Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait

Imanne El Maachi[a], Guillaume-Alexandre Bilodeau[a], Wassim Bouachir[b]

[a]*Polytechnique Montreal,*
2900 boul. Edouard Montpetit, Montreal (Qc), H3T 1J4, Canada
{`imanne.el-maachi, guillaume-alexandre.bilodeau`}`@polymtl.ca`

[b]*TÉLUQ University,*
5800 rue Saint-Denis, bur. 1105, Montreal (Qc), H2S 3L5, Canada
`wassim.bouachir@teluq.ca`

Abstract

Diagnosing Parkinson's disease is a complex task that requires the evaluation of several motor and non-motor symptoms. During diagnosis, gait abnormalities are among the important symptoms that physicians should consider. However, gait evaluation is challenging and relies on the expertise and subjectivity of clinicians. In this context, the use of an intelligent gait analysis algorithm may assist physicians in order to facilitate the diagnosis process. This paper proposes a novel intelligent Parkinson detection system based on deep learning techniques to analyze gait information. We used 1D convolutional neural network (1D-Convnet) to build a Deep Neural Network (DNN) classifier. The proposed model processes 18 1D-signals coming from foot sensors measuring the vertical ground reaction force (VGRF). The first part of the network consists of 18 parallel 1D-Convnet corresponding to system inputs. The second part is a fully connected network that connects the concatenated outputs of the 1D-Convnets to obtain a final classification. We tested our algorithm in Parkinson's detection and in the prediction of

the severity of the disease with the Unified Parkinson's Disease Rating Scale (UPDRS). Our experiments demonstrate the high efficiency of the proposed method in the detection of Parkinson disease based on gait data. The proposed algorithm achieved an accuracy of 98.7%. To our knowledge, this is the state-of-the-start performance in Parkinson's gait recognition. Furthermore, we achieved an accuracy of 85.3 % in Parkinson's severity prediction. To the best of our knowledge, this is the first algorithm to perform a severity prediction based on the UPDRS.

These results show that the model is able to learn intrinsic characteristics from gait data and to generalize to unseen subjects, which could be helpful in a clinical diagnosis.

Keywords: 1D-Convnet, Parkinson, gait, Classification, Deep learning

1. Introduction

Today, over 10 millions people suffer from Parkinson's disease. Unfortunately, no cure exists to heal this disorder. That is why early diagnosis is important to improve the patient's treatment. Currently, physicians evaluate symptoms such as shaking, difficulty to initiate movements, slowness, and difficulty to walk (Jankovic, 2008). One of the most used tools in Parkinson clinical evaluation is the Unified Parkinson's Disease Rating Scale (UPDRS). This scale consists of 42 criteria/questions that cover different aspects of Parkinson's disease. These aspects comprise motor symptoms (including gait features), behavioral characteristics, and daily activities (Fahn et al., 1987). For the most part, doctors evaluate the severity of the criteria on a scale of 0 (normal) to 5 (severe). The total score is the sum over all the

criteria (maximum is 176).

Gait analysis is an important step in the diagnosis process of Parkinson's disease (Jankovic, 2008)(Reich and Savitt, 2019), as gait abnormalities have been documented to occur at the early stages (Pistacchi et al., 2017). The Parkinsonian gait is mainly characterized by small steps, a slower gait cycle, an increase in stride variability, a shorter swing phase, a longer stance phase and a flat foot strike instead of toe-to-heel strike (Morris et al., 2001) (Perumal and Sankar, 2016). Physicians evaluate these features in their diagnosis process to confirm the presence of the Parkinson disease (Jankovic, 2008). However, gait's evaluation can be challenging since it can be affected by several factors such as age and health condition.

Despite the significant interest in parkinsonian gait analysis, there is no objective tool to assist physicians for gait evaluation. Since changes in the gait are among the first symptoms of this disease (Pistacchi et al., 2017), a powerful gait classifier would be helpful for physicians. In this clinical context, the objective of our research work is to develop an intelligent tool to detect Parkinson's disease symptoms and predict the Parkinson severity rate (based on the UPDRS) from gait data.

To detect these characteristics, feature extraction methods have been widely used (Ertuğrul et al., 2016; Daliri, 2013; Sarbaz et al., 2012; Xia et al., 2015; Mannini et al., 2016; Wang, 2017). Previous studies used temporal (Ertuğrul et al., 2016; Xia et al., 2015; Wang, 2017; Mannini et al., 2016) or frequential (Daliri, 2013; Sarbaz et al., 2012) tools to get differentiable patterns between normal and parkinsonian gait. However, gait is a physiological characteristic that differs for each person according to age, health, and other

intrinsic factors. Therefore, manual preprocessing and feature extractions will always be limited in their capacity. To avoid hand-crafted signal processing, we propose a novel gait classifier based on deep learning without explicit feature extraction. The proposed Deep Neural Network (DNN) is composed of two parts. The first one consists of 18 parallel 1D-Convnets, each processing a vertical ground reaction force (VGRF) signal coming from a foot sensor. Each 1D-Convnet extracts deep features that are subsequently concatenated together. The second part is a fully connected network that processes the concatenated vector to output the final decision. Thanks to its ability to extract relevant gait features from different input signals, the proposed model outperforms state-of-the-art methods by achieving an accuracy of 98.7%.

The main contribution of this work is threefold. Firstly, we developed a 1D-Convnets that extracts relevant deep features for accurate gait classification, which avoids manual feature extraction. Secondly, we show that our proposed method achieves state-of-the-art accuracy (98.7 %) in Parkinson's detection. Thirdly, we present the first algorithm in UPDRS severity prediction, which is extremely valuable for clinical decision support systems, and achieves an accuracy of 85.3 %.

The rest of the paper is organized as follows. In section 2, we present a review of previous works on gait classification, as well as background concepts. Section 3 introduces the proposed method. Section 4 describes the experimental setup. Section 5 presents results and discussion, and we finally conclude in section 6.

2. Related work and background

2.1. Previous gait classification methods

Various feature extraction and classification approaches have been explored in previous work for gait analysis. Some of them extracted temporal patterns, while others used frequential features. In the temporal domain, Ertuğrul et al. (2016) proposed an algorithm based on shifted 1D local binary patterns (1D-LBP) with machine learning classifiers. They used 18 VGRF input signals coming from foot sensors of Parkinson patients and control subjects. For each signal, they applied shifted 1D-LBP to construct 18 histograms of the 1D-LBP patterns from which they extracted statistical features, such as entropy, energy and correlation. Finally, they concatenated the features from all the 18 histograms and used various supervised classifiers, such as Random Forest and Multi-Layer Perceptron (MLP), to classify the feature vectors. In the frequential domain, Daliri (2013) applied a Short-Time Fourier transform (STFT) on each input signal using the same VGRF input data. Then, they extracted the mean frequency and the variance of the frequency to construct a histogram representing the distribution of features extracted from all input signals. This histogram was processed with a feature discriminant ratio (FDR) in order to select the most significant bins. The final bins were then classified by a Support Vector Machine (SVM) with a chi-square kernel.

For other gait disorders, Mannini et al. (2016) first trained a Hidden Markov Model (HMM) on acceleration signals recorded with sensors from elderly, post-stroke and Huntington's disease patients. Then, they constructed a feature vector using the HMM log-likelihood of each patient with each class,

combined with various temporal and frequency features. The final vector was sent to an SVM classifier. Wang (2017) proposed a similar algorithm, but instead of using an HMM model to extract features, they trained a K-Nearest Neighbor (KNN) classifier and then used the probability given by that classifier for each class as a feature. An SVM was later trained to classify feature vectors containing those probabilities with other statistics. Xia et al. (2015) used spatiotemporal features of neurodegenerative disease patients and control subjects. For each time series, they computed various statistics, such as the fuzzy entropy, skewness, and kurtosis. They used machine learning classifiers, such as Random Forest, SVM, MLP, and KNN. After an optimization with a features selection algorithm, the best result was obtained with an SVM.

Recently, Zhao et al. (2018) developed a deep learning algorithm to detect Parkinson disease. Their model was composed of two parallel networks. The first network analyzed the spatial distribution of forces with a 2D-Convnet. The second network analyzed the temporal distribution with a recurrent neural network (RNN). The final classification was decided by the average of both channels.

2.2. Background concepts

The method that we propose is based on deep learning. To be self-explanatory, some basic concepts are briefly explained in this section. The learning in DNN relies on intermediate layers, commonly called hidden layers. The input is processed at each layer in order to reach higher abstraction levels. During the training stage, the DNN increases its accuracy by using stochastics optimizers that decrease the loss at each iteration. Then, with

6

enough data, the DNN learns how to extract features from the input (Good-fellow et al., 2016). Each layer is composed of neurons. The neuron computes a weighted summation of the inputs to which it is connected. A nonlinear activation function is then applied to obtain the neuron output. The learning is based on the adaptation of the weights W: The deep network learns the optimal W values to obtain the closest estimate to the real value (Goodfellow et al., 2016). In order to avoid over-fitting, regularization techniques, such as dropout, are commonly used. Dropout reduces the over-fitting by shutting off some neurons chosen randomly at each iteration during the training (Srivastava et al., 2014). In our model, we use convolutional layers, max-pooling layers, and fully-connected layers. Each type of layer is briefly described:

- **Convolutional layers**: they perform a spatial convolution between the inputs and the filters. The filters contain weights which are the learned elements of the layer.

- **Max-pooling layers**: they sample the input to form an output with smaller dimensions by selecting the maximum value element.

- **Fully connected layers (FC)**: they connect the output of the convolution layers to the final output. They allow a convergence of the decision at a higher level.

3. Methods

3.1. Method overview

The objective of our algorithm is to classify each subject's walk into one of the two categories: Parkinson and control. Each recorded walk was
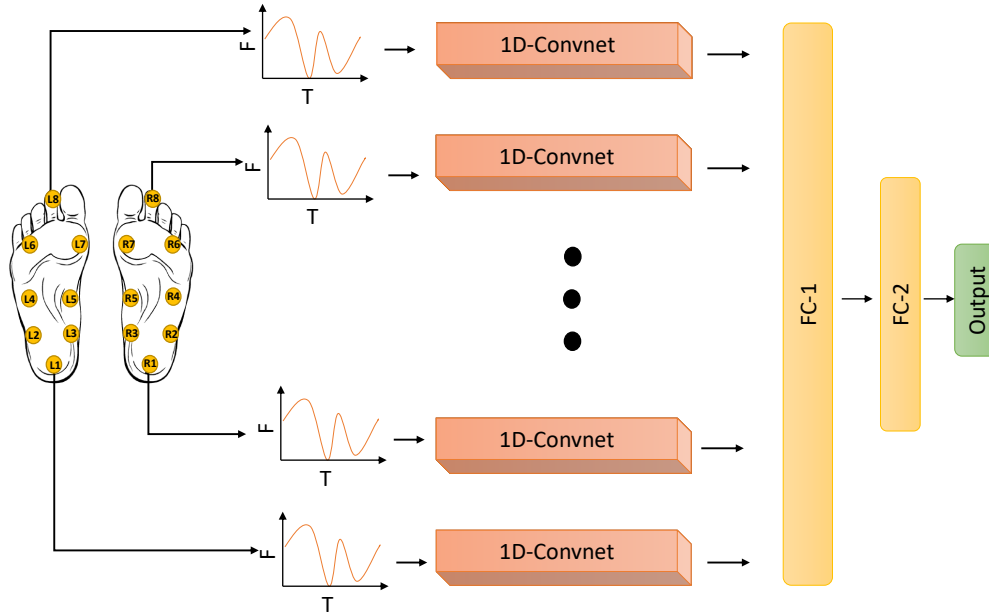
7

Figure 1: The architecture of the DNN: The first part is composed of 18 parallel 1D-Convnets, each processing a VGRF signal (F) coming from foot sensors. The 1D-Convnets are followed by fully connected layers that produce the final classification

documented with 18 VGRF 1D signals, measured in Newtons as a function of time, from several foot sensors. These signals correspond to 8 sensors placed underneath each foot (figure 1), in addition to two more signals that represent the sum of the 8 VGRFs for each foot. The VGRF signals comprise relevant information for gait analysis and characterization. In fact, important clinical spatio-temporal gait features, such as swing phase, stance phase and stride time, can be derived from VGRF signals. Feeding our deep learning model with raw data representing vertical ground reaction force records will allow to go beyond the hand-crafted features, by implicitly extracting relevant features for the studied application.
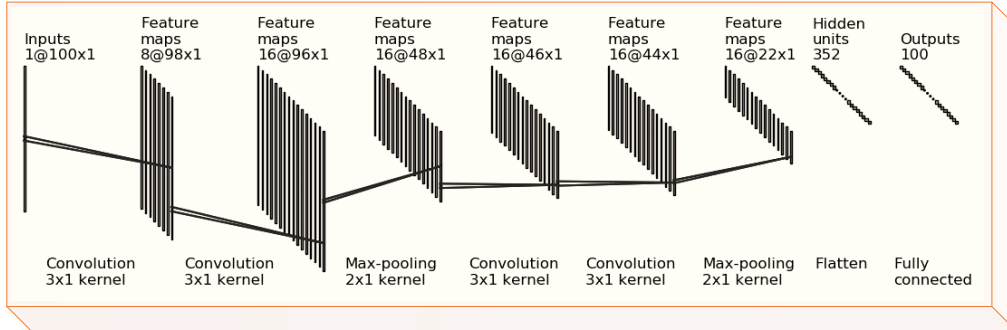
Figure 2: 1D-Convnet: Each 1D-Convnet block is composed of 4 convolutional layers, 2 max-pooling layers and 1 fully-connected layer.

We divided VGRF signals into m segments labeled with the subject category. These segments are the input samples of our DNN. The DNN is composed of two parts (figure 1). The first part consists of 18 parallel 1D-Convnets, while the second part is a fully connected network that operates on the concatenation of the features extracted by the 18 1D-CNN. The final walk classification was decided according to the majority classification of all the subject segments.

3.2. Parallel 1D-Convnet branches

The first part of the network is composed of 18 parallel 1D-Convnets (figure 2). Each network takes as input a VGRF signal going through 4 convolutional layers, that lead to a fully connected layer. Every two convolutional layers is followed by a max-pooling layer. Each filter in the convolutional layers extracts a specific pattern that allows differentiating between the Parkinson and control group. This 1D-Convnets parallelization allows the treatment of each signal independently. In fact, since each sensor records a specific data from a specific point, each time series has its own deep features

9

and is analyzed separately with the 1D-Convnets parallelization.

3.3. Fully connected bloc

The second part of the network is designed to learn the relationship between the spatial feature extracted from the 1D-Convnets and the final output classification. First, the outputs of those 18 parallel 1D-Convnets are all concatenated into one single vector. This deep feature vector passes through two fully connected layers leading to the output layer. For the Parkinson's detection, the output layer is composed of one neuron that predicts the classification probability. For the severity prediction, the output layer is composed of 5 neurons (for 5 classes) to predict the final classification. This design has two main advantages:

- Selection: the network selects the most significant time series that contribute the most to the performance of the algorithm.

- Abstraction: time series from different sensors are merged together only after passing through the 1D-Convnets. This allows getting significant features at the concatenation step. In fact, at this hidden state, each feature vector is obtained after multiple convolution steps. Therefore it would contain more abstract/global features. In our case, we believe that merging global features together is more efficient than merging raw data at the input level, since it is more robust to noise and local changes.

4. Experiments

4.1. Database

We used the public database collected by Physionet [1], which includes data reported by Frenkel-Toledo et al. (2005a,b), Yogev et al. (2005) and Hausdorff et al. (2007). In total, 93 parkinsonian patients and 73 control subjects participated in those studies. Each subject walked on a flat floor for two minutes at their natural pace. For each subject, eight sensors were placed under each foot (figure 1). Each sensor measures the VGRF (in Newton) under its specific point. The sampling frequency was 100 samples per second. The dataset also includes double-task time series, where subjects were walking while doing another activity. This results in an unbalanced 305 walks recorded (70% Parkinson and 30% control walks). For each walk, 18 time series signals are available: 16 (8×2) VGRF recorded from 8 sensors on each foot and 2 total VGRFs under each foot. For each subject (Parkinson or control), the UPDRS total score is reported in the database. The scores are distributed in a range between 0 and 70.

4.2. Parkinson's severity prediction

We have also tested our algorithm to predict the severity of the Parkinson's disease. We used the Unified Parkinson's Disease Rating Scale (UPDRS), which is the most used scale for Parkinson severity rating scale (Ramaker et al., 2002). We segmented the continous scale into 5 levels :

- class 1: UPDRS < 5

---

[1]https://physionet.org/content/gaitpdb/1.0.0/

- class 2: $5 \leq$ UPDRS < 15

- class 3: $15 \leq$ UPDRS < 25

- class 4: $25 \leq$ UPDRS < 35

- class 5: $35 \leq$ UPDRS

We kept the same architecture of the DNN and the same training hyper-parameters. We only modified the last layer. We replaced it with a fully connected of 5 neurons, with a softmax activation layer.

4.3. Evaluation metrics

We used cross-validation with 10 folds to test our algorithm on 300 walks. We divided each of the Parkinson and the control group into 10 folds at the subject-level. This allowed us to keep the same dataset balance (70 % Parkinson - 30% Control) for each fold. We report the evaluation metrics over the combined validation predictions.

4.3.1. Parkinson and control subject classification

The control group is identified as the negative (N) group and the Parkinson group is the positive (P) group. We use the following notations:

- Number of true positive (TP): Parkinson subject correctly classified

- Number of true negative (TN): control subject correctly classified

- Number of false positive (FP): control subject misclassified

- Number of false negative (FN): Parkinson subject misclassified

The specificity (Sp), sensitivity (Se), and accuracy (Acc) are calculated as follows:

$$Se = \frac{TP}{TP + FN} \tag{1}$$

$$Sp = \frac{TN}{TN + FP} \tag{2}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

4.3.2. Parkinson severity prediction

Since this is a multiclassification experiment, we report the precision, the recall, and F1 score for each class. For each class, true positives (TP) are the patients with the corresponding label who are correctly classified.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

4.4. Training

Since deep learning algorithms require large datasets, each walk was divided into smaller segments of 100 time steps with 50% overlap. Each segment was labeled with the subject category for the training. This segmentation was done inside each fold. Thus, the segments of a given subject are never divided between the training and validation set. The whole dataset contained 64468 segments.

The deep network was trained to classify these segments. For Parkinson's disease detection, the output probability of each segment was binarized (control or Parkinson). Then, the preponderant class (>50%) for the given segmented whole walk determined the final subject classification. For the severity prediction, the mode class over all the segments defined the subject's UPDRS class.

The final hyperparameters of the DNN are presented in table 1. A Scaled Exponential Linear Unit (SeLu) activation function (Klambauer et al., 2017) was used for all the hidden layers and a sigmoid activation (threshold = 0.5) was used for the final output. The algorithm was trained using Nesterov Adam optimizer (Dozat, 2016) with a batch size of 800 and an initial learning rate of 0.001. We used early stopping to avoid over-fitting: if the validation accuracy did not improve for 10 epochs, we stopped the training and went back to the weights that gave the best validation accuracy. From there, we decreased the learning rate by a factor of 2 and continued the training. This process was repeated 4 times to complete the training of the model. Also, in order to avoid over-fitting, some dropout was applied to the DNN to regularize the training.

4.5. Models used for comparaison

For Parkinson's detection, we compared our method with the work of Ertuğrul et al. (2016) and Zhao et al. (2018). Ertuğrul et al. (2016) work was chosen for comparison because they used a simple manual feature extraction method that achieved accuracy results comparable with other machine learning algorithms. Since Ertuğrul et al. (2016) used the same evaluation setting (10-folds cross-validation), we reported results from their paper. They

Table 1: Layer descriptions for Pakinson's disease detection: the first part of the DNN is composed of 18 1D-Convnet, and the second part is a fully connected network.

	Layer no	Layer type	Number of units	Kernel size	Dropout
1D-Convnet ×18	1	Convolutional	8	3	0
	2	Convolutional	16	3	0
	3	Max-pooling	-	2	0
	4	Convolutional	16	3	0
	5	Convolutional	16	3	0
	6	Max-pooling	-	2	0
	7	Flatten	-	-	-
	8	FC	100	-	0.5
FC	9	Concatenate	-	-	0.5
	10	FC	100	-	0.5
	11	FC	20	-	0.5
	12	Output	1	-	-

had employed a feature extraction algorithm summarized as follows:

- For each signal (18 in total):

  - Extract 1D-LBP with 8 neighbors. The output has the same length as the input with values between 2^0 and 2^8.

  - Construct a histogram of the frequency of the obtained LBP.

  - Extract statistical features from the histogram such as skewness, entropy, energy, a coefficient of variation, kurtosis and correlation.

- Concatenate the features of all the histograms.

- Train a machine learning classifier with the feature vectors created.

We report their results for three classifiers: Naive Bayes (NB), Random Forest (RF) and the Multi Layer Perceptron (MLP).

Zhao et al. (2018) work was chosen because they also use deep learning algorithms, but with a different network architecture. We reproduced Zhao et al. (2018) algorithm (according to the authors' paper description) in order to compare the algorithm in the same evaluation set-up with 10-fold cross-validation. As described before, they have employed a two-channel network: a 2D-Convnet and an RNN. The 2D-Convnet was built with 2 convolutional layers, 2 max-pooling layers, an FC layer, and an output classification layer. The RNN was built with two Long short-term memory (LSTM) layers, an FC layer, and an output classification layer. The final classification was made by the average of both channels. More implementation details are available in their paper (Zhao et al., 2018). They used the same 18 VGRF features, to which they added a time vector. Each input sample was an array of 19 features × 100 time steps. Their final output is a classification of the sample input. In order to evaluate their algorithm on a subject base, we took the majority classification of all the segmented walk samples to determine the subject final classification (as we do with our algorithm).

We used Keras with the Tensorflow backend and the sklearn library for model implementation.

5. Results & Discussion

5.1. Evaluation of the algorithm

The training curve illustrated in figure 3 shows that our algorithm can converge quickly. Furthermore, the accuracy of the validation set increases as
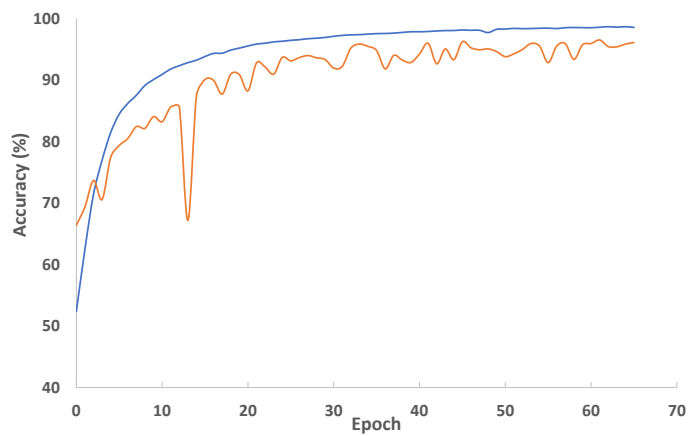
16

Figure 3: Training curve for Parkison's detection. The accuracy of the training set is represented by the blue line and the accuracy of the validation set is represented by the orange line. The accuracy corresponds to the proportion of segments correctly classified over all the walking segments in the training or the validation set.

17

Table 2: Cross-validation results for Parkison's detection. Best results are in bold and the second best results are in italic. Sp: Specificity, Se: Sensibility, Acc: Accuracy, SD: Standard deviation. Not available values for results reported in (Ertuğrul et al., 2016) are indicated by n/a.

Algorithm	TP	FN	TN	FP	Sp ± SD (%)	Se ± SD (%)	Acc ± SD(%)
Proposed DNN	206	4	90	0	**100.0 ± 0.0**	**98.1 ± 3.3**	**98.7 ± 2.3**
DNN (Zhao et al., 2018) (Reproduction)	202	8	69	21	76.7 ± 8.2	*96.2 ± 3.8*	*90.3 ± 2.9*
MLP (Ertuğrul et al., 2016)	n/a	n/a	n/a	n/a	*82.2*	88.9	88.9
NB (Ertuğrul et al., 2016)	n/a	n/a	n/a	n/a	n/a	n/a	76.1
RF (Ertuğrul et al., 2016)	n/a	n/a	n/a	n/a	n/a	n/a	86.9

the training accuracy increases, which shows that the algorithm is learning meaningful features without over-fitting. The performance of our method is compared with other studies in table 2. Recall that for deep learning algorithms, the result for a subject is obtained by a majority vote over the classification of gait segments. At the segment level, we have an accuracy of 98.3%, a specificity of 99.2% and a sensitivity of 97.8%. Our proposed DNN showed an accuracy of 98.7% at the subject-level, which clearly outperforms previous algorithms. Compared to other methods, our algorithm has the advantage of processing multiple input signals differently. This allows it to extract the most meaningful and specific features from each signal.

However, we notice a discrepancy between the results obtained for Zhao et al. (2018) algorithm and the results reported in their paper (accuracy of 98,6%). This could be explained by the different evaluation method (cross-validation vs 1 test fold). Our algorithm is different from Zhao et al. (2018) by the independence between input signals. In our method, the independence between the input 1D signals allows the method to be generalized to other experimental settings. Our method can be easily adapted to more or fewer

input signals. Thus, in a clinical context, our algorithm can be easily adapted to different gait clinical studies. In our research, we used this property of our algorithm to perform an ablation study between the input signals.

Compared to classical machine learning models, our algorithm is more adapted to the gait classification problem. First, gait time series are non-linear and noisy signals. Our model is adapted to this type of data since it uses a deep learning approach. In fact, compared to classic machine learning algorithms, deep neural networks are known for their ability to analyze complex signals because they have successive layers with non-linear activation functions. The deepness of the network has been correlated with the capacity of the network to analyze very complex problems (Simonyan and Zisserman, 2014; Srivastava et al., 2015). Second, hand-crafted methods are mostly inefficient in extracting discriminative gait features for a specific recognition problem. For example, in the work of Ertuğrul et al. (2016), LBP patterns were extracted and analyzed starting from VGRF signals. In contrast, the proposed model learns specific spatial kernels based on the data that allows the best distinction between the Parkinson and the control group. It is known that by increasing the model capacity, deep learning algorithms may make the classification vulnerable to overfitting. The algorithm may, in this case, memorize individual subject gait characteristics instead of memorizing discriminative patterns between the Parkison and control gaits. Nevertheless, our algorithm showed a high capacity to generalize to the validation set, which suggests that our 1D-Convnet has learned discriminative features between the parkinsonian and normal gait.

5.2. Ablation study: VGRF signal selection

In order to determine which VGRF time series has the most important impact on Parkinson's disease detection, we deleted each symmetric right and left VGRF pairs in turn (R1 & L1 for example) and we trained the model with the remaining 16 VGRF signals using the same hyperparameters as before.

Table 3: Impact of input signals on Parkinson's detection at segment level. In each row, our DNN was retrained without (w/o) two symmetric inputs Right (R) and Left (L) VGRF signals. Results corresponding to the most significant features are in bold. Sp: Specificity, Se: Sensibility, Acc: Accuracy.

VGRF inputs	Sp (%)	Se (%)	Acc (%)
w/o L1 & R1	97.4	93.7	94.8
w/o L2 & R2	96.7	93.4	94.4
w/o L3 & R3	99.3	98.6	98.8
w/o L4 & R4	96.0	94.0	94.6
w/o L5 & R5	**95.1**	95.0	95.1
w/o L6 & R6	97.3	97.3	94.8
w/o L7 & R7	96.2	94.7	95.2
w/o L8 & R8	96.5	94.9	95.4
w/o Total VGRF (R & L)	96.8	**92.9**	**94.1**

The results of the VGRF signal selection are presented in table 3. Here,

Table 4: Cross-validation results for Parkinson severity prediction (n: number of subjects)

Class	Precision (%)	Recall (%)	F1 (%)	n
1	77.6	100.0	87.4	90
2	100	75.0	85.7	8
3	100.0	76.3	86.6	76
4	91.8	80.0	85.5	70
5	78.0	82.1	80.0	56
Weighted average	87.3	85.3	85.3	300

we present the results at the segment-level to facilitate the differentiation between the ablations. The deletion of features L2 & R2, L4& R4 are the signal points that impacted the most the performance of our algorithm. Thus, we can hypothesize that the force measured at those points is different between the control and Parkinson group. However, the feature that seems the most important is the total VGRF. This feature is important in Parkinson's gait analysis, since most of the clinical features are generally extracted from the total VGRF. The deletion of features L3 & R3 seems to not affect the performance of the algorithm. We can conclude then that this input signal was not relevant for the gait classification.

5.3. Parkinson severity prediction

Table 4 presents the results per class for the severity level prediction. The overall achieved accuracy is 85.3 % (equivalent to the average recall). We can see that the F1 score is relatively similar between classes, so the

Table 5: Confusion matrix for Parkinson's severity prediction. All values are normalized with number of subjects in each category.

		Predicted				
		1	2	3	4	5
Ground truth (%)	1	100	0.0	0.0	0.0	0.0
	2	12.5	75.0	0.0	0.0	12.5
	3	15.8	0.0	76.3	2.6	5.3
	4	8.6	0.0	0.0	80.0	11.4
	5	12.5	0.0	0.0	5.3	82.1

algorithm is relatively consistent in his prediction between levels. Class 1 is the class where the algorithm was the most successful. This can be explained by the number of data available. In fact, this class contains all the control subjects. Nevertheless, the algorithm performed adequately and reaches a global accuracy of 85.3%, even if the dataset is unbalanced.

From the confusion matrix presented in table 5, we can also see that the vast majority of subjects for each class are correctly classified. However, the preponderant errors come from a confusion with class 1. Once again, this can be explained by the amount of data available in class 1 compared to other classes.

6. Conclusions

Parkinson diagnosis is still a very challenging problem in medicine. A Parkinson diagnosis is theoretically impossible to confirm, and doctors can detect the disorder by analyzing several symptoms through physical examination. Since gait perturbation is among the important motor symptoms,

we proposed an algorithm to recognize the Parkinsonian gait and predict the severity of the disease based on gait data. Our algorithm uses deep learning techniques, which avoids the drawbacks of hand-crafted feature extraction. The proposed DNN reached an accuracy of 98.7% in Parkinson's gait recognition. To the best of our knowledge, this represents the state-of-the-art performance. Furthermore, the proposed algorithm is the first to predict the UPDRS severity of a subject, with an accuracy of 85.3 %. This system can serve as a practical tool to screen a population in order to detect potential Parkinson patients in a clinical context. Moreover, we are witnessing an impressive democratization of biomedical sensors, that are becoming more and more present in our life. In the long-term, we believe that the proposed algorithm would be useful for the elderly, by monitoring and analyzing gait features during daily life activities. Such AI tools coupled with increasingly powerful biometric sensors would allow detecting gait abnormalities at the first stages of Parkinson's disease.

For future work, it would be interesting to get inside the DNN layers and analyze what they have learned. Such a study would allow a deeper understanding of the parkinsonian gait and its characteristics.

Aknowledgement

References

Daliri, M.R., 2013. Chi-square distance kernel of the gaits for the diagnosis of parkinson's disease. Biomedical Signal Processing and Control 8, 66–70.

Dozat, T., 2016. Incorporating nesterov momentum into adam .

Ertuğrul, Ö.F., Kaya, Y., Tekin, R., Almalı, M.N., 2016. Detection of parkinson's disease by shifted one dimensional local binary patterns from gait. Expert Systems with Applications 56, 156–163.

Fahn, S., Elton, R., et al., 1987. Updrs program members. unified parkinsons disease rating scale. Recent developments in Parkinson's disease 2, 153–163.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005a. Effect of gait speed on gait rhythmicity in parkinson's disease: variability of stride time and swing time respond differently. Journal of neuroengineering and rehabilitation 2, 23.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005b. Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. Movement disorders: official journal of the Movement Disorder Society 20, 1109–1114.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep learning. volume 1. MIT press Cambridge.

Hausdorff, J.M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C.,

Giladi, N., 2007. Rhythmic auditory stimulation modulates gait variability in parkinson's disease. European Journal of Neuroscience 26, 2369–2375.

Jankovic, J., 2008. Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry 79, 368–376.

Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S., 2017. Self-normalizing neural networks, in: Advances in neural information processing systems, pp. 971–980.

Mannini, A., Trojaniello, D., Cereatti, A., Sabatini, A.M., 2016. A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients. Sensors 16, 134.

Morris, M.E., Huxham, F., McGinley, J., Dodd, K., Iansek, R., 2001. The biomechanics and motor control of gait in parkinson disease. Clinical biomechanics 16, 459–470.

Perumal, S.V., Sankar, R., 2016. Gait and tremor assessment for patients with parkinson's disease using wearable sensors. ICT Express 2, 168–174.

Pistacchi, M., Gioulis, M., Sanson, F., De Giovannini, E., Filippi, G., Rossetto, F., Marsala, S.Z., 2017. Gait analysis and clinical correlations in early parkinson's disease. Functional neurology 32, 28.

Ramaker, C., Marinus, J., Stiggelbout, A.M., Van Hilten, B.J., 2002. Systematic evaluation of rating scales for impairment and disability in parkinson's disease. Movement disorders: official journal of the Movement Disorder Society 17, 867–876.

Reich, S.G., Savitt, J.M., 2019. Parkinson's disease. Medical Clinics of North America 103, 337 – 350. URL: `http://www.sciencedirect.com/science/article/pii/S002571251830138X`, doi:`https://doi.org/10.1016/j.mcna.2018.10.014`. neurology for the Non-Neurologist.

Sarbaz, Y., Towhidkhah, F., Gharibzadeh, S., Jafari, A., 2012. Gait spectral analysis: an easy fast quantitative method for diagnosing parkinson's disease. Journal of Mechanics in Medicine and Biology 12, 1250041.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15, 1929–1958.

Srivastava, R.K., Greff, K., Schmidhuber, J., 2015. Training very deep networks, in: Advances in neural information processing systems, pp. 2377–2385.

Wang, X., 2017. Machine Learning for Gait Classification. Ph.D. thesis. Universität Bremen.

Xia, Y., Gao, Q., Ye, Q., 2015. Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models. Biomedical Signal Processing and Control 18, 254–262.

Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E.S., Hausdorff, J.M., 2005. Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding? European journal of neuroscience 22, 1248–1256.

Zhao, A., Qi, L., Li, J., Dong, J., Yu, H., 2018. A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data. Neurocomputing 315, 1–8.