



Sparse dynamical features generation, application to Parkinson's disease diagnosis

Houssem Meghnojdj^{*}, Bogdan Robu, Mazen Alamir

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

ARTICLE INFO

Keywords:

Parkinson's Disease
Dynamical system
Electroencephalogram
Sparse features
Machine Learning

ABSTRACT

This study focuses on the diagnosis of Parkinson's Disease (PD) based on electroencephalogram (EEG) signals. A novel approach inspired by the functioning of the brain is proposed, which uses the dynamics, frequency, and temporal content of EEGs to extract new discriminant features of the disease. The generated sparse dynamic features (SDFs) allow, through a transformation, to change the point of view on the data giving access to more informative features that are more faithful to the concept of EEG generation. Nevertheless, the method remains generic and can be applied to any signal but for this application it was evaluated on a publicly available dataset containing EEG signals recorded during a 3-oddball auditory task involving $N = 50$ subjects, of whom 25 suffer from PD. Given the adequate perspective on the data, it comes out that by using only two extracted features from the generated SDFs the healthy and unhealthy subjects are separated using a linear classifier. The classification yields an accuracy of 90.0% ($p < 0.03$) using a single channel. By aggregating the information from three channels and making them vote, an accuracy of 94%, a sensitivity of 96% and a specificity of 92% is obtained. The evaluation was carried out using a nested Leave-One-Out cross-validation procedure, thus preventing data leakage problems and giving a less biased evaluation. Several tests were carried out to assess the validity and robustness of our approach.

1. Introduction

Parkinson's Disease (PD) is a chronic neurodegenerative disorder affecting more than 6 million persons worldwide as reported by the World Health Organization (WHO, 2006; Dorsey et al., 2018). It is primarily caused by the lack of dopamine in the brain, due to the slow death of the dopaminergic cells (WHO, 2006; Balestrino and Schapira, 2020). PD is known by the general public for its motor symptoms such as: tremor at rest, rigidity, bradykinesia, akinesia, etc. (WHO, 2006; Balestrino and Schapira, 2020), however, non-motor symptoms may accompany or precede the onset of motor symptoms, sometimes even arriving 20 years before the onset of the latter (Chaudhuri et al., 2005; Kalia and Lang, 2015). Affecting patients on a daily basis, non-motor symptoms are various: pain, fatigue, sleep disturbances, bradyphrenia, communication issues, etc., only to cite a few (Pfeiffer, 2016; Witjas et al., 2002).

The diagnosis of PD is entirely clinical and is usually based on the manifestation of the motor symptoms (Berardelli et al., 2013). Before the appearance of the latter, patients suffer, raising the urge to look for new biomarkers allowing the early diagnosis of PD. The clinical diagnosis of PD is generally based on a pathological diagnosis or on

a clinical diagnosis criterion (as the one from the United Kingdom PD Society Brain Research Center (Gibb and Lees, 1988)). The overall clinical diagnosis accuracy is in the order of 75% according to the World Health Organization (WHO, 2006) or around 79% following (Rizzo et al., 2016). It is very important to note that the clinical diagnosis accuracy did not significantly improve during the last years particularly in the early stages of the disease where the response to dopaminergic treatment is not clear and less prominent (Rizzo et al., 2016). Actually, during the early disease manifestation (< 5 years of disease duration) the clinical diagnosis accuracy is around 53% and even lower, around 26% accuracy, for patients with < 3 years disease duration (Adler et al., 2014).

Electroencephalography (EEG) is a non-invasive method to record the electrical activity on the scalp which has been shown to represent the macroscopic activity of the brain underneath. It is used by several studies to assess individuals health conditions and to study brain function in healthy individuals as well as to diagnose various diseases that alter the brain electrical activity such as: Parkinson's Disease, epilepsy, Alzheimer's, sleep disorders, schizophrenia, etc (Soufneyestani et al., 2020).

^{*} Correspondence to: 11 Rue des Mathématiques, 38400 Saint-Martin-d'Hères, France.

E-mail addresses: houssem.meghnojdj@gipsa-lab.fr (H. Meghnojdj), bogdan.robun@univ-grenoble-alpes.fr (B. Robu), mazen.alamir@gipsa-lab.inpg.fr (M. Alamir).

EEG signals are known to have a low signal-to-noise ratio and present many difficulties. EEG noise is defined by any measured signal whose source is not the coveted brain activity (Urigüen and Garcia-Zapirain, 2015). Unfortunately, in most cases the EEG signal is contaminated by various unwanted artefacts, even though we try to limit their occurrence during the recording session. These artefacts are entangled with the desired brain activity and can have an amplitude up to 100 times that of the brain activity. In most EEG we encounter the following undesired artefacts: ocular, muscular, cardiac, perspiration, line noise, etc. (Luca, 2021; Urigüen and Garcia-Zapirain, 2015 give more details). Another difficulty that we may encounter during EEG analysis is the volume conduction, i.e. the transmission of electric fields from a primary current source through biological tissue towards the recording electrodes (Olejniczak, 2006). Because of volume conduction, unwanted artefacts will impact a broader region and therefore will contaminate more electrodes. In addition, we lose the ability to study a single source or brain region of interest; information is diluted and a signal recorded at one electrode is a combination of all the electrical activities present elsewhere (Urigüen and Garcia-Zapirain, 2015).

Parkinson's disease diagnosis using EEG has been studied in several works. Cavanagh et al. (2018) uses a selection of Fourier transform coefficients to achieve a maximum accuracy of 82%. It is to be noted that in our study we use the same data as the former. Oh et al. (2020) proposes a fully automated approach based on a 1-Dimensional Convolutional Neural Network (1-D CNN). The model directly classifies the temporal EEG epochs achieving an accuracy of 88.2%. To perform the diagnosis, Bhurane et al. (2019) relies on correlation coefficients calculated between channels as well as the coefficients of an AR model identified on the EEG to yield a presumable accuracy of 99.1%. Yuvaraj et al. (2018) uses high-order spectra to perform the diagnosis by extracting thirteen features from the EEG frequency spectrum, he achieved a presumable accuracy of 99.25%. Han et al. (2013) uses the coefficients of an AR model and the wavelet packet entropy to analyse and investigate whether there is a difference between the parkinsonians and the healthy individuals with no attempt to separate the subjects. Finally, Liu et al. (2017) utilizes entropy-based features of 10 channels and a three-way decision model to obtain a classification accuracy of 92.9%. This last study would have been more relevant if the author addressed the problem of unbalanced data-set. We note that the majority of studies are based only on the frequency features of the EEG and that few studies focus on the temporal features while the two domains should complement each other. Only a few of the features used are explainable and we can understand their design basis to derive conclusions for future work.

We strongly believe that some of the above mentioned methods (Cavanagh et al., 2018; Oh et al., 2020; Bhurane et al., 2019; Yuvaraj et al., 2018) are subject to data leakage problems. Data leakage is defined as the use of information in the model training process that is not supposed to be available at the time of prediction (Kaufman et al., 2012). This would not be possible in a real life scenario, where we receive new samples of unlabelled data that we need to categorize. This data leakage will bias the evaluation of the model, which will perform better on the available data used for training, but will perform poorly on the new data. The first type of data leakage that some of the proposed methods suffer from is group leakage, where correlated data from the same subject are present in both the training and the test sets (Ayotte et al., 2021). In this case, and using limited amounts of data, a complex model such as the 1-D CNN can even identify the subject's signature. The second type of data leakage is the fact of optimizing hyper-parameters and perform feature selection directly on the test-set (absence of a validation set) (Kaufman et al., 2012).

The aim of this paper is to propose a method for PD diagnosis using EEG signals recorded during a 3-oddball auditory task. The data at our disposal are composed of $N = 50$ subjects, of which 25 patients suffering from Parkinson's disease. Our main focus is not to have the highest accuracy at any cost, but rather to develop a valid method

with minimal bias. We aim to identify new biomarkers that go beyond traditional EEG statistics and spectral content as found in the literature, but instead consider the combination of frequency content, dynamics, and temporal aspects of the EEG. By changing the point of view on the data and by not considering them only as a time series, we propose a novel method to generate new features that are much more truthful to the EEG generation mechanism, in addition to be more informative than the classical features. What distinguishes our work from the existing ones is that we try to bring explainability to the results by using only few simple and explainable features, rather than using many features with complex models which are mostly black boxes like neural networks. Moreover, knowing that EEGs are very noisy, we propose a methodology with a hyper-parameter that allows to fit with different levels of fitting to the EEG signal, which allows to manage the trade-off between signal information retrieval and noise retrieval.

The proposed method has several notable advantages:

- It is inspired by the current understanding of how cognitive processes and brain works.
- It involves explainable features that may lead to future studies.
- It has the potential to work for the early and late stage disease diagnosis.
- It utilizes a simple and interpretable model with low computational demands.
- It has been rigorously constructed to avoid data leakage issues.
- It has been validated on a publicly available database, with transparent and open access implementation code and clearly described execution steps.

In the present paper, the data we used and the pre-processings we applied are presented in Section 2. In Section 3, the basic concepts of our method and its resemblance to the mechanism underlying cognitive processes are described, furthermore, how the proposed idea can be put into practice from a mathematical point of view is also described in this section. In Section 4, the results obtained are presented along with the various validity and robustness tests, moreover, an evaluation in a more constrained settings is provided. Finally, the conclusion of our work is outlined in Section 5.

2. Dataset

First of all it is to note that our method is agnostic to the dataset selection as long as it contains EEG data. Moreover it is straightforwardly applicable if the EEG was recorded during a 3-oddball auditory experiment. Several EEG datasets dealing with the diagnosis of PD exist, we examined these before selecting the one on which we can evaluate and test our method. As we want a significantly large number of patients as well as a large amount of data, the following dataset: <http://predict.cs.unm.edu> (ID: d001) (Cavanagh et al., 2017) was chosen. For clarity and reproducibility, we tested our method on a publicly available dataset and made our implementation code accessible to the public through the link: https://github.com/HoussemMEG/SDF_PD. Additionally, an explanatory animation is included.

2.1. Data-set description

The experimental EEG data available was recorded from $N = 50$ participants, 25 of whom were suffering from PD and an equal number of sex and age matched participants serving as a control group (CTL). The PD group were subject to the same experiment twice, once on-medication and the other time off-medication. In this document, we only consider the off-medication sessions as they showed a noticeable separability from the CTL group in comparison to the on-medication sessions.

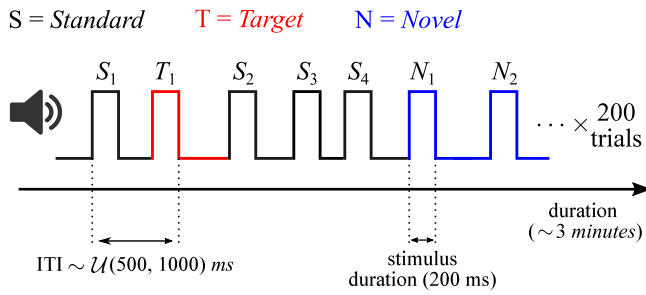


Fig. 1. Example of a sequence of auditory stimuli.

The PD group were subject to a Unified Parkinson's Disease Rating Scale (UPDRS) assessing the severity of their disease which was scored by neurologists, the mean UPDRS score is (24.80 ± 8.66) . All participants underwent a Mini Mental State Exam (MMSE), and all obtained a score above 26 (PD: 28.68 ± 1.03 , CTL: 28.76 ± 1.05) confirming their ability to comprehend the task they would be subjected to. Complete details and informations regarding the subjects and the experimental procedure can be found in Cavanagh et al. (2018).

The experiment consisted of a 3-Oddball auditory task, during which the subjects were presented with a series of 200 repetitive auditory stimuli (trials) infrequently interrupted by a deviant stimulus. Three types of stimuli can be distinguished:

1. *Standard* (70% of the trials).
2. *Target* (15% of the trials).
3. *Novel / Distractor* (15% of the trials).

During this task, the subjects had to count the number of *target* stimuli they had heard throughout the whole experiment. The auditory stimuli were presented for a period of 200 ms and were separated by a random Inter-Trial Interval (ITI) drawn from a uniform distribution of (500–1000) ms preventing subjects habituation and anticipation. Fig. 1 draws an example of an auditory stimuli sequence.

2.2. Data analysis and pre-processing

Throughout the experiment, the EEG signal was continuously recorded at a sampling rate of $f_s = 500$ Hz by the mean of 64 electrodes (channels). Very ventral temporal sites were removed by Cavanagh et al. (2018) as they tend to be unreliable, leaving at the end 60 channels. The data were then re-referenced to an average reference.

As mentioned in the introduction part, EEG signals are known to be very noisy and present many practical difficulties. Indeed, the coveted brain activity is of a low amplitude and is often drowned out by ambient noise, making the pre-processing stage mandatory. Despite the intrinsic complexity of EEGs and their noise content, the pre-processing steps we have applied are very mild due to the fact that our method is robust to noise. Firstly, to separate and disentangle the unwanted, high-amplitude ocular activity from the coveted cerebral activity, we conducted an Independent Component Analysis (ICA) on the data (Tharwat, 2018). We analysed each independent component (IC) of each subject individually, the ICs that contained eye blinking were removed by projection following the guidelines and recommendation of Luca (2021) and Cavanagh et al. (2018). Secondly, the data were then bandpass filtered using a Hamming window, attenuating the frequencies outside the (1–30) Hz interval. This frequency interval was selected because many studies take 20 or 30 Hz as the upper filtering limit (Starkstein et al., 1989). We have taken the widest interval knowing that our method remains valid even if we widen this interval further.

Time windows (segments) starting from stimulus onset (0 ms) up to (+500 ms) post-stimulus were formed, resulting in 200 time-locked

segments, one for each stimulus (see Fig. 2 for a more detailed graphical representation). An event-related potential (ERP) was also calculated separately for each stimulus type by vertically averaging all the signal segments corresponding to the same stimulus type and channel (see Fig. 3) (Luck, 2005). The aim of this step is to filter the signal and sum up the events occurring at the same time to make them stand out from the ambient noise. Moreover, all the pre-processing steps were performed using MNE (An open-source Python package for exploring, visualizing, and analysing human neurophysiological data) version 0.24.1 (Gramfort et al., 2013).

3. Methodology

In this section, the idea that inspired this work will be described first. Then, our major contribution, which is the generation of Sparse Dynamic Features (SDF), will be presented. Afterwards, the feature selection performed on the generated SDFs will be discussed. Finally, the classification step will be addressed, wherein we will elaborate on the different Cross-Validation procedures employed along with their specificities. The Fig. 4 provides us with an overall idea of all the steps that were applied to the raw data up to the classification.

3.1. Idea and inspiration

To process, encode, retrieve and transmit information, biological neuronal networks oscillate (Ward, 2003; Buzsaki and Draguhn, 2004). The frequencies and timings (phase) of these oscillations are important as they are at the basis of the mechanism underlying cognitive processes (Başar et al., 2001; Fries, 2005). As suggested by many studies, oscillation frequencies are task dependent (Ward, 2003). The oscillation timing is of a great importance since it carries the information about the neuronal dynamics and it is also what makes neuronal synchronization possible. This latter plays a crucial role in cognitive processes (Ward, 2003; Buzsaki and Draguhn, 2004). It should be noted that the EEG mainly measures the electrical potential of a group of neurons oscillating in synchrony. Due to volume conduction, the recorded EEG is the result from the combined activity of different electrical sources distributed in several regions of the brain. The mechanism of synchronization and desynchronization of a group of neurons suggest that the rhythms contributing to the EEG occur in a *pulsatory* manner (Olejniczak, 2006).

We assume that the ERP associated to a stimulus can be decomposed into two types of rhythms:

1. An ongoing activity: it refers to the oscillatory state of the brain when no stimulus is perceived (spontaneous background activity).
2. An evoked activity: it refers to the brain rhythms in response to a stimulus. It starts around 200 ms after stimulus onset and then vanishes in a time comprised between 500 ms–700 ms.

By having a set of m harmonic oscillators (pendulums) with distinct natural pulsations $\{\omega_1, \omega_2, \dots, \omega_m\}$, we can represent our signal as a combination of these oscillators where each one of them contributes with its own amplitude according to the Fourier decomposition. However, the temporal information about when an oscillator is turned on or off cannot be recovered by the Fourier transform. Since several rhythms can coexist at the same time, at the same or at different places and moreover interact with each other, the temporal information that cannot be recovered using the Fourier transform, is of great importance. For a given ERP signal, the design objectives of our method are the following:

- Disentangle and separate the contribution of each harmonic oscillator.
- Indicate which oscillators are involved as well as their amplitude contribution.

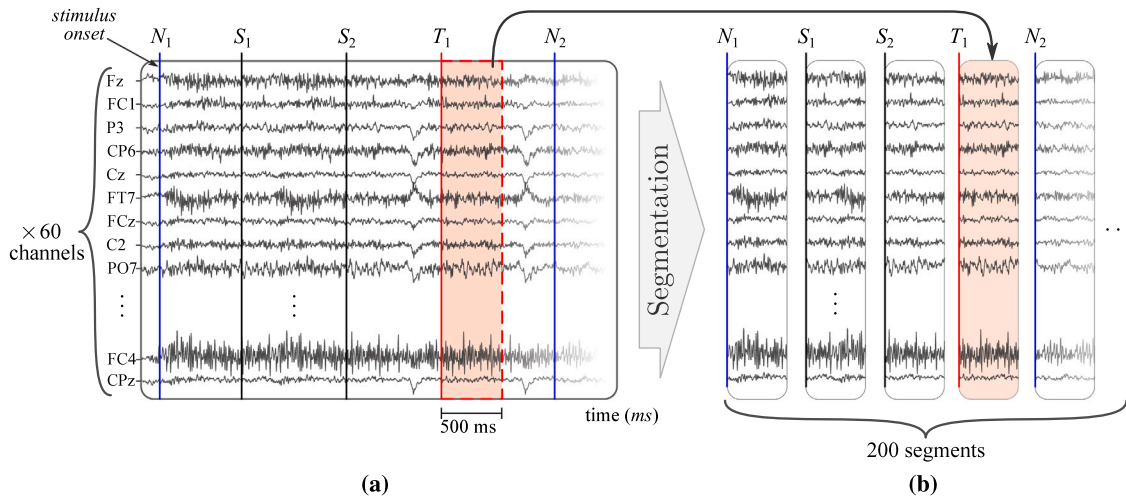


Fig. 2. EEG segmentation. (a) Raw and continuous 60-channel EEG, recorded for a duration of ~ 3 minutes. Stimuli are time-locked to their arrival instant. Segments of 500 ms were taken starting from the stimulus onset. (b) The result of the segmentation: a 60-channel signal divided into 200 time windows, each time window corresponds to one stimulus response.

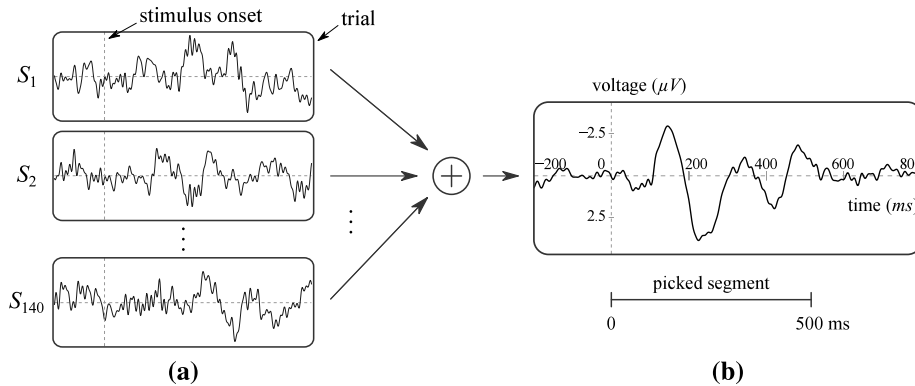


Fig. 3. Example of ERP creation process for the *Standard* stimulus on one specified channel. (a) Vertical averaging of the corresponding time-locked EEG segments. (b) Result of the vertical averaging process. The signals length used in this example is only demonstrative, the true window length used in our method is indicated at the bottom.

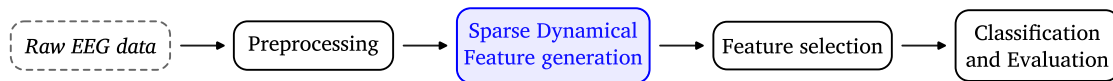


Fig. 4. Flowchart showing all the important steps that the raw data underwent until the classification.

- Indicate when an oscillator is turned on or off.
- Separate the ongoing activity from the stimulus response.

The pulsatory nature of brainwaves guided us towards using the principle of parsimony in our method. A high level of parsimony leads to the use of fewer oscillators and they will be excited less frequently (pulsatory fashion). By adjusting the parsimony level we can control the number of oscillators that can be awakened/used by our method and the frequency of their usage.

Studies have shown that patients with PD tend to be significantly slower than healthy individuals (latency observed in auditory ERP) (Hansch et al., 1982; Ferrazoli et al., 2022). They also tend to have lower ERP amplitude over certain brain regions compared to CTL individuals (Polich, 2007; Ferrazoli et al., 2022). However, even if this difference is observed between PD and CTL groups, the last two biomarkers extracted directly from the ERP do not permit the discrimination between the two groups with good separability and accuracy.

This is mainly due to the large variance in latency and amplitude within the two groups, as well as the high sensitivity of these two indicators to the disease duration and severity (Ferrazoli et al., 2022). We believe that the intra-group variance is due to the high noise content of the EEG, which makes the extracted measures very sensitive and variable among patients of the same group.

We believe that if we change our point of view on the data, and we do not look directly at the raw EEG but rather try to break it down into its basic components could lead to the development of new biomarkers. By basic components we are referring to oscillators that are turned on and off at a specific timing. Among these new biomarkers, we can find: the latency and activation time of the oscillators, the amplitude or energy content of these oscillators, etc. We believe that

these new biomarkers can significantly separate, with good accuracy and robustness¹, the PD group from the CTL group.

3.2. Sparse dynamical features

In order to put the idea described in the previous section into practice, we model the EEG response by the mean of a battery of pendulums with distinct angular frequencies. These pendulums are turned on and off at a specific time with the appropriate excitation amplitude, so that our model's prediction $\hat{y}(k)$ fits the EEG signal of interest² $y(k)$.

Consider m oscillators with distinct angular frequencies $\{\omega_1, \omega_2, \dots, \omega_m\}$. The system combining the m decoupled harmonic oscillators can be described by the following discrete state space representation given a sampling period $T_s = 1/f_s$:

$$\Sigma : \begin{cases} x_{k+1} = Ax_k + Bu_k; & x_k \in \mathbb{R}^{2m}, u_k \in \mathbb{R}^m, \hat{y}_k \in \mathbb{R}, \\ \hat{y}_k = Cx_k \end{cases} \quad (1)$$

where x_k is the state vector, u_k is the input vector and \hat{y}_k is the output of the system Σ , and:

$$A = \text{diag}(a_1, \dots, a_m), \quad B = \text{diag}(b_1, \dots, b_m), \quad C = (c_1 \ \dots \ c_m).$$

It is to note that $\text{diag}(a_1, \dots, a_m)$ is a block diagonal matrix with the elements $\{a_1, \dots, a_m\}$ being on the main diagonal and zero entries elsewhere.

$$\text{More precisely: } a_i = \begin{pmatrix} 1 & T_s \\ -T_s\omega_i^2 & 1 \end{pmatrix}, \quad b_i = \begin{pmatrix} 0 \\ T_s \end{pmatrix}, \quad c_i = (f_s \omega_i \quad 0), \quad (2)$$

are the matrices of a single harmonic oscillator: a_i contains the pendulum dynamics, b_i is the input matrix, c_i is the output matrix.

The m oscillators are merged throughout the matrices A , B and C such that the m oscillators remains decoupled but all their contribution is summed to form the signal $\hat{y}(k)$. In the control theory field, the oscillators described by the matrices (2) are called modes. The input vector u_k contains the excitation force of all the m modes at instant k , $u_k = (v_1(k) \ v_2(k) \ \dots \ v_m(k))^T$. The form of the matrix b_i has been chosen such that the i th pendulum is controllable by its corresponding control input $v_i(k)$.

The role of the scaling term ω_i present in the matrix c_i is to ensure that an oscillator at rest, excited by a force $v_i(k) = z$, starts oscillating at time instant k with the same amplitude as the excitation, i.e. z . This gives a physical meaning to the values of u_k as they will be directly proportional to the oscillations amplitude of the EEG/ERP. This also opens up the possibility of comparison between the v_i values since they are now on the same scale. We can now compare the modes in terms of their amplitude contribution, but also compare their energy, activity duration, starting time, number of times they have been excited, etc. From this representation it can be noted that the analysis of the excitation forces u can be conducted on a single specific oscillator or a desired set of oscillators.

The system (1) can be then rewritten in the explicit form as:

$$x_k = A^k x_0 + \sum_{i=0}^{k-1} A^i B u_{(k-1-i)}, \quad k = 1, 2, \dots, \quad (3)$$

$$\hat{y}_k = CA^k x_0 + \sum_{i=0}^{k-1} CA^i B u_{(k-1-i)}, \quad k = 1, 2, \dots, \quad (4)$$

where x_0 represents the initial oscillating state of our model (initial position and velocity of each pendulum). If no external force is applied on these pendulums as in the case of the non-forced regime, i.e. $u_k = 0$

¹ By robustness we are referring to: consistency, statistical validity, the fact that changing a hyperparameter does not change the result, and simplicity (low variance).

² Not the entire EEG signal, but a single signal over a specified channel and window. It can also be, as we have used, an ERP.

for $k = 1, 2, \dots$, the pendulums will keep swinging in the same manner. Therefore, the information about the background activity of the brain in the case where no stimulus is perceived is carried by x_0 .

For a given signal of interest Y of length L , which we want our model Σ to fit, we define:

$$Y = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{L-1} \end{pmatrix} \in \mathbb{R}^L, \quad \hat{Y} = \begin{pmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \vdots \\ \hat{y}_{L-1} \end{pmatrix} \in \mathbb{R}^L, \quad U = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{L-2} \end{pmatrix} \in \mathbb{R}^{m(L-1)}, \quad (5)$$

where, for our application, Y is an ERP signal of length $L = 250$ from a specific channel and a specific stimulus. \hat{Y} is the predicted output signal of the system Σ and U is the control sequence (how our pendulums are swung and excited over time). By forming the vector U in this manner, the information of which of the m oscillators is activated is embedded in the element u_k . The temporal information of when the excitation arrives is indicated by the subscript k .

Given (5), Eq. (4) can be written in a matrix form as:

$$(\phi_1 \ \phi_2) \cdot \begin{pmatrix} x_0 \\ U \end{pmatrix} = \phi \cdot \beta = \hat{Y}, \quad (6)$$

$$\text{where: } \phi_1 = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{pmatrix} \in \mathbb{R}^{L \times 2m},$$

$$\phi_2 = \begin{pmatrix} 0 & 0 & \dots & 0 \\ CB & 0 & \dots & 0 \\ CAB & CB & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B & CA^{L-3}B & \dots & CB \end{pmatrix} \in \mathbb{R}^{L \times m(L-1)}.$$

3.2.1. Virtual modal stimuli generation

A simple and cheap primary solution that finds β and minimizes the squared error $\epsilon = \|Y - \phi\beta\|_2^2$ from Eq. (6) is given by the least square method (Hastie et al., 2009, chap. 2). Despite its simplicity, it does not meet our expectations due to the fact that the resulting β will be full (not sparse). As we mentioned in the Section 3.1, we want to reproduce the pulsating nature of the brain rhythms, which should be achieved by a sparse β . Moreover, it should be noted that the minimization of the squared error will not necessarily yield the lowest classification error and the highest classification robustness.

Several subset selection and shrinkage methods exist (for more details see Hastie et al. (2009, chap. 3) for example). To induce sparsity in the β solution, we have chosen Lasso-LARS algorithm (Least Absolute Shrinkage and Selection Operator — Least-Angle Regression) (Efron et al., 2004). Briefly, instead of minimizing only the squared error, Lasso-LARS penalizes the number of non-zero entries of β through a weighted l_1 norm. The corresponding cost function is:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \frac{1}{2} \|Y - \beta_0 - \phi\beta\|_2^2 + \alpha \|\beta\|_1, \quad (7)$$

where $\beta_0 \in \mathbb{R}$ is the intercept³ term, $\hat{\beta}$ is the estimated β solution and α is a weighting constant (more details will be given below).

We have opted for the Lasso-LARS method over others for its inherent ability to compute the full solution path as α varies. Moreover, it is a very efficient algorithm to compute the solution of the Lasso problem, particularly when $\dim(\beta) \gg L$, which corresponds to our case. We used the Lasso-LARS implementation of scikit-learn⁴ version

³ Mean value of the response variable when all of the predictor variables in the model are equal to zero. In this case it is equal to \bar{Y} .

⁴ Open source python library for data analysis and machine learning.

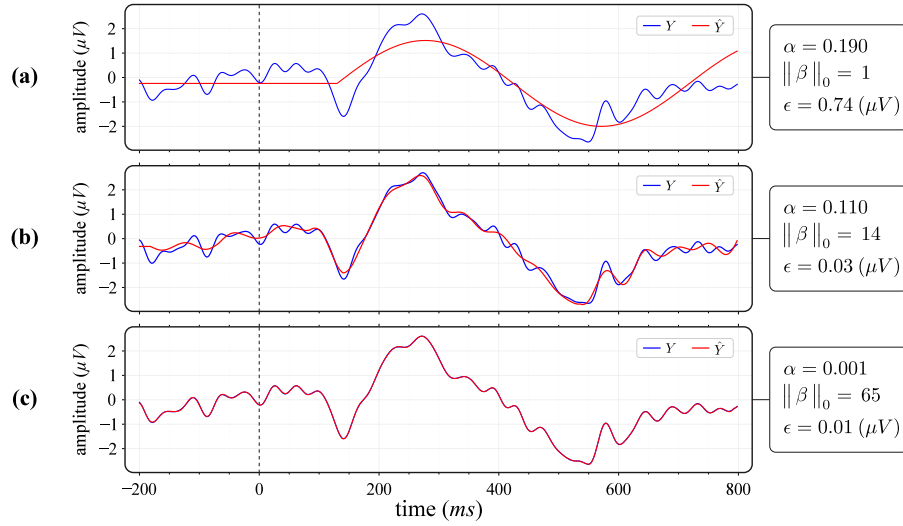


Fig. 5. Effect of α values on the predicted signal \hat{Y} and on the number of VMS mobilized by our model. $\|\beta\|_0$ indicates the number of non-zero entries of β . (a) Shows the activation of a single mode, selected with the appropriate frequency and excited at the right time. (b) Utilization of more modes and its effect on the \hat{Y} prediction. (c) Shows the ability of our model to fit the Y signal by exciting more modes and more frequently (65 excitations).

1.0.2 (Pedregosa et al., 2011). Throughout the rest of this paper, the vector β in Eq. (6) is named VMS for Virtual Modal Stimuli.

In our case, the best solution that yields the best result does not necessarily have to fit the signal Y perfectly as there is a trade-off between fitting the signal Y tightly and risking to capture the noise as well, and not fitting the signal Y tightly at the risk of not capturing the important information. We are looking for something in between, to do so, we adjust the value of α to control the sparsity of the VMS and thus controlling the fitting level of our model (see Fig. 5). For a large value of α , the sparsity is high and only a few modes are excited, while for a small α value, the sparsity is low, so more modes are utilized by the model.

At the initial stage of the algorithm no modes are excited, then, at each iteration the Lasso-LARS starts to excite and awaken modes. The α value at this step corresponds to a sparsity level of 100 % and is calculated on the basis of the correlations (see Hastie et al. (2009) for more details). After several iterations, the algorithm will utilize sufficient number of modes to reach the final α_f value provided to the algorithm. This low α_f corresponds to a low sparsity level of 0 %. Between the initial value and the final value of α , the Lasso-LARS generates the entire solution path of Eq. (7) for decreasing level of sparsity. Path length depends on the specified α_f , signal complexity and the ϕ matrix (number of oscillators m that we included in the model).

The value of the hyper-parameter α_f that yields the best results is not known beforehand. Thus, we selected several solutions along that path corresponding to several levels of sparsity by slicing the interval (100 — 0)% linearly with $level\%$ steps. Finally, we evaluated the classification results for each of these β solutions with varying levels of parsimony. It should be noted that due to this selection, the exact value of α_f is not as important as its magnitude. The Lasso-LARS algorithm is stated in terms of correlations, so if the input features (columns of ϕ) are standardized, the algorithm will run faster as it will utilize inner-products. The standardization serves also to mitigate the effect of the differing magnitude of the input features on the correlations, therefore the entries of β have an equal chance of being selected a priori.

Instead of standardizing the ϕ matrix in the conventional way, we have made a small modification. Let us denote ϕ_1 and ϕ_2 as the vector-wise standardized version of the matrices ϕ_1 and ϕ_2 respectively. We have introduced a weighting constant w so that the final standardized

ϕ matrix is created by concatenation as follows:

$$\phi = \left(w \phi_1 \quad (1 - w) \phi_2 \right). \quad (8)$$

This step is done in order to manage the trade-off between a priori likelihood and favouring/guiding the choice of modes that are chosen first (either those of x_0 or U). Therefore, the factor w serves to adjust whether we put more emphasis into the ongoing activity or the evoked activity. A value of $w = 1$ will place all the predicted VMS only in x_0 while a value of $w = 0$ will place all the VMS in U . It also allows to balance between the number of entries that will be distributed either in x_0 or in U of the final solution. This variation was made to favour adjusting for the x_0 features first since they affect the whole prediction unlike the U which are active only on a specific small time span (see Fig. 6). This step will make the algorithm more stable and give better results.

The Lasso-LARS algorithm is mainly used to induce sparsity in the solution of β . However, the main objective of the algorithm is to minimize the cost function (7) and to find its estimated optimal solution. As the Lasso-LARS algorithm is shaped, the final resulting β is not completely fitted, i.e. to minimize the cost function (7) β must be small in magnitude so that $\|\beta\|_1$ is small. As stated previously, we are only interested in finding the few most important entries of β that are used in fitting as much as possible the signal Y . Once β is computed, we will only pick its non-zero entries and used them to fit the Y signal using the Least Square algorithm. This reduces the estimation error ϵ without increasing the number of modes and their frequency of use. The temporal information of when a mode is excited is also preserved.

From Eq. (6), we can observe that the predicted signal \hat{Y} can be decomposed into two parts: (1) an ongoing activity contribution \hat{Y}_{x_0} (heavily dependent on the environment and not necessarily on the patient) and (2) an evoked activity contribution \hat{Y}_U . Fig. 6 illustrates how a signal Y is decomposed into a background rhythm and a stimulus response. For this figure, Y is an ERP associated with *standard* stimuli over the channel C_2 , and \hat{Y} was obtained using the same values as in the results i.e. $\alpha_f = 8 \times 10^{-4}$, $m = 40$, and $w = 0.55$ (more details are given in Section 4).

$$\hat{Y} = \underbrace{\phi_1 x_0}_{\hat{Y}_{x_0}} + \underbrace{\phi_2 U}_{\hat{Y}_U}. \quad (9)$$

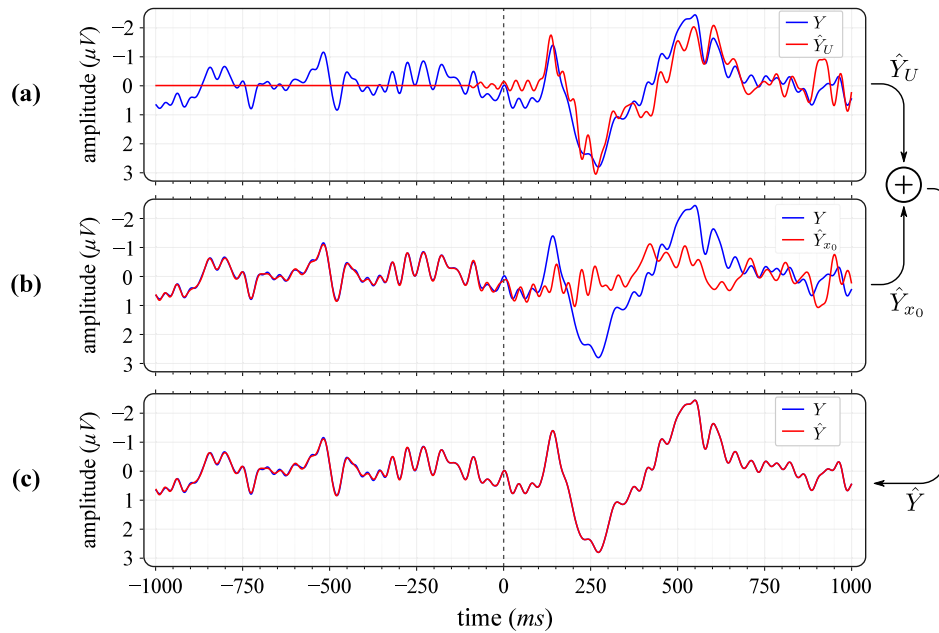


Fig. 6. Example illustrating how a signal of interest Y (in blue) is fitted by our model (1) (in red). (a) Estimated contribution of the forced regime alone \hat{Y}_U . As shown in the sub-figure, the modes are triggered near the stimulus onset and the shape of this response matches the shape of the ERP. (b) Estimated background rhythm \hat{Y}_{x_0} , the oscillations match the ERP response prior to stimulus arrival, and keep oscillating in the same fashion. (c) The contribution of (a) and (b) are summed to give the fitted signal \hat{Y} .

For the rest of the method, we will only consider the forced regime response (the second part of the equation) and therefore only \hat{Y}_U . Thus, we will not analyse β in its entirety but only the part regarding U .

3.2.2. Features selection

Given that we have a limited number of data instances, and for validity purposes of our method, we extracted new features from the generated VMS (concerning U only). The model will rely on these extracted Sparse Dynamical Features to classify the subjects as healthy or non-healthy. The advantages of conducting feature extraction are:

1. Improve the accuracy of the model on the test set.
2. Give physical meaning to our model and facilitate its explicability.
3. Extract only the useful and important information that the model should utilize to perform the classification.
4. Reduce the complexity of the model as we no longer need a very rich model that should address a complex task, but rather have a simple model for a simpler problem.
5. Give new directions for further studies.
6. Enhance model stability, robustness⁵ and increase the confidence placed on the model. The accuracy will not vary much for new, unseen data, which for an application based on a small number of data instances is crucial. Concerning our case, we can even afford to have a test-set size of 50% without impacting significantly the classification accuracy (see Table 2(c)).
7. Reduce the over-fitting, especially as in our case where we have a small number of data instances. If we over-fit, it may make our method impractical and therefore unusable.

As previously noted in Section 3.1, parkinsonians are mentally slower, with observed latency in the ERP and lower amplitude compared to healthy individuals. We will therefore mould the information

⁵ Lower accuracy variance under other novel conditions such as the arrival of a new subject.

contained in U (activation timing and amplitude) to fit these discriminative features. The operating time intervals I_1 and I_2 of the below SDF are shown in Fig. 7:

- F_1 : the instant of the lowest activation amplitude of all oscillators defined by: $\text{argmin}\{U(I_1)\}$.
- F_2 : the average excitation forces occurring between 180 ms and 500 ms defined by: $\text{mean}\{U(I_2)\}$.

3.3. Classification

The aim of the classification is to use a model, also called classifier, to categorize the input data into different classes. For our case, the input data are the SDF and the output is whether the subject belongs to PD or CTL group. Different classification algorithms exist with different levels of complexity and use cases. The less complex models are often appreciated because of their simplicity of use, their low computational costs and especially for their explainability and the simple interpretation of the results.

As we only have two features and a straight line or a quadratic curve is sufficient to separate the two classes (see Fig. 11(b)), we only tried three simple classification algorithms: Linear Discriminant Analysis (LDA), Quadratic discriminant Analysis (QDA) and Linear Support Vector Machine (linear SVM). It is to be noted that these methods do not have hyper-parameters that we need to tune (more details about these algorithms can be found in Hastie et al., 2009).

Among these 3 algorithms, QDA showed the lowest accuracy for our application, while LDA yielded the highest accuracy. Furthermore, the linear SVM had a slightly lower accuracy than the LDA classifier. Throughout the rest of this article, we will only present the results of the LDA classifier.

3.4. Nested cross-validation

To evaluate the generalization performance of our method on new unseen data, we first used the simplest sampling method, namely a

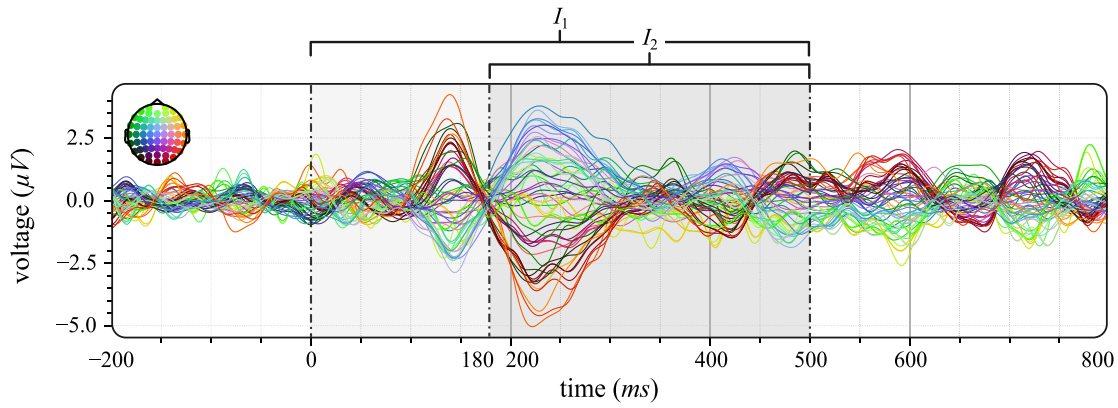


Fig. 7. Features operating intervals, example on a standard ERP.

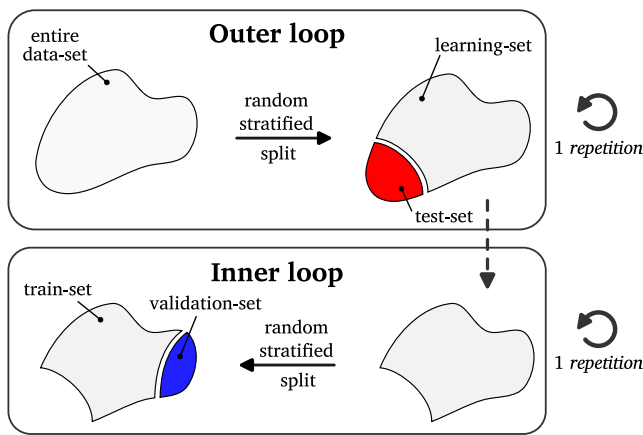


Fig. 8. Nested train-test-validation splitting for a single trial.

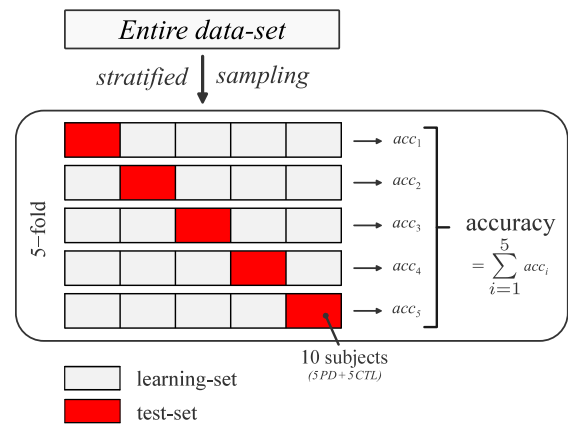


Fig. 9. Example of 5-fold Cross Validation.

Nested Train-Test-Validation split (NTTV) (Molinaro et al., 2005). This procedure consists first of all of splitting the data-set into a holdout test-set and a learning-set, this step takes place in the outer-loop and the splitting is done in a random and stratified manner (keeping the same proportion of the classes as they appear in the initial population if possible). The learning-set is then randomly stratified split into a training-set and a validation-set within the inner-loop (see Fig. 8). This sampling method allows us to have a validation set in order to select on the latter the best hyper-parameters, but at the cost of reducing the size of the training-set. It is important to recall that we cannot select the hyper-parameters that perform the best on the test-set as this is considered as data-leakage. The latter will lead to a biased model with over-optimistic performances where the model performs very well on the available data but poorly on new, unseen data as is the case in a real-life scenario (see Varma and Simon, 2006 for a study on this bias).

Firstly, the model is trained on the training-set, then the hyper-parameters⁶ are fine tuned and selected on the validation-set. The model is then retrained on the learning-set, using the best hyper-parameter obtained. Finally, we evaluate the trained model performances on the test-set. This assessment, will result in only one performance estimate, which will vary considerably and will depend heavily on the splits (which subject is in which group). One solution to cope

⁶ In our case they are the number of oscillators m , their angular frequency ω_i and most importantly the sparsity level α .

with this variance problem is to repeat the splitting procedure several times with different randomization in each repetition, giving several performance estimates, one for each given split. The overall performance will then be the average performance obtained on these different splits. However, even if we randomly split the subjects into three sets, some subjects will be chosen much more frequently in one set than in the others. This creates a selection bias that significantly affects the results (the effect is much more pronounced when the dataset is small) (Molinaro et al., 2005). This selection bias is reduced as the number of repeated splits is increased, but at the expense of computational load and time.

K-fold Cross Validation (CV) is another solution to reduce the selection bias that works well on small data sets (Stone, 1976). It consists of randomly partitioning the data set into K-folds (K partitions). For K times, we keep a fold that was not previously chosen as a holdout test-set and use the remaining $K - 1$ folds as a learning-set (see Fig. 9 for $K=5$ example). K may be set to the total number of data instances so that each observation is the holdout once; This procedure is called Leave-One-Out (LOO) Cross-Validation (Molinaro et al., 2005). This is the best way to reduce the selection bias and improve the learning performance of the model, but at the expense of the computational load (Varma and Simon, 2006; Molinaro et al., 2005). In our case, this is practicable because our model is simple and computationally inexpensive, moreover, we are working on a small data set.

To take advantage of the benefits of the two sampling methods presented above (NTTV and LOO) and to compensate for their drawbacks, we combined them to obtain the Nested Leave-One-Out (NLOO)

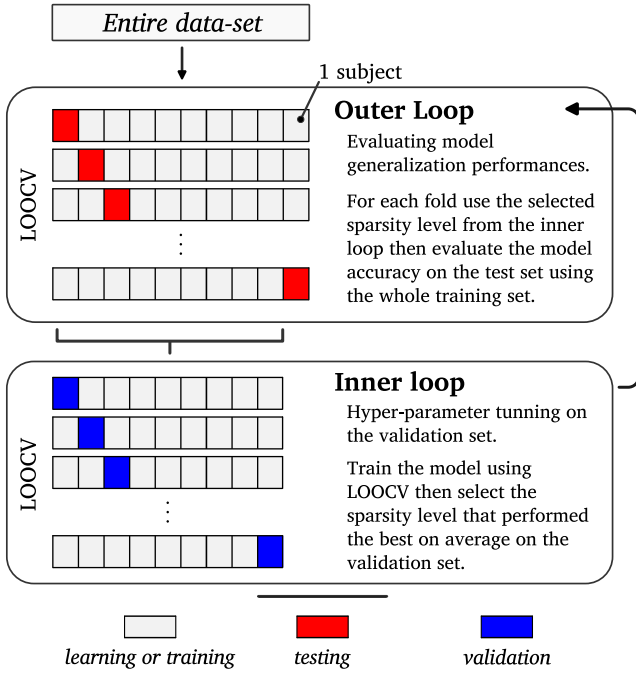


Fig. 10. Schematic display of the nested cross-validation.

Cross-Validation. This combination enables us to do a hyper-parameters tuning, reduce the selection bias and compensate for the small training-set bias present in NTV. The NLOO cross-validation procedure is composed of two nested loops: an outer-loop and an inner-loop, in each loop there will be a LOO procedure. The outer-loop splits the entire data-set into a learning-set and a test-set while the inner-loop splits the learning-set into a train-set and a validation-set (for more details, see Fig. 10). For the rest of this article, the internal loop always corresponds to a LOO CV procedure. However, with regard to the outer-loop, the case where we have a K-fold CV instead of a LOO CV will be mentioned. Moreover, the accuracies presented in this paper are the average accuracies obtained on the test-set over all the repetitions, the case where the accuracy concerns the train-set or validation-set will also be mentioned.

4. Results and discussions

For each subject, we computed the ERPs corresponding to each stimulus type and channel. The window width used ranged from 0 ms to 500 ms post-stimuli⁷, resulting in an ERP of length $L = 250$. The system Σ was created using $m = 40$ pendulums with an evenly spaced angular frequency taken from the interval $2\pi \times (1 \text{ to } 30)$ Hz. The weighting constant w in has been set to $w = 0.55$. We have set this value so as to obtain a separation between the forced and steady-state regime as shown in Fig. 6 while ensuring that the information is not completely carried by x_0 or by U alone. The evaluation metric used is the accuracy, which is defined as the number of true hits divided by the number of decisions to be made. In the case of multiple repetitions, the stated accuracy is the mean value of the accuracy obtained over all folds and all repetitions.

The weighting constant α_f was set to $\alpha_f = 8 \times 10^{-4}$ by visual inspection, so that this value is just small enough to obtain a very good fit. We selected the β solutions with different sparsity levels by taking the resulting solutions in the interval $(100 - 0)\%$ using a regular

Table 1

Results of the highest accuracy obtained for each stimulus type with the corresponding channel.

Stimuli type	Standard	Target	Novel
Channel*	AF_z	CP_z	CP_3
Accuracy	74%	90%	74%

spacing level = 2%, resulting for each ERP in 50β variants. Finally, once the solution path was computed, we extracted the features F_1 and F_2 from the generated excitation forces U . We used these extracted features to perform the classification using LDA.

The form and complexity of the ERP vary according to the type of stimulus and channel position. This variation is even more pronounced when the channels are far from one another. Our model utilizes fewer modes excitations to fit a less complex ERP, while it will utilize more modes excitations to fit a more complex ERP. Therefore, the best working level of sparsity that yields the best classification accuracy varies according to the channel location. Since the best working sparsity level cannot be determined in advance and since we are evaluating our method on all channels and stimuli type, we have left the sparsity level as a hyper-parameter to be tuned during the classification.

As we test a large number of combinations (stimuli type, channels and sparsity level) and have a small number of data-instances, in some cases for a specific condition and at a given sparsity level $\alpha_{k\%}$, we observe a good classification accuracy (over 80%) due to sampling noise (Frost, 2020). Fortunately, in this case, we can observe that the surrounding sparsity levels $\alpha_{k-1\%}$ and $\alpha_{k+1\%}$ give much poorer results (an accuracy jump is observed). We have not considered the level of sparsity $\alpha_{k\%}$ as a valid result since it is mainly due to luck and is not consistent. We have been careful to take into consideration only the sparsity levels that yielded good classification accuracy while the surrounding levels of sparsity being also good. This accuracy smoothness is expected since the variation in sparsity levels is small. This ensures us that our results are not due to luck, that they are sufficiently consistent and contain valuable information.

4.1. Stimulus choice

We started by evaluating the performance of our method for the different stimuli. For each stimulus, we evaluated the accuracy obtained by our model, for all the channels using the NLOO CV. Table 1 shows the results obtained. As we are dealing with 60 channels we have only indicated the channel that yielded the highest accuracy (channel*) with its corresponding accuracy.

As indicate the Table 1 the stimulus *target* showed the highest separation using the features F_1 and F_2 . These results, need to be interpreted with caution as other features can lead to different results. Furthermore, the results obtained for the stimuli *standard* and *novel* are not statistically significant. For the rest of this paper we only consider the *target* stimulus.

4.2. Luck or informative features ?

To test the significance, truthfulness and consistency of the results obtained for the CP_z channel we performed three tests:

1. *Permutation test* (Nichols and Holmes, 2003): We define the null hypothesis H_0 as: the features F_1 and F_2 do not allow to differentiate between the PD and CTL group. Under this hypothesis we evaluated the probability that the obtained result was simply due to chance and sampling noise as we have a small data set and we carried out a large number of runs (each per stimuli type, channel and parsimony level) (Frost, 2020). After randomly permuting the labels $n = 500$ times, and assessing the best accuracy obtained, we obtain a p -value of ($p < 0.03$) which

⁷ The upper limit of 500 ms was chosen due to the minimal interval between stimuli $\min(\text{ITI}) = 500$ ms.

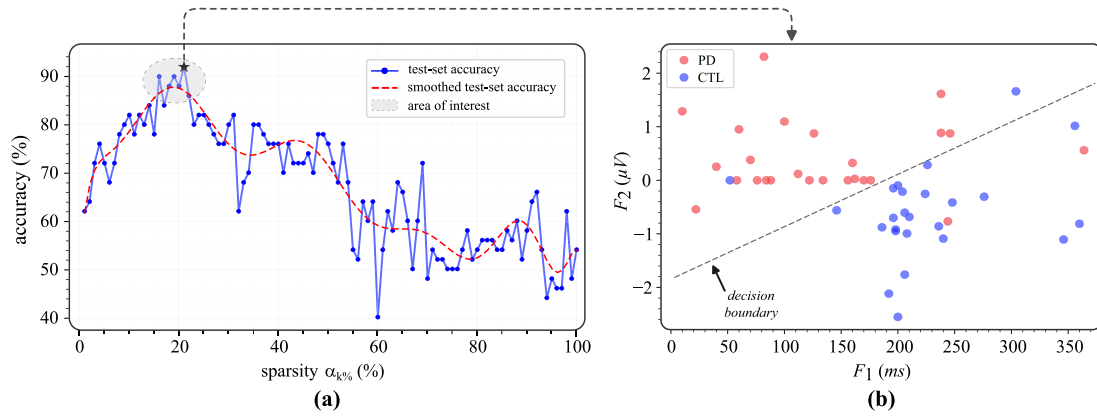


Fig. 11. (a) Model’s accuracy for a varying level of sparsity using LOOCV. (b) Scatter plot of the SDFs with the corresponding decision boundary taken for the sparsity level indicated by a star in (a).

- indicates and favours the fact that the results are significant. To recall, for statistical significance, the p -value should typically be $p < 0.05$ (Fisher, 1936, Chapter 1).
- Parametric consistency:** For this part, unlike the other parts, the parsimony level is fixed (it is no longer a hyper-parameter to be tuned). We used LOOCV for each parsimony level to assess the model’s accuracy. Moreover, we set the new spacing $level = 1\%$ to have a finer sparsity grid. Fig. 11(a) shows the accuracies obtained at each sparsity level. The small variation of the accuracy in the parsimony zone that yielded the highest accuracy (see area of interest in Fig. 11(a)) may suggest the presence of information and that indeed, by using F_1 and F_2 the subjects are separable. It should be noted that this area of interest is the area selected in 100% of the times during the model selection (in the inner-loop of the NLOO CV). To illustrate how the features F_1 and F_2 are scattered on the plane, we have plotted these in Fig. 11(a) for the sparsity level indicated by a star in Fig. 11(b). In addition, we have also plotted the decision boundary.
 - Spatial consistency:** We have evaluated the spatial evolution of the accuracy for the channels surrounding CP_z . If these channels also show good results, this may indeed indicate the presence of information over this region, moreover, it coincides to the suggestions of Fig. 7 in Polich (2007). Fig. 12 shows the accuracies obtained on the test-set after a NLOO CV procedure. We observe that the channels CP_1 and CP_2 located on the same horizontal line to CP_z showed great results individually, while the other surrounding channels did not yield significant results. We can observe an increasing accuracy as we get closer to CP_z , which may suggest the presence of a discriminative information over that region.

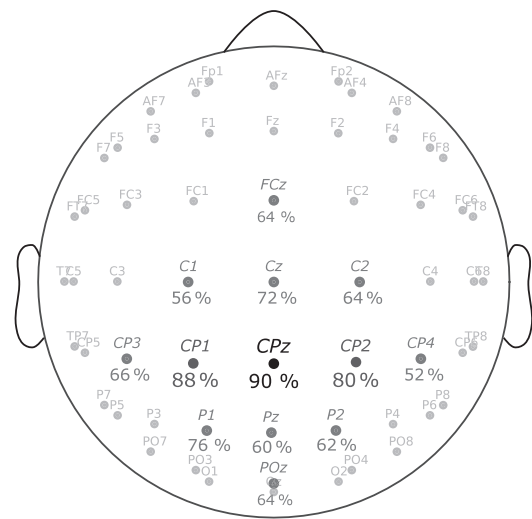


Fig. 12. Model performances over the region surrounding CP_z .

4.3. Increasing the accuracy through voting

Observing in Fig. 12 that the channels CP_1 , CP_z and CP_2 performed well individually, we aggregated the information of these three channels by making them vote on whether the subject is healthy or non-healthy. All the three channels have the same voting power, so if at least two voters agree on a subject’s condition, we consider the resulting vote as the output of this new model. Table 2(a) reports the accuracies obtained after a NLOO CV procedure. The validation set is non-existent for the voting strategy as we have no parameters to tune.

Fig. 13 represents the resulting confusion matrix of the voting strategy for the NLOO CV procedure. We can observe that the classification is balanced as the model mislabels the both classes almost equally.

		Predicted label	
		PD	CTL
True label	PD	24	1
	CTL	2	23

Fig. 13. Voting strategy confusion matrix.

Indeed, we can see that 1 subject out of 25 PD was wrongly categorized and that 2 subjects out of 25 CTL were wrongly categorized.

Regarding the early diagnosis, as presented in Section 1, in most cases patients suffer from non-motor symptoms before motor symptoms develop. The onset of the latter is slow and in some cases can take decades. The main difficulty of early clinical diagnosis is that it is

Table 2

Resulting accuracies obtained for the voting strategy and the three individual channels CP_1 , CP_2 and CP_3 using: (a) the nested cross-validation. (b) 5-fold CV (80% learning data and 20% test data). (c) 2-fold CV (50% learning data and 50% test data).

	Accuracy	Individual channels			Voting strategy
		CP_1	CP_2	CP_3	
(a) NLOO	Learning	88.0%	90.0%	84.9%	95.4%
	Validation	87.3%	89.9%	84.1%	–
	Testing	88.0%	90.0%	80.0%	94.0%
(b) 5-fold	Learning ($\pm std$ %)	88.7% (± 1.3)	89.7% (± 0.7)	86.5% (± 1.5)	94.1% (± 1.1)
	Validation ($\pm std$ %)	87.8% (± 1.2)	88.8% (± 0.7)	85.7% (± 1.4)	–
	Testing (5-th percentile %)	82.7% (74.0)	85.4% (79.7)	77.7% (70.2)	87.6% (81.9)
(c) 2-fold	Learning ($\pm std$ %)	89.4% (± 2.5)	90.5% (± 2.1)	86.6% (± 2.1)	94.5% (± 2.1)
	Validation ($\pm std$ %)	88.3% (± 2.3)	88.9% (± 2.1)	85.7% (± 2.0)	–
	Testing (5-th percentile %)	78.0% (67.9)	81.7% (72.7)	73.9% (65.7)	83.8% (75.5)

mainly based on the manifestation of the motor symptoms that have not yet developed sufficiently. To recall, the True Positive (TP) rate⁸ of clinical diagnosis is about 26% (9 of 34) for a duration since the first diagnosis <3 years, and of 53% (8 of 15) for a duration <5 years (Adler et al., 2014). Now regarding our data-set, we have 12 patients who have been diagnosed for <3 years and 2 patients for <5 years. For these 14 patients, we obtain a TP = 100% strongly suggesting the utility of our method and its ability to work for early diagnosis cases.

4.4. Results with less learning data

For this part, we used K-fold CV instead of the usual LOOCV to evaluate the model performances⁹ under the constraint of having fewer learning data. This step also gives an indication about the ability of our model to generalize to new, unseen data. We believe that if the model trained only by using half of the data performs almost the same as the model trained on the entire data-set it may suggest that the model trained on the entire data-set may obtain similar results on a data-set that are double the size of ours (Under the assumptions that we have no covariate shift, for more details see Quinonero-Candela et al., 2022).

Table 2(b) reports the accuracies obtained using a 5-fold CV procedure where the learning-set is about 80% of the entire data-set size. Table 2(c) reports the accuracies obtained using a 2-fold CV procedure where the learning-set is about 50% of the data-set size. For both 5-fold and 2-fold CV the random partitioning were repeated 30 000 times. From the Tables 2(a), 2(b), and 2(c) we can observe that the accuracies obtained on the learning and validation sets have not changed significantly, but their variance increases as the learning-set gets smaller. Regarding the test accuracy, it decreased for the case of individual channels and for the voting strategy as the learning-set gets smaller which is expected as less information is present in the learning-set. However, the decrease in accuracy is not substantial, which demonstrates the robustness of our approach, which is mainly due to the utilization of only two well-separable features.

The test accuracy of the 30 000 repetitions is shown in Fig. 14 with the 5-th percentile marked. This indicates that the accuracy of the model is above the indicated value in 95% of the cases.

4.5. Omitting the feature generation step

In order to test the contribution of the Sparse Dynamical Feature generation on the results (blue block in Fig. 4), we tried to omit it. Therefore, the features we used in Section 3.2.2 will be used as they are defined and will be applied directly on the pre-processed signal Y and this for the same time intervals I_1 and I_2 , as a reminder the extracted features will now be defined by:

- $F_1(Y)$: the instant of the lowest amplitude of the signal Y defined by: $\text{argmin}\{Y(I_1)\}$.
- $F_2(Y)$: the mean of the signal Y between 180 ms and 500 ms defined by: $\text{mean}\{Y(I_2)\}$.

The evaluation of the accuracy will be carried out in the same way and the results obtained for a LOO procedure are indicated on the Table 3(a). As a comparison the results keeping the SDF generation block, are shown for the case of a NLOO procedure in the Table 3(b).

Note that for the Table 3(a), there is no validation as there are no hyper-parameters to select. We can observe that the accuracy with SDF generation performed better than without the SDF generation, both in training and in testing, for individual channels or the voting strategy.

We observe that in the case where there is no SDF generation, the accuracy during voting is quite low, and we did not benefit from the voting strategy. Indeed, the more confident voters are in their decision, the better the accuracy of the vote, as we can take advantage of everyone's good knowledge.

We also observe that the accuracy remains reasonably high. This shows first and foremost, that the selected features (F_1 and F_2) are not specific to our method and that they can be used on preprocessed signals directly without further manipulation. This demonstrates the ability of our SDF generation to preserve the physical meaning of the signal it was applied to (as it was designed for), adding an element of explainability to our method.

We believe that our method has yielded better results because our transformation provides different levels of parsimony, offering varying levels of fit to the noise and relevant information. The method thus acts as a filter that retains only the necessary information. The intrinsic functioning of the method, which resembles the functioning of the brain, provides a new perspective (a new point of view) on the data, which allows us to separate patients from non-patients with good accuracy.

5. Conclusion

In this paper a novel approach for the diagnosis of Parkinson's disease using EEG signals is presented. The proposed method is inspired by the functioning of the brain and combines the frequency content, the dynamics and the temporal aspect of the EEG. By means of a set of oscillators, we generated Sparse Dynamical Features (SDF), which are the excitation forces applied on the decoupled oscillators and are calculated in a parsimonious and optimal way to fit the ERP signal of interest. The manner in which these features are generated allows to give a physical meaning to the SDFs while allowing to change the point of view on the data, thereby providing access to more informative features. The method remains generic and can be applied to any real signal, however we have opted for this application for demonstrative purposes.

⁸ Ratio of PD patients identified correctly.

⁹ This only concerns the outer-loop, we kept LOOCV for the inner-loop.

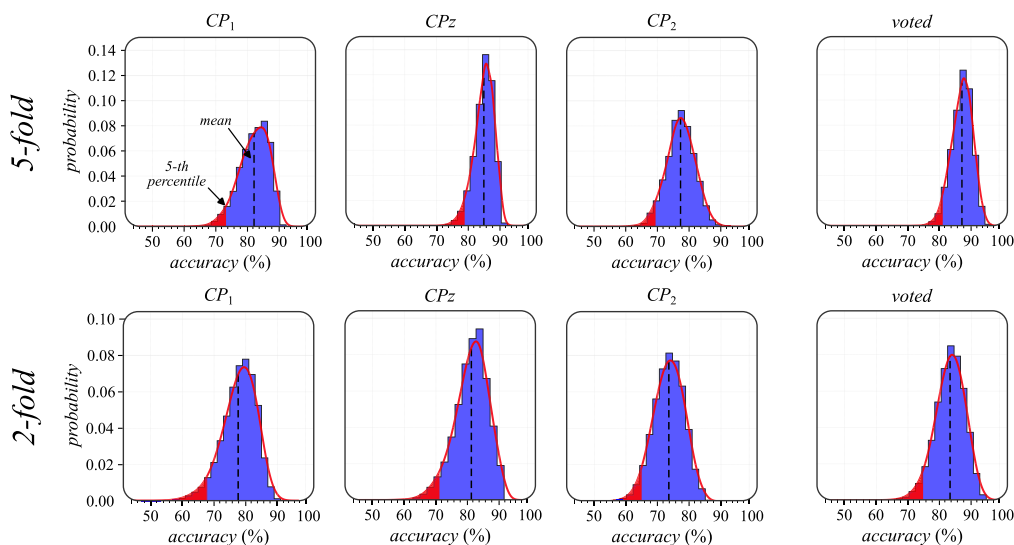


Fig. 14. Resulting test accuracies obtained for the 30000 random partitioning.

Table 3

Resulting accuracies obtained for the voting strategy and the three individual channels CP_1 , CP_z and CP_2 . (a) without the SDF generation in a LOO procedure. (b) with SDF generation in a NLOO procedure.

	Accuracy	Individual channels			Voting strategy
		CP_1	CP_z	CP_2	
(a) Without SDF generation	Learning	73.6%	81.9%	77.2%	78.0%
	Validation	–	–	–	–
	Testing	72.0%	82.0%	74.0%	78.0%
(b) With SDF generation	Learning	88.0%	90.0%	84.9%	95.4%
	Validation	87.3%	89.9%	84.1%	–
	Testing	88.0%	90.0%	80.0%	94.0%

The novel method was evaluated on a publicly available dataset containing $N = 50$ subjects of which 25 have Parkinson’s disease and the remaining individuals serve as a control group. The evaluation of the model was carried out with attention to induce the least bias possible so that we do not have an overly optimistic model that only works on this specific data base, but rather have a simple model with a strong generalization capability that can be used for new data samples. We were able to separate the healthy individuals from the unhealthy ones with an accuracy of 94% using only two features extracted from the generated SDF and thus, by making the models obtained for the channels CP_1 , CP_z and CP_2 vote. The different validity tests conducted suggest that the method and the results are correct and reliable. The proposed new biomarkers worked also for the case where clinicians face the most problems i.e. the early diagnosis of the disease.

The genericity of the method remains the most important aspect, however the method presents other advantages as well, such as: The simplicity and interpretability of the extracted features. Indeed, the latter are easily associated to the disease in question, since they retain the physical meaning of the signal from which they were generated. Moreover, there is a hyper-parameter that allows to manage the degree of fit of the model to the signal of interest, allowing to adjust the compromise between fitting very closely to the signal and risking to capture the noise, or not fitting very closely risking of not capturing the signal’s information. This step acts as a filter, which is very advantageous for this application based on EEGs as they are known to be noisy. The method and classifiers used are simple and computationally inexpensive.

The method also presents some disadvantages such as: The large number of hyper-parameters that were set certainly with the problem

in mind, but are still subject to human error. Moreover, the manual extraction of new features from the generated SDFs may not be straightforward in some cases.

We believe it would be interesting to test the genericity of the method on other problems, either based on EEG signals or other types of signal. A first work was carried out subsequent to this paper, using the same methodology to diagnose Schizophrenia from EEG suggests encouraging results (Meghnojdj et al., 2023). Regarding the algorithm, automating the selection of hyper-parameters can be a possible extension to the framework proposed in this work. An important point that remains unexplored is the evaluation of the method’s effectiveness in a useful but completely different task, consisting of separating ongoing and event-related activities. Finally, exploring alternative function basis for the model can offer intriguing possibilities. Rather than relying on the sinusoidal basis, it is perhaps worth considering an alternative function basis that is better suited to the specific problem at hand.

CRedit authorship contribution statement

Houssef Meghnojdj: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Bogdan Robu:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision. **Mazen Alamir:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Data availability

I have shared the link to the data and the implementation code within the paper.

Acknowledgements

This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

References

- Adler, C.H., Beach, T.G., Hentz, J.G., Shill, H.A., Caviness, J.N., Driver-Dunckley, E., Sabbagh, M.N., Sue, L.I., Jacobson, S.A., Belden, C.M., Dugger, B.N., 2014. Low clinical diagnostic accuracy of early vs advanced parkinson disease: Clinicopathologic study. *Neurology* 83 (5), 406–412. <http://dx.doi.org/10.1212/wnl.0000000000000641>.
- Ayotte, B., Banavar, M.K., Hou, D., Schuckers, S., 2021. Group leakage overestimates performance: A case study in keystroke dynamics. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. pp. 1410–1417. <http://dx.doi.org/10.1109/CVPRW53098.2021.00156>.
- Başar, E., Başar-Eroglu, C., Karakaş, S., Schürmann, M., 2001. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *Int. J. Psychophysiol.* 39 (2–3), 241–248. [http://dx.doi.org/10.1016/S0167-8760\(00\)00145-8](http://dx.doi.org/10.1016/S0167-8760(00)00145-8).
- Balestrino, R., Schapira, A.H., 2020. Parkinson disease. *Eur. J. Neurol.* 27 (1), 27–42. <http://dx.doi.org/10.1111/ene.14108>.
- Berardelli, A., Wenning, G.K., Antonini, A., Berg, D., Bloem, B.R., Bonifati, V., Brooks, D., Burn, D.J., Colosimo, C., Fanciulli, A., Ferreira, J., Gasser, T., Grandas, F., Kanovsky, P., Kostic, V., Kulisevsky, J., Oertel, W., Poewe, W., Reese, J.P., Relja, M., Ruzicka, E., Schrag, A., Seppi, K., Taba, P., Vidali, M., 2013. EFNS/MDS-ES recommendations for the diagnosis of Parkinson's disease. *Eur. J. Neurol.* 20 (1), 16–34. <http://dx.doi.org/10.1111/ene.12022>.
- Bhurane, A.A., Dhok, S., Sharma, M., Yuvaraj, R., Murugappan, M., Acharya, U.R., 2019. Diagnosis of Parkinson's disease from electroencephalography signals using linear and self-similarity features. *Expert Syst.* (June), 1–12. <http://dx.doi.org/10.1111/exsy.12472>.
- Buzsaki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. *Science* 304 (5679), 1926–1929. <http://dx.doi.org/10.1126/science.1099745>.
- Cavanagh, J.F., Kumar, P., Mueller, A.A., Richardson, S.P., Mueen, A., 2018. Diminished EEG habituation to novel events effectively classifies Parkinson's patients. *Clin. Neurophysiol.* 129 (2), 409–418. <http://dx.doi.org/10.1016/j.clinph.2017.11.023>.
- Cavanagh, J.F., Napolitano, A., Wu, C., Mueen, A., 2017. The patient repository for EEG data + computational tools (PRED+CT). *Front. Neuroinform.* 11 (November), 1–9. <http://dx.doi.org/10.3389/fninf.2017.00067>.
- Chaudhuri, K.R., Yates, L., Martinez-Martin, P., 2005. The non-motor symptom complex of Parkinson's disease: A comprehensive assessment is essential. *Curr. Neurol. Neurosci. Rep.* 5 (4), 275–283. <http://dx.doi.org/10.1007/s11910-005-0072-6>.
- Dorsey, E.R., Elbaz, A., Nichols, E., Abbasi, N., Abd-Allah, F., Abdelalim, A., Aducci, J.C., Ansha, M.G., Brayne, C., Choi, J.-Y.J., Collado-Mateo, D., Dahodwala, N., Do, H.P., Edessa, D., Endres, M., Fereshtehnejad, S.-M., Foreman, K.J., Gankpe, F.G., Gupta, R., Hamidi, S., Hankey, G.J., Hay, S.I., Hegazy, M.I., Hibstu, D.T., Kasaeian, A., Khader, Y., Khalil, I., Khang, Y.-H., Kim, Y.J., Kokubo, Y., Logroscino, G., Massano, J., Ibrahim, N.M., Mohammed, M.A., Mohammadi, A., Moradi-Lakeh, M., Naghavi, M., Nguyen, B.T., Nirayo, Y.L., Ogbo, F.A., Owolabi, M.O., Pereira, D.M., Postma, M.J., Qorbani, M., Rahman, M.A., Roba, K.T., Safari, H., Safiri, S., Satpathy, M., Sawhney, M., Shafieesabet, A., Shiferaw, M.S., Smith, M., Szoek, C.E.I., Tabarés-Seisdedos, R., Truong, N.T., Ukwaja, K.N., Venketasubramanian, N., Villafaina, S., Gidey weldegweres, K., Westerman, R., Wijeratne, T., Winkler, A.S., Xuan, B.T., Yonemoto, N., Feigin, V.L., Vos, T., Murray, C.J.L., 2018. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 17 (11), 939–953. [http://dx.doi.org/10.1016/s1474-4422\(18\)30295-3](http://dx.doi.org/10.1016/s1474-4422(18)30295-3).
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *Ann. Statist.* 32 (2).
- Ferrazoli, N., Donadon, C., Rezende, A., Skarzynski, P.H., Sanfins, M.D., 2022. The application of P300-long-latency auditory-evoked potential in Parkinson disease. *Int. Arch. Otorhinolaryngol.* 26 (1), e158–e166. <http://dx.doi.org/10.1055/s-0040-1722250>.
- Fisher, R.A., 1936. Design of experiments. *Br. Med. J.* 1 (1), 554.
- Fries, P., 2005. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences* 9 (10), 474–480. <http://dx.doi.org/10.1016/j.tics.2005.08.011>.
- Frost, J., 2020. Hypothesis Testing. Statistics by Jim Publishing.
- Gibb, W.R., Lees, A.J., 1988. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J. Neurol., Neurosurg. Psychiatry* 51 (6), 745–752. <http://dx.doi.org/10.1136/jnnp.51.6.745>.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7 (7 DEC), 1–13. <http://dx.doi.org/10.3389/fnins.2013.00267>.
- Han, C.X., Wang, J., Yi, G.S., Che, Y.Q., 2013. Investigation of EEG abnormalities in the early stage of Parkinson's disease. *Cogn. Neurodyn.* 7 (4), 351–359. <http://dx.doi.org/10.1007/s11571-013-9247-z>.
- Hansch, E.C., Syndulko, K., Cohen, S.N., Goldberg, Z.I., Potvin, A.R., Tourtellotte, W.W., 1982. Cognition in Parkinson disease: An event-related potential perspective. *Ann. Neurol.* 11 (6), 599–607. <http://dx.doi.org/10.1002/ana.410110608>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. In: *Springer Series in Statistics*, Springer, New York, NY.
- Kalia, L.V., Lang, A.E., 2015. Parkinson's disease. *Lancet* 386 (9996), 896–912. [http://dx.doi.org/10.1016/s0140-6736\(14\)61393-3](http://dx.doi.org/10.1016/s0140-6736(14)61393-3).
- Kaufman, S., Rosset, S., Perlich, C., Stitelman, O., 2012. Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data* 6 (4), <http://dx.doi.org/10.1145/2382577.2382579>.
- Liu, G., Zhang, Y., Hu, Z., Du, X., Wu, W., Xu, C., Wang, X., Li, S., 2017. Complexity analysis of electroencephalogram dynamics in patients with Parkinson's disease. *Parkinson's Disease* 2017, 1–9. <http://dx.doi.org/10.1155/2017/8701061>.
- Luca, P.-T., 2021. ICLabel tutorial: EEG independent component labeling, Swartz Center for Computational Neuroscience, University of California, San Diego, URL <https://labeling.ucsd.edu/tutorial>.
- Luck, S.J., 2005. An Introduction to the Event-Related Potential Technique. In: *Cognitive neuroscience*, Bradford Books, Cambridge, MA.
- Meghroudj, H., Robü, B., Alamir, M., 2023. Sparse control-inspired features generation for Schizophrenia diagnosis. In: *IFAC 2023, 22nd World Congress, Jul 2023, Yokohama, Japan*. [Accepted].
- Molinari, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 21 (15), 3301–3307. <http://dx.doi.org/10.1093/bioinformatics/bti499>.
- Nichols, T.E., Holmes, A.P., 2003. Nonparametric permutation tests for functional neuroimaging. In: *Human Brain Function: Second Edition*. 25, (July), pp. 887–910. <http://dx.doi.org/10.1016/B978-012264841-0/50048-2>.
- Oh, S.L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., Acharya, U.R., 2020. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput. Appl.* 32 (15), 10927–10933. <http://dx.doi.org/10.1007/s00521-018-3689-5>.
- Olejniczak, P., 2006. Neurophysiologic basis of EEG. *J. Clin. Neurophysiol.* 23 (3), 186–189. <http://dx.doi.org/10.1097/01.wnp.0000220079.61973.6c>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pfeiffer, R.F., 2016. Non-motor symptoms in Parkinson's disease. *Parkinsonism Rel. Disord.* 22, S119–S122. <http://dx.doi.org/10.1016/j.parkrelidis.2015.09.004>.
- Polich, J., 2007. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118 (10), 2128–2148. <http://dx.doi.org/10.1016/j.clinph.2007.04.019>.
- Quinero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D. (Eds.), 2022. Dataset shift in machine learning. In: *Neural Information Processing series*, MIT Press, London, England.
- Rizzo, G., Copetti, M., Arcuti, S., Martino, D., Fontana, A., Logroscino, G., 2016. Accuracy of clinical diagnosis of Parkinson disease. *Neurology* 86 (6), 566–576. <http://dx.doi.org/10.1212/WNL.0000000000002350>.
- Soufneyestani, M., Dowling, D., Khan, A., 2020. Electroencephalography (EEG) technology applications and available devices. *Appl. Sci. (Switzerland)* 10 (21), 1–23. <http://dx.doi.org/10.3390/app10217453>.
- Starkstein, S.E., Esteguy, M., Berthier, M.L., Garcia, H., Leiguarda, R., 1989. Evoked potentials, reaction time and cognitive performance in on and off phases of Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 52 (3), 338–340. <http://dx.doi.org/10.1136/jnnp.52.3.338>.
- Stone, M., 1976. Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* 38 (1), 102. <http://dx.doi.org/10.1111/j.2517-6161.1976.tb01573.x>.
- Tharwat, A., 2018. Independent component analysis: An introduction. *Appl. Comput. Inform.* 17 (2), 222–249. <http://dx.doi.org/10.1016/j.aci.2018.08.006>.
- Urigüen, J.A., Garcia-Zapirain, B., 2015. EEG artifact removal - State-of-the-art and guidelines. *J. Neural Eng.* 12 (3), <http://dx.doi.org/10.1088/1741-2560/12/3/031001>.

- Varma, S., Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* 7 (1), <http://dx.doi.org/10.1186/1471-2105-7-91>.
- Ward, L.M., 2003. Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences* 7 (12), 553–559. <http://dx.doi.org/10.1016/j.tics.2003.10.012>.
- WHO, 2006. *Neurological Disorders : Public Health Challenges*. World Health Organization, Geneva.
- Witjas, T., Kaphan, E., Azulay, J.P., Blin, O., Ceccaldi, M., Pouget, J., Poncet, M., Ali Chérif, A., 2002. Nonmotor fluctuations in Parkinson's disease: Frequent and disabling. *Neurology* 59 (3), 408–413. <http://dx.doi.org/10.1212/WNL.59.3.408>.
- Yuvaraj, R., Rajendra Acharya, U., Hagiwara, Y., 2018. A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals. *Neural Comput. Appl.* 30 (4), 1225–1235. <http://dx.doi.org/10.1007/s00521-016-2756-z>.