Signal and Visual Approaches for Parkinson's Disease Detection from Spiral Drawings

Manuel Gil-Martín*[1], Cristina Luna-Jiménez*[1], Fernando Fernández-Martínez[1], and Rubén San-Segundo[1]

[1]Grupo de Tecnología del Habla y Aprendizaje Automático (T.H.A.U. Group), Information Processing and Telecommunications Center, E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid, 28040, Madrid, Spain

Abstract

The development of medical decision-support technologies that provide accurate biomarkers to physicians is an important research area. For example, in the case of Parkinson's Disease (PD), the current supervisions of patients become intrusive, occasional, and subjective. However, new technologies such as wearable devices, signal processing, computer vision, and deep learning could offer a non-intrusive, continuous, and objective solution to help physicians with patient monitoring. The Parkinson's Disease Spiral Drawings public dataset was selected to face PD detection in this work by comparing four representation methods of the X, Y, and the pressure time series: signal, visual, hand-crafted, and fusion. The signal approach uses the Fast Fourier Transform of recording windows and a Convolutional Neural Network for modeling; the visual strategy employs visual transformer features from gray-scale images; the hand-crafted technique utilizes statistics calculated from temporal signals, and the fusion combines the information from the previous approaches. In these procedures, a Random Forest classifier was used for PD detection using the attributes extracted from each type of representation. The best results showed an F1 score of 93.33% and 93.06% at the user level using a signal approach with the three signals for the Static Spiral Task and an image-based proposal with X and Y coordinates for the Dynamic Spiral Task, respectively.

*Corresponding Authors (equal contribution): manuel.gilmartin@upm.es; cristina.lunaj@upm.es

1 Introduction

Deep learning has been evolving in recent years until achieving state-of-the-art performance in different research fields such as time series prediction [5] or image classification [11] in different applications. In particular, clinical applications that involve healthcare [6; 13] could help physicians through medical-decision support tools.

Parkinson's Disease (PD) is a neuro-degenerative condition caused by the loss of dopamine hormones activity that could provoke symptoms such as tremors, bradykinesia, micrographia, and postural instability [9]. An accurate analysis of these main symptoms could offer medical support to physicians in an early PD diagnosis. Current procedures are based on visits to the hospital where the expert checks the patient to adjust the drug dosage. However, these visits are short and occasional and the patient could not suffer symptoms during these medical appointments. In this context, a continuous, remote, and non-intrusive supervision of the disease could complement these on-site visits and provide accurate information on the evolution of the symptoms. Moreover, an early diagnosis could enable the development of customized therapies for PD patients [1].

By now, the most employed medical tests for examining patients' motor systems functioning consist of drawing tests of spirals or holding something at a certain distance to estimate their tremor levels, which are tasks that could be automatized for helping in the final diagnosis. Following this aim of studying assistant tools for doctors, Gallicchio et al. [3] proposed the use of the time series of the draw-

https://doi.org/10.7557/18.6809

ings of spirals, the pen pressure and the grip angle as inputs to the Deep Echo State Networks obtaining an accuracy of 89.3%, Khatamino et al. [10] designed a Convolution Neural Network (CNN) based on a LeNet architecture obtaining an accuracy of 72.5%. These two previous works were focused on detecting PD at the user level using the raw data extracted from the whole drawing. Gil-Martín et al. [4] used a CNN approach employing the spectrum information from fractions of the drawing to detect the PD at the window level obtaining an accuracy of 96.5%. In this work, we proposed and compared four different approaches using the same dataset to detect PD at the user level. These previous works used a subject-wise cross-validation strategy where the data from the same subject was included in the same subset during training, validation, and test procedures.

In contrast to signal-based approaches, Das et al. [2] proposed visual solutions to detect PD from drawings. The first approach considered handcrafted features obtained from a histogram of oriented gradients, which counts the occurrences of orientation in the gradient in localized areas of an image. The second approach evaluated features extracted from a pre-trained VGG16 model. Both procedures used different machine learning algorithms to perform the PD classification. They obtained an accuracy of 98% for the mixture image set, but they did not specify the evaluation strategy followed or if they used data from the same subject in both training and testing subsets.

As a result of the analysis of previous work, this paper proposes and compares four different approaches to face PD detection using handwriting recordings following a Leave-One-Subject-Out (LOSO) evaluation. The signal approach consists of feeding a deep learning network with motion information, the visual method employs a Vision Transformer (ViT) on drawn images, the handcrafted approach by extracting statistics characteristics from raw data, and a fusion approach where features from previous methods are combined. The aim of this comparison is the deepening into these techniques and into the effects of combining XY coordinates together with the pressure information, as well as studying how features extracted from a transformer model could compete with a fine-tuned CNN model trained with the spectrum of the hand-written signals and with a more classic hand-crafted feature approach.

2 Material and methods

This section describes the dataset used in this study, the cross-validation process, and the signal, visual, hand-crafted and fusion approaches followed for PD detection from drawing movements.

2.1 Dataset

The Parkinson's Disease Spiral Drawings Using Digitized Graphics Tablet dataset [8; 12] was chosen for this study. This public dataset contains spiral drawings from 77 people: 15 healthy people and 62 with PD. It was recorded using the Wacom Cintiq 12WX graphics tablet [7], which allows interactions with a digital pen. The dataset contains information about X-Y-Z coordinates, pressure, and grip angle for each drawing as shown in Figure 1.

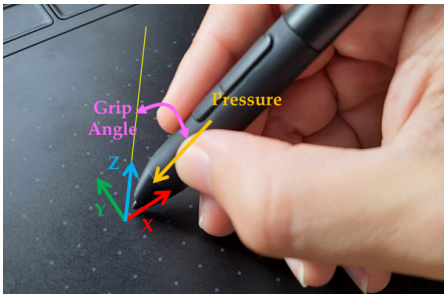


Figure 1: Recorded signals from tablet: X-Y-Z coordinates, pressure, and grip angle.

Three types of handwriting tasks were performed by all the subjects. First, the users completed the Static Spiral Test (SST) [14]. In this task, subjects retraced Archimedean spirals that appeared on the screen. Second, the subjects performed the Dynamic Spiral Test (DST). In this case, the Archimedean spiral appeared and disappeared for a few seconds, forcing people to keep the image in their mind while drawing. Finally, they followed the Stability Test on Certain Point where the subjects had to hold the pen on a red point for several seconds without touching the screen. In this study, we only used the data of the X and Y coordinates, and the pressure for both SST and DST tasks.

2.2 Leave-One-Subject-Out Cross-Validation

We followed a LOSO cross-validation. In the validation set, data from one subject was used for adjusting the main parameters of each approach, samples from another subject for testing, and the remaining subjects were employed as the training set. The experiments were repeated by modifying the training, validation, and testing subsets in each fold and ensuring that data from healthy and PD patients were included in all the subsets. Results show the average of the cross-validation process, using the F1 score as the evaluation metric.

2.3 Signal approach

The signal approach followed a previous work [4] methodology: windowing the signal recordings, computing the magnitude bins of the Fast Fourier Transform, and feeding a CNN with the spectra corresponding to the 0-25 Hz frequency band. In this case, we divided the sample sequences of X, Y, and pressure signals into 3-second windows separated by 0.5 s. Since the sampling frequency of the signals was 110 Hz and the final frequency of the spectrum was 25 Hz, the neural architecture was fed with examples of 75 magnitude bins per signal. The neural network was composed of two convolutional layers (using 16 filters of 1x5 dimensions) and an intermediate Maxpooling layer between the convolutional ones for feature learning and three fully connected layers (128, 32, and 1 neuron respectively) for classification. Dropout layers were included to avoid overfitting. The network architecture was optimized over the validation subset in the previous work [4]. Figure 2 shows the neural network architecture. The inputs were assembled in a 2 D matrix with N $\times$ 75 dimensions, where N is the number of signals considered in the CNN: two when considering only the X and Y coordinates or three when also including the pressure. In the optimization process, the validation subset was used to adjust some parameters of the network: 25 epochs, batch size of 100, and ReLU as the activation function in all intermediate layers. The optimizer was fixed to the root-mean-square propagation method as in previous works [10]. The network was implemented in Python using the Keras library.

This network focused on learning from recording

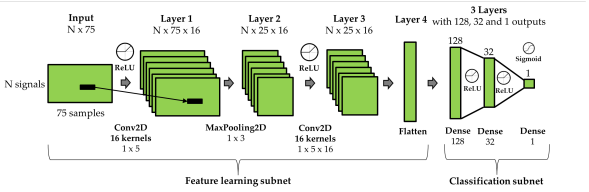


Figure 2: Neural Network architecture of the signal approach.

windows and classified these examples into healthy (labeled as class 0) and PD people (labeled as class 1). The last layer of the architecture used a sigmoid function to classify between these two classes and the binary cross-entropy as the loss metric. As a difference from the previous work, we conducted three methods to offer performance at the user level to compare with the other approaches that used the whole drawing. The first method was based on a voting scheme where a user was diagnosed with PD when more than a specific percentage of the analyzed windows were predicted as PD. This percentage (around 30%) was optimized over validation. The second method consisted in averaging or computing the standard deviation of the 32 or 128-dimensional embeddings of each subject's window, extracted from the last dense layers of the CNN. These embeddings were used as input to a Random Forest (RF) classifier to perform the detection of Parkinson's. The last method consisted in generating a normalised distribution based on the scores of the analyzed windows. Figure 3 shows the distribution of scores for a control subject, where it is possible to observe that most of the windows obtained a predicted score lower than 0.5. This scores distribution was used as input to a RF classifier to perform PD detection.

2.4 Visual approach

The visual approach consisted in generating grayscale images of the drawn Archimedean spirals by the participants, using the OpenCV library. The default images generated contained a white stroke on the annotated X and Y components, in which subjects moved the digital pen, over a black background.

To include pressure information on the images, we compressed the pressure levels into the 0-255

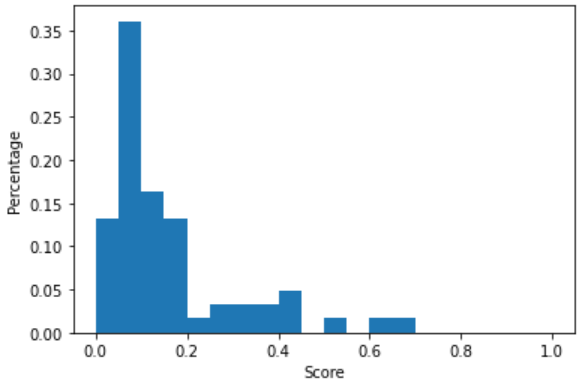


Figure 3: Normalised Scores Distribution for a Control Subject.

range of gray-scale values, where 0 represented values with the minimum pressure and 255 those where the subject did more strength while drawing. Some examples of the resultant images can be observed in Appendix A. In all cases, the spiral was centered into the image and the final images were re-sized to 775x775 square images to satisfy the input requirements of the Visual Transformer (ViT) that received them.

After that, the obtained set of images was passed to a ViT of the HuggingFace library ('google/vit-base-patch16-224-in21k'), pre-trained on the ImageNet dataset. From this model, we extracted 768-dimensional embeddings from its last pooling layer. Subsequently, these embeddings were fed into an RF to perform the PD detection.

2.5 Hand-crafted approach

An analysis of hand-crafted features was also included to compare more elaborated strategies against standard strategies based on statistics.

This method consists of computing six statistics (mean, standard deviation, minimum, maximum, skewness, and kurtosis) from each subject's drawing and their time series (i.e. X, Y, and pressure). This procedure returns 12 features for the non-pressure version and 18 attributes when pressure is added into the experimentation set-up.

In addition, the time required for completing the task was included in the set, considering that this feature may contain also relevant information for the task to solve because it could influence the precision of the drawing. After that, an RF classifier was fed with this feature set.

Task	Time Series	Input features to RF	Val. F1
SST	X, Y	distribution of scores	90.67
		mean of 32D-embs.	85.70
		std of 32D-embs.	88.00
	X, Y, P	distribution of scores	93.33
		mean of 32D-embs.	93.33
		std of 32D-embs.	93.33
DST	X, Y	distribution of scores	91.67
		mean of 32D-embs.	87.33
		std of 32D-embs.	90.15
	X, Y, P	distribution of scores	92.97
		mean of 32D-embs.	92.97
		std of 32D-embs.	91.17

Table 1: LOSO-Average F1-score comparison for the signal approach per task, time series and input features to RF for PD detection.

2.6 Fusion approach

To combine information from different representations an early fusion strategy was tested

The features fed into the models in the early fusion mechanism contained the distribution of the scores obtained from the CNN trained in the signal approach, the embeddings obtained from the pooling layer of the ViT in the visual approach, and the statistical features calculated from the X, Y and pressure components per user in the hand-crafted method. These attributes were concatenated and provided to several RF models from the sklearn library, varying the value of the number of forests parameter by 10, 20, 30, 50, 100, 150, and 200. After testing all the parameters, the optimized value was decided based on the F1 score obtained in the validation set. This RF training procedure was also employed in single-modality experiments with each feature set to compare the differences in performance when these features were combined.

3 Results and Discussion

Using the spectrum at signal approach is crucial because the tremor, which is one of the most prevalent PD symptoms, becomes more visible in the

Approach	Model	Task	Time Series	Test F1-score
Signal	CNN + Voting of windows	SST	X, Y	88.83
			X, Y, P	93.13
		DST	X, Y	87.90
			X, Y, P	89.86
	CNN + scores distribution + RF	SST	X, Y	89.18
			X, Y, P	93.33
		DST	X, Y	83.70
			X, Y, P	91.44
Visual	ViT + RF	SST	X, Y	90.79
			X, Y, P	90.79
		DST	X, Y	93.06
			X, Y, P	90.28
Hand-crafted	Features + RF	SST	X, Y	88.40
			X, Y, P	84.36
		DST	X, Y	91.44
			X, Y, P	91.17
Early Fusion	Features + RF	SST	X, Y	92.89
			X, Y, P	93.13
		DST	X, Y	91.44
			X, Y, P	91.44

Table 2: LOSO-Average F1-score comparison per modality, task, and time series for PD detection. X and Y represent the coordinates where subjects moved the pen and P the pressure applied while drawing.

frequency domain. For this reason, initially, we trained a CNN model able to handle spectrum images. We used the learned information at different levels of the network (scores distribution or embeddings) to make the final decision. Table 1 compares the applied strategies per task (SST and DST), time series (only X and Y coordinates or adding pressure), and input features. Only 32D embeddings were included in the table for simplicity. As the Table displays, the top F1 score was obtained for the strategy based on the distribution of scores for the two tasks and feature combination options. Moreover, it seems that the standard deviation could offer similar or higher performance than computing the mean of the embeddings. Since tremors could be noticeable in some parts of the drawings, calculating the average of the temporal series could be counterproductive to detect PD, while the standard deviation could inform about the peaks of tremors.

Apart from the signal, other transfer-learning approaches could be considered from the pre-trained network. Regarding the comparison between different representations, Table 2 summarizes the top F1 scores achieved by the RF trained with the referenced time series for each task, except for the voting strategy of the signal approach, where no RF was used.

In the signal approach, there is a performance improvement when incorporating pressure for both tasks, SST (4.30) and DST (1.96) when applying the voting strategy. The same pattern can be observed using as inputs the distribution of scores and training an RF on SST (4.15) and DST (7.74) tasks. In the case of using pressure, we observed that including the distribution of the scores to the RF could offer higher performance than directly using the voting scheme.

Unlike the signal approach, visual and handcrafted methods do not present such enhancement when examining experiments using X, Y against X, Y, P. Drawn spirals with coded pressure obtained the same F1 score compared to using spirals without pressure information for the SST task, and a decrement of 2.78 in the DST task. These results may suggest two explanations: first, pressure data

is not correctly exploited by the models or not relevant for solving the task when spirals are drawn; or second, the pressure codification on gray-scale levels was not appropriate for letting the models learn from this information as it was done in the signal approach.

In an attempt to combine different representations of the patients' performance of spiral drawing tasks, an early fusion was performed. Table 2 displays the results of the fusion after combining features and applying an RF algorithm to them, which obtained top F1 of 93.13 and 91.44 on SST and DST tasks, respectively. Results show that the fusion approaches proposed do not overpass the performance of the models trained with a unique modality, such as the signal approach on the SST task (93.33) or the visual proposal on the DST task (93.06). These results suggest that although different modalities could carry complementary or distinct information, simple models such as an RF could not be able to discover and employ these differences to improve the final recognition performance.

4 Conclusions

This paper was focused on detecting the alteration in the kinematics of the handwriting of PD patients because it is one of the first symptoms of the disease. Non-intrusive methods through analyzing drawing motions via signal of visual approaches could easily distinguish between PD patients and healthy people.

Comparing performances by task and modality, the signal approach based on spectrum revealed better performance when the pressure information is introduced as an additional channel into the trained CNN. Especially relevant is the reached F1 score of 93.33 in the SST task when employing the proposed method based on windows and the distribution of scores as features to an RF.

In the DST task, results achieved by the RF on features extracted from Visual Transformers pre-trained on the ImageNet dataset reached a top performance of 93.06 without including pressure information. This outcome emphasizes the flexibility of embeddings of transformers-type models pre-trained on other domains to solve problems in which the amount of data available is limited. ViT captured successfully shape variations in images, although more research should be done to encode pressure information in inputs to let the model handle it for increasing detection rates.

Regarding hand-crafted features, they reported relatively high performance but not reaching the results obtained by the CNN embeddings of the signal proposal or the ViT features of the visual approach. However, they still could be employed in scenarios where real-time processing is required or with low-computational resources due to their simplicity.

In future work, it would be interesting to develop a system that could classify the recordings into different clusters based on the number of patients' tremors, according to the disease severity. In addition, tested approaches on this dataset will be extended to other corpora to corroborate the tendencies discovered in this article and obtain significance ranges. Finally, new fusion designs will be proposed and tested to combine information from different modalities.

5 Acknowledgements

The work leading to these results was supported by the Spanish Ministry of Science and Innovation through the projects GOMINOLA (PID2020-118112RB-C21 and PID2020-118112RB-C22, funded by MCIN/AEI/10.13039/501100011033), AMIC-PoC (PDC2021-120846-C42, funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR") and BeWord (PID2021-126061OB-C43, funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR"). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

References

[1] E. Ammenwerth, P. Nykänen, M. Rigby, and N. de Keizer. Clinical decision support systems: Need for evidence, need for evaluation. *Artificial Intelligence in Medicine*, 59(1): 1–3, Sept. 2013. doi: 10.1016/j.artmed.2013.

05.001. URL https://doi.org/10.1016/j.artmed.2013.05.001.

[2] A. Das, H. S. Das, A. Neog, B. Bharat Reddy, A. Choudhury, and M. Swargiary. Detection of parkinson's disease from hand-drawn images using machine learning algorithms. In V. S. Reddy, V. K. Prasad, J. Wang, and K. T. V. Reddy, editors, *Soft Computing and Signal Processing*, pages 241–252, Singapore, 2021. Springer Singapore. ISBN 978-981-33-6912-2.

[3] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep echo state networks for diagnosis of parkinson's disease, 2018. URL https://arxiv.org/abs/1802.06708.

[4] M. Gil-Martín, J. M. Montero, and R. San-Segundo. Parkinson's disease detection from drawing movements using convolutional neural networks. *Electronics*, 8(8):907, Aug. 2019. doi: 10.3390/electronics8080907. URL https://doi.org/10.3390/electronics8080907.

[5] M. Gil-Martín, R. San-Segundo, R. de Córdoba, and J. M. Pardo. Robust biometrics from motion wearable sensors using a d-vector approach. *Neural Processing Letters*, 52(3):2109–2125, Sept. 2020. doi: 10.1007/s11063-020-10339-z. URL https://doi.org/10.1007/s11063-020-10339-z.

[6] M. Gil-Martin, R. San-Segundo, A. Mateos, and J. Ferreiros-Lopez. Human stress detection with wearable sensors using convolutional neural networks. *IEEE Aerospace and Electronic Systems Magazine*, 37(1):60–70, 2022. doi: 10.1109/MAES.2021.3115198.

[7] U. Hahne, J. Schild, S. Elstner, and M. Alexa. Multi-touch focus+context sketch-based interaction. In *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, SBIM '09, page 77–83, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605586021. doi: 10.1145/1572741.1572755. URL https://doi.org/10.1145/1572741.1572755.

[8] M. E. Isenkul, B. E. Sakar, and O. Kursun. Improved spiral test using digitized graphics tablet for monitoring parkinson's disease. In *The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014)*, 2014.

[9] J. Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, Apr. 2008. doi: 10.1136/jnnp.2007.131045. URL https://doi.org/10.1136/jnnp.2007.131045.

[10] P. Khatamino, I. Cantürk, and L. Özyılmaz. A deep learning-CNN based system for medical diagnosis: An application on parkinson's disease handwriting drawings. In *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*. IEEE, Oct. 2018. doi: 10.1109/ceit.2018.8751879. URL https://doi.org/10.1109/ceit.2018.8751879.

[11] C. Luna-Jiménez, J. Cristóbal-Martín, R. Kleinlein, M. Gil-Martín, J. M. Moya, and F. Fernández-Martínez. Guided spatial transformers for facial expression recognition. *Applied Sciences*, 11(16), 2021. ISSN 2076-3417. doi: 10.3390/app11167217. URL https://www.mdpi.com/2076-3417/11/16/7217.

[12] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4):828–834, July 2013. doi: 10.1109/jbhi.2013.2245674. URL https://doi.org/10.1109/jbhi.2013.2245674.

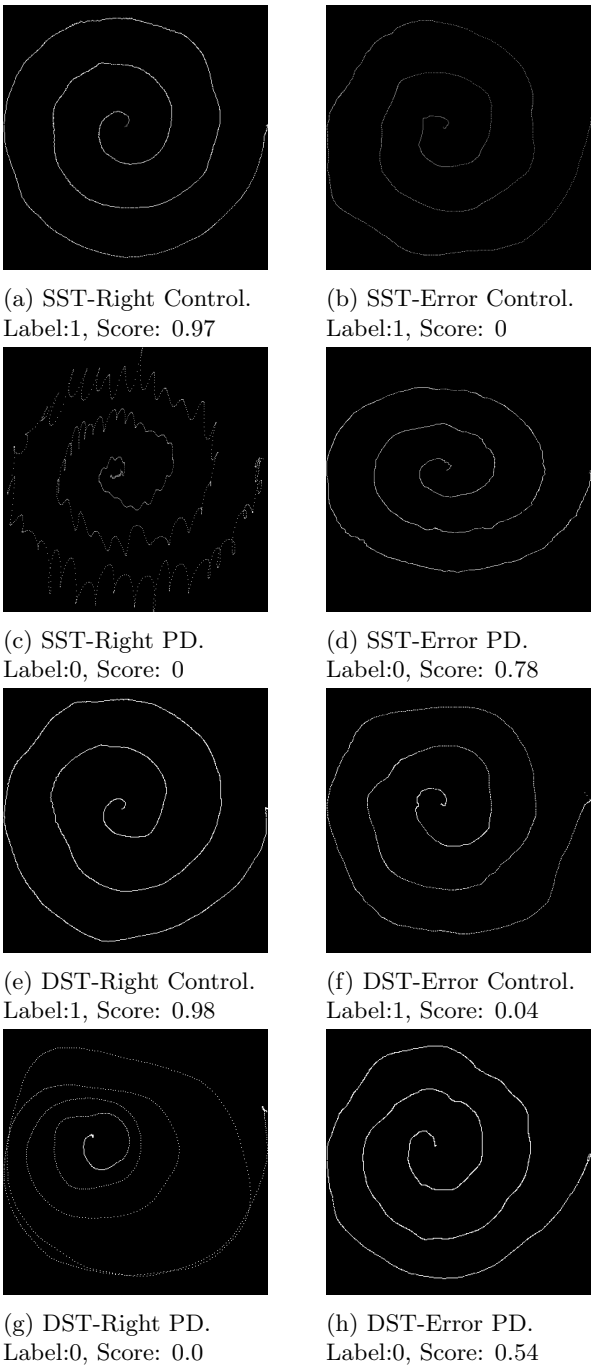
[13] R. San-Segundo, M. Gil-Martín, L. F. D'Haro-Enríquez, and J. M. Pardo. Classification of epileptic eeg recordings using signal transforms and convolutional neural networks. *Computers in Biology and Medicine*, 109:148–158, 2019. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2019.04.031. URL https://www.sciencedirect.com/science/article/pii/S0010482519301398.

[14] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis,

and S. L. Pullman. Validity of spiral analysis in early parkinson's disease. *Movement Disorders*, 23(4):531–537, 2008. doi: 10.1002/mds.21874. URL `https://doi.org/10.1002/mds.21874`.

A Samples of drawn spirals

Figure 4 includes samples of drawn spirals. The first two rows of the figure contain samples of outputs of the test subset of the signal approach that reached an F1-score of 93.33% on the SST task using X, Y, and P channels. In contrast, the third and fourth rows contain samples of outputs of the test subset of the visual approach that reached an F1-score of 93.06% on the DST task using X and Y channels. In this figure, it is possible to observe that control subjects that are correctly classified (4a and 4e) drew the spirals in a steady way, as the PD patients that were incorrectly classified (4d and 4h). In the same way, PD patients that were correctly classified (4c and 4g) drew the spirals with noticeable tremors. This effect could be observed in some healthy patients incorrectly classified (4b and 4f) in some parts of the drawing.



(a) SST-Right Control. Label:1, Score: 0.97

(b) SST-Error Control. Label:1, Score: 0

(c) SST-Right PD. Label:0, Score: 0

(d) SST-Error PD. Label:0, Score: 0.78

(e) DST-Right Control. Label:1, Score: 0.98

(f) DST-Error Control. Label:1, Score: 0.04

(g) DST-Right PD. Label:0, Score: 0.0

(h) DST-Error PD. Label:0, Score: 0.54

Figure 4: Error analysis of Parkinson detection tasks. The first two rows contain 4 samples of outputs of the test set of the signal approach that reached an F1=93.33 on the SST task using X, Y, P. The third and fourth rows contain 4 samples of outputs of the test set of the visual approach that reached an F1=93.06 on the DST task using X, Y.