	90.26
Naïve Bayes	83.94	86.57	86.39
CNN with 2-s frame	89.22	91.29	90.83
CNN with 5-s frame	91.78	93.88	93.07
CNN with 10-s frame	94.89	95.78	95.52

We can note an improvement in classification performance when using the CNN-based architectures. Also, results show that wider data frame implied better performance. A larger input frame allowed the convolutional layer filters to extract more representative features from the data, contributing to the classifier precision.

The CNN model with the 10-s input vector reached the best performance. However, this option would not be suitable for systems that require a fast response. In addition, an activity transition will take longer to be identified by the system. For a faster response, a window of 2 s would be more suitable. Nevertheless, this configuration did not present a satisfactory result, although better than the techniques that use handcrafted features.

In order to improve the performance of the 2-s input vector configuration, it is necessary to use a larger frame as input, i.e., use a higher acquisition rate, in this case. Thus, the convolutional layers will be able to extract more specific features implicit in the signal.

6 Conclusion

This work presented the development of a HAR system based on CNN to recognize six activities of daily living. The obtained results demonstrated the CNN-based models reached a state-of-art performance, as we could analyze from related work, with better results, in some cases, than those classifiers that used handcrafted feature-extraction techniques, demonstrating the efficiency of convolutional neural networks on extracting the most representative features from data.

In addition, the performance of the CNN classifier was analyzed according to the input vector dimension, and we verified that the configuration using input data with larger dimension had the best performance.

In future work, we intend, at first, to evaluate the proposed model in available datasets of activities. Later, we will work on the development of a new wearable device, able to embed the CNN classifier algorithm, to run the HAR system in real time.

References

1. Majumder, S., Mondal, T., Deen, M.: Wearable sensors for remote health monitoring. Sensors **17**(1), 1–45 (2017)
2. Mukhopadhyay, S.: Wearable sensors for human activity monitoring: a review. IEEE Sens. **15**(3), 1321–1330 (2015)
3. Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., Amirat, Y.: Physical human activity recognition using wearable sensors. Sensors **15**(12), 31314–31338 (2015)
4. Gao, L., Bourke, A., Nelson, J.: Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. Med. Eng. Phys. **36**(6), 779–785 (2014)
5. Lara, O., Labrador, M.: A survey on human activity recognition using wearable sensors. IEEE Commun. Surv. Tutor. **15**(3), 1192–1209 (2013)
6. Lee, S., Yoon, S., Cho, H.: Human activity recognition from accelerometer data using Convolutional Neural Network. In: IEEE International Conference on Big Data and Smart Computing, pp. 131–134. IEEE, Jeju (2017)
7. Cornacchia, M., Ozcan, K., Zheng, Y., Velipasalar, S.: A survey on activity detection and classification using wearable sensors. IEEE Sens. **17**(2), 386–403 (2017)
8. Ha, S., Yun, J., Chon, S.: Multi-modal convolutional neural networks for activity recognition. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 3017–3022. IEEE, Kowloon (2015)
9. Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., Having, P.: Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: a survey. In: 23rd International Conference on Architecture of Computing Systems, pp. 1–10. VDE, Hannover (2010)
10. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: a survey. Pattern Recogn. Lett. 1–9 (2018)
11. Pham, C., Diep, N., Phuong, T.: e-Shoes: smart shoes for unobtrusive human activity recognition. In: 9th International Conference on Knowledge and Systems Engineering, pp. 269–274. IEEE, Hue (2017)
12. Ronao, C., Cho, S.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst. Appl. **59**, 235–244 (2016)
13. Zebin, T., Scully, P., Ozanyan, K.: Human activity recognition with inertial sensors using a deep learning approach. In: IEEE Sensors Conference, pp. 1–3. IEEE, Orlando (2016)
14. Goodfellow, I., Benfio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
15. Ignatov, A.: Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. Appl. Soft Comput. **62**, 915–922 (2018)