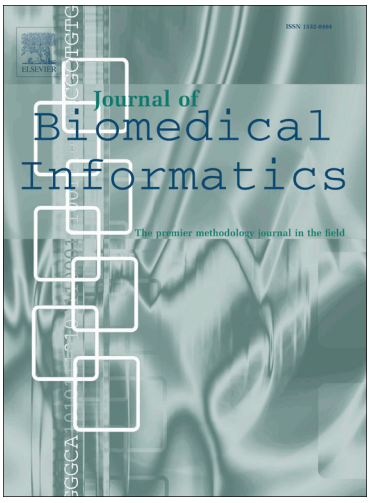
Journal Pre-proofs



Alzheimer's Disease Stage Identification Using Deep Learning Models

Santos Bringas, Sergio Salomón, Rafael Duque, Carmen Lage, José Luis Montaña

PII: S1532-0464(20)30142-8
DOI: https://doi.org/10.1016/j.jbi.2020.103514
Reference: YJBIN 103514

To appear in: *Journal of Biomedical Informatics*

Received Date: 28 February 2020
Revised Date: 15 July 2020
Accepted Date: 16 July 2020

Please cite this article as: Bringas, S., Salomón, S., Duque, R., Lage, C., Montaña, J.L., Alzheimer's Disease Stage Identification Using Deep Learning Models, *Journal of Biomedical Informatics* (2020), doi: https://doi.org/10.1016/j.jbi.2020.103514

Alzheimer's Disease Stage Identification Using Deep Learning Models

Santos Bringas[a,*], Sergio Salomón[b], Rafael Duque[c], Carmen Lage[d], José Luis Montaña[c]

[a] *Fundación Centro Tecnológico de Componentes CTC, 39011 Santander, Spain*
[b] *Axpe Consulting Cantabria SL, 39600 Camargo, Spain*
[c] *Department of Mathematics, Statistics and Computer Science, Universidad de Cantabria, 39005 Santander, Spain*
[d] *Cognitive Disorders Unit, Department of Neurology, Marqués de Valdecilla University Hospital (HUMV), Valdecilla Biomedical Research Institute (IDIVAL), 39008 Santander, Spain*

Abstract

Objective: The aim of this research is to identify the stage of Alzheimer's Disease (AD) patients through the use of mobility data and deep learning models. This process facilitates the monitoring of the disease and allows actions to be taken in order to provide the optimal treatment and the prevention of complications.
Materials and methods: We employed data from 35 patients with AD collected by smartphones for a week in a daycare center. The data sequences of each patient recorded the accelerometer changes while daily activities were performed and they were labeled with the stage of the disease (early, middle or late). Our methodology processes these time series and uses a Convolutional Neural Network (CNN) model to recognize the patterns that identify each stage.
Results: The CNN-based method achieved a 90.91 % accuracy and an F1-score of 0.897, greatly improving the results obtained by the traditional feature-based classifiers.
Discussion and conclusion: In our research, we show that mobility data can be a valuable resource for the treatment of patients with AD as well as to study the progress of the disease. The use of our CNN-based method improves the accuracy of the identification of AD stages in comparison to common supervised learning models.

*Corresponding author
Email addresses: sbringas@centrotecnologicoctc.com (Santos Bringas), ssalomong@axpecantabria.com (Sergio Salomón), rafael.duque@unican.es (Rafael Duque), clage@idival.org (Carmen Lage), joseluis.montana@unican.es (José Luis Montaña)

Keywords: Convolutional Neural Network, Alzheimer's disease, Accelerometer, Deep learning.

1. Introduction

In our progressively aged societies, dementia has become a major public health priority. In 2015, it was estimated that almost 47 million people worldwide were suffering from dementia, with global associated costs of more than $818 billion [1]. Given that increasing age represents the first risk factor for dementia, the number of people affected is estimated to reach 151 million by 2050 [1]. Dementia is characterized by progressive cognitive impairment and disability which implies an increasing need of supervision and assistance for daily life activities, entailing a huge social and economic burden in addition to the dramatic impact on the person and on their families. Alzheimer's disease (AD) constitutes the leading cause of dementia, implicated in more than 80% of cases. Over the last few decades, considerable progress has been made towards a greater understanding of the biological mechanisms implicated in the generation of the disease, which has made it possible to know that the pathological brain changes begin up to 15-20 years before the onset of symptoms, a period termed "preclinical AD" [2]. After clinical onset, AD presents a progressive course which is divided into different stages, clinically defined depending on the cognitive status of the patient and especially on their degree of disability. In clinical practice, specific interviewing tools, such as the Global Deterioration Scale [3] or the Clinical Dementia Rating [4], are used in periodic medical assessments to track the progression of the disease. However, a way of detecting a clinical worsening between medical appointments would be desirable, in order to evaluate the need for treatment adjustments or to take preventive measures, such as increasing surveillance or promoting engagement in stimulating activities. Furthermore, AD has been associated with changes in motor activity patterns from the early stages. Since this can only be assessed through the collection of 24h-data, sensor-based wearable systems offer an accurate solution.

Deep learning methods are widely used to perform multidimensional data analysis, to recognize images and to classify time series [5]. These methods have been commonly applied to detect Alzheimer's disease (AD) from neuroimaging data [6]. The potential genetic biomarkers of AD have also been explored by these deep learning techniques [7]. Thus Zhang et al. [8] approach the diagnosis of AD with a multi-modal deep learning model that combines neuroimaging data

and the results of clinical and neuropsychological assessments. The proposal of Spasov et al. [9] highlights the usefulness of these deep learning architectures to prevent the evolution of mild cognitive impairment in patients who are at risk of AD. Large-scale gene expression and DNA methylation information data have been also used by deep learning-based models to predict AD [10] and so enabling the prevention of the usual undesired effects of the symptoms (falls, memory loss, etc.).

Everyday activities and gait analysis are other sources of information used by classical and deep learning techniques to detect AD and to study how physical activity can offer resistance to the pathology [11]. Convolutional Neural Networks (CNNs) provide effective support for gait recognition [12]. The effectiveness of CNNs has enabled performing gait recognition processes that use 3D images and identify the subject [13]. CNNs can approximate complex non-linear functions that perform an effective gait recognition, even if the subjects adjust their clothing or the viewing angle varies [14]. Our work offers an evolution of these proposals (neuroimaging and clinical analysis) with a method that observes the daily activities of patients with a smartphone accelerometer to study the relationship between their mobility patterns and the stage of AD.

The remainder of this paper is organized as follows. Section 2 reviews previous works on the analysis of the relation between mobility patterns of patients and Alzheimer's disease. Section 3 describes our CNN-based proposal to recognize the stage of Alzheimer's disease by the mobility of the patients. Section 4 shows the experimental results of applying this method. Section 5 discusses the results of the work. Conclusions are given in Section 6.

2. Related work

Recent advances in technology, especially in terms of availability and usability, are broadening healthcare to a more ubiquitous paradigm, complementing the traditional hospital-centered approach with the possibility of collecting a great amount of information from the user's daily living or making personalized interventions. Specially related to dementia, sensor-based devices have the advantage of not requiring any interaction effort from the patient and therefore its application at home offers a wide range of possibilities, including assistance in basic daily living activities (as medication reminder systems) or safety monitoring, as the detection of falls [15] or leaving-bed episodes during night time [16]. In this way, wearable sensors have been explored for many different purposes.

For instance, approaches based on tri-axial accelerometers have been proved to be useful not only for the detection of the occurrence of falls, which represent a major source of medical complications and disability in patients with dementia, but also for the estimation of risk of falls. In the work conducted by Gietzelt in a nursing home [17], data collected by a tri-axial accelerometer allowed to differentiate between patients with dementia who subsequently suffered falls from those who did not, raising the possibility of making estimations of individual fall risk. Similar results were found by Van Schooten [18] in a sample of 319 non-demented older people. Another interesting potential use of wearable accelerometers is the monitoring of behavioral disturbances. Several works have shown that actigraphy is able to identify apathy, which constitutes one of the most frequent neuropsychiatric symptoms of AD [19], [20], [21]. Additionally, Goerss et al. [22] evaluated patients with dementia in severe stages in two nursing homes with wearable sensors for 7 days. Collected data were used to calculate an accelerometric motion score (AMS), which was positively correlated to the total intensity of challenging behaviors, measured by a clinical scale. Furthermore, AMS was significantly correlated with specific abnormal behaviors as pacing, mannerisms or apathy. These results suggest that accelerometry can be helpful to make individual predictions of behavioral changes, as well as to evaluate the response to potential treatments. Since there are no current curative treatments for AD, sensor-based systems directed at supporting home care and safety aids constitute a relevant field of research. However, the investigation of novel therapies for AD is moving forward and the application of treatment capable of modifying the course of the disease must be accompanied by an accurate and early diagnosis; ideally, at the preclinical stage, when clinical symptoms have not yet appeared. In this sense, some studies have evaluated the clinical utility of accelerometry for diagnostic purposes through different approaches. Some authors have used accelerometric signals to make estimations of gait parameters and this enables to differentiate healthy controls from patients with dementia [23] and also from mild cognitive impairment (MCI), a condition that precedes dementia but does not necessarily evolve into it [24]. Additionally, recent work showed the utility of accelerometry to distinguish dementia subtypes, including AD, Dementia with Lewy Bodies and Parkinson's Disease dementia, identifying significant differences in 7 estimated gait features [25]. Changes in gait parameters have also been explored as a risk factor for conversion to dementia in MCI patients. Gillain et al. [26] evaluated MCI patients were evaluated with a tri-axial accelerometer and followed up over time. Parameters

such as gait speed, symmetry and regularity showed significantly lower values in those patients who subsequently developed dementia, suggesting that it could be considered as a marker of a worse prognosis. Apart from gait characteristics, changes in daily activity patterns are under actual research as a plausible AD biomarker and, in this sense, wearable devices offer the advantage of continuous and non-invasive monitoring. One of the first works to show this was the work of Kirste et al. [27], who evaluated everyday motion behavior in 23 AD patients and their cognitively unimpaired partners through the data obtained by ankle-mounted tri-axial accelerometers. Both groups could be differentiated with a classification accuracy of 91%. Interestingly, the scores in the Mini-Mental State Examination [28], a brief cognitive test frequently used to assess clinical progression in AD patients, were significantly correlated with motion features, which suggest a potential utility of this approach to identify the stage of the disease. Which exact features of daily motion behavior are characteristic of AD have not been elucidated yet, but intra-individual variability of physical activity was found to be significantly different between AD patients with mild-stage dementia and cognitively unimpaired elders in a similar work [29]. Furthermore, recent studies point to possible usefulness of accelerometry for preclinical AD detection. Circadian rhythm disturbances are a frequent finding in AD dementia and have also been evaluated with accelerometry [30], but they are not well known in the early stages of the disease. A recent work analyzed circadian rhythms using actigraphic data from 189 cognitively unimpaired participants who also underwent a study of AD biomarkers (cerebrospinal fluid analysis or amyloid PET), which allow the existence of preclinical AD to be defined [31]. Increased intradaily variability, a marker of rest-activity rhythm fragmentation, was associated with preclinical AD, independently of age or sex. Moreover, Li et al. [32], described a cohort of 1097 cognitively unimpaired elders who underwent an extensive cognitive study, including the evaluation of daily motor activity by a wearable wrist actigraph for 10 days, and then they were followed up over a period of up to 11 years to assess the conversion to dementia. The incidence of AD dementia was highly associated with degraded motor fractal regulation (FR), a number of mechanisms that regulate patterns of daily motor activity fluctuations (HR 1.31, 95%CI 1.15-1.49, $P < 0.0001$). This supposed a 1.8-fold higher risk of developing AD for those subjects who displayed the greatest degradation, a magnitude equivalent to being 5.2 years older. Additionally, more degraded fractal regulation was associated with a faster cognitive decline over time, suggesting that it can anticipate a rapid clinical progression. Interestingly,

these findings preceded the diagnosis of AD 4.6 years on average, which implies that perturbations in FR could be detected by actigraphy years before clinical onset. In summary, AD seems to be accompanied by changes in motor activity features, and accelerometry has proved to be a feasible approach to assess it. Some previous studies have shown how to take advantage of accelerometers to perform an energy expenditure analysis [33], even exploiting those included in smartphones[34]. Until now, accelerometry has been investigated as a way to detect the presence of AD, either in clinically affected subjects or in the preclinical state, but rarely to recognize different clinical stages, which would be of capital importance to identify clinical progression. Some previous works describe the classification of AD patients into three functional stages through the use of sensor-based devices [35] [36], but Convolutional Neural Networks have not been applied with Dementia staging purposes until now, having some reference examples that use similar systems to detect different diseases, like in El Maachi et al. [37] that propose a CNN-based method to detect Parkinson's disease analyzing gait data. Improving the results of the work of Nieto-Reyes et al. [35], Bringas et al. [38] gives a first approach to this problem using convolutional neural networks. In addition to introducing a preprocessing system for this data, the paper uses a fixed architecture and optimizes its parameters based solely on the accuracy, obtaining better and more balanced results than the previous methods used. On this basis, our work will improve the results obtained so far, testing different architectures of several Machine Learning models and making an exhaustive search of hyperparameters. Particular emphasis is placed on 1D-Convolutional Neural Networks, as they gave the best results of all the tested models.

3. Method

This section shows the data source used to fit a CNN model to identify AD stages. Moreover, the layers of the architecture of the CNN model are described. Finally, the process of training and evaluating the CNN model is specified.

3.1. Data source

The main objective of our proposal is to establish a methodological approach that allows us to recognize the stage of Alzheimer's disease from the data on patient mobility captured by the accelerometer. The methodology considers the use of the accelerometer smartphone to capture the mobility data. The aim is

avoiding the use of wearable sensors that can be more unfamiliar and intrusive for the patient. For this reason, the methodology should prevent problems arising from accidentally changing the orientation of the smartphone in the pocket of the patient.

The accelerometer sensor of the smartphone generates a data sequence for each patient. These data sequences register the acceleration changes in the three axes (X, Y and Z) over time. Our methodological proposal considers these three data features along the temporal dimension to predict the AD stage. Data of about 6 hours duration were obtained from a total of 35 patients with the three different AD stages: 7 early, 18 moderate and 10 severe. These patients moved freely, without any prior indication, so that no initial bias was introduced. Data sequences are preprocessed to divide them into shorter segments of the same length. This partition aims to obtain a larger number of data samples instead of having only a data sequence for each patient. The generation of samples of the same size also tries to homogenize periods between consecutive points. For this homogenization, the methodology considers a time resolution parameter and calculates the average value of all the data recorded under each period (for example, every 0.1 seconds), the same for all the samples. The high frequency of sampling of the accelerometer sensor generates a large amount of data to be processed and only considering the average value allows us to reduce this data size. Additionally, in order to homogenize the size of all samples, each one was extended with zeros to match the largest one.

The methodology proposes a supervised learning process to build a CNN that uses the accelerometer data to classify each patient according to the Alzheimer's disease stage. Therefore, it is necessary to know information about the AD stage of a group of patients in order to fit a CNN model. The information that will characterize the stage of the disease of each participant is based on the Global Deterioration Scale (GDS). This scale establishes the following seven different stages:

- GDS 1-2: No objective cognitive impairment.
- GDS 3: Mild cognitive impairment (MCI) but there is no significant functional impact.
- GDS 4: Mild dementia. Dementia stages implicate that cognitive impairment is severe enough to cause functional decline.
- GDS 5: Moderate dementia.

- GDS 6: Moderately severe dementia.
- GDS 7: Severe dementia.

The methodology characterizes the stage of the AD of each patient with one of the three following *labels*: (i) *early* (corresponding to GDS 2 and 3), (ii) *middle* (GDS 4 and 5) and (iii) *late* (GDS 6 and 7). In this way, a supervised learning process is fostered with two kinds of data sources: (i) accelerometer data and (ii) labels that specify the AD stage.

3.2. CNN-based approach

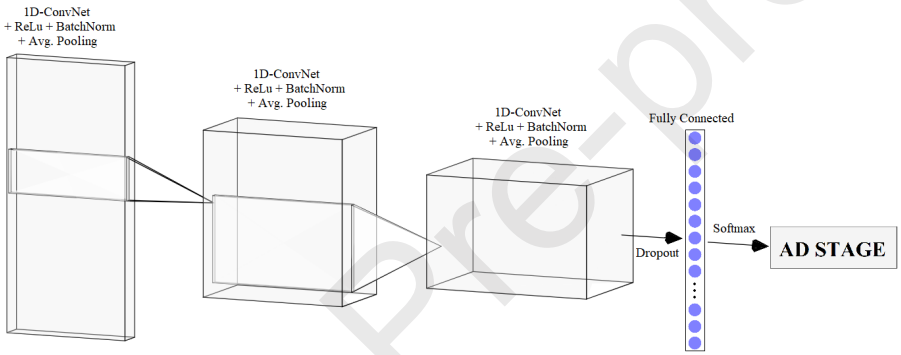


Figure 1: Overview of the architecture of the network

Convolutional Neural Networks are models that are specialized in the analysis of multidimensional data, such as time series, images, etc. They generate consecutive feature maps, obtaining simple characteristics of the data (e.g., vertices, edges in images) in the first layers and grouping them in more complex patterns (e.g., geometric shapes). These feature maps are obtained by applying convolutional operations with trainable kernels to the input of the layer. Non-linear transformations and pooling are complementary functions that help the network to converge. After that, a prediction is generated (usually by fully connected layers) based on the feature maps processed.

For this problem, we have used several convolutional layers with ReLU transformations and average-pooling. Moreover, in order to work with time-series, we have used a 1-dimensional architecture for the CNN. We detail these components in the next sections.

3.2.1. Convolutional layer

These are the core layers of this class of networks. The convolutional layer obtains its output by applying the convolutional operation with different trainable kernels to the entire input, using a sliding window method, to produce several feature maps containing different characteristics of the input. In this work, we have used 1-dimensional convolutions, since we only have the time dimension. For this case, the convolution operation is described as:

$$C(n) = (\sum_{i=1}^{k} I_{n+i} \cdot f(n)) + b \tag{1}$$

where I is the input channel, f the filter, k the size of the filter, and b the bias.

The convolutional operation is applied for every channel separately, learning the weights of the filters and the bias independently (e.g., if the first layer has 3 channels, the total number of parameters to adjust per filter will be $(3 \cdot n) + 1$). Additional parameters can alter the output of the layer, like the zero-padding, that extends the input with zeros, and stride, that controls how much the filter is moved after each application. In this case, we have used a stride of 1 and no padding. In order to define the next layers, it is useful to know the sizes of the inputs/outputs of all of the layers. In this case, the size of the result output is calculated with the following formula:

$$|output| = \frac{l - k + 2 \cdot p}{s} + 1 \tag{2}$$

where l is the input size, k the filter size, p the padding, and s the stride.

These layers analyze the input data and capture relevant features about the behavior of the patient, extracting some relevant points in the first layers and then composing complex patterns (for example, a strange or sudden movement) in the last layers. Thus, the original information is encoded to a few characteristics that can be analyzed with more direct layers like Fully Connected.

3.2.2. ReLU (non-linear function)

The non-linear functions introduce non-linearity in the model to allow faster learning, ensuring the output of the convolutional operation is not a linear

combination of the inputs. For this work, we have selected the Rectified Linear Unit (ReLU), commonly used in the CNNs. The ReLU operation is computed for every point in the input tensor $\vec{x}$ as:

$$R(\vec{x}) = \max(0, \vec{x}) \tag{3}$$

3.2.3. Batch normalization

This component [39] applies a normalization function for each mini-batch of the training process. This helps the model to learn faster and to generalize better. The normalization applies regularization to the output of the previous activation layer for each mini-batch by subtracting the mean of the batch and dividing it by its standard deviation.

3.2.4. Average-Pooling

The pooling operation allows the network to regularize and to reduce the dimension of the data between layers at the cost of losing some information. It divides the size of the received input by the size of the selected filter. In this case, we have selected the average-pooling function, that keeps the mean value of each cluster of inputs. The average pooling is more suitable since we are working with time-series instead of images (where max-pooling obtains better results). The average-pooling operation for a given point n in the output is defined as:

$$P(n) = \frac{\sum_{i=n}^{n+k-1} I_i}{k} \tag{4}$$

where I is the input, and k the size of the filter.

Table 1: Full description of the network

		Operation	Number of filters	Size of filters	Output	Number of params
	Input Layer				10804 x 3	
Convolutional Network	Convolutional Layer	1-D Convolution	50	8	10797 x 50	1250
		ReLU + Batch Norm	-	-	10797 x 50	200
		Average Pooling	5	-	2159 x 50	-
	Convolutional Layer	1-D Convolution	100	16	2144 x 100	80100
		ReLU + Batch Norm	-	-	2144 x 100	400
		Average Pooling	10	-	214 x 100	-
	Convolutional Layer	1-D Convolution	200	32	183 x 200	640200
		ReLU + Batch Norm	-	-	183 x 200	800
		Average Pooling	10	-	18 x 200	-
FC Network	Dropout	Dropout (p=0.75)	-	-	18 x 200	-
	FC	Fully Connected + ReLU	-	-	500	1800500
	FC (output)	Fully Connected + Softmax	-	-	3	1503

3.2.5. Fully Connected (FC)

These are Feedforward networks that give the resulting prediction based on the features identified in the convolutional layers. In this type of network or layer every neuron receives all the inputs. Each neuron weights the inputs, sums all of them and then applies the activation function. These weights are learned in the training process. The number of neurons in the output layer corresponds with the number of classes of the problem (three). Here, a Softmax activation function scales the outputs to represent the probability of the input of being in a class, summing them up to one. The Softmax function is described in the following formula:

$$S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{m} e^{y_i}} \text{ for } i = 1...m \tag{5}$$

where y is the output vector, and m the number of outputs/classes.

With this, we obtain a vector that indicates the probability that the model assigns to each of the classes: the n-th output represents the probability of the input of being in the class n.

In a complementary way to these convolutional layers, dropout has been used before the fully connected layers. Dropout layer randomly drops out neurons

of the network at training process, ignoring them for that particular epoch or training stage. The probability of dropping out a neuron is defined as a hyperparameter. This helps the network to generalize and thus learn faster.

3.2.6. Model description

In order to process the different data from the patients, we design the CNN architecture shown in Figure 1. The input for the network is a 10804x3 tensor, corresponding to the time dimension and axes (x, y, z) respectively. The network has several convolutional layers combining batch normalization and ReLU activation, followed by a pooling layer. The acceleration values are processed by the first convolutional layer, where the different used filters combine them. The set of convolutional layers reduces the dimensionality of the data while they extract relevant features of the input. The Dropout layer was added after the convolutional layers of the network to generalize better and increase the accuracy of the model. Finally, a Fully Connected network processes the extracted features and provides the prediction: the probability of the input of being in each one of the three different classes. The highest value will determine the predicted class.

Table 1 shows the architecture of the network in detail, describing the layers, operations, size and number of filters used in the Convolutional Layers, and the output of each layer. Moreover, the parameters for each layer are also shown. Every parameter is a trainable one (fitted in the backpropagation process), except for half of the parameters of the Batch Normalization layers, that correspond to the mean and the standard deviation of each input. The total number of parameters is 2,524,953, with 2,524,253 trainable by backpropagation and 700 non-trainable.

Different variations are assessed in order to test the suitability of the different layers and some hyperparameters used in the network. In these variations, we analyze the batch normalization, different Dropout rates, and different pooling operations.

3.3. Training and Evaluation

For the optimization process of the CNN parameters, we select the Adam algorithm [40], which is suitable for these types of models. The loss function used in this process is the categorical cross-entropy, as outlined in the next equation:

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij} \cdot \log(\hat{y}_{ij})) \tag{6}$$

where y is the true value, $\hat{y}$ is the predicted value, N is the number of examples and M the number of classes.

Each model configuration is trained using a 10-fold cross-validation process, with a partition of 80% for training and 20% for testing. This way we avoid incorrect measurements of the learning capacity of the models because of the training/testing division. The average of all of the 10 folds is taken as the final metric value of the model.

The metrics used to evaluate the models are accuracy, precision, recall, and F1-Score. Accuracy is defined as follows:

$$Acc = \frac{Correct\ Predictions}{All\ Predictions} \tag{7}$$

The function of F1-Score groups together precision and recall values, as it computes the harmonic mean between both metrics. Precision, recall and F1-score are defined only for binary classification. Hence, since we have to deal with multiple classes, we have used the macro-average and weighted-average for obtaining a single metric for all the classes. First, we obtain the metrics for every class in a One vs All approximation and then, obtain the mean of all the obtained metrics. Thus, single-class F1-score is defined as:

$$\begin{gathered} F1\text{-}score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \\ \textbf{with } precision = \frac{TP}{TP + FP}\ ,\ recall = \frac{TP}{TP + FN} \\ \textbf{alternatively } F1\text{-}score = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \end{gathered} \tag{8}$$

where TP are the true positives, FP the false positives, and FN the false negatives

For the staging problem, perfect precision would indicate that the model does not identify incorrectly any data sequence in a certain stage (no false positives), and perfect recall would indicate that it does identify correctly all data sequences in a certain stage (no false negatives).

4. Results

This section describes the experimental results after applying the proposed methodology. The dataset description shows the data used to build the CNN. Finally, the accuracy of the CNN to predict the AD stage is compared with other well-known classifiers.

All the deep learning models were developed with Python language, using the Keras library [41] with TensorFlow[42] as backend. These models were trained with the support of an RTX 2070 graphic card that greatly accelerates the optimization process of the networks. The total time it took each of the models to train was between 3 and 4 minutes per fold in the cross validation process. Scikit-learn library [43] provided the feature-based models tested and the functions to compute the evaluation metrics.

4.1. Dataset

The data used were obtained from the AFAC daycare center of Santander (Spain) for patients with Alzheimer's disease [36]. In this study, 35 patients participated and their activity was monitored in the center for a week. It is estimated that each patient stayed on average 6 hours a day in the daycare center. During this time, patients had absolute freedom to develop their daily activities in the daycare center.

Table 2: Details of the experiments' setting.

Place	Daycare Center
Participating patients	35
Duration	Daily samples for 7 days
Patients activity	Total freedom of movement and daily activities

Patients were not asked to do any specific task. Each patient carried an Android smartphone in their pocket that incorporated an accelerometer, all with the same type and characteristics. This Android accelerometer sensor works with a sampling rate of 8 Hz in a range of $\pm 3.28g$. A neuropsychologist from the center placed the phones in their pockets, independent of the pocket it was in. Changes in the orientation of the device should not be a problem either, as they would be caused by the activities and movements made by the patient, which is something the system must be able to deal with correctly by learning these variations.

Other proposals indicate that similar studies have been conducted but with the devices in a fixed location (such as at the hip/wrist like in [44] and some examples of [45]), which allows for more consistent data but with less variance. In this case, we prefer to have a higher variance in the data, since the system can learn more characteristics without restrictions on the movement of the device. In addition to demonstrating that these intelligent systems can work under poorly controlled environments, as is often the case within the healthcare field.

All participants had previously been diagnosed with Alzheimer's disease and at a specific GDS stage. Table 3 shows the number of participants associated with each label. As shown in this table, there are more data sequences of the middle stage than the other two stages. Therefore, the data is not balanced, which could cause the appearance of overfitting.

Table 3: Labels assigned to the participants.

Label	**Number of participants**	**Samples after 1 week**	**GDS stage**
Early	7	41	2 and 3
Middle	18	100	4 and 5
Late	10	46	6 and 7

Table 4 shows a short example extracted from a sample of a random patient. It can be clearly seen that the data is unevenly distributed over time, with very different gaps between timestamps, which shows that a preprocessing is necessary.

Table 4: Example of raw data with the accelerometer data and their timestamp

X	0.063	0.145	0.29	0.265	0.266	0.25	0.213	0.097	0.186
Y	0.905	0.905	0.976	0.98	1.0	0.966	0.856	0.858	0.897
Z	0.395	0.393	0.343	0.394	0.409	0.352	0.234	0.322	0.31
Time	6.331	6.508	6.517	6.825	6.828	6.832	6.835	6.839	6.847

4.2. CNN evaluation

Since the data from the accelerometers were irregular and scattered, we preprocessed it to transform all 187 sequences to the same length. The first step of the preprocessing is the data homogenization, performed by summarizing values within 0.1-second intervals. The points within each interval are replaced by the average. We apply this because the devices take measurements at a higher and often irregular resolution (i.e., many points with very short periods). Also, this process avoids outliers and redundant information (e.g., peaks, repeated points).

Next, since data sequences were too long (approx. 1 hour per sample, more than 40k points in some cases), each one is divided into 5 different samples, which also serves to increase the sequences per class in the dataset. Because of this, all samples have the same length (10,804 points, around 18 minutes of data per sample) and the total number of samples increases to 935 (205 early, 500 middle and 230 late stage). To carry out the experiments, the entire dataset is divided into a training set and a test set: 80% of the dataset is for training and the remaining 20% for testing. This results in a total of 750 examples in the training set and 185 in the test set. Samples from the same patient were sent to the same split (either training or test).

Figure 2 shows a pre-processed sample of each of the three types of data on each of the three accelerometer axes, which are the input of the convolutional network. It can be observed that each sequence is very distinct, with different

Table 5: Description of the configurations for the different models tested.

Type	Model	Architecture	Training configuration
Convolutional Network	Base Model	CNN model, Avg.Pool., B.Norm., Dropout(0.75)	Adam Optimizer Categorical cross-entropy 300 epochs Batch size 32
	No B. Norm.	CNN model, Avg.Pool., Dropout(0.75)	
	Dropout 0.5	CNN model, Avg.Pool., B.Norm., Dropout(0.5)	
	Dropout 0.25	CNN model, Avg.Pool., B.Norm., Dropout(0.25)	
	Max-pooling	CNN model, Max.Pool., B.Norm., Dropout(0.75)	
Feature-based Classifier	AB	AdaBoost model	50 estimators Decision stumps
	DT	Decision Tree model	Entropy criterion No maximum depth
	KNN	k-Nearest Neighbors model	$k = 3$ Minkowski distance, $p = 1$
	LR	Logistic Regression model	No regularization
	MLP	Feedforward N.N., 2 hidden layers (100), tanh activation	Adam Optimizer Categorical cross-entropy 1000 epochs
	RF	Random Forest model	Gini criterion 100 estimators Maximum depth = 15
	SVM	Support Vector Machine model, RBF kernel	Regularization $C = 10$

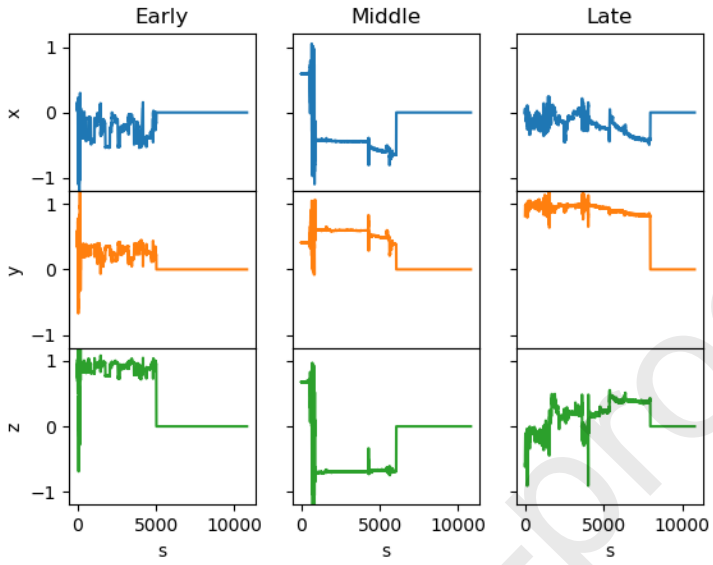


Figure 2: Preprocessed data of the accelerometers, with a example of each stage

variability in movement and even of diverse length, seen when the data flattens to zeros (made by the preprocessing).

With the preprocessed data, the different CNN models have been tested using the specifications shown in Section 3.3. No extra computation has been performed, being the preprocessed data of the accelerometers the direct input of the network. Each model is trained for 300 epochs. A batch-size of 32 is used due to the large length of the data and the amount of training data (750 samples). Table 5 shows the different configurations used for all the models tested in this research.

Table 6: Results obtained with the different configurations of the network. Average and standard deviation for each metric were reported from 10 trials per configuration.

Configuration	Loss	Accuracy	F1-Score
Base model	0.4509(±0.2161)	90.91%(±4.95)	89.7%(±5.58)
No B. Norm.	0.5227(±0.1749)	85.75%(±3.26)	83.89%(±3.55)
Dropout 0.5	0.4728(±0.2285)	87.80%(±4.28)	85.40%(±4.39)
Dropout 0.25	0.6188(±0.2117)	85.44%(±3.01)	84.25%(±3.86)
Max-pooling	0.4919(±0.2264)	87.18%(±4.40)	86.17%(±4.54)

In Table 6 we can observe the results obtained by the different configurations selected. The best result is given by the Base model, obtaining 90.91% of mean accuracy. From the table we can retrieve several conclusions:

- Batch normalization greatly assists the training step by diminishing the overfitting in order to obtain a better outcome.
- A high dropout also helps the model to achieve a better result, improving the generalization of the model and (again) avoiding overfitting.
- In this case, as explained in Section 3.2, average-pooling gives slightly better results than max-pooling since we are working with time-series and not with images.

We can also see that the models created with CNNs are accurate, with the worst of them having 85.44% accuracy. This indicates us that the selected architecture and other hyperparameters learn to fit the data well. Furthermore, it can be seen that even with these variations the overall accuracy of the model does not change much (about 5%).

Figure 3 shows the accuracy and loss curves for a fitting process of the Base model. We can clearly observe in both graphs that the process is not leading to overfitting since test curves increase (accuracy)/decrease (loss) at the same rate as training does. This means that the model is learning how to classify the data correctly, extracting the features that represent each of the different stages and using them to identify the classes, both in training and test phases. Also, the model does not overfit to the most frequent class (middle) despite the data imbalance.

4.3. Comparison with other classifiers

Intending to assess the performance of our CNN method, we have also compared the results with other common models for time series classification. For this problem, the models used are feature-based, distance-based or generative [46]. In our case, the sequences are large in size and each point is a 3-tuple with the real values of the acceleration changes (in X, Y and Z axes), so the feature-based category is the most suitable one. Therefore, we first generate a feature vector for each of the preprocessed sequences and then we train the other classifiers with these vectors.

For the generation of the feature vectors, we select several sample statistics as features for each data sequence: mean, median, variance, maximum, minimum

(a) Accuracy training curve

(b) Loss training curve

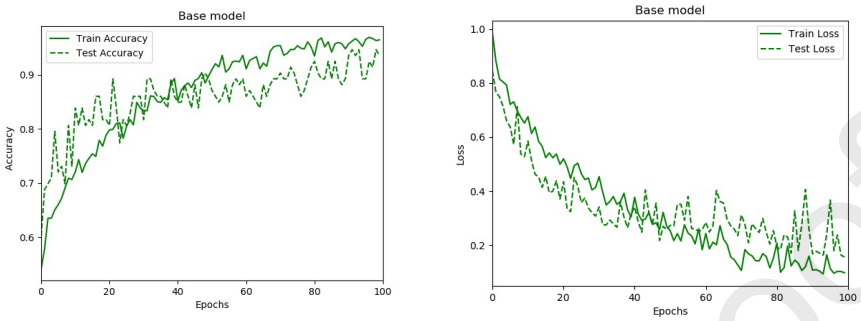


Figure 3: Graphs of the training process for the Base model

and sum. These statistics are computed for each axis, so we obtain each one for the X, Y and Z:

$$(\text{mean}_X, ..., \text{sum}_X, \text{mean}_Y, ..., \text{sum}_Y, \text{mean}_Z, ..., \text{sum}_Z)$$

The chosen feature-based models for the comparison are Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Multi-layer Perceptron (MLP) and AdaBoost (AB). The hyperparameters for these classifiers are shown in Table 5. These are selected after a grid search among several parameter configurations where the optimal performance is measured by the resulting accuracy.

Table 7 shows the results of these models along with the CNN model. Here we can observe the improvement of our model as it obtains a higher success

Table 7: Classification results for the feature-based models and the best CNN architecture (Base model).

	Precision				**Recall**					
Model	*Early*	*Middle*	*Late*	Weight.Avg	*Early*	*Middle*	*Late*	Weight.Avg	**F-Score**	**Acc.**
AB	0.51	0.85	0.72	0.76	0.73	0.85	0.50	0.74	0.74	74%
DT	0.68	0.84	0.62	0.76	0.70	0.84	0.61	0.76	0.76	76%
KNN	0.85	0.83	0.80	0.83	0.70	0.93	0.70	0.83	0.82	83%
LR	0.55	0.76	0.68	0.70	0.36	0.89	0.57	0.72	0.70	72%
MLP	0.76	0.84	0.77	0.81	0.76	0.90	0.65	0.81	0.81	81%
RF	0.87	0.88	0.89	0.88	0.79	0.99	0.67	0.88	0.87	88%
SVM	0.66	0.86	0.84	0.82	0.76	0.94	0.57	0.81	0.81	81%
CNN	0.88	0.92	0.90	0.91	0.86	0.94	0.86	0.90	0.90	91%

rate for the classification of all stages, while the others do not recognize the early or late stages with the same success. Especially, we observe from the recall values of the other models that the proportion of false negatives in the early or late stages is noticeably higher. It is the case even with the best of the feature-based models (Random Forest), which misclassifies some late and early sequences. Thus, in comparison with the CNN model, all other classifiers seem weaker, less accurate and more sensible to data imbalance.

5. Discussion

In this article, we have proposed a novel methodology to relate the daily activity of Alzheimer's patients with the stage of the disease. The methodology is made up of two phases: (i) a preprocessing phase which transforms accelerometer data into shorter sequences of the same length and homogenizes periods between data points; and (ii) a supervised learning phase which builds a CNN that predicts the AD stage. The methodology has been applied to study the activity of 35 Alzheimer's patients in a daycare center for a week. In contrast to previous studies, no additional hardware or resource (computers, motion analysis hardware, cameras, etc.) is necessary to apply this methodology.

Our study generated a daily data sequence for each patient in the daycare center. This hinders the CNN learning process because very few sequences are obtained for each patient. Moreover, the data sequence of each patient has a different size depending on the time he/she spent in the daycare center. For this reason, the preprocessing phase of the methodology divided the data sequence of each patient into segments of the same size. A large amount of data is collected because of the high frequency of sampling. Low variations are registered in these short periods and, therefore, information is not lost.

After preprocessing the data sequences, the main goal of our work was to build mechanisms that automatically predict the stage of Alzheimer's disease based on information on patient mobility. For this purpose, we designed a CNN architecture based on three 1-Dimensional Convolutional layers. These layers process fixed-length data segments to detect local patterns in the x-y-z axes where acceleration changes. The results obtained allow considering that the CNN effectively achieves this objective (91% success rate), while other common supervised classifiers (AdaBoost, k-Nearest Neighbors, Logistic Regression, Multilayer Perceptron, Random Forest, Decision Tree, Support-Vector Machine) achieve a lower success rate. Moreover, the CNN model is the only one that

can successfully recognize all three stages even with data imbalance, while the other classifiers fail to classify early or late stages.

The proposal is based on the smartphone accelerometer that is a non-intrusive and widely used device. For this reason, the methodology can be considered as easily applicable to analyze the daily activities of patients. Moreover, the predictive outputs of the CNN can be useful to detect an aggravation of the disease and to prevent the progression of the disease to a more severe stage and thus to prevent arising complications as falls, spatial disorientation, etc.

In the future, we will work to exploit this methodology in a software system that follows a cloud-computing architecture to collect the accelerometer data and a service which users can subscribe to for the monitoring of changes in the AD stage. This architecture will integrate the mobile devices of the patients as distributed nodes. In parallel, we will seek to increase the size of the original dataset, so that a second validation of the results achieved in this article can be made. Additionally, it will be possible to consolidate and reinforce the capacity of the system created, being able to retrain the network with a greater amount of data, which will improve the reliability of this system.

6. Conclusions

Mobility disorders are one of the earliest symptoms that Alzheimer's Disease patients exhibit. We have developed a methodology to identify the disease stage and the evolution of a patient in order to apply convenient measures. Our method processes mobility data obtained from accelerometer sensors and makes use of a deep learning model based on CNN to recognize patterns in patient movement. In the results obtained, the CNN model achieved 91 % accuracy and a F1-score of 0.897, improving the results of the standard feature-based models. Our research shows that this methodology can be very valuable to facilitate the monitoring of AD progression without requiring specialized devices or healthcare supervisors.

Acknowledgments

The authors want to acknowledge the financial support from the ISCIII (Instituto de Salud Carlos III) and Ministerio de Economía y Competitividad, Gobierno de España for the project PI17/00936; as well as acknowledge the "Asociación de Familiares de Enfermos de Alzheimer en Cantabria" for their participation in the various studies.

References

[1] M. Prince, A. Wimo, M. Guerchet, G. Ali, Y. Wu, M. Prina, World alzheimer report 2015. london, uk, Alzheimer's Disease International (2015) 1–92doi:10.1111/j.0963-7214.2004.00293.x.
URL https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf

[2] R. J. Bateman, C. Xiong, T. L. S. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, D. S. Marcus, N. J. Cairns, X. Xie, T. M. Blazey, et al., Clinical and biomarker changes in dominantly inherited alzheimer's disease, The New England journal of medicine 367 (9) (2012) 795–804. doi:10.1056/NEJMoa1202753.

[3] B. Reisberg, S. H. Ferris, M. J. de Leon, T. Crook, The global deterioration scale for assessment of primary degenerative dementia, The American journal of psychiatry 139 (9) (1982) 1136–1139. doi:10.1176/ajp.139.9.1136.

[4] J. C. Morris, The clinical dementia rating (cdr): current version and scoring rules, Neurology 43 (11) (1993) 2412–2414. doi:10.1212/wnl.43.11.2412-a.

[5] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8) (2013) 1798–1828. doi:10.1109/TPAMI.2013.50.

[6] M. A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review, Computer Methods and Programs in Biomedicine 187 (2020) 105242. doi:https://doi.org/10.1016/j.cmpb.2019.105242.
URL http://www.sciencedirect.com/science/article/pii/S0169260719310946

[7] D. Pan, Y. Huang, A. Zeng, L. Jia, X. Song, Early diagnosis of alzheimer's disease based on deep learning and gwas, in: A. Zeng, D. Pan, T. Hao, D. Zhang, Y. Shi, X. Song (Eds.), Human Brain and Artificial Intelligence, Springer Singapore, Singapore, 2019, pp. 52–68.

[8] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, X. Zhang, Multi-modal deep learning model for auxiliary diagnosis of alzheimer's

disease, Neurocomputing 361 (2019) 185–195. doi:https://doi.org/10.1016/j.neucom.2019.04.093.
URL http://www.sciencedirect.com/science/article/pii/S092523121930921X

[9] S. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease, NeuroImage 189 (2019) 276–287. doi:https://doi.org/10.1016/j.neuroimage.2019.01.031.
URL http://www.sciencedirect.com/science/article/pii/S105381191930031X

[10] C. Park, J. Ha, S. Park, Prediction of alzheimer's disease based on deep neural network by integrating gene expression and dna methylation dataset, Expert Systems with Applications 140 (2020) 112873. doi:https://doi.org/10.1016/j.eswa.2019.112873.
URL http://www.sciencedirect.com/science/article/pii/S0957417419305834

[11] M. Raza, M. Awais, W. Ellahi, N. Aslam, H. Nguyen, H. Le-Minh, Diagnosis and monitoring of alzheimer's patients using classical and deep learning techniques, Expert Systems with Applications 136 (2019) 353–364. doi:https://doi.org/10.1016/j.eswa.2019.06.038.
URL http://www.sciencedirect.com/science/article/pii/S0957417419304385

[12] H. Wu, J. Weng, X. Chen, W. Lu, Feedback weight convolutional neural network for gait recognition, Journal of Visual Communication and Image Representation 55 (2018) 424–432. doi:https://doi.org/10.1016/j.jvcir.2018.06.019.
URL http://www.sciencedirect.com/science/article/pii/S1047320318301445

[13] A. R. Hawas, H. A. El-Khobby, M. Abd-Elnaby, F. E. Abd El-Samie, Gait identification by convolutional neural networks and optical flow, Multimedia Tools and Applications 78 (18) (2019) 25873–25888. doi:10.1007/s11042-019-7638-9.
URL https://doi.org/10.1007/s11042-019-7638-9

[14] M. Alotaibi, A. Mahmood, Improved gait recognition based on specialized deep convolutional neural network, Computer Vision and Image Understanding 164 (2017) 103–110. doi:https://doi.org/10.1016/j.cviu.2017.10.004.
URL http://www.sciencedirect.com/science/article/pii/S1077314217301674

[15] S. R. Lord, J. C. Close, New horizons in falls prevention, Age and Ageing 47 (4) (2018) 492–498. doi:10.1093/ageing/afy059.

[16] Y. Higami, M. Yamakawa, K. Shigenobu, K. Kamide, K. Makimoto, High frequency of getting out of bed in patients with alzheimer's disease monitored by non-wearable actigraphy, Geriatrics & gerontology international 19 (2) (2019) 130–134. doi:10.1111/ggi.13565.

[17] M. Gietzelt, F. Feldwieser, M. Gövercin, E. Steinhagen-Thiessen, M. Marschollek, A prospective field study for sensor-based identification of fall risk in older people with dementia, Informatics for Health and Social Care 39 (3-4) (2014) 249–261. doi:10.3109/17538157.2014.931851.

[18] K. S. Van Schooten, M. Pijnappels, S. M. Rispens, P. J. Elders, P. Lips, A. Daffertshofer, P. J. Beek, J. H. Van Dieën, Daily-life gait quality as predictor of falls in older people: A 1-year prospective cohort study, PLoS ONE 11 (7) (2016) 1–13. doi:10.1371/journal.pone.0158623.

[19] R. David, E. Mulin, L. Friedman, F. Le Duff, E. Cygankiewicz, O. Deschaux, R. Garcia, J. A. Yesavage, P. H. Robert, J. M. Zeitzer, Decreased daytime motor activity associated with apathy in alzheimer disease: an actigraphic study, The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry 20 (9) (2012) 806–814. doi:10.1097/JGP.0b013e31823038af.

[20] A. Kuhlmei, B. Walther, T. Becker, U. Muller, T. Nikolaus, Actigraphic daytime activity is reduced in patients with cognitive impairment and apathy, European psychiatry : the journal of the Association of European Psychiatrists 28 (2) (2013) 94–97. doi:10.1016/j.eurpsy.2011.04.006.

[21] J. M. Zeitzer, R. David, L. Friedman, E. Mulin, R. Garcia, J. Wang, J. A. Yesavage, P. H. Robert, W. Shannon, Phenotyping apathy in individuals with alzheimer disease using functional principal component analysis, The

American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry 21 (4) (2013) 391–397. doi:10.1016/j.jagp.2012.12.012.

[22] D. Goerss, A. Hein, S. Bader, M. Halek, S. Kernebeck, A. Kutschke, C. Heine, F. Krueger, T. Kirste, S. Teipel, Automated sensor-based detection of challenging behaviors in advanced stages of dementia in nursing homes, Alzheimer's & dementia : the journal of the Alzheimer's Association (Oct. 2019). doi:10.1016/j.jalz.2019.08.193.

[23] M. Gietzelt, K.-H. Wolf, M. Kohlmann, M. Marschollek, R. Haux, Measurement of accelerometry-based gait parameters in people with and without dementia in the field, Methods of Information in Medicine 52 (4) (2013) 319–325. doi:10.3414/me12-02-0009.

[24] J. M. Hausdorff, I. Hillel, S. Shustak, S. Del Din, E. M. J. Bekkers, E. Pelosin, F. Nieuwhof, L. Rochester, A. Mirelman, Everyday stepping quantity and quality among older adult fallers with and without mild cognitive impairment: Initial evidence for new motor markers of cognitive deficits?, The journals of gerontology. Series A, Biological sciences and medical sciences 73 (8) (2018) 1078–1082. doi:10.1093/gerona/glx187.

[25] R. Mc Ardle, S. Del Din, B. Galna, A. Thomas, L. Rochester, Differentiating dementia disease subtypes with gait analysis: feasibility of wearable sensors?, Gait & posture 76 (2019) 372–376. doi:10.1016/j.gaitpost.2019.12.028.

[26] S. Gillain, M. Drame, F. Lekeu, V. Wojtasik, C. Ricour, J.-L. Croisier, E. Salmon, J. Petermans, Gait speed or gait variability, which one to use as a marker of risk to develop alzheimer disease? a pilot study, Aging clinical and experimental research 28 (2) (2016) 249–255. doi:10.1007/s40520-015-0392-6.

[27] T. Kirste, A. Hoffmeyer, P. Koldrack, A. Bauer, S. Schubert, S. Schröder, S. Teipel, Detecting the effect of alzheimer's disease on everyday motion behavior, Journal of Alzheimer's Disease 38 (1) (2014) 121–132. doi:10.3233/JAD-130272.

[28] M. F. Folstein, S. E. Folstein, P. R. McHugh, "mini-mental state". a practical method for grading the cognitive state of patients for the clinician, Journal of psychiatric research 12 (3) (1975) 189–198.

[29] A. Watts, R. W. Walters, L. Hoffman, J. Templin, Intra-individual variability of physical activity in older adults with and without mild alzheimer's disease, PloS one 11 (4) (2016) e0153898. doi:10.1371/journal.pone.0153898.

[30] K. Weissova, A. Bartos, M. Sladek, M. Novakova, A. Sumova, Moderate changes in the circadian system of alzheimer's disease patients detected in their home environment, PloS one 11 (1) (2016) e0146200. doi:10.1371/journal.pone.0146200.

[31] E. S. Musiek, M. Bhimasani, M. A. Zangrilli, J. C. Morris, D. M. Holtzman, Y.-E. S. Ju, Circadian rest-activity pattern changes in aging and preclinical alzheimer disease, JAMA neurology 75 (5) (2018) 582–590. doi:10.1001/jamaneurol.2017.4719.

[32] P. Li, L. Yu, A. S. P. Lim, A. S. Buchman, F. A. J. L. Scheer, S. A. Shea, J. A. Schneider, D. A. Bennett, K. Hu, Fractal regulation and incident alzheimer's disease in elderly individuals, Alzheimer's & dementia : the journal of the Alzheimer's Association 14 (9) (2018) 1114–1125. doi:10.1016/j.jalz.2018.03.010.

[33] Á. Ruiz-Zafra, E. O. Gonzalez, M. Noguera, K. Benghazi, J. M. H. Jiménez, Energy expenditure analysis: A comparative research of based on mobile accelerometers, in: International Workshop on Ambient Assisted Living, Springer, 2014, pp. 38–45.

[34] A. Ruiz-Zafra, E. Orantes-González, M. Noguera, K. Benghazi, J. Heredia-Jimenez, A comparative study on the suitability of smartphones and imu for mobile, unsupervised energy expenditure calculi, Sensors 15 (8) (2015) 18270–18286.

[35] A. Nieto-Reyes, R. Duque, J. L. Montaña, C. Lage, Classification of alzheimer's patients through ubiquitous computing, Sensors (Switzerland) 17 (7) (2017). doi:10.3390/s17071679.

[36] R. Duque, A. Reyes, C. Martinez, J. Montaña, Detecting human movement patterns through data provided by accelerometers. a case study regarding alzheimer's disease, in: Ubiquitous Computing and Ambient Intelligence - 10th International Conference, UCAmI 2016, San Bartolomé de Tirajana, Gran Canaria, Spain, November 29 - December 2, 2016, Proceedings, Part

I, 2016, pp. 56–66. doi:10.1007/978-3-319-48746-5_6.
URL https://doi.org/10.1007/978-3-319-48746-5_6

[37] I. E. Maachi, G.-A. Bilodeau, W. Bouachir, Deep 1d-convnet for accurate parkinson disease detection and severity prediction from gait, Expert Systems with Applications 143 (2020) 113075. doi:https://doi.org/10.1016/j.eswa.2019.113075.
URL http://www.sciencedirect.com/science/article/pii/S0957417419307924

[38] S. Bringas, S. Salomón, R. Duque, J. L. Montaña, C. Lage, A convolutional neural network-based method for human movement patterns classification in alzheimer's disease, in: Multidisciplinary Digital Publishing Institute Proceedings, Vol. 31, 2019, p. 72.

[39] S. Ioffe, C. Szegedy, 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, ICML'15, JMLR.org, 2015, pp. 448–456.

[40] D. Kingma, J. Ba, Adam: A method for stochastic optimization, International Conference on Learning Representations abs/1412.6980 (2014).

[41] F. Chollet, et al., Keras, https://keras.io (2015).

[42] M. Abadi, A. Agarwal, P. Barham, et al., TensorFlow: Large-scale machine learning on heterogeneous systems (2015).
URL https://www.tensorflow.org/

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[44] M. E. Rosenberger, W. L. Haskell, F. Albinali, S. Mota, J. Nawyn, S. Intille, Estimating activity and sedentary behavior from an accelerometer on the hip or wrist, Medicine and science in sports and exercise 45 (5) (2013) 964.

[45] S. L. Murphy, Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct, Preventive medicine 48 (2) (2009) 108–114.

[46] Z. Xing, J. Pei, E. Keogh, A brief survey on sequence classification, SIGKDD Explor. Newsl. 12 (1) (2010) 40–48. doi:10.1145/1882471.1882478.

Highlights

- Movement data were used to identify severity stage of Alzheimer's Disease
- Smartphone accelerometers are useful and accessible tools to obtain meaningful data
- Machine Learning models can serve as an effective diagnostic tool
- 1D-CNNs outperform other models on identifying the AD stage using accelerometer data
- Different settings (Batch-Norm, Dropout, Pooling operations) were tested for the CNN

CRediT author statement

Santos Bringas: Conceptualization, Methodology, Data Curation, Software, Writing- Original draft preparation.

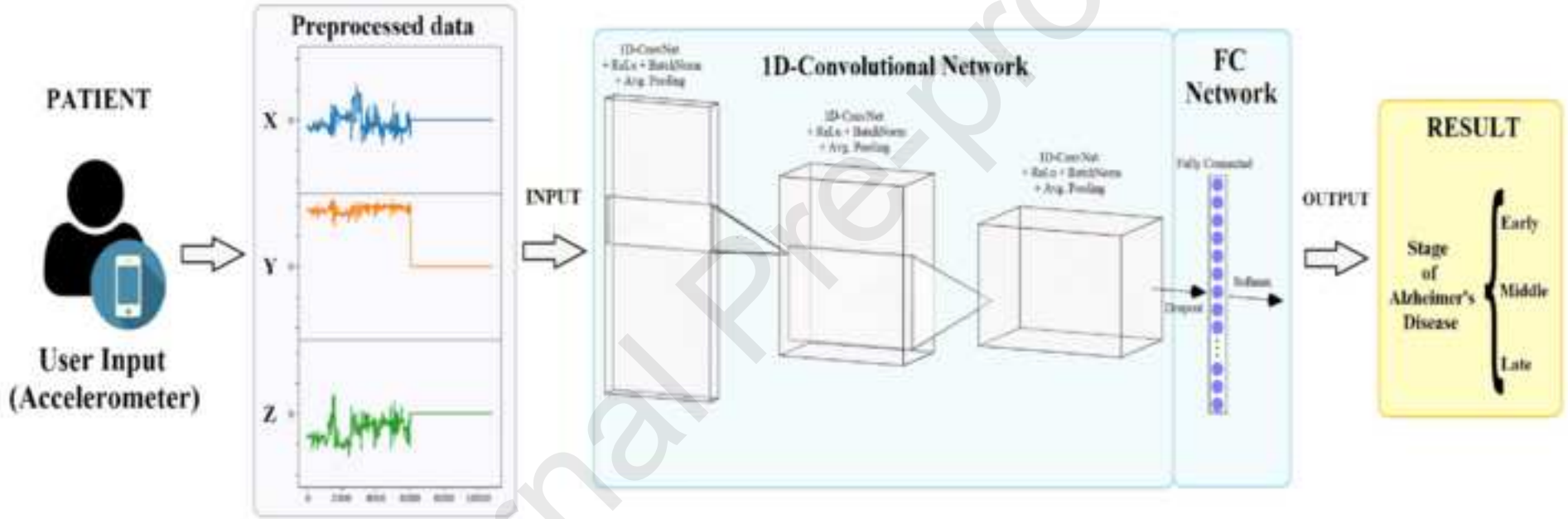
Sergio Salomón: Software, Methodology, Formal Analysis, Writing- Original draft preparation.

Rafael Duque: Conceptualization, Supervision, Project Administration, Writing- Original draft preparation.

Carmen Lage: Investigation, Writing - Review & Editing.

José Luis Montaña: Writing- Reviewing and Editing, Supervision.

Graphical Abstract



Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: