

## Article

# Classification of Parkinson's Disease Patients—A Deep Learning Strategy

Helber Andrés Carvajal-Castaño <sup>1,2,\*</sup> , Paula Andrea Pérez-Toro <sup>1,3</sup>  and Juan Rafael Orozco-Arroyave <sup>1,3,\*</sup> 

<sup>1</sup> GITA Lab, Electronics Engineering and Telecommunications Department, Faculty of Engineering, Universidad de Antioquia, Medellín 050010, Colombia

<sup>2</sup> GIBIC Lab, Bioengineering Department, Faculty of Engineering, Universidad de Antioquia, Medellín 050010, Colombia

<sup>3</sup> Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany

\* Correspondence: helber.carvajal@udea.edu.co (H.A.C.-C.); rafael.orozco@udea.edu.co (J.R.O.-A.)

**Abstract:** (1) Background and objectives: Parkinson's disease (PD) is one of the most prevalent neurodegenerative diseases whose typical symptoms include bradykinesia, abnormal gait and posture, shortened strides, and other movement disorders. In this study, we present a novel framework to evaluate PD gait patterns using state of the art deep learning algorithms. A comparative analysis with three different approaches is presented and evaluated upon three groups of subjects: PD patients, Young Healthy Controls (YHC), and Elderly Healthy Controls (EHC). (2) Methods: The three approaches used in the study include: (i) The energy content of the gait signals in the frequency domain is captured with spectrograms that are used to feed a CNN model, (ii) Temporal information is incorporated by creating GRU networks, (iii) Temporal and spectral information is simultaneously captured by creating a new architecture based on CNNs and GRUs. (3) Results: Accuracies of up to 83.7% and 92.7% are found for the classification between PD vs. EHC and PD vs. YHC, respectively. According to our observations, the proposed approach based on the combination of temporal and spectral information, yields better results than others reported in the state of the art. (4) Conclusions: The results obtained in this study suggest that the combination of temporal and spectral information is more accurate than individual approaches used to classify and evaluate gait patterns in PD patients. To the best of our knowledge, this is the first study in gait analysis where temporal and spectral information is combined in an architecture of deep learning.

**Keywords:** Gait analysis; Parkinson's disease; convolutional neural networks; gate recurrent units; deep learning



**Citation:** Carvajal-Castaño, Helber Andrés; Pérez-Toro, Paula Andrés.; Orozco-Arroyave, Juan Rafael Classification of Parkinson's Disease Patients—A Deep Learning Strategy. *Electronics* **2022**, *11*, 2684. <https://doi.org/10.3390/electronics11172684>

Academic Editor: Byung-Gyu Kim

Received: 26 July 2022

Accepted: 24 August 2022

Published: 27 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Parkinson's Disease (PD) is a neurodegenerative disease that produces movement disorders including tremor, rigidity, postural instability and lack of coordination which affect patients' gait [1–3]. PD patients are characterized by abnormal gait patterns associated with bradykinesia (slowness of movement), less steady walk, reduced stride length and shuffling steps or impaired gait initiation [4–6]. The symptoms of PD may appear about 10 years prior to the clinical manifestations [7], besides, several studies show that PD mainly impacts elderly people [8]. An important fact is that the prevalence of the disease is increasing with age worldwide [9,10]. Neurologists usually use clinical scales to evaluate and quantify the neurological state of the patient. The most used is the Movement Disorder Society—Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [11]. This scale allows neurologists to evaluate the patient's state and it is useful to follow up on therapies. The MDS-UPDRS scale is composed of four sections. The third one is called MDS-UPDRS-III and corresponds to the assessment of routine motor activities including 33 tasks, therefore it ranges from 0 to 132.

Gait patterns allow to obtain information about different movement disorders that are sometimes associated to Parkinson's disease (PD) symptoms. In the literature, the use of Inertial Movement Units (IMU) has increased considerably since they allow to capture gait patterns to study the movement dynamics of patients. This includes the study of kinematic characteristics of PD patients [12,13], nonlinear dynamics [14,15], stability and deep learning approaches [16], among others.

Computer vision methods and force platforms are used in laboratories to evaluate gait disorders [17,18]; however, they are expensive and difficult to access. Conversely, wearable sensors allow for designing low-cost and unobtrusive solutions that enable continuous monitoring of patients [19]. The most common wearable sensors for gait analyses are those based on plantar pressure systems [20–22] and IMU sensors [23,24]

Gait analysis is of great interest for the research community due to its suitability to perform unobtrusive automatic and continuous evaluation of motor symptoms of PD patients.

### 1.1. State of the Art

The study of gait patterns is related to the human locomotion and includes the study of people move while walking. Models can be created considering different gait features, related to kinematics, such as: stride length, stride velocity, turning angle, swing phase, and others [25,26]. The analysis of abnormal gait patterns have been typically performed considering Inertial Movement Units (IMU). An IMU is an electronic device usually consisting of accelerometer and gyroscope sensors, and in some cases also include magnetometer sensors. In [27] the authors presented a complete study related to the use of IMU sensors. The aim of the authors is to model abnormal gait patterns. According to the authors IMU sensors in gait analysis are used due to their low cost and the potential for designing wearable devices for continuous monitoring. Even though the study was presented about seven years ago, the same claim continues to be valid today. In [28] the authors proposed the use of one IMU sensor in each foot to analyse patients with different neurological conditions. The authors extracted several kinematic gait measures like stride length, stance time, swing time, and cycle time. The proposed method was tested with a dataset comprised of 22 healthy control (HC) subjects recorded with a camera-based system. A clinical discussion using a dataset of 17 subjects with different neurological disorders was also presented. According to the authors, it is possible to obtain relevant information on different neurodegenerative diseases, even outside clinical settings. In [20] the classification of PD patients and HC subjects was performed by using several spatial-temporal measures like stride length, cadence, stance time, and swing time. Different classifiers were tested including Random Forest (RF), Support Vector Machine (SVM) and Kernel Fisher Discriminant. The best result was found with a RF classifier (92.6%). A multi-modal study for the discrimination between PD patients and HC subjects, considering information of three bio-signals: speech, handwriting, and gait, was presented in [16]. To merge the information of each bio-signal a Convolutional Neural Networks (CNN) was implemented. The authors reported the highest accuracy with the combination of the three bio-signals (97.6%). Another approach in gait analysis is based on non-linear dynamics (NLD) measures. In [14] the authors extracted several NLD and Entropy measures. Three classifiers were compared: SVM, RF and k-nearest neighbours (KNN). Accuracies up to 92% were reported In [15] the authors proposed a new strategy considering Poincaré sections. Accuracies up to 89% in the classification of PD vs. HC were reported, besides, the authors' proposal includes experiments with PD patients in three different stages of the disease: mild, moderate, and severe where accuracies up to 67.2% were reported. Recently, in [13] the authors computed three sets of features named kinematics, NLD, and stability, and proposed a clinical interpretation based on the most discriminant feature per subset. The authors reported accuracies of up to 92% when using only three of the features.

In this paper, we use raw gait signals captured using IMU sensors to assess the ability of different deep learning architectures to classify PD patients vs. HC subjects.

Three architectures were considered: Convolutional Neural Networks (CNN), Gate Recurrent Units (GRU) and a new approach that considers energy information at the input of a CNN and temporal information with a GRU. In order to consider the effect of age three groups of subjects are examined: Young Healthy Control (YHC), Elderly Healthy Control (EHC), and PD patients. The EHC group and the PD group are matched in age. Accuracies up to 85.3% were reported in the PD vs. EHC scenario, and accuracies of up to 92.7% were found in PD vs. YHC.

### 1.2. Contributions of this Study

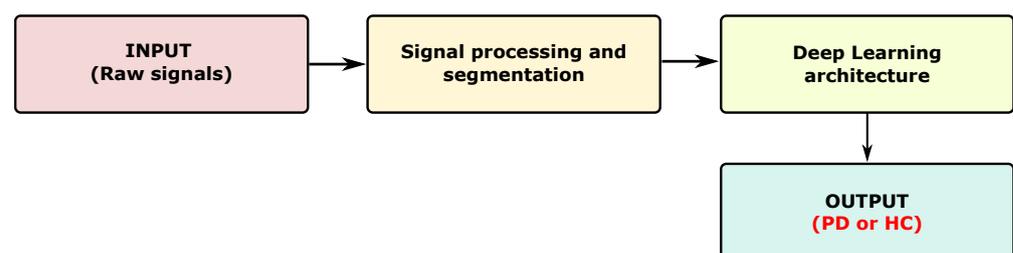
Three different deep learning architectures, namely CNN, GRU and CNN + GRU were evaluated in this study to classify between PD patients and HC subjects. Models based on CNNs yielded good results but did not consider temporal information, therefore we decided to evaluate an architecture based on GRUs to incorporate relevant information possibly encoded in the evolution of the patterns, i.e., temporal information. The combination of CNNs and GRUs in the same model was introduced to take advantage of incorporating temporal and frequency information in the same model, which potentially enables clinical interpretation. We believe that the CNN+GRU model did not show better results due to the small amount of data available for the present experiments.

In this study two tasks were considered:  $2 \times 10$  m task corresponds to a 10 m walk performed twice and  $4 \times 10$  m task corresponds to a 10 m walk which is performed 4 times. In general terms, the  $4 \times 10$  m task is better than the  $2 \times 10$  m one. We think that this is because longer tasks allow to collect more information and therefore increase the chances to find abnormal patterns in the gait signal.

## 2. Materials and Methods

### 2.1. Methodology

The general methodology proposed in this study is summarized in Figure 1. Gait signals are collected using wearable IMU sensors. Note that the main characteristic of the proposed methodology is that there is no a sophisticated feature extraction stage. The segmentation process is based on sliding windows of fixed length and, in the case of the CNN architecture, we compute the spectrogram that is used as input. Information of each foot and their combination are considered. In the following subsections, the stages of this methodology are explained.

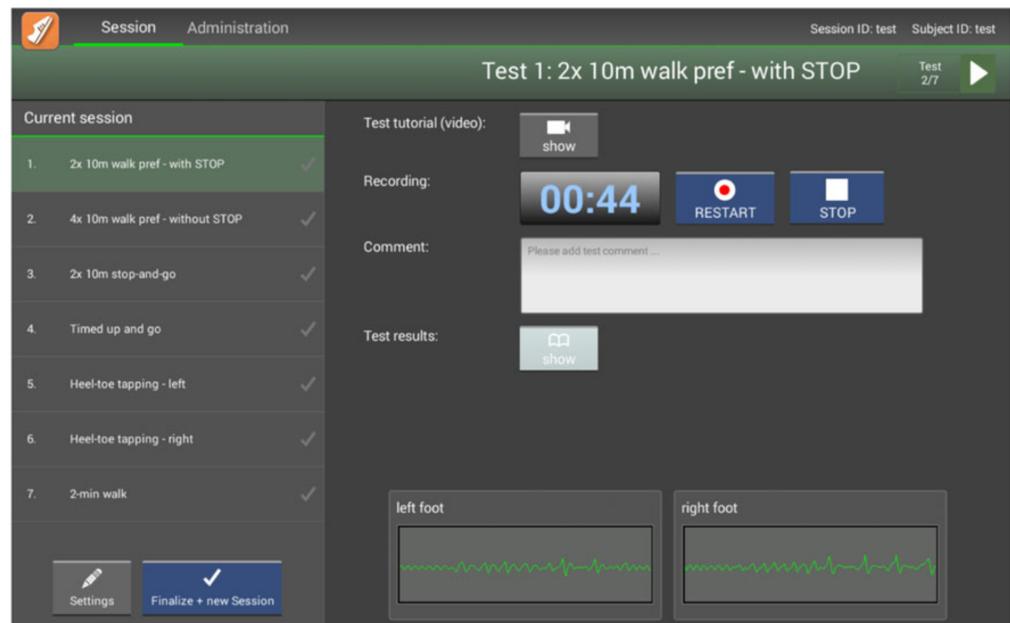


**Figure 1.** Scheme of the general methodology addressed in the study.

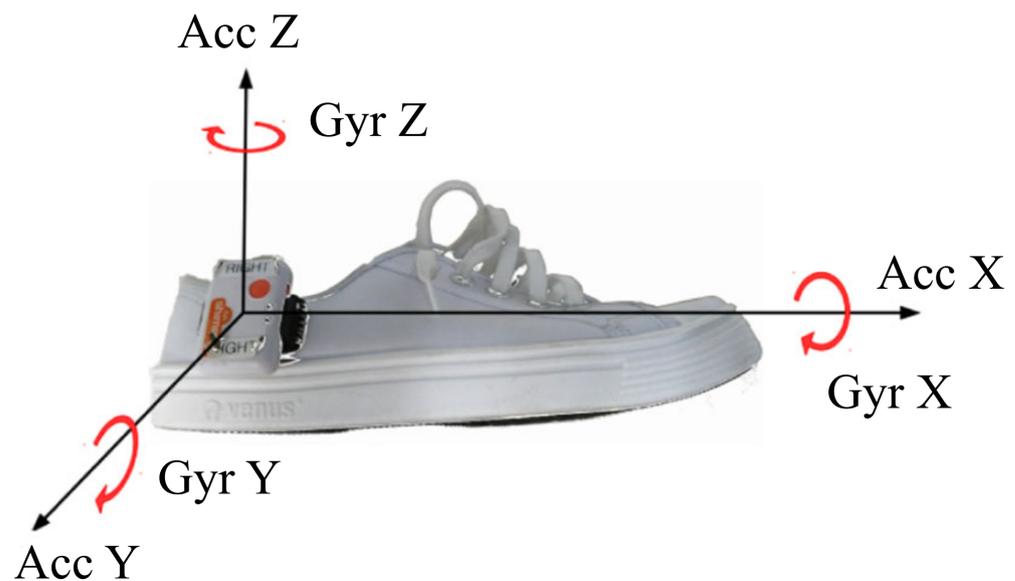
### 2.2. Data Collection and Participants

The eGaIT system (Embedded Gait analysis using Intelligent Technology), was used to record gait signals. eGaIT consists of a 6 degrees of freedom sensor to capture accelerometer and gyroscope signals. The accelerometer allows to measure the acceleration in a range of  $\pm 6$  g and 200 mV/g of sensibility. Gyroscope allows to measure rotational velocities in a range of  $\pm 500$  /s and  $\pm 2$  mV/g of sensitivity. A representation of the eGait system and the position of the sensor in the shoe is shown in Figure 2. Signals are collected using an Android Application.

The sensor used captures motion patterns at a sampling frequency ( $F_s$ ) of 102.4 Hz with 12 bits of resolution. Besides, this value enables capturing information with a resolution good enough to model low-frequency patterns such as those related to the patient's gait.



(a)



(b)

**Figure 2.** (a) Interface of the eGait software, (b) Location of the eGait sensor in the shoe.

In this study, two tasks were considered:

- i.  $2 \times 10$  m task:
  - The subject starts standing.
  - The subject walks 10 m in a straight line.
  - The subject stop.
  - The subject turns right and returns to the starting point.
- ii.  $4 \times 10$  m task:
  - The subject starts standing
  - The subject walks 10 m straight.
  - The subject turns right and returns to the starting point.
  - The subject turns right walks 10 m.
  - Finally, the subject turns right, again, and returns to the starting point.

The dataset used in this study consists of 134 recordings where 45 are PD patients and 89 HC subjects. The HC group is divided into two groups: 44 YHC subjects under 45 years old and 45 EHC subjects of people older than 45 years. In Table 1 the information of the dataset is presented. The age of the EHC group is balanced with respect to the PD participants.

**Table 1.** Details about the participants.

Group	Gender	# Subjects	Age & Range	MDS-UPDRS-III & Range
PD	Male	17	65.0 ± 10 & [41–82]	37.6 ± 21 & [8–82]
	Female	28	58.9 ± 11 & [29–75]	33.5 ± 21 & [9–106]
EHC	Male	23	63.3 ± 11 & [49–85]	-
	Female	22	58.9 ± 10 & [45–83]	-
YHC	Male	26	25.3 ± 5 & [21–42]	-
	Female	18	22.9 ± 3 & [19–32]	-

Age and MDS-UPDRS-III score are presented in terms of mean ± standard deviation. There are no significant differences in the age of PD vs EHC (t-student test, p-value << 0.05). The last column includes the MDS-UPDRS-III values associated to PD patients.

### 2.3. Convolutional Neural Network (CNN)

A CNN is a deep learning architecture typically used for image analysis where convolution and pooling layers are used with the aim to obtain relevant information of the input [29]. The main advantage of a CNN is that it requires minimal or sometimes no pre-processing for the input to implement the architecture. Let's define the input of a CNN as a tensor as follows:

$$\mathbf{X} \in \mathbb{R}^{p \times q \times r} \quad (1)$$

where  $p$ ,  $q$  and  $r$  correspond to the number of vertical pixels, horizontal pixels and channels of the image, respectively. The convolution process is performed between the input tensor  $\mathbf{X}$  and a convolutional filter, named kernel, represented as follow:

$$\mathbf{W} \in \mathbb{R}^{n \times n \times d} \quad (2)$$

where  $n$  is the size of the kernel and  $d$  is the number of kernels in the convolutional layer. The result of the convolution between  $\mathbf{X}$  and  $\mathbf{W}$  per channel produces a hidden representation  $\mathbf{H}$  as follows:

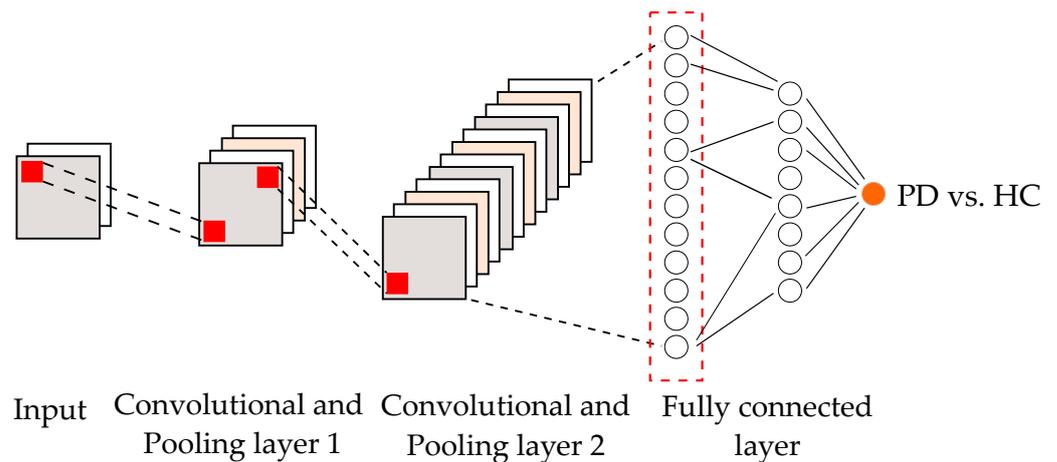
$$\mathbf{H} = \mathbf{X} * \mathbf{W} \quad (3)$$

where:

$$\mathbf{H} \in \mathbb{R}^{(p-n+1) \times (q-n+1) \times d} \quad (4)$$

Note that tensor  $\mathbf{H}$  represents the extracted features obtained from the input  $\mathbf{X}$ . A pooling layer is implemented after each convolution step. The pooling layer reduces the size of the hidden representation  $\mathbf{H}$ . One of the aims of the pooling layer is to reduce the computational cost required to process the information, in addition, it is useful to remove some invariant features [29]. Finally, a fully connected layer with  $h$  hidden units followed by an activation function is implemented to obtain the final decision of the classification process.

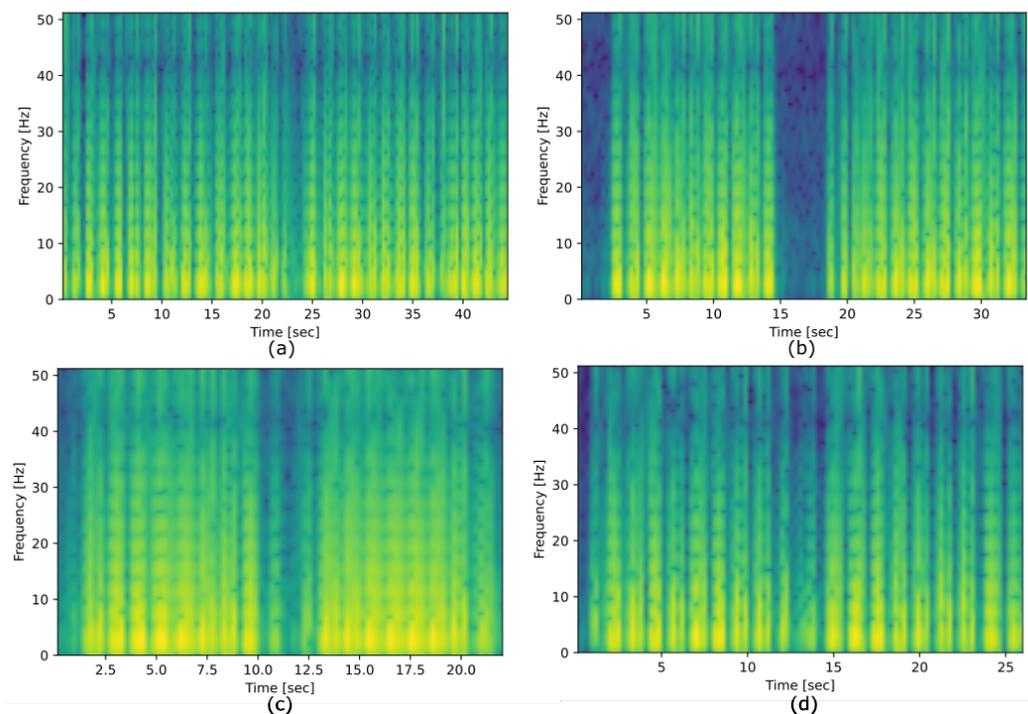
It is important to note that different CNN architectures can be created depending on the problem. Figure 3 presents an illustration of a CNN architecture with two convolutional hidden layers.



**Figure 3.** Illustration of a CNN with two convolutional layers.

For the case of gait signals, the CNN architecture corresponds to a two-dimensional (2D) CNN. The input to the CNN consists of  $r = 12$  channels when the two feet are considered. The channels have information of the accelerometer and gyroscope signals in the x, y, and z-axes.

With the aim to guarantee at least 3 quasi-periods in the gait signal, segments of 3 s are considered. The Short Time Fourier Transform (STFT) is computed to create the input to the CNN. Figure 4 shows four examples of STFT computed upon two PD patients (a and b), one EHC subject (c), and one YHC subject (d). In the four cases images are extracted from gyroscope signals (z-axis) of the left foot during the 2 × 10 task.



**Figure 4.** Resulting STFT computed to: (a) PD female patient, Lower limbs score: 50, Age: 75; (b) PD female patient, Lower limbs score: 10, Age: 65; (c) EHC female patient, Age: 50; (d) YHC female patient, Age: 20.

The CNN was trained using the stochastic gradient descent (SGD) algorithm. The loss function is the cross-entropy between the label of the training data  $y$  and the prediction  $\hat{y}$ . An Exponential Linear Unit (elu) is used as activation function for the convolutional layer. Dropout is included to avoid over-fitting in the training process. The architecture

of the CNN for this study includes two convolutional layers with max-pooling, dropout for regularization, and five fully connected hidden layers. A sigmoid activation function is used at the output. Figure 5 summarizes the details of the architecture.

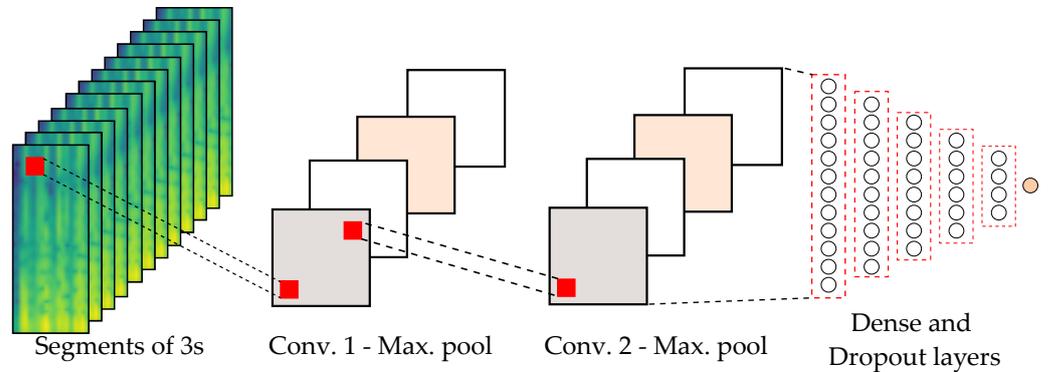


Figure 5. Architecture of the CNN implemented in this study.

2.4. Gate Recurrent Network, GRU

The paradigm of GRU was proposed in [30,31] as a variation to recurrent neural networks (RNN). A GRU is composed of two gates: update and reset, whose objective is to only pass relevant information through the network to improve the predictions. Among the advantages of the GRU over other recurrent networks are the fact that they require less memory, therefore their training process is faster. Figure 6 illustrates a single GRU unit.

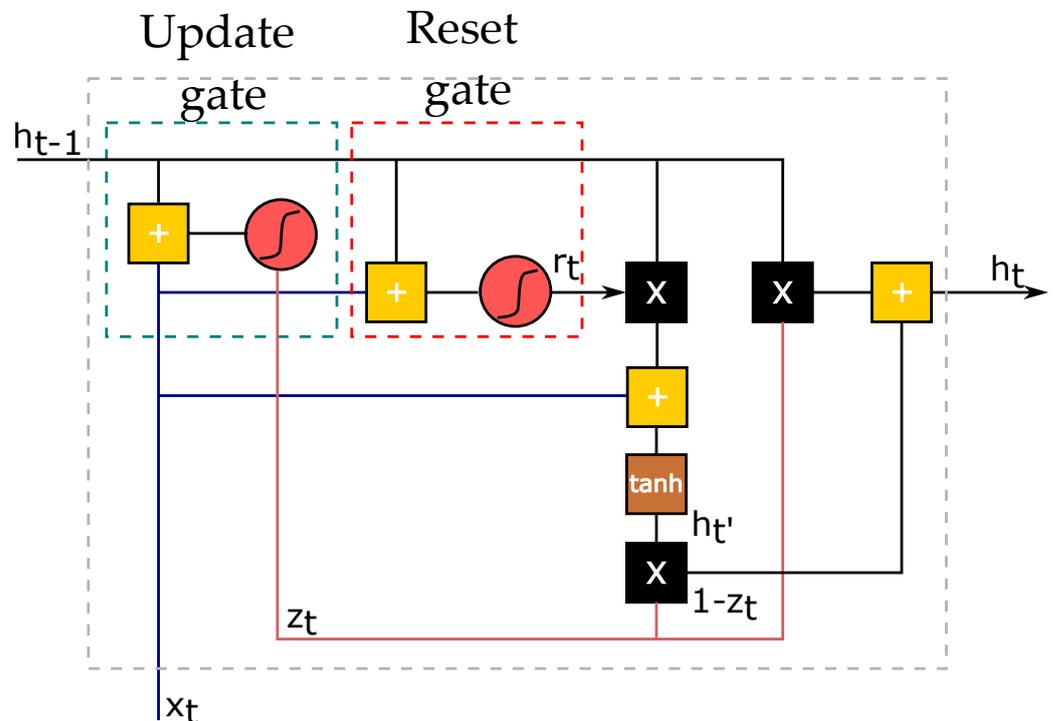


Figure 6. Single GRU unit.

The computation of a GRU starts with the calculation in the step time  $t$  for the update gate  $z_t$ , as follow:

$$z_t = \sigma(W^{(z)} x_t + U^{(z)} h_{t-1}) \tag{5}$$

$x_t$  is multiplied by  $W^{(z)}$ , which is its own weight. The same process is performed with  $h_{t-1}$ , which has information of the previous step time  $t - 1$  and is multiplied by its own weight  $U^{(z)}$ . A sigmoid activation function is applied to the sum of both products. The aim of the update gate is to define the information to be considered in the future.

The reset gate intends to find the information to be forgotten, in this case it is called  $r_t$ , defined as follows:

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1}) \quad (6)$$

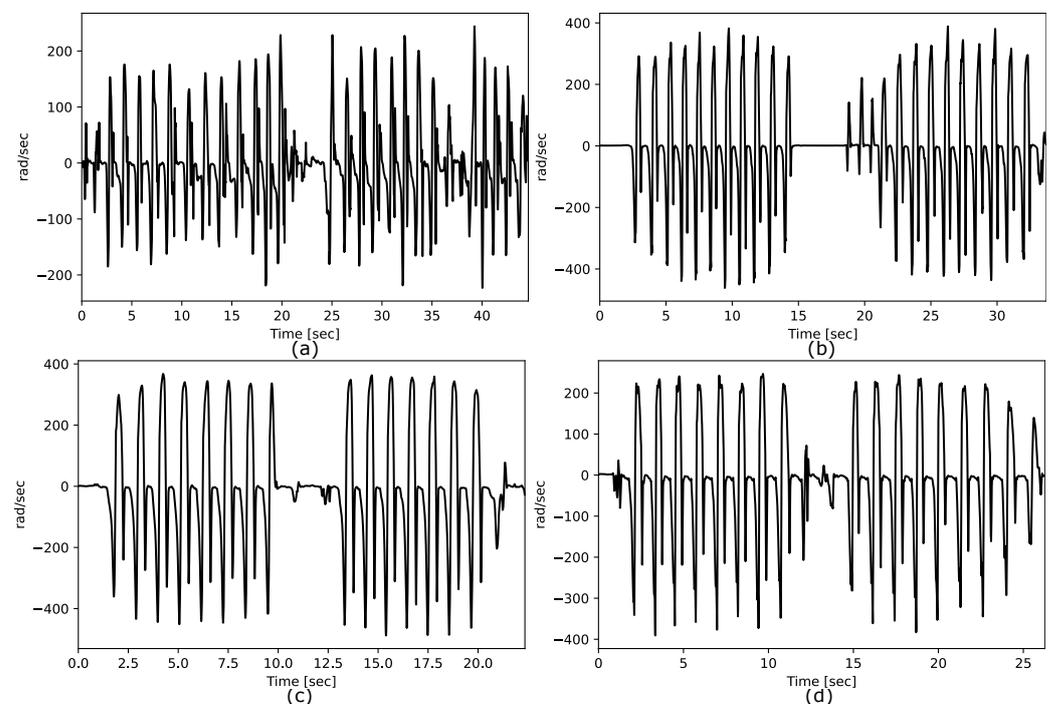
Which is similar to the equation of the update gate except for the weights. The current memory content  $h'_t$  is calculated as follow:

$$h'_t = \tanh(Wx_t + r_t \odot Uh_{t-1}) \quad (7)$$

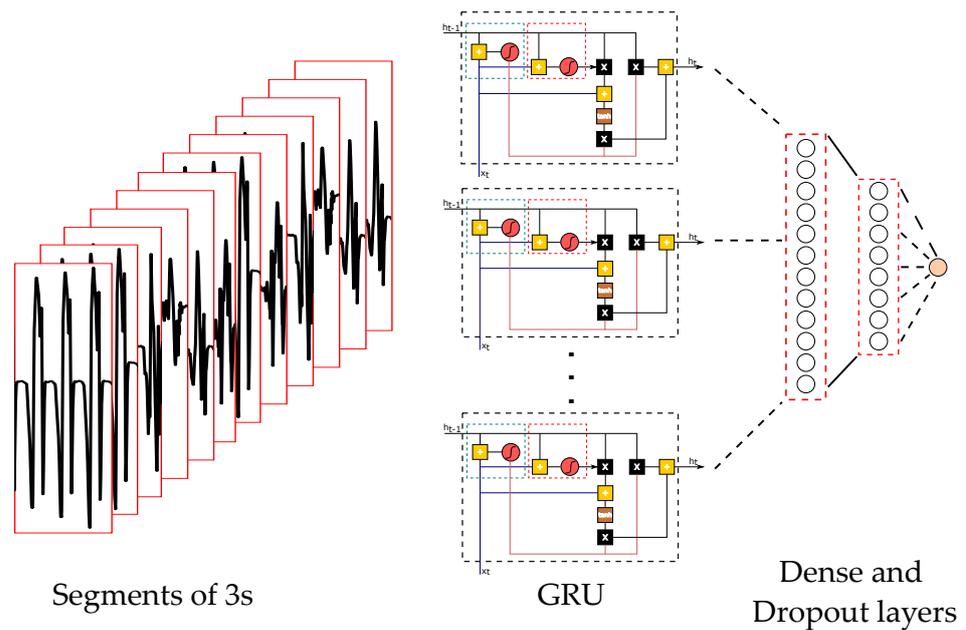
where  $\odot$  is the Hadamard matrix product. The final memory at the time step  $t$  is:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h'_t \quad (8)$$

A GRU architecture is able to process information of time series such as the one existing in raw gait signals. In this work, the input to the GRU consists of 12 raw signals captured with the IMU sensor. Figure 7 shows four examples of signals collected from two PD patients (a and b), one EHC subject (c), and one YHC subject (d). The GRU architecture implemented in this study is presented in Figure 8.



**Figure 7.** Comparison between the raw time series of: (a) PD female patient, Lower limbs score: 50, Age: 75; (b) PD female patient, Lower limbs score: 10, Age: 65; (c) EHC female patient, Age: 50; (d) YHC female patient, Age: 20.



**Figure 8.** Architecture of the GRU implemented in this study.

### 2.5. Training Process and Classification

A 5-fold cross-validation strategy was used to evaluate the proposed approach along the experiments. Four folds were used for training and one-fold for testing. Each experiment was repeated ten times and the reported results correspond to the average over those repetitions. Adam optimizer [32] and binary cross-entropy were used in the classification stage of all experiments.

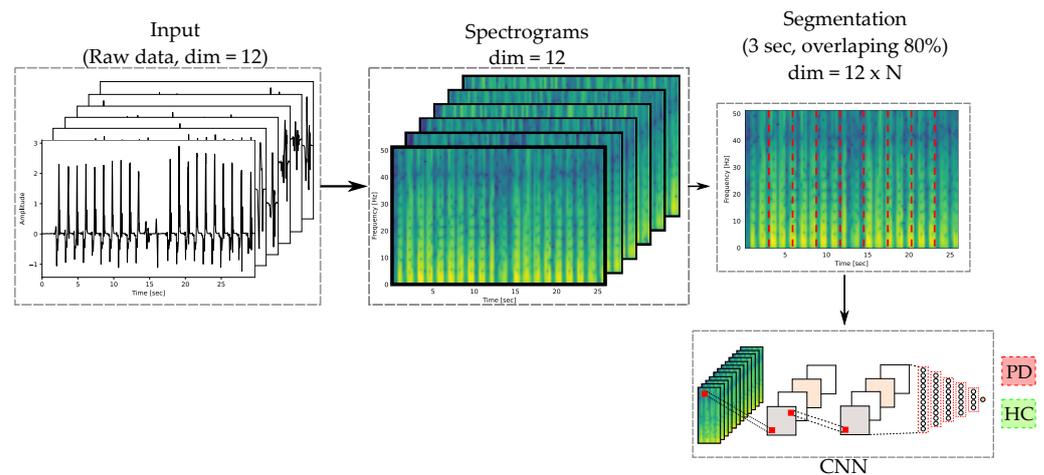
The acquisition of the acceleration data performed in this work is by-default normalized between  $-6$  g and  $+6$  g and the gyroscope signals between  $+500$   $^{\circ}/s$  and  $-500$   $^{\circ}/s$ . Therefore, it is not necessary to perform any additional normalization. Besides, the architecture of the Neural Network by itself performs an “internal batch” normalization according to the patterns that it is observing during the training process. Further details of the batch normalization can be found in [33].

## 3. Experiments and Results

Three different experiments were considered: only with the CNN, only with the GRU, and with the combination of both architectures. Each experiment considers two scenarios: PD vs. EHC and PD vs. YHC. The two gait tasks were considered independently. Results are reported in terms of accuracy (Acc), sensitivity (Sen), specificity (Spe), and Area Under ROC Curve (AUC) [34]. Two different accuracy values are reported in each experiment, accuracy in development refers to the result obtained within the 4 folds considered during the training process and accuracy in test refers to the result obtained in the external fold that did not participate in the optimization process. Standard deviation values appear because the experiments were repeated ten times independently to perform a fair evaluation of the proposed approach.

### 3.1. Classification with CNN

The general scheme of the proposed CNN architecture is presented in Figure 5. This approach includes two convolutional layers and five fully connected hidden layers, besides Max-pooling and dropout layers are included to avoid overfitting. Details of the implemented architecture are presented in Appendix A, Table A1. Figure 9 shows details of the pre-processing stages applied in this experiment before feeding the CNN architecture. Notice that the raw input contains 12 channels, therefore there is the same number of spectrograms before the segmentation step. The STFT is computed upon segments of 3s per channel with an overlap of 80%.



**Figure 9.** Methodology for the classification based on CNNs.

Note that when both feet are considered the dimension of the input is  $r = 12$  which corresponds to three accelerometer signals and three gyroscope signals per foot. Table 2 shows the results obtained in the classification with the  $2 \times 4$  task. The highest accuracy in test is 82.4% for the PD vs. EHC scenario, while 87.5% for PD vs. YHC. Table 3 presents the results obtained with the  $4 \times 10$  task. Notice that in this case the results are higher compared to those obtained in the previous task. In the PD vs. EHC scenario, the highest accuracy in test is 82.7% while in PD vs. YHC it is 92.1%. This improvement could be associated to the fact that this task is longer than the previous one, therefore there are more chances to observe abnormal patterns in the gait signals. Also, a longer task likely produces more fatigue in the participants, especially the patients.

**Table 2.** Results using CNN and  $2 \times 10$  m task.

Scenario	Foot	Acc. Dev. [%]	Acc. Test [%]	Sen [%]	Spe [%]	AUC
PD vs. EHC	Left	79.1 ± 3	81.3 ± 4	74.3 ± 4	84.3 ± 4	0.87
	Right	79.5 ± 4	80.1 ± 4	79.3 ± 4	80.3 ± 4	0.88
	Both	80.2 ± 3	82.4 ± 3	76.2 ± 3	87.3 ± 4	0.88
PD vs. YHC	Left	83.9 ± 5	85.4 ± 5	87.2 ± 3	88.0 ± 4	0.90
	Right	81.7 ± 4	83.8 ± 4	88.1 ± 3	85.0 ± 4	0.90
	Both	84.3 ± 5	87.5 ± 4	88.7 ± 3	87.1 ± 4	0.91

Acc. Test: accuracy in test, Acc. Dev.: accuracy in development, Sen: Sensitivity, Spe: Specificity, AUC: Area under the ROC curve.

**Table 3.** Results using CNN and  $4 \times 10$  m task.

Scenario	Foot	Acc. Dev. [%]	Acc. Test [%]	Sen [%]	Spe [%]	AUC
PD vs. EHC	Left	83.2 ± 5	85.3 ± 4	83.0 ± 6	91.3 ± 3	0.92
	Right	83.3 ± 4	79.5 ± 6	83.3 ± 5	82.9 ± 6	0.91
	Both	83.5 ± 6	82.7 ± 4	85.6 ± 2	87.9 ± 2	0.92
PD vs. YHC	Left	87.4 ± 4	88.5 ± 4	88.3 ± 6	90.5 ± 5	0.95
	Right	85.3 ± 6	86.5 ± 4	91.4 ± 4	87.9 ± 5	0.94
	Both	88.5 ± 5	92.1 ± 5	91.2 ± 3	88.4 ± 5	0.95

Acc. Test: accuracy in test, Acc. Dev.: accuracy in development, Sen: Sensitivity, Spe: Specificity, AUC: Area under the ROC curve.

### 3.2. Classification with GRU

The general scheme of the GRU architecture used in this work is presented in Figure 8. In this case, the raw input with 12 channels is first segmented into windows of 3s with 80%

overlap. Each window is segmented into  $N$  number of steps. Details of the implemented architecture are presented in Appendix A, Table A2. Notice that since every person can produce a different number of steps during the time window, the number of steps needs to be variable in order to make it the method robust and flexible. This segmentation procedure is shown in Figure 10. Table 4 shows results of the two classification scenarios: PD vs. EHC and PD vs. YHC when the  $2 \times 10$  task is considered. Similarly, Table 5 includes results obtained with the  $4 \times 10$  m task. Note that the GRU architecture yields better results in most of the experiments, compared to those obtained with the CNN. Similar to what we observed in the previous experiment, the  $4 \times 10$  task yields better results. In the case of the classification between PD patients and EHC subjects the highest accuracy was 82.7% and in the case of PD vs. YHC the best result was 92.5%. Signals of both feet provided the best results in both scenarios, as it was also observed in the experiment with CNN.

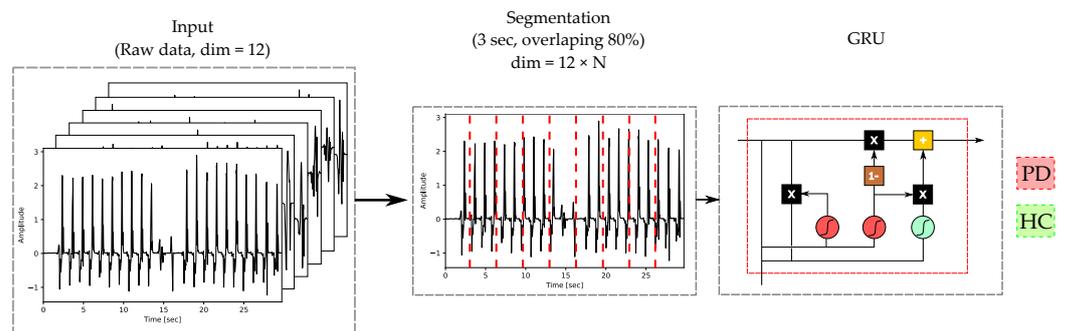


Figure 10. Methodology used for the classification with a GRU architecture and both feet.

Table 4. Results using GRU and  $2 \times 10$  m task.

Scenario	Foot	Acc. Dev. [%]	Acc. Test [%]	Sen [%]	Spe [%]	AUC
PD vs. EHC	Left	75.7 ± 4	75.2 ± 4	70.6 ± 6	74.3 ± 4	0.82
	Right	74.9 ± 4	73.3 ± 5	75.7 ± 4	78.8 ± 6	0.81
	Both	78.5 ± 6	78.7 ± 5	72.8 ± 3	83.3 ± 5	0.84
PD vs. YHC	Left	84.7 ± 4	83.4 ± 5	88.4 ± 4	87.2 ± 5	0.89
	Right	81.3 ± 6	82.6 ± 4	87.7 ± 5	84.4 ± 4	0.88
	Both	86.7 ± 5	84.5 ± 4	89.1 ± 3	87.9 ± 5	0.90

Acc. Test: accuracy in test, Acc. Dev.: accuracy in development, Sen: Sensitivity, Spe: Specificity, AUC: Area under the ROC curve

Table 5. Results using GRU and  $4 \times 10$  m task.

Scenario	Foot	Acc. Dev. [%]	Acc. Test [%]	Sen [%]	Spe [%]	AUC
PD vs. EHC	Left	77.5 ± 4	76.6 ± 3	75.1 ± 2	80.8 ± 3	0.88
	Right	75.6 ± 4	74.5 ± 3	77.3 ± 5	82.1 ± 5	0.87
	Both	84.1 ± 4	82.7 ± 6	83.8 ± 4	86.3 ± 6	0.92
PD vs. YHC	Left	91.8 ± 6	90.3 ± 4	89.3 ± 6	90.1 ± 6	0.94
	Right	89.5 ± 5	88.6 ± 6	90.4 ± 4	92.4 ± 5	0.94
	Both	93.7 ± 4	92.5 ± 5	92.6 ± 3	94.1 ± 6	0.96

Acc. Test: accuracy in test, Acc. Dev.: accuracy in development, Sen: Sensitivity, Spe: Specificity, AUC: Area under the ROC curve.

### 3.3. Classification with CNN + GRU

To consider temporal and frequency information of the gait signals simultaneously, a novel strategy is proposed in this work. Details of the implemented architecture are presented in Appendix A, Table A3. The input to the proposed architecture are the spectrograms and also the raw signals. Figure 11 shows how the two approaches can be

considered simultaneously to perform the final decision of whether a subject belongs to the PD or HC group.

Results presented in Tables 6 and 7 show that this methodology yields results slightly better than those obtained with the GRU model. When observing the  $4 \times 10$  m task, the highest accuracy in the PD vs. EHC scenario was 83.7%, while in PD vs. YHC the result was 92.7%.

Although the results of the CNN + GRU model are not much higher than those obtained with the GRU architecture, we believe that this is due to the small amount of data considered in this work. We are currently working on the collection of more data to validate whether these kinds of architectures yield results significantly better than others where only temporal or frequency information is considered separately.

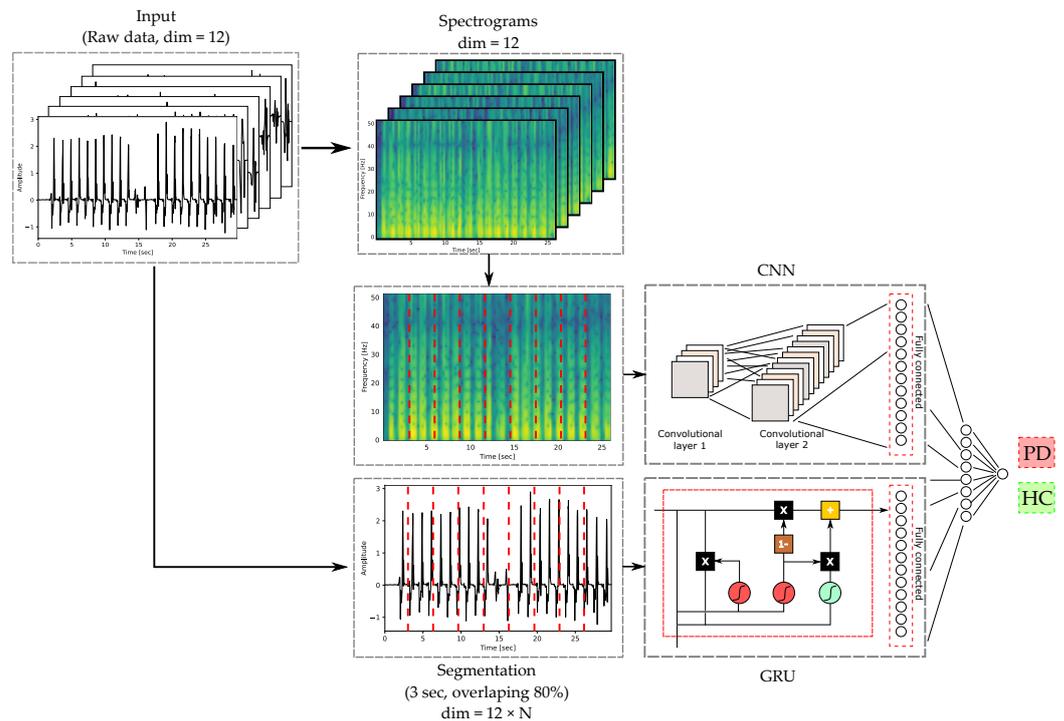


Figure 11. Proposed methodology considering a CNN + GRU architecture

Table 6. Results using CNN + GRU and  $2 \times 10$  m task.

Scenario	Foot	Acc. Dev. [%]	Acc. Test [%]	Sen [%]	Spe [%]	AUC
PD vs. EHC	Left	$76.7 \pm 5$	$74.6 \pm 6$	$71.3 \pm 5$	$75.2 \pm 4$	0.83
	Right	$75.2 \pm 3$	$73.6 \pm 5$	$74.9 \pm 6$	$79.3 \pm 4$	0.81
	Both	$80.4 \pm 5$	$78.7 \pm 4$	$73.3 \pm 5$	$84.1 \pm 5$	0.85
PD vs. YHC	Left	$85.4 \pm 5$	$83.7 \pm 4$	$89.2 \pm 5$	$86.9 \pm 4$	0.88
	Right	$83.5 \pm 6$	$82.0 \pm 6$	$88.4 \pm 4$	$85.2 \pm 6$	0.87
	Both	$88.5 \pm 4$	$85.7 \pm 5$	$90.2 \pm 3$	$86.5 \pm 5$	0.89

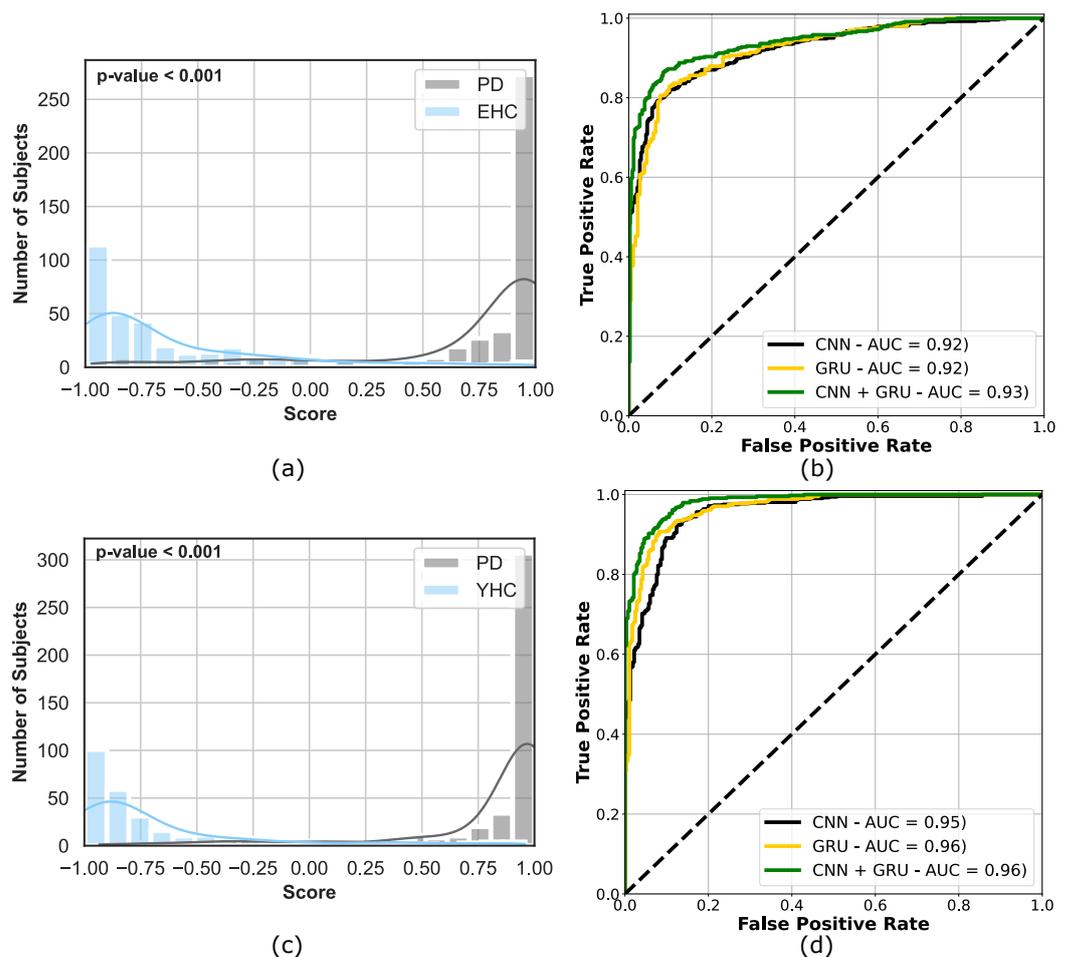
Acc. Test: accuracy in test, Acc. Dev.: accuracy in development, Sen: Sensitivity, Spe: Specificity, AUC: Area under the ROC curve.

**Table 7.** Results using CNN + GRU and  $4 \times 10$  m task.

Scenario	Foot	Acc. Dev. [%]	Acc. Test [%]	Sen [%]	Spe [%]	AUC
PD vs. EHC	Left	$76.7 \pm 6$	$74.7 \pm 4$	$75.2 \pm 5$	$77.8 \pm 5$	0.84
	Right	$74.2 \pm 7$	$72.4 \pm 6$	$74.2 \pm 4$	$78.5 \pm 5$	0.82
	Both	$85.7 \pm 3$	$83.7 \pm 4$	$84.7 \pm 5$	$85.9 \pm 4$	0.93
PD vs. YHC	Left	$90.9 \pm 6$	$91.1 \pm 3$	$89.5 \pm 6$	$91.1 \pm 5$	0.94
	Right	$90.3 \pm 5$	$87.9 \pm 5$	$90.6 \pm 3$	$92.2 \pm 6$	0.95
	Both	$93.3 \pm 4$	$92.7 \pm 4$	$91.9 \pm 3$	$94.4 \pm 5$	0.96

Acc. Test: accuracy in test, Acc. Dev.: accuracy in development, Sen: Sensitivity, Spe: Specificity, AUC: Area under the ROC curve.

Figure 12 presents the best results of the  $4 \times 10$  m task in the two scenarios and the three experiments. The distribution of the scores/posteriors obtained in the classification stage of each scenario (PD vs. EHC and PD vs. YHC) are included in Figures 12a,c. Although both scenarios are clearly separable due to the robustness of the proposed approach based on a GRU+CNN architecture, it can be observed that there is more overlap in the first scenario. Regarding Figures 12b,d, they include the ROC curves resulting from the three experiments (CNN, GRU, and CNN + GRU) in each scenario. Notice that in both cases the CNN + GRU architecture yields the highest AUC values, which confirms its superiority compared to other approaches.



**Figure 12.** Comparison of the best results considering the  $4 \times 10$  m task and both feet. (a) Distribution of the scores in PD vs. EHC using the GRU + CNN architecture. (b) Comparison of the ROC curves in the PD vs. EHC scenario. (c) Distribution of the scores in PD vs. YHC using the GRU + CNN architecture. (d) Comparison of the ROC curves in the PD vs. YHC scenario.

Table 8 summarizes the results obtained with the different architectures with the  $4 \times 10$  task.

**Table 8.** Summary of the accuracy in test for the best results considering the  $4 \times 10$  m task.

Scenario	Foot	CNN	GRU	CNN + GRU
PD vs. EHC	Left	$85.3 \pm 4$	$76.6 \pm 3$	$74.7 \pm 4$
	Right	$79.5 \pm 6$	$74.5 \pm 3$	$72.4 \pm 6$
	Both	$82.7 \pm 4$	$82.7 \pm 6$	$83.7 \pm 4$
PD vs. YHC	Left	$88.5 \pm 4$	$90.3 \pm 4$	$91.1 \pm 3$
	Right	$86.5 \pm 4$	$88.6 \pm 6$	$87.9 \pm 5$
	Both	$92.1 \pm 5$	$92.5 \pm 5$	$92.7 \pm 4$

Data correspond to the accuracy in test by each architecture.

#### 4. Discussion

Three different deep learning architectures were considered for the classification of PD vs. HC subjects. Two subgroups of healthy healthy subjects were included, elderly (EHC) and young (YHC). The architectures evaluated in this work correspond to the state of the art in gait analysis and are based on CNNs and GRUs. Previous works suggest that CNN architectures are a suitable approach when considering the STFT of gait signals [16,35]. In [16] the use of a CNN is introduced to classify gait signals and the authors reported an accuracy of 88%. We found comparable results in our present study, where the best accuracy obtained with the CNN is 82.7% considering the YHC group and the best result with the EHC group is 82.4%. The algorithm performed better always when the  $4 \times 10$  m task was considered. We believe that this is because a longer task allows to capture more information about possible abnormal gait patterns which provides better classification results and also improves the generalization of the algorithms. The GRU model presented here allowed modelling information of gait signals without any pre-processing. In [36] the authors explored the use of RNNs to predict gait phases. They showed that these kinds of architectures are promising for the analysis of gait signals. In our experiments, we could observe that the GRU architecture improved the accuracy in most of the experiments. Similar to the CNN case, better results were obtained with the  $4 \times 10$  m task. The classification of PD vs. EHC yields an accuracy of 82.7%, while with the YHC group, the accuracy is 92.5%. Besides CNN and GRU architectures evaluated individually, in this paper we proposed a model where CNN and GRU architectures are considered together. We hypothesized that results could improve when temporal and frequency information were combined in a single model. We found that, in general, results were similar to those obtained with the GRU architecture, with accuracies of 83.7% and 92.7% in the PD vs. EHC and PD vs. YHC scenarios, respectively. We believe that the results of the CNN + GRU model were not higher due to the small amount of data that we could consider here. Further research with a larger group of participants is required to validate whether this could lead to better results. We consider that this work is a step forward in the development of deep learning models for the automatic classification of PD patients.

In addition, it is necessary to consider other deep learning architectures seeking to improve the results, such as transfer learning, data argumentation and combinations of classifiers. Perhaps the most realistic approach would be to do transfer learning based on existing datasets, however, it is necessary to perform the experiments to raise strong conclusions.

#### 5. Conclusions

GRU architectures clearly yielded better results than the CNN ones and this is likely due to the fact that temporal information is incorporated when the first approach is considered. Besides, we validated that the combination of CNN and GRU methods are suitable and provide similar results to those observed with GRUs only. Although we expected to find higher accuracies with the combination of methods, it was not possible to prove our hypothesis. We believe that it was due to the small amount of data considered in this study.

Regarding the comparison of gait tasks, we could validate that the  $4 \times 10$  m task is more suitable and systematically yields better results than the  $2 \times 10$  m task. Very likely this is because longer tasks allow us to collect more information and also give more chances to observe abnormal patterns in the gait signals.

Besides the evaluation of state-of-the-art deep learning architectures, the results obtained in this paper are comparable to others reported in the literature, so we believe that this study is a contribution to the topic gait analysis in PD patients.

We are aware that one of the limitations of this study is the small amount of data. We expect to perform experiments with more participants in the near future to make it possible to validate further hypotheses.

**Author Contributions:** Conceptualization, H.A.C.-C. and J.R.O.-A.; methodology, H.A.C.-C. and J.R.O.-A.; software, H.A.C.-C. and P.A.P.-T.; validation, H.A.C.-C. and J.R.O.-A.; formal analysis, H.A.C.-C. and J.R.O.-A.; investigation, H.A.C.-C. and J.R.O.-A.; resources, J.R.O.-A.; data curation, H.A.C.-C., P.A.P.-T. and J.R.O.-A.; writing—original draft preparation, H.A.C.-C., P.A.P.-T. and J.R.O.-A.; visualization, H.A.C.-C.; supervision, J.R.O.-A.; project administration, J.R.O.-A.; funding acquisition, J.R.O.-A. All authors have read and agreed to the published version of the manuscript.

**Funding:** H.A.C.-C. is under grants of the Colombian Ministry of Science “Programa de Becas de Excelencia Doctoral del Bicentenario—Cohorte 1” grant # 20230017-19-20. This work was also funded by CODI at UdeA grant # PRG2020-34068.

**Institutional Review Board Statement:** This study was approved by the Ethical Research Committee of the University of Antioquia and according to the Helsinki declaration (1964) and its later amendments.

**Informed Consent Statement:** Informed consent was obtained from all the participants of the study.

**Data Availability Statement:** Not applicable.

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

## Appendix A

Details of the architectures used in the models.

**Table A1.** Details of the CNN architecture considering both feet (12 channels).

Layer	Output Shape	Number of Parameters
Input	(12, 257, 12)	0
Convolution 2D	(10, 255, 8)	872
Convolution 2D	(8,253,16)	1168
Max pooling	(4,126,16)	0
Flatten	2016	0
Dense 1	256	2,064,640
Dropout	256	0
Dense 2	128	32,896
Dropout	128	0
Dense 3	64	8256
Dropout	64	0
Dense 4	32	2080
Dropout	32	0
Dense 5	16	528
Dropout	16	0
Output	1	17

**Table A2.** Details of the GRU architecture considering both feet (12 channels).

Layer	Output Shape	Number of Parameters
Input	(12, 306)	0
GRU	128	148,608
Dense 1	256	8256
Dropout	256	0
Dense 2	128	1040
Dropout	128	0
Output	1	17

For the CNN + GRU architecture we combined the dense 5th layer of the CNN architecture with the dense 2nd layer of the GRU architecture as follow:

**Table A3.** Details of the CNN + GRU architecture considering both feet (12 channels).

Layer	Output Shape	Number of Parameters
Input	Dense 5th of CNN and dense 2nd of GRU	0
Dense	4	36
Output	1	5

## References

1. Cacabelos, R. Parkinson's disease: From pathogenesis to pharmacogenomics. *Int. J. Mol. Sci.* **2017**, *18*, 551. <https://doi.org/10.3390/ijms18030551>.
2. Pahwa, R.; Lyons, K.E. *Handbook of Parkinson's Disease*, 5 ed.; CRC Press: Boca Raton, FL, USA, 2013.
3. Jankovic, J.; Tolosa, E. *Parkinson's Disease & Movement Disorders*, 6 ed.; Wolters Kluwer: Alphen aan den Rijn, Netherlands 2015.
4. Kressig, R.W.; Gregor, R.J.; Oliver, A.; Waddell, D.; Smith, W.; O'Grady, M.; Curns, A.T.; Kutner, M.; Wolf, S.L. Temporal and spatial features of gait in older adults transitioning to frailty. *Gait Posture* **2004**, *20*, 30–35. [https://doi.org/10.1016/S0966-6362\(03\)00089-4](https://doi.org/10.1016/S0966-6362(03)00089-4).
5. Ellis, T.; Cavanaugh, J.T.; Earhart, G.M.; Ford, M.P.; Foreman, K.B.; Dibble, L.E. Which measures of physical function and motor impairment best predict quality of life in Parkinson's disease? *Park. Relat. Disord.* **2011**, *17*, 693–697. <https://doi.org/10.1016/j.parkreldis.2011.07.004>.
6. Hannink, J.; Kautz, T.; Pasluosta, C.F.; Gasmann, K.G.; Klucken, J.; Eskofier, B.M. Sensor-Based Gait Parameter Extraction with Deep Convolutional Neural Networks. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 85–93. <https://doi.org/10.1109/JBHI.2016.2636456>.
7. Sheinerman, K.S.; Umansky, S.R. Early detection of neurodegenerative diseases: Circulating brain-enriched microRNA. *Cell Cycle* **2013**, *12*, 1–2. <https://doi.org/10.4161/cc.23067>.
8. De Rijk, M.C.; Launer, L.J.; Berger, K.; Breteler, M.M.; Dartigues, J.F.; Baldereschi, M.; Fratiglioni, L.; Lobo, A.; Martinez-Lage, J.; Trenkwalder, C.; et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. In *Neurologic Diseases in the Elderly Research Group; Neurology* **2000**; *54*, pp. 21–23. <https://doi.org/10.1212/WNL.54.11.21A>.
9. Pringsheim, T.; Jette, N.; Frolkis, A.; Steeves, T. The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Mov. Disord.* **2014**, *29*, 1583–1590. <https://doi.org/10.1002/mds.25945>.
10. Ou, Z.; Pan, J.; Tang, S.; Duan, D.; Yu, D.; Nong, H.; Wang, Z. Global trends in the incidence, prevalence, and years lived with disability of parkinson's disease in 204 countries/territories from 1990 to 2019. *Front. Public Health* **2021**, *9*, 776847.
11. Goetz, C.G.; Tilley, B.C.; Shaftman, S.R.; Stebbins, G.T.; Fahn, S.; Martinez-Martin, P.; Poewe, W.; Sampaio, C.; Stern, M.B.; Dodel, R. et al. Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov. Disord.* **2008**, *23*, 2129–2170. <https://doi.org/10.1002/mds.22340>.
12. Roth, N.; Küderle, A.; Ullrich, M.; Gladow, T.; Marxreiter, F.; Klucken, J.; Eskofier, B.; Kluge, F. Hidden Markov Model based Stride Segmentation on Unsupervised Free-living Gait Data in Parkinson's Disease Patients. *J. Neuroeng. Rehabil.* **2021**, *18*, 1–15 <https://doi.org/10.21203/rs.3.rs-269965/v1>.
13. Carvajal-Castaño, H.A.; Lemos-Duque, J.D.; Orozco-Arroyave, J.R. Effective detection of abnormal gait patterns in Parkinson's disease patients using kinematics, nonlinear, and stability gait features. *Hum. Mov. Sci.* **2022**, *81*, 102891. <https://doi.org/10.1016/j.humov.2021.102891>.
14. Pérez-Toro, P.A.; Vásquez-Correa, J.C.; Arias-Vergara, T.; Garcia-Ospina, N.; Orozco-Arroyave, J.R.; Nöth, E. A Non-linear Dynamics Approach to Classify Gait Signals of Patients with Parkinson's Disease. *Commun. Comput. Inf. Sci.* **2018**, *916*, 268–278. <https://doi.org/10.1007/978-3-030-00353-1-24>.

15. Pérez-Toro, P.A.; Vásquez-Correa, J.C.; Arias-Vergara, T.; Nöth, E.; Orozco-Arroyave, J.R. Nonlinear dynamics and Poincaré sections to model gait impairments in different stages of Parkinson's disease. *Nonlinear Dyn.* **2020**, *100*, 3253–3276. <https://doi.org/10.1007/s11071-020-05691-7>.
16. Vásquez-Correa, J.C.; Arias-Vergara, T.; Orozco-Arroyave, J.R.; Eskofier, B.; Klucken, J.; Nöth, E. Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1618–1630. <https://doi.org/10.1109/JBHI.2018.2866873>.
17. González, I.; López-Nava, I.H.; Fontecha, J.; Muñoz-Meléndez, A.; Pérez-SanPablo, A.I.; Quiñones-Urióstegui, I. Comparison between passive vision-based system and a wearable inertial-based system for estimating temporal gait parameters related to the GAITRite electronic walkway. *J. Biomed. Inform.* **2016**, *62*, 210–223. <https://doi.org/10.1016/j.jbi.2016.07.009>.
18. Taborri, J.; Palermo, E.; Rossi, S.; Cappa, P. Gait partitioning methods: A systematic review. *Sensors* **2016**, *16*, 40–42. <https://doi.org/10.3390/s16010066>.
19. Channa, A.; Popescu, N.; Ciobanu, V. Wearable solutions for patients with parkinson's disease and neurocognitive disorder: A systematic review. *Sensors* **2020**, *20*, 2713. <https://doi.org/10.3390/s20092713>.
20. Wahid, F.; Begg, R.K.; Hass, C.J.; Halgamuge, S.; Ackland, D.C. Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1794–1802. <https://doi.org/10.1109/JBHI.2015.2450232>.
21. Ramirez-Bautista, J.A.; Huerta-Ruelas, J.A.; Chaparro-Cárdenas, S.L.; Hernández-Zavala, A. A Review in Detection and Monitoring Gait Disorders Using In-Shoe Plantar Measurement Systems. *IEEE Rev. Biomed. Eng.* **2017**, *10*, 299–309. <https://doi.org/10.1109/RBME.2017.2747402>.
22. Wang, F.; Yan, L.; Xiao, J. A new hidden markov model algorithm to detect human gait phase based on information fusion combining inertial with plantar pressure. *Sensors Mater.* **2019**, *31*, 2637–2655. <https://doi.org/10.18494/SAM.2019.2431>.
23. Ghassemi, N.H.; Hannink, J.; Martindale, C.F.; Gaßner, H.; Müller, M.; Klucken, J.; Eskofier, B.M. Segmentation of gait sequences in sensor-based movement analysis: A comparison of methods in Parkinson's disease. *Sensors* **2018**, *18*, 145. <https://doi.org/10.3390/s18010145>.
24. Renggli, D.; Graf, C.; Tachatos, N.; Singh, N.; Meboldt, M.; Taylor, W.R.; Stieglitz, L.; Schmid Daners, M. Wearable Inertial Measurement Units for Assessing Gait in Real-World Environments. *Front. Physiol.* **2020**, *11*, 90. <https://doi.org/10.3389/fphys.2020.00090>.
25. Hubble, R.P.; Naughton, G.A.; Silburn, P.A.; Cole, M.H. Wearable sensor use for assessing standing balance and walking stability in people with Parkinson's disease: A systematic review. *PLoS ONE* **2015**, *10*, e0123705. <https://doi.org/10.1371/journal.pone.0123705>.
26. Vienne, A.; Barrois, R.P.; Buffat, S.; Ricard, D.; Vidal, P.P. Inertial sensors to assess gait quality in patients with neurological disorders: A systematic review of technical and analytical challenges. *Front. Psychol.* **2017**, *8*, 817. <https://doi.org/10.3389/fpsyg.2017.00817>.
27. Sprager, S.; Juric, M.B. Inertial sensor-based gait recognition: A review. *Sensors* **2015**, *15*, 22089–22127. <https://doi.org/10.3390/s150922089>.
28. Tunca, C.; Pehlivan, N.; Ak, N.; Arnrich, B.; Salur, G.; Ersoy, C. Inertial sensor-based robust gait analysis in non-hospital settings for neurological disorders. *Sensors* **2017**, *17*, 825. <https://doi.org/10.3390/s17040825>.
29. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: <http://www.deeplearningbook.org> (accessed on 30 June 2022).
30. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder–decoder approaches. In Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 103–111. <https://doi.org/10.3115/v1/W14-4012>.
31. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1724–1734. <https://doi.org/10.3115/v1/D14-1179>.
32. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. <https://doi.org/10.48550/ARXIV.1412.6980>.
33. Han, J.; Pei, J.; Tong, H. *Data Mining*, 4th ed.; Morgan Kaufmann Publishers, Burlington, Massachusetts, United States, 2017; pp. xxiii–xxxii. <https://doi.org/10.1016/B978-0-12-804291-5.00017-9>.
34. Orozco, J.R.; Vasquez, J.; Noeth, E. Current methods and new trends in signal processing and pattern recognition for the automatic assessment of motor impairments: The case of Parkinson's disease. *Neurol. Disord. Imaging Phys.* **2020**, *5*, 8-1 to 8-57. <https://doi.org/10.1088/978-0-7503-2723-7ch8>.
35. Ashfaq Mostafa, T.; Soltaninejad, S.; McIsaac, T.L.; Cheng, I. A Comparative Study of Time Frequency Representation Techniques for Freeze of Gait Detection and Prediction. *Sensors* **2021**, *21*, 6446. <https://doi.org/10.3390/s21196446>.
36. Sarshar, M.; Polturi, S.; Schega, L. Gait phase estimation by using lstm in imu-based gait analysis—Proof of concept. *Sensors* **2021**, *21*, 5749. <https://doi.org/10.3390/s21175749>.