Assessing Physical Activity and Functional Fitness Level Using Convolutional Neural Networks

Alejandro Galán-Mercant[a], Andrés Ortiz[b,c,*], Enrique Herrera-Viedma[c,d], Maria Teresa Tomas[e], Beatriz Fernandes[e], Jose A. Moral-Munoz[a,f]

[a]*Department of Nursing and Physiotherapy, Universidad de Cádiz, Cádiz, Spain*
[b]*Communication Engineering Department, University of Málaga, Málaga, Spain*
[c]*Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI)*
[d]*Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology), University of Granada, Granada, Spain*
[e]*Escola Superior de Tecnologia da Saúde de Lisboa (ESTeSL), Instituto Politécnico de Lisboa, Lisboa, Portugal*
[f]*Institute of Research and Innovation in Biomedical Sciences of the Province of Cádiz (INiBICA), Universidad de Cádiz, Cádiz, Spain*

Abstract

Older adults are related to a reduction in physical functionality, as a result of a musculoskeletal system degeneration. In that way, physical exercise has been stated as a suitable intervention to prevent such health problems. Therefore, an adequate assessment of the physical activity and functional fitness levels is needed to plan the individualized intervention. A broad test used to assess the functional fitness level is the 6-minutes walk test (6MWT). It has been previously measured using accelerometer sensors. In views of this background, the main aim of the present study is to use deep learning to extract automatically and to predict the physical activity and functional fitness levels of the older adults through the acceleration signals recorded by a smartphone during the 6MWT. A total of 17 participants were recruited. Anthropometric measurements (weight, height, and body mass index), physical activity, and functional fitness levels from each participant were recorded. Consecutively, two deep learning-based methods were applied to

*Corresponding Author: e-mail: aortiz@ic.uma.es Tel: +34 952133353
Email addresses: alejandro.galan@uca.es (Alejandro Galán-Mercant), aortiz@ic.uma.es (Andrés Ortiz), viedma@decsai.ugr.es (Enrique Herrera-Viedma), teresa.tomas@estesl.ipl.pt (Maria Teresa Tomas), beatriz.fernandes@estesl.ipl.pt (Beatriz Fernandes), joseantonio.moral@uca.es (Jose A. Moral-Munoz)

determine the prediction. According to the results, the proposed method can predict the physical activity and functional fitness levels with high accuracy, even using only one cycle. Thus, the approach described in the present work could be implemented in future mobile health systems to identify the physical activity profile of older adults.

Keywords: Physical activity, Functional fitness, Deep Learning, Inertial signal, Deep Convolutional Autoencoder/sep Convolutional Network

1. Introduction

One of the most problematic conditions in older adults population is the fragility (Clegg et al., 2013). According to a relatively recent study (Choi et al., 2015), the global prevalence of fragility was found from 4.9% to 27.3%, and prefragility was established from 34.6% to 50.9%. In views of this framework, it is logical that fragility has been considered as a prior public health problem (Cesari et al., 2016, Curtis et al., 2017). In that way, older adults are associated with a reduction in the functionality of the musculoskeletal system. It is related to a bone mass and strength loss, and a decrease in hormone production (Hong et al., 2015). Therefore, all these physiological changes involve an increase in the risk of developing different clinical conditions (Cruz-Jentoft et al., 2017).

Besides, physical activity-based programs seem to be an effective intervention to prevent fragility (Bullo et al., 2018, Karinkanta et al., 2015, Lagerros et al., 2017). Nevertheless, to know the functional level of the older adults, and consequently plan the intervention, there are several tests. One of the most used tests to evaluate the functional exercise capacity is the 6-minutes walking test (6MWT). It is considered as a simple, non-invasive, and reproducible test that reflects the physical condition status of the tested subject through an objective measurement (ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories, 2002, Lima et al., 2018). In that sense, the use of accelerometers, such as Actigraph, Mini-Mitter or IM systems, to assess the physical activity is widely employed in the research literature (Murphy, 2009). Hence, the revolution of using smartphone applications is providing new opportunities for physical activity and functional fitness assessment.

Smartphones incorporate different sensors that allow them to perform different tasks (Banos et al., 2015, Gaikwad et al., 2016). Among these sensors, they are provided with accelerometers, making available the physical activity quantification. Furthermore, smartphone applications, also well-known as apps, makes their use more suitable for a broader population, allowing them to monitor and manage several chronic conditions (Salazar et al., 2018). Therefore, these devices can get the sensors information and store and share it (by WiFi, 3G, LTE, etc.) with another remote device to submit it to a post-processing stage (Moral-Munoz et al., 2018).

Classification of time series data such as signals generated by inertial sensors are difficult to be classified in a *sample-as-features* fashion: the high dimensionality of the feature space along with the limited number of samples (as it is usual in biomedical problems) produces the so-called *curse of the dimensionality* problem, limiting the generalization capabilities of the model. In that way, a post-processing stage to compute relevant information to classify inertial signals by extracting statistical temporal and spectral descriptors reduces the dimensionality of the feature space improving the generalization capability of the model and reducing the computational burden. While temporal features are based on statistics computed directly over the time signal, spectral features require computing the Power Spectrum (PSD) of the signal to extract information (i.e., power) in different sub-bands. This has been traditionally carried out by Fourier (Aggarwal and Ryoo, 2011) or Wavelet Analysis (Ayachi et al., 2016, Lockhart et al., 2013, Aggarwal and Ryoo, 2011). However, these analysis methods present some drawbacks related to the non-stationary nature of the inertial signals.

Furthermore, the use of features computed from the time signal (such as mean, variance, or amplitude-related features) as well as features derived from the PSD spectrum (peak power, spectral centroid, spectral kurtosis, etc.) could neither be descriptive enough or capture the pattern related to class discrimination. Thus, our hypothesis is that the physical activity and functional levels of older adults could be predicted using two complementary methods that extract disciminative features from the inertial signals recorded during the 6MWT.

In this paper, we present two different approaches based on deep learning architectures. The first consist on using a Convolutional Autoencoder (CAE) to extract features from the acceleration data. These features are then classified using a support vector machine. The second approach presented in this paper consist on using a convolutional neural network (CNN) that firstly extract features from the acceleration data and then, uses a perceptron-like network (fully connected layers) to implement a classifier. In this case, the same network extracts features and classify them. Moreover, we compare the results obtained using the deep learning-based approaches presented in this paper, to the previously presented in (Galán-Mercant et al., 2018), which uses only signal processing techniques (specifically, Empirical Mode Decomposition) to decompose the acceleration signals and then computes classical time and frequency statistical descriptors. These descriptors eventually classified using a support vector machine.

Thus, The novelty of this paper is twofold. Firstly, we used the 6 minutes walking test to explore human movement patters related to fragility. Secondly, we proposed a method that avoids either to use of classical statistical signal processing techniques or the computation of predefined statistics as features to describe the signals. Instead, we propose two deep learning architectures that automatically compute specific features through a learning process. In this way, we present two methods that can be seen as complementary. The CAE approach uses unsupervised learning to compute representative features that can be eventually classified. The CNN consist in a convolutional stage that extracts features and a classification network. These two approaches perform equally (in view of the statistical analysis), but provides two ways using different learning paradigms to extract features from inertial signals using machine learning.

The rest of the paper is organized as follows. Section 2 describes the current studies available in which the 6MWT is measured using accelerometers. Next, Section 3 presents the database used to assess the method proposed in this work along with a description of the methodology, including the feature extraction process from the original inertial signals. Then, the results obtained are shown in Section 4 and finally, conclusions and future work is presented in Section 5.

2. Related work

As stated above, smartphones incorporate several sensors that make them suitable to measure different physical parameters. Furthermore, several new devices have appeared that also include this technology, such as smartwatches and smartbands. Therefore, several papers can be found in the current literature in which physical condition was predicted using the accelerometer sensor. In that way, Drover et al. (2017) used Fourier transform to classify older adults according to their fall risk, using the 6MWT acceleration data obtained by a commercial device. Furthermore, Similä et al. (2017) also employed Fourier transform to detect early signs of balance deficits using acceleration data obtained during the Berg Balance Scale, Timed-Up-and-Go, and 4-meters walk test. Besides, Vervoort et al. (2016) applied Wavelet analysis to classify the population according to the aging effects, through the inertial sensor data obtained during the Timed-Up-and-Go test. A study to predict the physical activity and functional fitness levels applying Empirical Mode Decomposition (EMD) features to the acceleration data obtained during the 6MWT using a smartphone is yet to be undertaken. According to the previously addressed, this approach takes into account the non-stationary nature of the inertial signals.

Therefore, this kind of analysis allows us to think of the future application for new devices, such as smartwatches and smartbands. The reliability and validity of different fitness tracker (smartbands) to measure the step count during the 2-minutes walking test (2MWT) in older adults have been assessed, reporting acceptable outcomes. Nevertheless, new devices more accurate to measure physical activity in a free-living environment are needed. In this sense, a study (Duncan et al., 2018) recommends taking precautions considering the walk-related data obtained by iPhone accelerometer in free-living condition; they can reach a mean bias of 21.5% or an imbalance of 1340 steps/day. On another note, they consider that iPhone accelerometer is suitable in laboratory conditions, with a mean bias of less than 5% (acceptable for pedometers). Therefore, current user-available technologies need to be designed for controlled scenarios.

Previous works such as (Lockhart et al., 2013, Yang et al., 2012), are focused on the classification of human movement, using classical signal processing techniques to preprocess the inertial signals and to extract statistical descriptors. Specifically, in (Lockhart et al., 2013), spatio-temporal features are derived from wavelet analysis. Other works such as (Yang et al., 2012) present a tool to compute temporal, spectral and spatio-temporal features to describe the inertial signals aiming to model the gait. These works aims to classify inertial signals, focused on Human Activity Recognition (HAR) but not to predict clinical labels related to the physical status or more specifically, to fragility risk. In this way, the previous work (Galán-Mercant et al., 2018) proposes the use of EMD as decomposition method to obtain the time signal on different frequency sub-bands. Then, temporal and spectral features are extracted and used to classify the time series. Beyond the use of classical signal processing techniques, machine learning approaches constitute an alternative to extract representative or discriminative features from inertial signals, without the previous knowledge or assumptions needed to address the problem from a classical signal processing point of view (stationarity or periodicity, for instance). Furthermore, current Deep Learning techniques have demonstrated their success in different complex classification problems (Bengio, 2009, Zhang et al., 2018) and their ability extracting features at different abstraction levels (Ortiz et al., 2016).

A survey paper for the use of different Deep Learning-based methods for HAR is presented in Wang et al. (2018), using public datasets. In this paper, the use of different architectures such as Deep Neural Networks, Deep Belief Networks and CNNs is slightly reviewed. The use of CNNs is also presented in Cho and Yoon

(2018) for the two-classes HAR problem. In addition, a method for 6-class HAR classification is presented in Zeng et al. (2014). Nevertheless, previous works using classical signal processing techniques, including multiresolution and wavelet analysis (Vishwakarma et al., 2015) or autoregressive models (He, 2010), show similar performance using computationally lighter methods. On the contrary, the classification of subjects according to clinical labels regarding their physical activity require models capable of extracting discriminative features and figuring out patterns (Galán-Mercant et al., 2018).

3. Materials and methods

3.1. Dataset

The dataset employed in the present study to apply CAE and CNN was obtained from a group of subjects. Information relative to anthropometric measures and physical activity was recorded in form of inertial signals. These signals can be seen as sequences of acceleration values sampled at a specific rate (namely, sampling rate). In this case, acceleration in the three axes (x,y,z) are sampled simultaneously. In our case, a sampling rate of 32 Hz was used, which means that 32 acceleration values are stored per second. This results in a sequence of acceleration values periodically sampled and temporarily sorted. This, in general is called time series.

3.1.1. Study subjects

In order to get the information into this cross-sectional study, a total of 17 older adults (14 women and 3 men), were recruited between May and July 2016 from a primary health center in Lisbon (Portugal). The Inclusion criteria were: older adults who could get up and down from a chair five times without at external-internal aid, and older adults who could complete over 6 minutes walking as fast as possible without at external-internal aid. Each participant received a detailed explanation of the study and gave written informed consent before participation. This study complied with the principles of the Declaration of Helsinki. All study procedures were approved by the institutional review boards of the participating institutions.

3.1.2. Protocol

Four different types of variables were recorded from our subjects: I) anthropometric measurements, II) physical activity levels, III) functional fitness levels and

IV) 6-Minutes Walking Test. In that way, the inertial signals was only obtained during the 6MWT.

I) *Anthropometric Measurements:* The anthropometrics procedures described were obtained following the guidelines of The International Society for the Advancement of Kinanthropometry (ISAK) (Ross et al., 1978). Height and weight were recorded with the participant barefoot and in light clothing. The subject, standing in anatomical position with the occipital region, back, gluteal region and heels in contact with the height rod, takes a deep breath for height measurement. Height is the distance from the vertex to the soles of the feet. Body mass index (BMI) was calculated by dividing weight in kilograms (kg) by height in square meters (m2). Calf circumference on the dominant side was measured at the point of the widest diameter of the calf. Mid-upper arm circumference on the dominant side was measured on the upper arm at the midpoint between the acromion and the olecranon.

II) *Physical Activity Level:* Physical activity level (PAL) were assessed and subjects were classified as sedentary/inactive, insufficiently active or active according to the classification of physical behavior (American College of Sports Medicine, 2013). An active participant was considered who perform 30 min of moderate activity at least 5 days a week and/or 20 min of vigorous activity 3 days per week or a combination of both. All the subjects were classified as sedentary or insufficiently active profiles. Thus, the variable was considered as dichotomous. From the 17 participants, 64.71% were sedentary/inactive and 35.29% insufficiently active.

III) *Functional Fitness Level:* The Senior Fitness Test battery was used to assess functional fitness level of the participants. The tests were performed according to guidelines and protocols for administration (Rikli and Jones, 2013). Lower limbs strength was measured with the 30s chair stand test (30-s CST); to measure aerobic endurance we used the 6-min walk test (6MWT) (ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories, 2002); and estimated distance walked was calculated with a validated equation (Troosters et al., 1999).

IV) *6 Minutes Walking Test:* The inertial data is acquired during the 6MWT, which consists of repetitive 30 meters short walks at his/her own pace with a point to point track (marked with small cones with 5-meters increment signalization). Two chairs may be available at the two points limits to rest in case the subject becomes symptomatic. Although the safety of this test is high, it is recommendable to perform by a qualified personal. In that way, the aim is to obtain different cycles in the inertial signal, each corresponding to a different walk. An iPhone 4 was snugly secured to the test subjects by a neoprene fixation belt over the sternum with the screen looking forward to obtain the signal data (Figure 1).

Figure 1: Kinematic data collection procedure.

Information about all the characteristics and measures recorded of the participants is shown in Table 1.

Table 1: Descriptive and clinical characteristic of the subjects (n=17)

	Mean	**SD**
Age (years)	83.26	6.56
Weight (Kg)	64.53	7.42
Height (m)	1.52	0.07
BMI (kg/m2)	28.03	2.74
6MWT (m)	359.26	107.50
30-s CST (repetitions)	11.37	4.89

3.2. Signal pre-processing

3.2.1. Signal filtering

Most of the information related to human movement is contained in frequencies below 20 Hz. This way, the resultant signal is firstly smoothed by low pass filtering using a Butterworth 5th order filter with a cut-off frequency of 16 Hz to comply with the Niquist theorem. This aims to remove spikes with a high-frequency content that only contributes to the final features as noise.

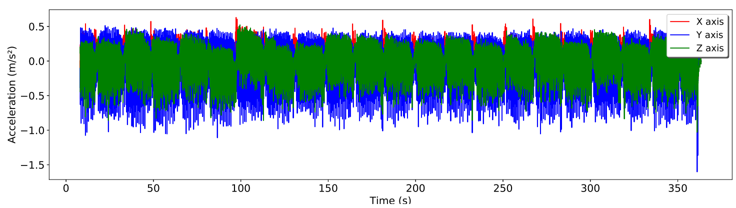
3.2.2. Automatic signal segmentation

As explained above, inertial data is acquired during the 6MWT. As a result, different cycles appear in the inertial signal, each corresponding to a different walk. A previous protocol (Galán-Mercant and Cuesta-Vargas, 2014) was used to identify the kinematic variables of the 6MWT. In this protocol, the linear acceleration was measured along three orthogonal axes using the iPhone 4 accelerometer.

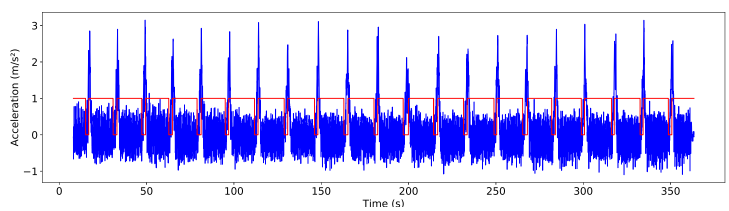
The application used to obtain kinematic data was xSensor Pro, Crossbow Technology, Inc. The data sampling rate was set to 32 Hz. A previous study showed an internal error (standard deviation of the difference between measurements by two different observers) of 4.0' for the iPhone and 3.4' for the protractor (Galán-Mercant et al., 2014).This way, the overall test signal has to be firstly segmented to extract the excerpts corresponding to each walk. This has been addressed by detecting activity periods on the x-axis of the gyroscope signal which indicates a rotation of the body around its axis. This determines (as shown in Figure 2b) abrupt changes in the acceleration corresponding to changes in the movement direction due to the specific orientation of the accelerometer used in the experiments. Figure 1 has been added in this revision to illustrate how direction changes mainly affect to the gyroscope x-signal. Then, this is treated as an activity detection problem, where non-activity periods corresponds to changes of direction (turn at the end of the corridor according to the 6 minutes walking test). This activity detection in the x-axis of the gyroscope is determined by computing its envelope, using the magnitude of the analytical signal obtained using the Hilbert transform. Then, the activity period is detected by thresholding the signal, so that it stays at least a minimum number of samples above the threshold. This provides a robust enough method for the automatic segmentation which worked correctly over all the available samples in our database.

Figure 2 (a) shows the original magnitude m of the resultant acceleration $m = \sqrt{x^2 + y^2 + z^2}$ over time. The method described above is used to segment the gyroscope x-axis signal, shown in 2 (b) and then, segmentation mask is applied onto the m signal. The result is shown in Figure 2 (c). In order to homogenize the segment length, each extracted segment is cut to keep 1200 center samples. Subsequently, a pool containing all the segments is composed as well as the corresponding indexes to identify the segments belonging to a specific subject. This way, each segment obtained from each subject is used as an input to the autoencoder. This deals with a twofold objective. Since segments belonging similar condition subjects are thought to contain similar information, these segments can be pooled and labelled to have more available data to train the classifier, reducing the overfitting effect and improving the generalization capabilities of the model. On the other hand, classification can be carried out on single segments that can be subsequently combined to improve the classification performance.

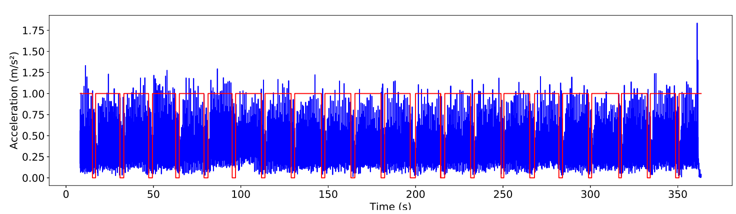
Figure 3 shows all the stages involved in the proposed method. This includes 1) the acquisition of the raw inertial signals, 2) preprocessing, 3) feature extraction and 4) classification and 5) combination. Figure 2 shows the general methodology proposed in this work. This consist of four stages:



(a)



(b)



(c)

Figure 2: Resultant acceleration (a), gyroscope x-axis segmented (b) and resultant acceleration segmented (c)

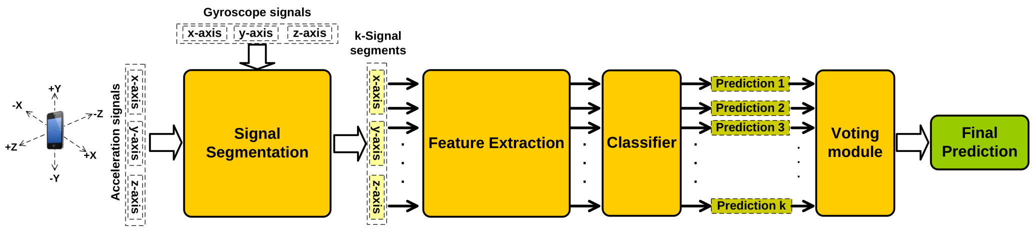


Figure 3: Overall method, including segmentation of the inertial signals, feature extraction, classification and ensemble

1. Preprocessing. After raw inertial signals are acquired using the IMU, sampled at 32 Hz, they are low-pass filtered, since information related to human movement is contained in low-frequency components. Specifically, a 15Hz, 5th order Butterworth filter is used.
2. Segmentation. As explained in the introduction, the 6 minutes walking test consists on a series walks along a 30m corridor. However, the entire time series (containing all the walks) is stored during the signal acquisition. This preprocessing step takes advantage of the gyroscope x-axis signal to detect turns around (see IMU orientation in Figure 1), and each of these walks is what we called segment. Each segment will be treated as a sample.
3. Feature extraction. This step aims to compute descriptors from each segment. These descriptors has to be informative enough to represent the signals and eventually, to classify them (discriminative information). Unlike previous approaches (Lockhart et al., 2013, Yang et al., 2012, Galán-Mercant et al., 2018) that use a priori known, statistical descriptors, we propose the use of deep learning architectures to learn features. Specifically, two methods (CAE and CNN) are presented, that uses unsupervised and supervised learning, respectively. Deep-learning based methods allow to compute the most discriminative and specific features from raw data, avoiding the use of signal decomposition methods (such as Fourier or Wavelet analysis) that use a predefined basis. On the other hand, supervised and unsupervised methods presented here can be used for complementary objectives. The unsupervised, CAE-based method is focused on the extraction of representative features in a lower dimensional space. These features can be further used for classification or regression once the network is trained. Supervised CNN-based method generates specific features during the learning process when the network is trained to differentiate between classes. The resulting discriminative features reveal specific samples and shapes of the

original signals that activates neurons at different layers for a specific class.

4. Classification and combination. The features extracted from each segment are then used to feed a supervised classifier. This results in a prediction for each segment of the same subject. Hence, these predictions have to be combined to produce an unique outcome. This is addressed by a majority-voting strategy, whose mathematical details are provided in Section 3.4.2

3.3. Convolutional Neural Networks

In the last years, deep learning architectures have beat the state-of-the-art results in classification problems. Different neural architectures, usually composed by a high number of layers have been proposed to this aim. In this way, CNN has become one of the most popular models due to the classification performance provided in the image analysis field (Baldi, 2011, Krizhevsky et al., 2012, Sabour et al., 2017, Ortiz et al., 2018). CNNs are bioinspired by the convolutional response of neurons whose combination in different layers allow to extract features in an increasing abstraction levels. The same principle can be applied to time series data to learn representative or discriminative features since neighbourhood samples are usually related and patterns can be found from that relationship.

Convolutional layers perform the convolution operation of their input $\mathbf{x}_{i-1}$ with a set of K filters $\mathbf{w}_i$. This way, the k^{th} convolution term for the k^{th} filter is

$$\mathbf{w}_{ik} * \mathbf{x}_{i-1} = \sum_{l=0}^{N-1} \left[\mathbf{w}_{ik}(N-l) \cdot \mathbf{x}_{i-1}(j+l) \right] \quad (1)$$

In CNNs, it is usual to reduce the size of the feature maps in subsequent layers. This downsampling effect is addressed by the so-called *Pooling*, which consist in computing the maximum or average value in a window (MaxPooling and MeanPooling, respectively). In our case, the feature reduction is performed by MaxPooling layers.

3.3.1. Batch Normalization

Batch normalization (BN) (Ioffe and Szegedy, 2015) layers aims to increase the stability of the network. This layer standarizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation for each feature. Let $\mathcal{B} = \{y_{1,\dots m}\}$ a batch containing the activation output for m samples. Then, the batch mean and batch variance can be respectively defined as

$$\mu_{\mathcal{B}} = \frac{1}{m} \sum_{i=1}^{m} y_i, \quad \sigma_{\mathcal{B}}^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \quad (2)$$

and samples in the batch $\mathcal{B}$ can be normalized as

$$\widehat{y_i} = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \tag{3}$$

where ϵ is a constant added to the batch variance for numerical stability Ioffe and Szegedy (2015).

Afterwards, the normalized activation $\widehat{y_i}$ is shifted and scaled as:

$$z_i = \gamma \widehat{y_i} + \beta \tag{4}$$

where (β, γ) are two parameters to be learnt.

Thus, BN adds noise to the activation values, producing a regularization effect that reduces the layer inter-dependence. As a consequence, BN reduces the overfitting and improves the generalization ability of the autoencoder.

3.4. Feature extraction by Deep Convolutional Autoencoder

Autoencoders (Bengio, 2009) are neural architectures which aim to compute a compact (low-dimensional) representation of high-dimensional samples through unsupervised learning. They are usually presented as symmetric structures composed of some layers with a bottleneck at the center. Thus, while the first part, also called encoder, is fed with the input data samples x, the second part, also called the decoder uses the output of the decoder as input. The entire architecture is trained by minimizing the reconstruction error between the output and the input. In other words, the autoencoder learns the best representation of the input data samples in a low-dimensional, latent space. Hence, an autoencoder can be seen as a combination of 1) an encoding function $e : \mathbb{R}^n \rightarrow \mathbb{R}^d$ that compresses the input data samples x into the latent space $z \in \mathbb{R}^d$ and 2) a decoding function $d : \mathbb{R}^d \rightarrow \mathbb{R}^n$ that reconstruct the original samples from their compressed version while minimizing the reconstruction error.

Autoencoders have traditionally been composed of fully connected layers, similarly that in multilayer perceptron networks. These architectures perform reasonably well for many applications. However, applications requiring the processing of data manifolds with complex structures and distributions, usually require to compute more complex features corresponding to a higher abstraction level. This can be addressed by constructing autoencoders containing convolutional layers, providing two main advantages. On the one hand, convolutional layers allow us computing features taking into account neighbourhood signal values. On the

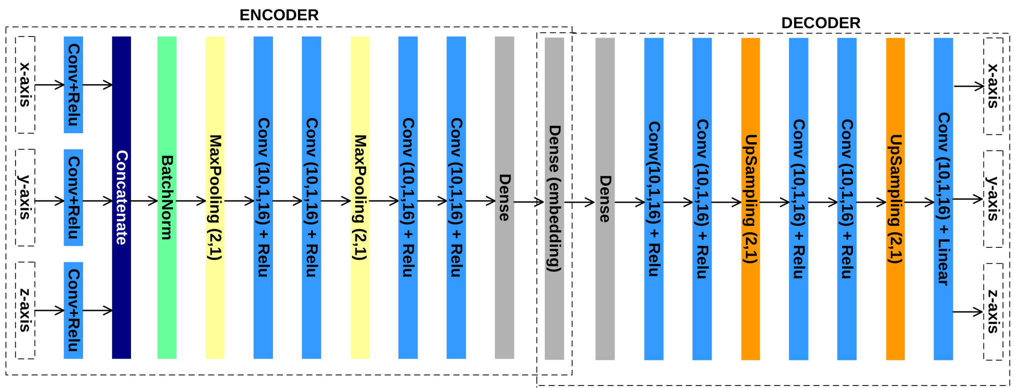


Figure 4: Deep Convolutional Autoencoder architecture used to compute a reduced set of features from the intertial signals.

other hand, the use of deep architectures provides the arena to compute features in a higher abstraction level.

Unlike methods referred in Section 2 in which a set of predefined temporal or spectral features are computed, we propose the use of CAE to automatically learn representative features from the raw inertial signals by unsupervised learning. The specific CAE architecture used in this work is shown in Figure 4. The encoder section input is composed of time series data from the three acceleration axes (x,y,z) separately, and separate convolutions are also performed at the first layer. The result of the individual convolutions is then concatenated to fed the subsequent encoder layer. After each convolution operation, an activation function that implements the non-linear behaviour of the layer is performed. In our case, the Rectified Linear Unit activation (ReLU) (LeCun et al., 2015) is applied in all layers but in the last one, where linear activation used. ReLU is a non-saturating activation function defined as $g(x) = max(0, x)$ that avoids negative activation values. On the contrary, the linear activation is simply the identity function $g(x) = x$. Subsequently, a *MaxPooling* operation with 2 samples filter and stride of 2 samples, downsamples the convolution result by a factor of 2. Overall, the encoder section is composed of four convolutional layers, two MaxPoolings using a stride of 2 (in order to reduce in a half the sample) as well as a batch normalization layer. The output of the encoder is a *Dense* layer to embed the output of the last convolutional layer in fewer dimensions. The decoder stage starts at the *Dense* layer and uses a structure composed of five convolutional layers to transform back into the original space. In this stage, upsampling layers counteract the effect of max-pooling layers used in the encoder stage Goodfellow et al. (2016).

Additionally, Figure 4 shows the use of the Scaled exponential Linear Unit (SeLU) (Klambauer et al., 2017), that is defined as:

$$SeLU(x) = \lambda \begin{cases} \alpha(exp(x) - 1) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \tag{5}$$

SeLU activation has advantages over the ReLU activation, due to its self-normalizing properties that tend to produce standarized activations. This improves the robustness of the learning process, especially in deep networks.

Moreover, the mean squared error is used as loss function to minimize the reconstruction error, along with the *RMSprop* optimizer

$$\mathcal{L} = \frac{1}{N} \sum_i (\mathbf{x}_i - \widetilde{\mathbf{x}}_i)^2 \tag{6}$$

where $\mathbf{x}_i$ is the i^{th} sample of the dataset and $\widetilde{\mathbf{x}_i}$ is its reconstructed version, obtained at the output of the CAE.

In order to visually assessing the discriminative properties of the extracted features, we used the T-SNE algorithm (van der Maaten and Hinton, 2008) to obtain a 2D view of each segment and the class distribution. Thus, Figure 5 shows the 2D representation of the features, labelled by different clinical criteria: Figures 5a, 5b and 5c show a 2D of the features for the PAL, 30s-CST and total distance travelled during the 6MWT, respectively.

3.4.1. Feature classification by Support Vector Classifier

Once the CAE is trained, the encoder Section is used to compute k features $(f_{1,\ldots k})$ from the inertial signals. Then, these features are classified using a linear Support Vector Classifier (SVC). A linear classifier can be defined as

$$g(f_i) = \mathbf{W}^\top f_i + b \tag{7}$$

The training of the classifier consists on calculating $\mathbf{W}$ and b, to define the best separating hyperplane as $f_i \mathbf{W}^\top + b = 0$. The computation of the hyperplane can be formulated in different ways. In Support Vector Machines, the hyperplane is computed by maximizing the margin $\frac{2}{\|\mathbf{W}\|}$ to the hyperplane as

$$\max_{\mathbf{W}} \frac{2}{\|\mathbf{W}\|} \quad subject\, to\, \mathbf{W}^\top f_i + b = \begin{cases} \geq +1, & \text{if } l_i = +1 \\ \leq -1, & \text{if } l_i = -1 \end{cases} \tag{8}$$

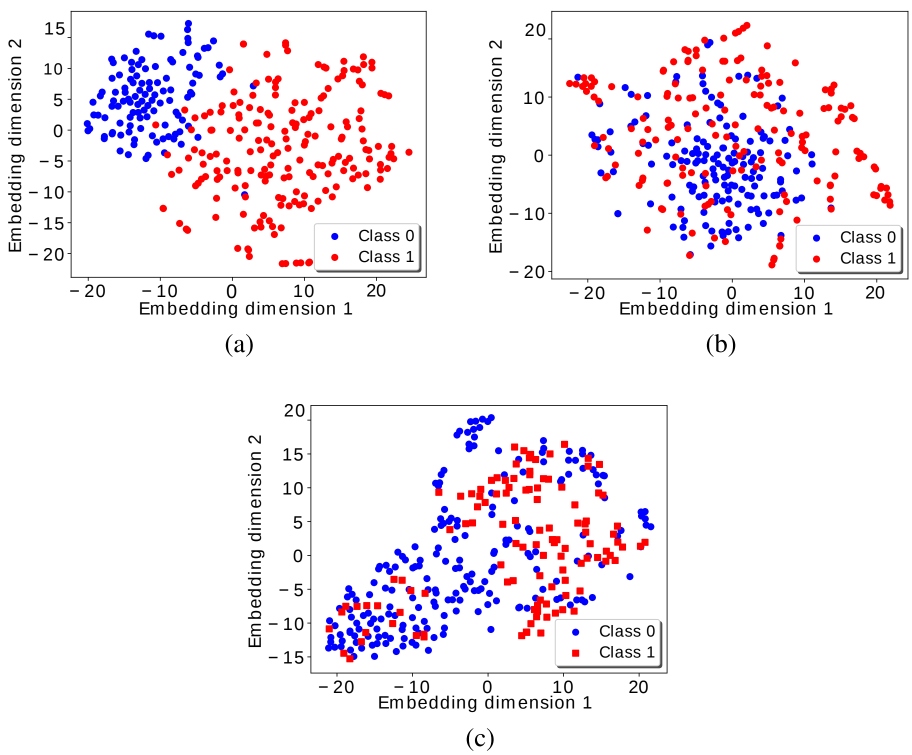


Figure 5: Embedings of the feature space obtained using the Deep Autoencoder. Different class distributions are considered according to different clinical labelling criteria (a) distance, (b) 30-s CST, (c) PAL

3.4.2. Combining segments in CAE-SVM classifier

As explained in Section 3.2.2, the autoencoder is trained using single segments extracted from the inertial signals (x, y and z axes) of each subject. Since segments extracted from subjects with similar condition (i.e. the same clinical label) are expected to be similar (i.e. similar features can be extracted from them), we used the same classification model. However, this method considerably augmented the available data, reducing the model overfitting. Otherwise, a combination of classifiers is usually considered to be more accurate and robust than the individual - also called *weak* - classifiers (Kittler et al., 1998, Liu et al., 2012). Hence, classification results based on single segments from a subject are combined using a majority voting strategy aiming to improve the classification performance and stability. Let $o_k, \{k = 1, ...K\}$ be the prediction based only on the $k - th$ segment and $C_j, \mathrm{J} = 0, 1$ the label assigned to $j - th$ class. These predictions can be combined by computing the number of segments N that predict the class j (Kim et al., 2003):

$$N_j = \#\{k|o_k = Cj\} \quad k = \{1, ..., K\} \tag{9}$$

and the combined prediction O can be computed as:

$$O = \underset{j}{\mathrm{argmax}}\, N_j \tag{10}$$

3.5. Classification using a Self-Normalizing Convolutional Neural Network

In this Section, we show the use of the CNNs shown in Figure 6 to classify the inertial signals. The overall method is the same as shown in Figure 3, but in this case, feature extraction and classification is performed by a single CNN, without the use of a support vector classifier.

Architecture in Figure 6 uses similar hyperparameters than the ones in the case of the autoencoder (Figure 4, determined by experimentation. Thus, convolutions using a kernel of 10x1 data points and stride of 1 point are used in all the cases. Moreover, MaxPooling layers reduce the size of the data at subsequent layers by a factor of 2. Besides, dense layers composing the fully connected stage contain 256, 256 and 2 units. The last one, corresponding to the number of classes. SeLU activation function is used in all cases but in the last layer which uses *softmax* activation. Finally, AlphaDropout (p=0.5) is used to regularize the network in order to improve the generalization ability.

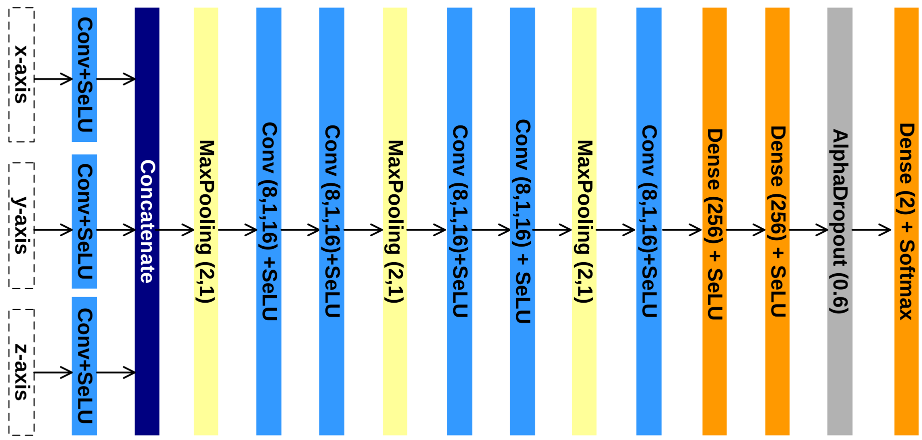


Figure 6: Architecture of the Convolutional Network used to in this Section

3.5.1. Combining segments in CNN classifier

In the case of CNN-based classifier, two output neurons are present in the last layer. Since a softmax function activates this layer, the prediction is represented by two values (the activation of each output neuron) corresponding to the probability of a sample to belong to a specific class. Indeed, a majority voting strategy is used to combine the predictions obtained for all the segments from a subject.

The score of the combination of different predictions corresponding to all segments from a subject is performed as

$$S = \sum_{i=1}^{K_0} s_i^0 - \sum_{j=1}^{K_1} s_j^1 \quad i \neq j \tag{11}$$

where K_0 and K_1 are the number of segments classified as class 0 and class1, respectively, and s_i^0 and s_i^0 are the scores obtained for the segments corresponding to class 0, and class 1, respectively.

4. Results and discussion

In this Section, we present the results obtained using the architectures described in previous Sections, namely CAE and CNN Deep Learning architectures, as well as the comparison to the EMD-based method proposed in Galán-Mercant et al. (2018). All the implementations have been developed in Python, and in the case of deep learning architectures, we used Tensorflow (Abadi et al., 2015) and

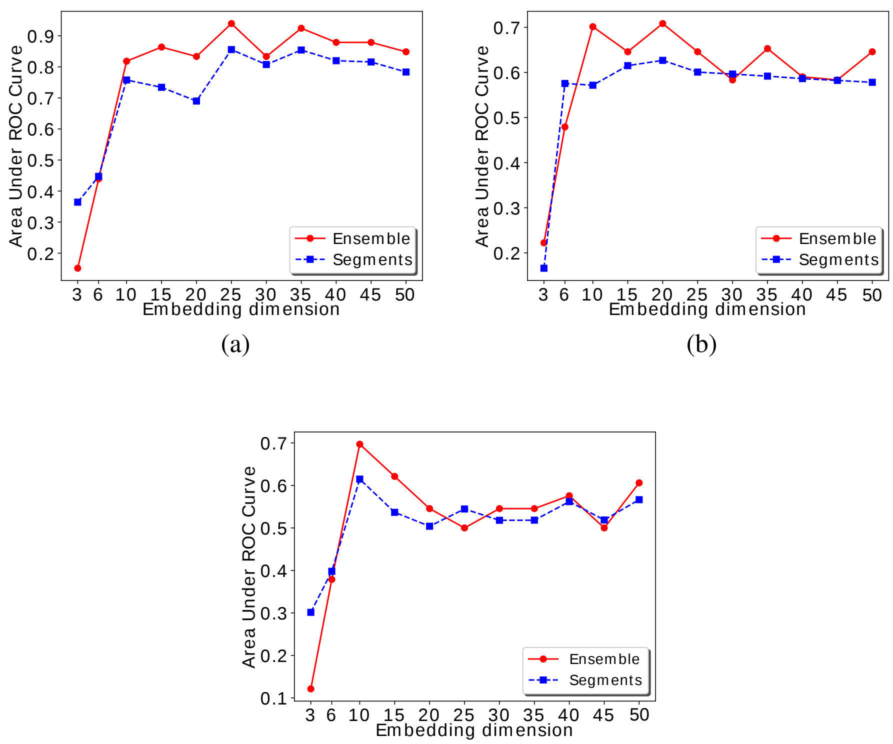
Keras (Chollet et al., 2015) along with the python Application Programming Interface. Classification performance is evaluated by means of the area under the Receiver Operating Curve (ROC), which provides a measure of the trade-off between sensitivity and specificity. In addition, ROC curves are also computed. All the experiments were carried out by cross-validation, specifically leave-one-out cross-validation. Cross-validation is of crucial importance in classification experiments to prevent double-dipping. Thus, in this work, the performance of the proposed methods have been assessed by Leave-One-Out to determine its generalization ability, taking special care of not using the test data in any of the previous training steps. Cross-validation is a general method to estimate the generalization error. It is worth noting that unfortunately, there is no way to determine the optimal network architecture. This requires performing many simulations to determine the optimal architecture (structure, type of layers and hyperparameters). This has been done by splitting the training data into train and validation subsets: the network is trained with the train subset whereas the performance metric to tune the hyperparameters is determined by the validation subset.

Figure 8 shows the classification performance in terms of the Area Under the Curve (AUC) and its dependence on the dimension of the embedding space. Moreover, this figure shows the ensemble-based method outperforms the classification results using single segments and can be used to determine the optimal embedding dimension. Hence, an embedding dimension of 25 has been chosen for classification using the distance travelled during 6MWT and 30-s CST, while the best results for PAL are obtained for a dimension of 20.

As shown in Table 2 the classifier based on CNN performs better than the CAE-SVM approach. However, the features obtained by CAE are more easy to interpret and pave the way to measure quantitative differences between subjects. Furthermore, the projection on a low dimensional space (i.e. 2D or 3D) of the inertial signals allows to measure the evolution of different clinical variables as a result of an intervention program.

Table 2: Classification results

Criteria	**EMD**				**CAE**				**CNN**			
	Acc	Sens	Spec	AUC	Acc	Sens	Spec	AUC	Acc	Sens	Spec	AUC
$Distance$	1.00	1.00	1.00	1.00	0.94	0.96	0.95	1.00	0.99	0.98	0.99	1.00
$30S-CST$	0.52	0.57	0.50	0.51	0.70	0.75	0.66	0.68	0.70	0.76	0.75	0.77
PAL	0.70	0.57	0.80	0.74	0.73	0.75	0.69	0.70	0.88	0.90	0.83	0.80



(a) (b)

(c)

Figure 7: Area Under ROC curve obtained according to different clinical labelling criteria (a) distance travelled during 6MWT, (b) 30-s CST, (c) PAL for single segments and by majority voting ensemble.

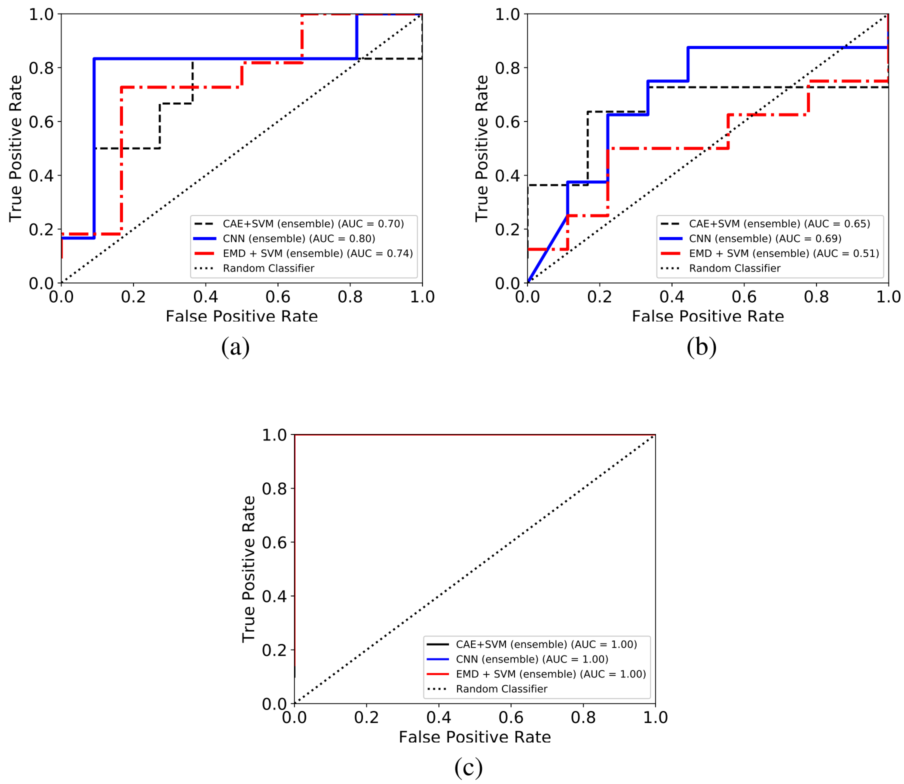


Figure 8: Comparison of ROC curves obtained by different classification methods according to different clinical labelling criteria (a) PAL, (b) 30-s CST, (c) distance travelled during 6MWT, for single segments and by majority voting ensemble.

4.0.1. Exploring CNN features

Exploring the network insights once it is trained can provide information regarding the samples the network is considering to extract discriminative features. Thus, tools directed to reveal the parts of the signal the network is focusing on, break with the black box view of the Deep learning architectures which are usually though as high accuracy classifiers. Analyzing the network activation on a specific input is a way to take a step beyond the classification results. One usual way to reveal the features computed during the training stage consist of exploring the activation that an input belonging to a specific class produces in different parts of the network. The most frequent procedure in CNNs is to analyze the raw features at the input layer that activate specific neurons at the output layer, which helps to understand the information the network is using to classify samples (i.e. to detect differences between subjects from different classes). This is addressed here by computing the *saliency* maps (Simonyan et al., 2013). Saliency maps are a representation of changes in the network output concerning small changes in the input, being able to highlight those regions of the image that play a more important role in the output. In classical CNN for classification, saliency is obtained by computing the gradient of an output category for the input. Moreover, the Class Activation Maps (CAM) Zhou et al. (2016), also aim to localize class-specific regions in the input signal that contain relevant discriminative information. Instead of using gradients concerning output as in saliency maps, CAM projects back the weights of the output layer on to the convolutional feature maps.

Figures 9a, 9b and 9c shows the mean CAM according to the corresponding clinical criteria for non-risk subjects. At the same time, Figures 9d, 9e and 9f show the mean CAM for risk subjects. It is worth noting that a segment on the center of the 6MWT is used to generate the activations. These maps, figure out the relative importance of different parts of a segment, have been computed by composing a 2D image containing the CAM values corresponding to each signal sample. Then, the resultant acceleration is overlapped.

CAM maps aims to show the parts of the signal where the neural network is extracting features for a specific class. Thus, CAM maps shown in Figure **??** depict the samples being used for classification by the neural network. As shown in these figures, different time samples are used for classification depending on the label being predicted, which suggest that information related to the clinical status determined by different labels are present at different time instants. Furthermore, the relative importance is shown using different colors according to the colorbar

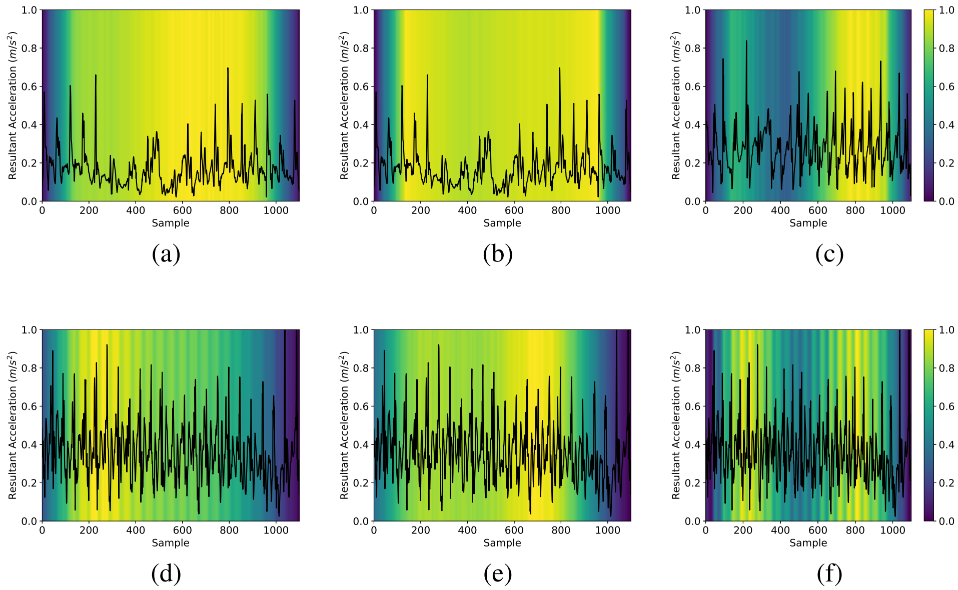


Figure 9: Mean Class Activation Maps obtained for (a,d) distance travelled during 6MWT, (b,e) 30-s CST, (c,f) PAL for non-risk (a,b,c) and risk (d,e,f) subjects. Colorbar indicates the relative importance of different regions in a normalized scale [0,1]

shown in this figure.

5. Conclusions and future directions

In this work, we present two methods to predict the physical activity and functional fitness levels through inertial data acquired through a simple method based on a smartphone during the 6MWT.

In previous works, temporal, spectral or spatio-temporal features are extracted from the IMU signals, using methods based on Fourier analysis, or Wavelet Analysis. However, the classical stationarity or periodicity assumptions needed for the validity of these methods are not always met. As an alternative to avoid these assumptions, EMD method is used in (Galán-Mercant et al., 2018) to decompose the acceleration signal into band-limited components (Intrinsic mode functions) to subsequently extract statistical time and frequency descriptors. Nevertheless, the calculation of a priori known statistical features is still needed. In this paper, we propose an unsupervised feature extraction method based on CAE that avoids the classical assumptions of Fourier or Wavelet Analysis, i.e., the signal stationarity or the use of a predefined basis of functions. Unlike these decomposition methods, CAE learns the best representation of the input in a lower dimensional space utilizing the minimization of the reconstruction error. The discriminative ability of the extracted features has been assessed using a support vector classifier. Additionally, a second Deep Learning model also based on a CNN is used to directly perform the classification of the subjects according to different clinical criteria. These two methods are complementary. The CAE method extract features that allow representing the samples in a lower dimensional space, and that information can be used to assess the evolution of a clinical variable such as PAL when a subject is under some clinical intervention. In future work, we plan to address the quantitative measure of clinical variables through the distance between samples in the embedding space. Complementarily, the CNN classifier aimed to directly classify the subjects dealing with the prediction of a clinical label employing patterns found in the intertial signals.

The proposed method outperforms previous approaches based on features extracted using classical signal processing methods, and in the future, similar architectures could be used to predict other clinical labels. In that way, the application of this kind of predictive methods supposes an important advance for future clinical assessment and monitoring. It could be implemented in m-health systems to automatically identify patients profiles, making the e-health management easier and also reducing the cost of health services.

Acknowledments

This work was partly supported by the MINECO/FEDER under TEC2015-64718-R, PSI2015-65848-R and PGC2018-098813-B-C32 projects; and the Erasmus+ Strategic Partnership for Higher Education Programme (Key Action 203) [Grant number: 2018-1-PL01-KA203-051055]. Furthermore, the mobility grant EST2018-090 was supported by the University of Cádiz. We gratefully acknowledge the support of NVIDIA Corporation with the donation of one of the GPUs used for this research.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. URL: `http://tensorflow.org/`. software available from tensorflow.org.

Aggarwal, J., Ryoo, M., 2011. Human activity analysis: A review. ACM Comput. Surv. 43, 16:1–16:43. doi:`10.1145/1922649.1922653`.

American College of Sports Medicine, 2013. ACSM's health-related physical fitness assessment manual. Lippincott Williams & Wilkins.

ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories, 2002. Ats statement: guidelines for the six-minute walk test.

Ayachi, F.S., Nguyen, H.P., Lavigne-Pelletier, C., Goubault, E., Boissy, P., Duval, C., 2016. Wavelet-based algorithm for auto-detection of daily living activities of older adults captured by multiple inertial measurement units (imus). Physiological measurement 37, 442.

Baldi, P., 2011. Autoencoders, unsupervised learning and deep architectures, in: Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning Workshop - Volume 27, JMLR.org. pp. 37–50.

Banos, O., Moral-Munoz, J.A., Diaz-Reyes, I., Arroyo-Morales, M., Damas, M., Herrera-Viedma, E., Hong, C.S., Lee, S., Pomares, H., Rojas, I., Villalonga, C., 2015. mDurance: A Novel Mobile Health System to Support Trunk Endurance Assessment. Sensors 15, 13159–13183. doi:10.3390/s150613159.

Bengio, Y., 2009. Learning deep architectures for ai. Found. Trends Mach. Learn. 2, 1–127. doi:10.1561/2200000006.

Bullo, V., Gobbo, S., Vendramin, B., Duregon, F., Cugusi, L., Di Blasio, A., Bocalini, D.S., Zaccaria, M., Bergamin, M., Ermolao, A., 2018. Nordic walking can be incorporated in the exercise prescription to increase aerobic capacity, strength, and quality of life for elderly: A systematic review and meta-analysis. Rejuvenation research 21, 141–161.

Cesari, M., Prince, M., Thiyagarajan, J.A., De Carvalho, I.A., Bernabei, R., Chan, P., Gutierrez-Robledo, L.M., Michel, J.P., Morley, J.E., Ong, P., et al., 2016. Frailty: an emerging public health priority. Journal of the American Medical Directors Association 17, 188–192.

Cho, H., Yoon, S.M., 2018. Divide and conquer-based 1d cnn human activity recognition using test data sharpening. Sensors 18.

Choi, J., Ahn, A., Kim, S., Won, C.W., 2015. Global prevalence of physical frailty by fried's criteria in community-dwelling elderly with national population-based surveys. Journal of the American Medical Directors Association 16, 548–550.

Chollet, F., et al., 2015. Keras. https://github.com/fchollet/keras.

Clegg, A., Young, J., Iliffe, S., Rikkert, M.O., Rockwood, K., 2013. Frailty in elderly people. The Lancet 381, 752–762.

Cruz-Jentoft, A.J., Kiesswetter, E., Drey, M., Sieber, C.C., 2017. Nutrition, frailty, and sarcopenia. Aging clinical and experimental research 29, 43–48.

Curtis, E.M., Moon, R.J., Harvey, N.C., Cooper, C., 2017. The impact of fragility fracture and approaches to osteoporosis risk assessment worldwide. Bone 104, 29–38.

Drover, D., Howcroft, J., Kofman, J., Lemaire, E., 2017. Faller Classification in Older Adults Using Wearable Sensors Based on Turn and Straight-Walking Accelerometer-Based Features. Sensors 17, 1321. doi:10.3390/s17061321.

Duncan, M.J., Wunderlich, K., Zhao, Y., Faulkner, G., 2018. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. Journal of Sports Sciences 36, 1695–1704. doi:10.1080/02640414.2017.1409855.

Gaikwad, B.D., Bobade, K.A., Jadhav, P.L., Deokate, K.B., 2016. Human Mobility Change of State Detection Using a Smartphone based on Accelerometer Sensor. International Journal of Engineering Science and Computing 6, 4569–4571. doi:10.4010/2016.1046.

Galán-Mercant, A., Barón-López, F.J., Labajos-Manzanares, M.T., Cuesta-Vargas, A.I., 2014. Reliability and criterion-related validity with a smartphone used in timed-up-and-go test. Biomedical engineering online 13, 156.

Galán-Mercant, A., Cuesta-Vargas, A.I., 2014. Differences in trunk accelerometry between frail and non-frail elderly persons in functional tasks. BMC research notes 7, 100.

Galán-Mercant, A., Moral-Muñoz, J., Ortiz, A., Herrera-Viedma, E., Tomás, M.T., 2018. Predicting physical activity and functional fitness levels through inertial signals and emd-based features in older adults, in: Fujita, H., Herrera-Viedma, E. (Eds.), New Trends in Intelligent Software Methodologies, Tools and Techniques. IOP Press Ebooks, pp. 954–966.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

He, Z., 2010. Activity recognition from accelerometer signals based on wavelet-ar model, in: 2010 IEEE International Conference on Progress in Informatics and Computing, pp. 499–502. doi:10.1109/PIC.2010.5687572.

Hong, W., Cheng, Q., Zhu, X., Zhu, H., Li, H., Zhang, X., Zheng, S., Du, Y., Tang, W., Xue, S., et al., 2015. Prevalence of sarcopenia and its relationship with sites of fragility fractures in elderly chinese men and women. PLoS One 10, e0138102.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, JMLR.org. pp. 448–456.

Karinkanta, S., Kannus, P., Uusi-Rasi, K., Heinonen, A., Sievänen, H., 2015. Combined resistance and balance-jumping exercise reduces older women's injurious falls and fractures: 5-year follow-up study. Age and ageing 44, 784–789.

Kim, H.C., Pang, S., Je, H.M., Kim, D., Bang, S.Y., 2003. Constructing support vector machine ensemble. Pattern Recognition 36, 2757–2767.

Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., 1998. On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. 20, 226–239. doi:10.1109/34.667881.

Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S., 2017. Self-normalizing neural networks, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 971–980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Curran Associates Inc., USA. pp. 1097–1105.

Lagerros, Y.T., Hantikainen, E., Michaëlsson, K., Ye, W., Adami, H.O., Bellocco, R., 2017. Physical activity and the risk of hip fracture in the elderly: a prospective cohort study. European journal of epidemiology 32, 983–991.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521. doi:10.1038/nature14539.

Lima, C.A., de Andrade, A.d.F.D., Campos, S.L., Brandão, D.C., Mourato, I.P., de Britto, M.C.A., 2018. Six-minute walk test as a determinant of the functional capacity of children and adolescents with cystic fibrosis: A systematic review. Respiratory medicine 137, 83–88.

Liu, M., Zhang, D., Shen, D., 2012. Ensemble sparse classification of alzheimer's disease. NeuroImage 60, 1106 – 1116. doi:https://doi.org/10.1016/j.neuroimage.2012.01.055.

Lockhart, T., Songra, R., Zhang, J., Wu, X., 2013. Wavelet based automated postural event detection and activity classification with single imu. Biomedical sciences instrumentation 49, 224–233.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of machine learning research 9, 2579–2605.

Moral-Munoz, J.A., Esteban-Moreno, B., Herrera-Viedma, E., Cobo, M.J., Pérez, I.J., 2018. Smartphone Applications to Perform Body Balance Assessment: a Standardized Review. Journal of Medical Systems 42, 119. doi:10.1007/s10916-018-0970-1.

Murphy, S.L., 2009. Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct. Preventive medicine 48, 108–114.

Ortiz, A., Munilla, J., Górriz, J.M., Ramírez, J., 2016. Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease. International Journal of Neural Systems 26, 1650025.

Ortiz, A., Murcia, F.J.M., Munilla, J., Górriz, J.M., Ramírez, J., 2018. Label aided deep ranking for the automatic diagnosis of parkinsonian syndromes. Neurocomputing doi:https://doi.org/10.1016/j.neucom.2018.10.074.

Rikli, R.E., Jones, C.J., 2013. Senior fitness test manual. Human Kinetics.

Ross, W., Hebbelinck, M., Faulkner, R., 1978. Kinanthropometry terminology and landmarks. Physical Fitness Assessment. Principles, Practice and Application. Springfield, Illinois: Charles C. Thomas , 44–50.

Sabour, S., Frosst, N., Hinton, G.E., 2017. Dynamic routing between capsules, in: NIPS.

Salazar, A., de Sola, H., Failde, I., Moral-Munoz, J.A., 2018. Measuring the quality of mobile apps for the management of pain: Systematic search and evaluation using the mobile app rating scale. JMIR mHealth and uHealth 6, e10718.

Similä, H., Immonen, M., Ermes, M., 2017. Accelerometry-based assessment and detection of early signs of balance deficits. Computers in Biology and Medicine 85, 25–32. doi:10.1016/J.COMPBIOMED.2017.04.009.

Simonyan, K., Vedaldi, A., Zisserman, A., 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 .

Troosters, T., Gosselink, R., Decramer, M., 1999. Six minute walking distance in healthy elderly subjects. European Respiratory Journal 14, 270–274.

Vervoort, D., Vuillerme, N., Kosse, N., Hortobágyi, T., Lamoth, C.J.C., 2016. Multivariate Analyses and Classification of Inertial Sensor Data to Identify Aging Effects on the Timed-Up-and-Go Test. PLOS ONE 11, e0155984. doi:10.1371/journal.pone.0155984.

Vishwakarma, D., Rawat, P., Kapoor, R., 2015. Human activity recognition using gabor wavelet transform and ridgelet transform. Procedia Computer Science 57, 630 – 636. doi:https://doi.org/10.1016/j.procs.2015.07.425. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L., 2018. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters .

Yang, M., Zheng, H., Wang, H., Mcclean, S., Newell, D., 2012. igait: An interactive accelerometer based gait analysis system. Comput. Methods Prog. Biomed. 108, 715–723.

Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P., Zhang, J., 2014. Convolutional neural networks for human activity recognition using mobile sensors, in: 6th International Conference on Mobile Computing, Applications and Services, pp. 197–205. doi:10.4108/icst.mobicase.2014.257786.

Zhang, Q., Yang, L.T., Chen, Z., Li, P., 2018. A survey on deep learning for big data. Information Fusion 42, 146 – 157. doi:https://doi.org/10.1016/j.inffus.2017.10.006.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929. doi:10.1109/CVPR.2016.319.