Deep learning on high-density EEG during a cognitive task distinguishes patients with Parkinson's disease from healthy controls

Yves Denoyer, Joan Duprez, Jean-François Houvenaghel, Fabrice Wendling, Pascal Benquet

▶ **To cite this version:**

Yves Denoyer, Joan Duprez, Jean-François Houvenaghel, Fabrice Wendling, Pascal Benquet. Deep learning on high-density EEG during a cognitive task distinguishes patients with Parkinson's disease from healthy controls. Journal of Neural Engineering, 2025, 22 (4), pp.046002. ⟨10.1088/1741-2552/ade6a9⟩. ⟨hal-05162720v2⟩

HAL Id: hal-05162720

https://hal.science/hal-05162720v2

Submitted on 15 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Journal of Neural Engineering

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

Deep learning on high-density EEG during a cognitive task distinguishes patients with Parkinson's disease from healthy controls

To cite this article: Yves Denoyer *et al* 2025 *J. Neural Eng.* **22** 046002

View the article online for updates and enhancements.

You may also like

- Reference layer adaptive filtering (RLAF) for EEG artifact reduction in simultaneous EEG-fMRI
David Steyrl, Gunther Krausz, Karl Koschutnig et al.
- Ballistocardiogram artifact removal from EEG signals using adaptive filtering of EOG signals
Myung H In, Soo Y Lee, Tae S Park et al.
- What can be found in scalp EEG spectrum beyond common frequency bands. EEG–fMRI study
R Marecek, M Lamos, M Mikl et al.

This content was downloaded from IP address 129.20.30.41 on 15/07/2025 at 10:51

Journal of Neural Engineering

PAPER

Deep learning on high-density EEG during a cognitive task distinguishes patients with Parkinson's disease from healthy controls

Yves Denoyer[1,*], **Joan Duprez**[1], **Jean-François Houvenaghel**[1,2], **Fabrice Wendling**[1] and **Pascal Benquet**[1]

[1] Univ Rennes, LTSI—U1099, F-35000 Rennes, France
[2] Neurophysiology Department, Rennes University Hospital, Rennes, France
[*] Author to whom any correspondence should be addressed.

E-mail: yt.denoyer@gmail.com, joan.duprez@univ-rennes1.fr, Jeanfrancois.HOUVENAGHEL@chu-rennes.fr, fabrice.wendling@univ-rennes1.fr and pascal.benquet@univ-rennes1.fr

Keywords: high density EEG, cognitive control, Simon task, Parkinson's disease, deep learning

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
4 November 2024

REVISED
30 May 2025

ACCEPTED FOR PUBLICATION
20 June 2025

PUBLISHED
3 July 2025

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Abstract

Objective. Parkinson's disease (PD) is a neurodegenerative disorder characterized by motor and non-motor symptoms, including cognitive impairment. Its diagnosis, which used to be based on clinical assessment, increasingly relies on biomarkers. While electroencephalography (EEG) biomarkers are still at an experimental stage, they have been studied using deep learning (DL) models. Our aim was to determine whether a cognitive task could improve the accuracy of EEG-based disease detection by activating cortical regions affected by the disease. *Approach.* We trained a DL model to discriminate PD patients from controls based on their high-density EEG recordings. Previous studies have employed a range of preprocessing techniques, models and, predominantly, resting state (RS) EEG. We also investigated different network architectures and hyperparameters, and the role of spatial and temporal resolution. *Main results.* The best model gave a classification accuracy of 83% on the cognitive task EEG dataset and 76% on the RS EEG dataset. Sensitivity analysis indicated that the model predominantly uses specific temporal and spatial components of the EEG in the cognitive task condition, differing from the RS. *Significance.* Our results suggest that cortical activation by the cognitive task unveils EEG features that are effective in distinguishing between PD and controls. These features can be used by the model, thereby improving its diagnostic accuracy.

Abbreviations

CNN	convolutional neural network
DL	deep learning
EEG	electroencephalography
HC	healthy control
HD	high density
ICA	independent component analysis
MRI	magnetic resonance imaging
PD	Parkinson disease
RS	resting state
SPECT	single photon emission computed tomography

1. Introduction

PD is a neurodegenerative disorder that primarily affects the midbrain and basal ganglia (Galvan and Wichmann 2008), leading to motor dysfunction due to the progressive loss of dopaminergic neurons. Non-motor features may precede the formal diagnosis of PD by several years (Fereshtehnejad *et al* 2017). Among them, cognitive impairment results from cortical damage (Postuma *et al* 2015) and/or dysregulations in the subcortical modulation of cortical functions (Ott *et al* 2023). Impairment of

© 2025 The Author(s). Published by IOP Publishing Ltd

executive functions (EF) is one of the most common and earliest cognitive disorders in PD (Dujardin *et al* 2015). Patients may have difficulties with task switching, planning, inhibitory control, and conflict resolution (Dirnberger and Jahanshahi 2013, Cavanagh *et al* 2022). The assessment of cognitive action control, a subtype of the EF, typically involves conflict tasks that require the inhibition of automatic responses in favor of goal-directed actions. The Simon task (ST) (Simon and Rudell 1967) is one such task in which participants respond with either hand to a stimulus color, regardless of its spatial location on a screen, with congruent (response on the same side as the stimulus location) or incongruent trials. In ST, PD patients show a decrease in the inhibition of incorrect responses and an increase in fast errors (Wylie *et al* 2010, Duprez *et al* 2017).

The diagnosis of PD, which is based on clinical criteria, increasingly includes biomarkers such as SPECT or F-DOPA positron emission tomography. There are other various candidates, including EEG (Waninger *et al* 2020). PD causes alterations in the surface patterns of the EEG, some of which respond to dopamine replacement therapy (Devos 2004, Geraedts *et al* 2021). Furthermore, within the RS-EEG, phase-amplitude coupling (PAC) between beta and gamma bands has been proposed as a diagnostic marker (de Hemptinne *et al* 2013, 2015, Swann *et al* 2015). RS-EEG is widely studied as a biomarker in clinical research. However, for certain neurological conditions, it could be more relevant to use a standardized paradigm aimed at activating specific brain circuits to improve accuracy.

The increasing use of DL in biomarker research for diagnostic purposes has sparked interest in the use of neural networks (NNs) to distinguish PD patients from HCs, especially using RS-EEG (Shi *et al* 2019, Oh *et al* 2020, Chu *et al* 2021, Khare *et al* 2021, Lee *et al* 2021, Loh *et al* 2021, Qiu *et al* 2022, Shaban and Amara 2022, Nour *et al* 2023). RS-EEG recordings provide information about spontaneous neuronal activity that reflects the default mode network (DMN), a fundamental state of the brain (Das *et al* 2022). During task performance, changes in alpha and beta band power are frequently observed in surface EEG recordings due to cortical desynchronization (Pfurtscheller and Lopes da Silva 1999). Especially in motor (Defebvre *et al* 1999) and visual tasks (Heida *et al* 2014) this desynchronization is disturbed by PD. Abnormalities in event-related desynchronization have been reported in the alpha (Defebvre *et al* 1999), beta (Brown and Marsden 1999, George *et al* 2013, Rosenblum *et al* 2022) and gamma (Waninger *et al* 2020) frequency bands. The correlates of cognitive control in the EEG were explored using cognitive-motor dual-tasks, which revealed changes in the alpha and theta bands (Singh *et al* 2018, Stuart *et al* 2021, Nwogo *et al* 2022). PD can thus affect the EEG in most of its components, which makes the problem suitable for a DL approach, as a NN is able to capture a wide range of available information to optimize the function it is modeling.

In this study, we used high-density EEG (HD-EEG) data recorded from PD patients and HC during ST (Duprez *et al* 2022) to train a DL model to classify PD and HC. The ST induces a dynamic cortical activation pattern (figure 4(A)). Firstly, the stimulus perception stage (from 0 to 150 ms) leads to desynchronization and reorganization of the functional networks of the occipital, parietal and temporal regions. From 150 ms onwards, this reorganization begins to affect frontal regions (Hassan *et al* 2015). The conflict task then modulates the EEG in beta and gamma bands across a large network involving parietal, prefrontal, cingulate and premotor cortex before reaction time, which occurs between 500 and 600 ms, and prefrontal dorsolateral and anterior cingulate cortex after reaction time (Duprez *et al* 2022). Motor execution induces a desynchronization of alpha and beta rhythms in somatosensory, motor and premotor cortex, and gamma synchronization in premotor and motor cortex (Babiloni *et al* 2016).

Given the results of previous studies on EEG changes in PD during cognitive (Schmiedt-Fehr *et al* 2007, Singh *et al* 2018, Rosenblum *et al* 2022) and motor (Defebvre *et al* 1999) tasks, we hypothesized that EEG data recorded during a ST would contain more information about PD than RS-EEG, leading to better classification performance. We also wanted to investigate different NN architectures and hyperparameters. Finally, we studied potential biases related to the database partitioning method by performing comparative analyzes between subject-wise and sample-wise datasets.

2. Methods

The pipeline applied to our EEG data and the key points of the study are summarized in figure 1. A HD-EEG was recorded during a ST preceded by a 10 min RS recording. The EEG data were preprocessed by filtering and noise reduction through interpolation of invalid channels, and ICA to remove artifact components. We performed trial-based segmentation into 1.9 s samples. The database was split 80%–20% subject-wise to test different architectures and hyperparameters of the CNN. The selected model was then trained with the dataset in the following paradigms. We performed a 10-fold testing with a 90%–10% subject-wise split; a spatial resolution test comparing 199 vs 60 vs 20 channels; a training process with a sample-wise split of the dataset; a sensitivity analysis to clarify the role of the cognitive task, whereby the influence of different signal components on classification was quantified in the time domain and in the electrode domain; a training process on the RS EEG dataset; finally, we tested the accuracy on an

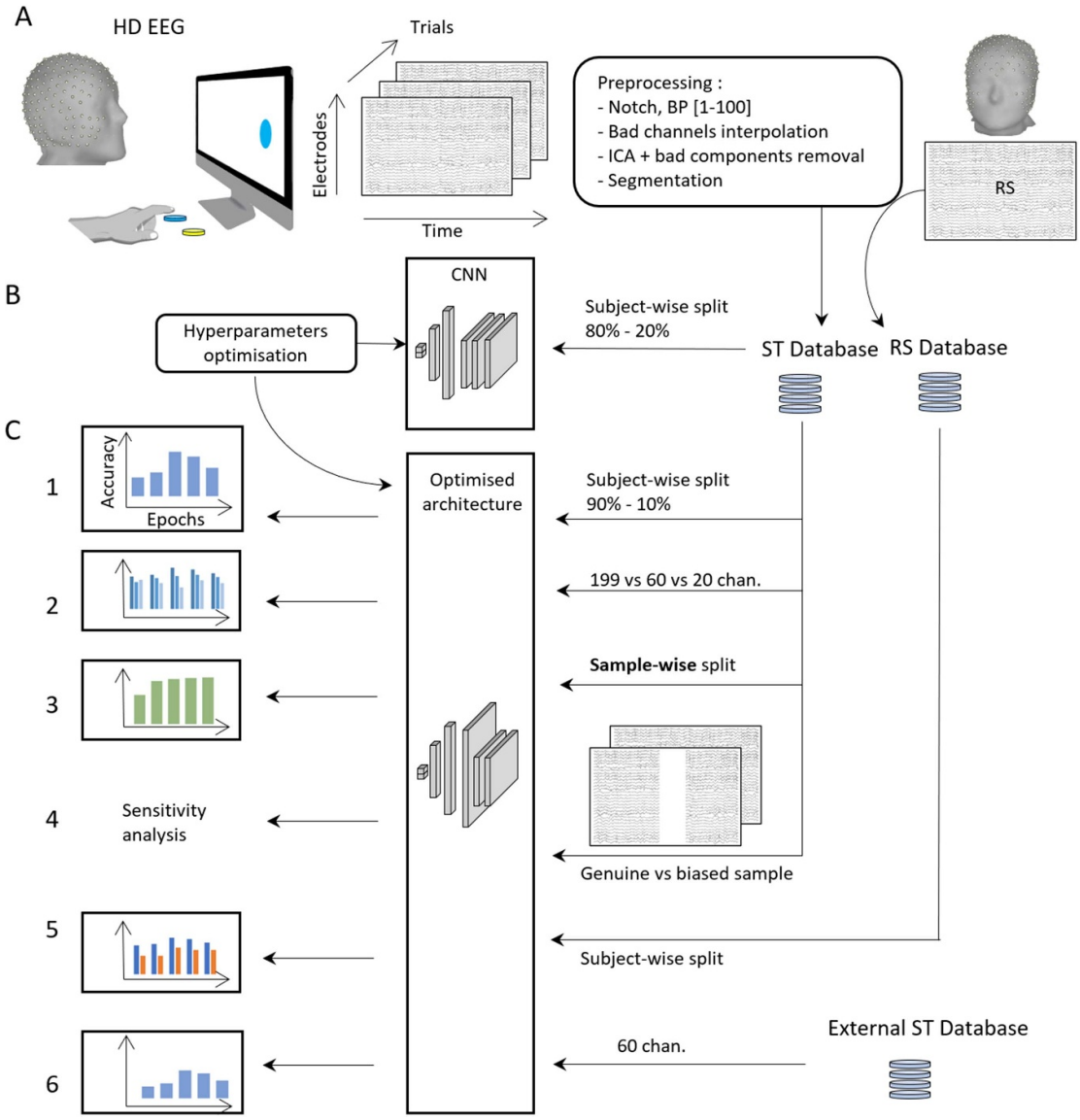

Figure 1. General overview of the study. (A): A high-density electroencephalogram (HD-EEG) was recorded during a Simon task (ST) preceded by a 10 min resting-state (RS) recording. The EEG data were preprocessed by filtering, manual selection and interpolation of bad channels, independent component analysis (ICA) and manual selection and interpolation of bad components, and trial-based segmentation into 1.9 s samples. (B): The dataset was split 80%/20% subject-wise to test different architectures and hyperparameters of the convolutional neural network. (C): During learning, an epoch refers to an entire pass on the training dataset (*x* axis). The selected model was trained with the dataset in the following paradigms: 1 (main result): 10-fold testing with a 90%–10% subject-wise split; 2: Test of spatial resolution: 199 vs 60 vs 20 channels; 3: Sample wise split vs subject-wise split of the dataset; 4: Sensitivity analysis; 5: Accuracy on RS vs ST datasets; 6: Test of an external database after training on the local database.

external database after training on the spatially down-sampled local database.

2.1. Participants

Thirty patients (15 men, 15 women) with a diagnosis of idiopathic PD, aged between 45 and 73 yr (mean = 60.4, sd = 7.3) and 30 HC (15 men and 15 women), aged between 45 and 70 yr (mean = 61.7, sd = 7.3) were included. Patients with major cognitive deficits (Montreal Cognitive Assessment (MoCA) <22) or with a previous or current neurological disease (except PD) or psychiatric disorder, as well as patients with deep brain stimulation, were not included in the study. The same criteria were used for HC. One patient was excluded from further analyses due to a technical problem during the recordings. Thus, the analyses were performed on 30 HC and 29 PD patients (table 1).

All HC were recruited from the general population at public conferences and through participation announcements. Patients were recruited from the University Hospital of Rennes, France, during hospitalization for their usual care. All participants gave informed consent to participate in the study, which was approved by a national ethics committee (CPP ID-RCB: 2019-A00608-49; approval

Table 1. Demographic and clinical data. Data are presented as mean (sd).

	HC ($n = 30$)	PD ($n = 29$)	t-test
Age (years)	61.7 (7.3)	60.4 (7.3)	$t = 0.77$; $p = 0.44$
Sex (M: F)	15:15	14: 15	—
Education (years)	13.5 (3.6)	12.5 (3.2)	$t = 1.19$; $p = 0.24$
MoCA (/30)	28 (1.9)	26.6 (2.0)	$t = 2.72$; $p < 0.01$
Disease duration (years)	—	12.1 (4.2)	—
Side of onset (L: R)	—	16: 13	—
LEDD (mg/day)	—	1348.0 (462.0)	—
UPDRS III On	—	10.0 (5.7)	—
UPDRS III Off	—	36.8 (12.5)	—
Hoehn & Yarh On	—	1.1 (0.8)	—
Hoehn & Yarh Off	—	2.0 (0.4)	—
Schwabb & England On	—	93.0 (5.6)	—
Schwabb & England Off	—	76.5 (10.3)	—

UPDRS: Unified Parkinson's Disease Rating Scale.
LEDD: Levodopa equivalent daily dose (Tomlinson *et al* 2010).
MoCA: Montreal Cognitive Assessment.

number: 19.03.08.63626). The patients provided written informed consent for all procedures, analysis, and publication of anonymized clinical data in accordance with the Declaration of Helsinki.

2.2. Experimental task

The Simon effect (Simon and Rudell 1967) refers to the difference in reaction time or accuracy between congruent and incongruent stimuli. At every trial, a circle randomly colored in blue or yellow appears on one random side of the screen. The subject is instructed to press the colored button corresponding to the color of the circle, regardless of its position. Reaction times are typically shorter and accuracy is better for congruent (stimulus and response button on the same side) than for non-congruent trials. For more details on the task, see Duprez *et al* (2022). Each recording included a first period of RS, consisting of the subject sitting still for about 10 min.

2.3. Data acquisition and preprocessing

The EEG signals were recorded on a HD-EEG system (EGI, Electrical Geodesic Inc), with 256-channel and a sampling rate of 1 kHz. A period of RS-EEG was followed by the ST. The facial electrodes were discarded, 199 channels distributed over the entire scalp were retained. The preprocessing software used was provided by EEGLab (Delorme and Makeig 2004). Preprocessing included notch filtering, bandpass filtering (1–100 Hz), and manual selection and interpolation of bad channels. These were distinguished by a markedly divergent signal from adjacent electrodes. Moreover, they exhibited a low-amplitude signal lacking the physiological frequencies of the EEG, with excessive sensitivity to mechanical artifacts and to the 50 Hz power. To improve the signal-to-noise ratio, an ICA was then performed, followed by suppression of bad components before inverse transformation. The identification of bad components was contingent upon the absence of physiological EEG rhythms and the presence of EMG or electrode artifacts. For ST recordings, EEG was segmented into single trials, with each trial lasting from the stimulus minus 700 ms to the stimulus plus 1200 ms. Trials with significant motion or myogram artifacts, despite filtering and ICA, were manually discarded. Each trial was labeled by the subject's disease status into a data sample, resulting in the ST database containing 43 706 samples, 22 594 from PD and 21 112 from HC, with a mean of 546 samples per subject.

2.4. Open source Iowa database

The Iowa Database, an open-source database by Singh *et al* (2018), was used to access EEGs from 29 PD and 30 HC subjects recorded during the ST. The EEG were recorded on a Brainvision 64 channel device according to the 10–10 EEG system, from which the authors subtracted the 4 most ventral channels FT9, FT10, TP9, and TP10. The samples lasted from -1500 to $+5000$ ms, with a sampling rate of 500 Hz. Each session averaged 856 samples. We used this data to test whether the classification could be applied to EEG recordings using with different techniques and devices. To match the Iowa database, we downsampled our own signals to 500 Hz and kept the 60 channels mentioned above. If the electrodes of the 60-channel cap did not exactly match with those of the 256-channel cap, the spatially closest electrode was selected.

2.5. Learning process

We used the entire database for each learning task, and did not conduct subgroup studies. For the choose of hyperparameters, the database was split 80%–20% between training and validation datasets. This was dictated by the need to repeat time-consuming learning processes with numerous parameter combinations. Once defined the optimal hyperparameters, the database was split 90%–10% between training and validation datasets, and all the learning processes were done according to a 10-fold schema.

2.6. ST-EEG vs RS-EEG

We applied the same preprocessing to the RS-EEG as to the ST-EEG. We then split the EEG into samples of the same duration as in the ST dataset (1.9 s), with an overlap of 0.7 s, to obtain a number of samples (on

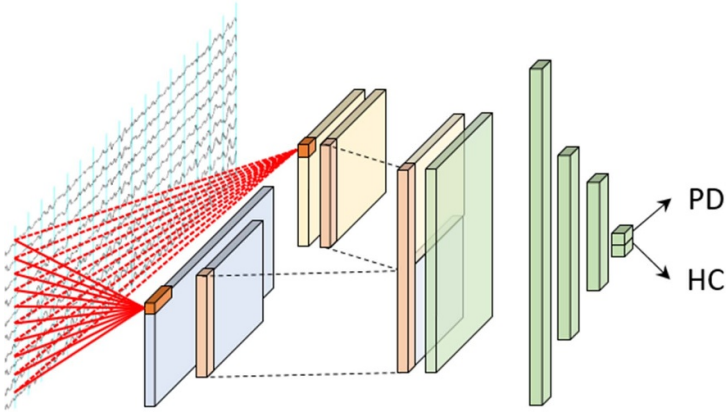

Figure 2. Architecture of the 2-scale convolutional neural network: EEG signal was downsampled (meanpool layer) to a sampling rate (SR) of 125 Hz (blue blocks) or 62 (yellow blocks). Signals with different SR were processed separately in a convolutional block, which consists of convolutional layers separated by maxpool layers if necessary. The blue blocks were intended to capture gamma band features (under 62 Hz), while the yellow blocks could only process beta or lower bands. The outputs were then concatenated along the electrode axis (green blocks), to facilitate the detection of patterns such as beta-gamma correlations, altered in Parkinson Disease.

average 539 per file) close to that of the ST dataset (546). After selecting the model that performed best on the ST database, we performed the same learning process on this RS database and compared the accuracies.

2.7. Spatial and temporal resolution

To evaluate the effects of temporal and spatial resolution on classification accuracy, we compared the performance and accuracy curves of our database with three different channel configurations: 20 channels (standard 10–20 EEG setup), 60 channels identically to the Iowa database, and all 199 channels of our database. For details about these electrode configurations, see figure 1 in supplementary data. We also applied different subsampling rates to the EEG before the first convolutional layer, ranging from 1:4 to 1:16, corresponding to sample rates from 250 to 62 Hz.

2.8. DL models

The NN and datasets were created using Tensorflow (Abadi *et al* 2016). We started with a classic CNN, alternating convolutional and pooling layers, followed by dense layers. We then introduced architectural variations and searched for optimal hyperparameters. Three types of architectures were tested with a set of hyperparameters based on our objectives and previous works on this topic (Khare *et al* 2021, Lee *et al* 2021, Qiu *et al* 2022, Shaban and Amara 2022, Li *et al* 2023). In all cases, the input layer performed a downsampling whose ratio was one of the model's hyperparameters listed below. We started with a basic CNN with 2–6 convolutional layers, separated individually or in pairs by 2–3 pooling layers, followed by 2–3 dense layers before an output layer; we tested a residual network with 2–3 residual cores followed by 2–3 dense layers, and finally a '2-scale' network intended to enhance the relationships between the gamma band and lower frequency bands. It was designed to work with a downsampling at the input of 1:8 or more, resulting in maximum frequencies in the gamma band when applied to a 1000 Hz input signal. Another 1:2 downsampling applied to the same signal, resulted in a maximum frequency in the upper beta band, that we used to feed a separate input branch of the network, processed independently before merging the two branches as shown in figure 2. We then assessed different CNN architectures known to provide improvements in specific learning tasks. These variations included a dropout layer after convolutional or dense layers, batch normalization layers after convolutional layers, and residual cores instead of single convolutional layers, as adopted by Qiu *et al* (2022).

2.9. Hyperparameter optimization

To compare the performance of the networks and hyperparameters, we executed five-epoch learning processes using combinations of hyperparameters by systematically varying: the rate of the first pool layer, which is somehow related to the sampling frequency, from 4 to 12 on the 500 Hz data; the stride length for first convolutional layer, from 1 to 3; the length of the convolution kernel in the time domain, from 3 to 5; the number of kernels per convolutional layer, from 24 to 128; the number of convolutional layers, 3 or 4, and dense layers, 2 or 3; if applicable, the dropout rate from 0.2 to 0.5 for dropout layers. Finally during the learning, we tested batch sizes of 16, 32, or 64. We then selected the optimal hyperparameters for the three types of architecture and compared their performance on a 10-fold test of the database.

2.10. 'Sample-wise' vs 'subject-wise' datasets

Each EEG sample in the database was labeled PD or HC. We assumed that the EEG data contains information related to the disease on one hand, and to the subject and the recording technique on the other hand. We referred to these as 'disease-related' and 'subject-related' information. DL algorithms are designed to exploit all available information in the data to optimize learning. Therefore, if the model identifies subject-related information to be useful to disease status classification, it should use it. This can happen here, since every subject generates a large subset of samples sharing the same label and common EEG characteristics, non-relevant for our purpose, such as features related to the anatomy or electrode impedances.

Our reference method to split the database involved the division of subjects into two sets: 90% for training and 10% for validation, ensuring the ratio PD:HC remained constant. Each set comprised all the samples from each of its subjects, and was therefore referred to as a 'subject-wise' dataset. Consequently, validation and training samples were systematically derived from different subjects. the accuracy was calculated by counting the number of correctly classified subjects, based on the average output across samples.

Then, the method of partitioning the data was modified. The sets were created by shuffling all samples of all subjects and distributing them into training and validation datasets. These datasets were referred to as 'sample-wise'. Finally, we constructed sample-wise datasets with a total number of subjects that we gradually reduced from 40 to 8 with steps of 4, maintaining a 1:1 ratio between PD and HC. The hypothesis was that reducing the number of subjects would make the classification task easier and improve accuracy if the model learned predominantly subject-specific features. Accordingly, in the sample-wise dataset, the model's performance for classifying a subject was the average model output across the samples derived from this subject.

2.11. Sensitivity analysis

NN make it difficult to explicit which features in the samples are the most relevant for classification. Sensitivity analysis methods measure the change in prediction resulting from a change in the samples, providing an indication of which part of the signal is most used. This analysis is performed after the model has been trained on all samples from all subjects.

We used an integrated gradient (IG) method. This method described by Sundararajan *et al* measures attribution of input features (Sundararajan *et al* 2017). It is widely used in image processing, and has been applied to EEG, as it is or in some variants (Kawai *et al* 2022).

After training the model, we embedded it into an IG algorithm. Then we selected the EEG samples with a prediction accuracy $>$ to 80% and we measured the average IG across all subjects. To highlight the role of the different time segments of the EEG samples, the result was visualized as a 199 $\times$ 1900 pixels image, the vertical axis representing the electrodes and the horizontal axis representing time. Then, to focus on the role of different cortical regions in the electrode domain, the result was visualized as the average of the gradients on the time axis for each electrode. The RS and the ST datasets were subjected to the same procedure, after which the results were compared in terms of electrodes and timing.

Additionally the temporal aspect of the ST dataset was examined with a segmentation *a priori*, whereby sensitivity to segments approximating the perceptual, pre-motor, and motor components of ST was investigated. In our samples, the 1900 ms time window includes a period of 700 ms before the stimulus, which can be considered the baseline, and extends to 1200 s after the stimulus, including around 600 ms after the motor response in the ST, which was considered sufficient time for a return to baseline. The duration of the window therefore exceeds that of the task. Biased samples were obtained by setting the signal to zero in one 200 ms time segment of the EEG sample, as illustrated in figure 4(B). The choice of 200 ms allowed us to define 8 temporal segments, three of which roughly corresponding to the visuospatial, cognitive and motor components of the task.

We compared the changes in model response between the original and the biased sample for each of the temporal segments. The data was averaged across all subjects then compared among the 200 ms segments. The average change in the response for a given segment reflects the relative quantity of relevant information found by the model in this segment (figure 4(C)).

3. Results

Briefly, the optimal architecture and hyperparameters resulted in a classification accuracy of 83% on the ST database. The models started to overfit after 3 epochs. We then tested the best model on the RS database, and its accuracy was 76%.

3.1. Type or network and hyperparameters

The optimal parameters yielded to an accuracy of 75% for a classical CNN and 76% for a residual network. The 2-scale network slightly outperformed them with an accuracy of 83%. The optimal hyperparameters were: a pool rate of 12 and a stride length of 1 for the first pool layer; a kernel length of 3; a number of convolutional layers of 3 with 48 kernels in each layer; a number of dense layers of 2 with sizes of 32 and 8. Additionally, we found that either adding a separate convolutional first layer, an initial normalization layer, dropout layers after dense or convolutional layers, and batch normalization after convolutional layers did not improve performance. Overall,

we observed that having a number of trainable parameters greater than 100 000 or less than 30 000 was associated with poorer performance in both the classical and 2-scale networks. The optimal batch size was 32.

3.2. Accuracy curves and overfitting

Overfitting occurred quickly when using a subject-wise dataset, with a peak in accuracy after 3 epochs (figure 3(A)). Further epochs did not improve it for any type of model.

3.3. Subject-wise vs sample-wise datasets

For sample-wise datasets consisting of an increasing number of subjects, equally divided between PD and HC, the speed of convergence and the maximum accuracy decreased as the number of subjects increased. Maximum accuracy was 98% for data sets containing all subjects, and up to 100% for smaller datasets (figure 3(C)).

3.4. ST vs RS-EEG

When comparing the ST and the RS datasets, learning curve followed a similar pattern over epochs, with a rapid peak around the third epoch, followed by overfitting. However, the maximum accuracy on the RS dataset was 76%, slightly lower than that of 81% on the ST dataset (figure 3(D)).

3.5. Spatial and temporal resolutions

When comparing the performances on data with 199, 60, and 20 channels, the accuracy decreased with the number of channels used, but not proportionally. On average, using 60 channels results in a 6% reduction in accuracy compared to 199 channels, while 20 channels lead to an 11% reduction. On the opposite, an increase in time resolution did not improve the accuracy beyond a threshold around 100 Hz. A 1/12 first layer pooling rate induced the highest maximal accuracy in the 2-scale network (figure 3(E)).

3.6. Sensitivity analysis

The contribution of EEG signals corresponding to different temporal segments and cortical regions was estimated and visualized following the method exposed in 2.7.6. Regarding the temporal aspect, the segments corresponding to the premotor and motor components of the task are associated with higher IG and thus are likely to be more discriminating than the others, as shown in figure 4(B). On the opposite, in the RS condition, the IGs exhibit a uniform intensity along time, indicating that no segment plays a predominant role. For the purpose of comparison, the relevant figure is provided in the supplementary data (supplementary figure 2). sensitivity analysis based on time segments corresponding to the three main components of the task confirmed this result. We measured and averaged the gap in the trained model's answer between the genuine sample and the biased samples obtained by setting the signal to zero in a time segment (figure 4(B)). The time segment [200–400 ms] produced the greatest variation in prediction, corresponding to the cognitive/premotor and motor execution (figure 4(C)), suggesting that these segments are the most useful for the model to discriminate between HC and PD.

3.7. Results on another HD-EEG during the ST

The model trained on the entire database achieved a mean accuracy of 61.8% in classifying the files from the Iowa database. However, it exhibited a bias towards false positives, correctly identifying 22 out of 27 patients but only 12 out of 28 controls. The mean scores for on vs off status during recordings were 72.9% and 74.9 %, respectively.

4. Discussion

We confirmed that a simple CNN architecture is able to discriminate PD patients from HC with 83% accuracy, based on HD-EEG data. Our study highlighted the potential information richness of EEG recordings obtained during a cognitive task compared to RS-EEG. In addition, we pointed out the central role of the dataset partitioning method, suggesting that EEG contains more subject-related than disease-related features.

4.1. EEG as a biomarker in PD

In the field of PD, a comprehensive review published in 2022 identified 143 studies employing CNN for classification purposes (Tanveer *et al* 2022). Only two used EEG as an input data source, a number that has since increased to ten (Shi *et al* 2019, Oh *et al* 2020, Chu *et al* 2021, Khare *et al* 2021, Lee *et al* 2021, Loh *et al* 2021, Qiu *et al* 2022, Shaban and Amara 2022, Li *et al* 2023, Nour *et al* 2023). Although when available, the information about the dataset construction indicated a sample-wise approach, they show that machine learning (ML) achieves high or very high performance, often exceeding accuracies of 98%. This figure should be compared to the 75%–95% of patients diagnosed with PD by human experts and whose diagnosis is confirmed at autopsy, according to the Movement Disorder Society (Postuma *et al* 2015). Although less represented, other ML methods than DL have been used to detect PD, with similar performances. Aljalal *et al* (2022) have attempted to identify PD from RS-EEG by extracting several features, including multiple entropy, energy and band power measures (Aljalal *et al* 2022). The authors compared ML classification methods that are not based on NNs with these features. They demonstrated that certain entropy measures, in combination with a suitable selection of EEG channels, can achieve performance comparable to NN methods while requiring less computational power.

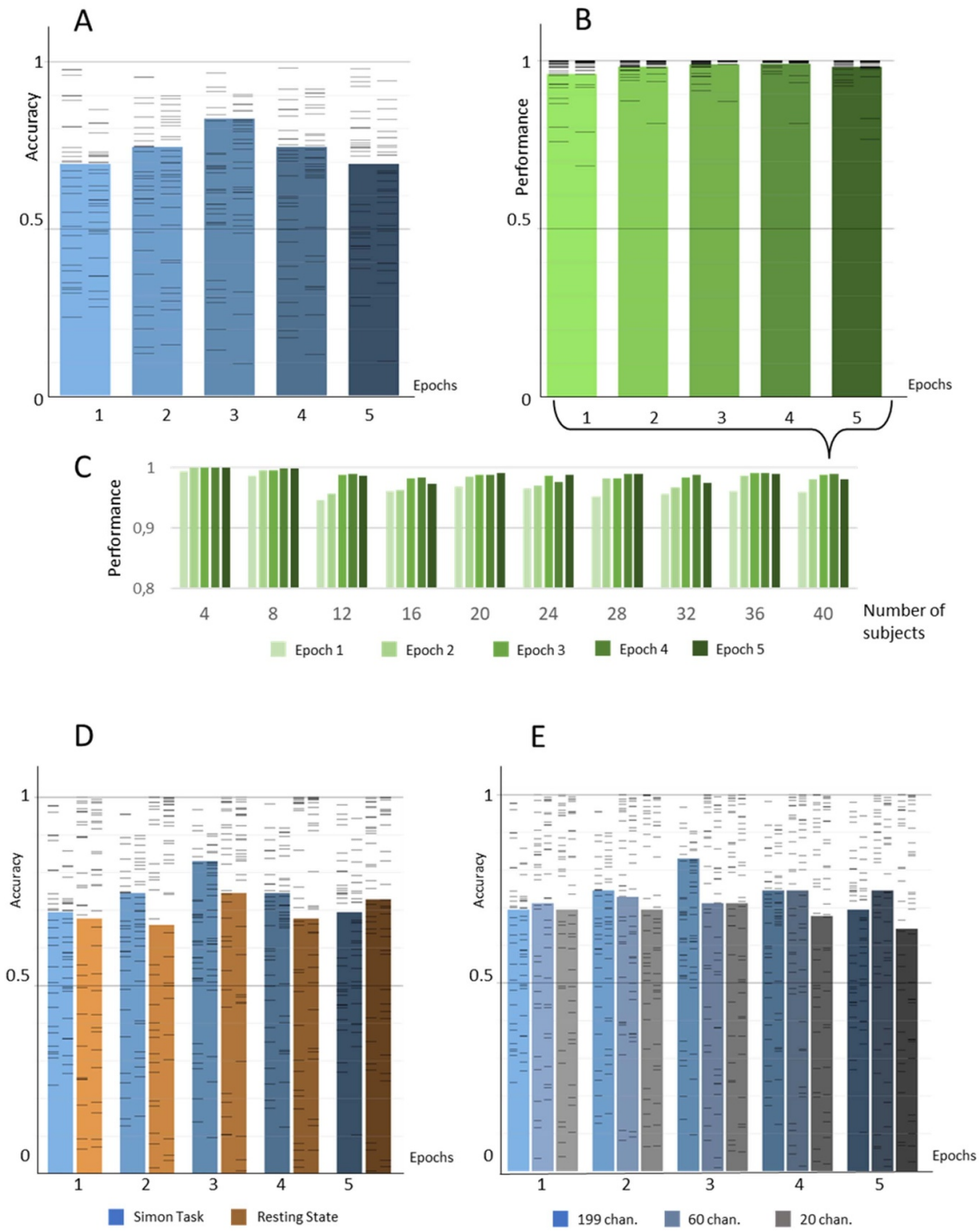

Figure 3. Results. (A): Accuracy in subject classification along epochs for 10-fold training on the subject-wise dataset with 199 channels. The bar chart shows the global classification accuracy (fraction of the subjects correctly classified). Horizontal dashes inside each bar represent the individual outputs for patients (left) and controls (right) averaged across samples. A dash above 0.5 is correctly classified. The best performance was obtained after 3 epochs, after which the model started to overfit. (B): Average performance in the sample classification along epochs on a sample-wise dataset for 10-fold training (bar chart). Horizontal dashes represent the individual subjects' average output as in (A). After 1 epoch, all subjects are correctly classified. The performances are much higher than those on the subject-wise dataset, possibly due to subject-specific information. (C): Average performance on sample-wise datasets with a progressive increase of the number of subjects. The performances were better with a smaller number of subjects, supporting the hypothesis of a bias induced by subject-specific information. (D): Accuracy in subject classification computed and displayed as in (A) compared along epochs between Simon task (ST) and resting state datasets. The accuracy was slightly better for ST. (E): Accuracy in subject classification compared along epochs between the 199 channel, 60 channel, and 20 channel subject-wise datasets. Although not proportional, the performances were better with more channels.

In addition to disease detection, other studies have investigated the prediction of cognitive impairment in PD patients based on EEG biomarkers. Singh *et al* (2023) studied cue-evoked delta and theta rhythms on the Cz electrode during cognitive control tasks including ST. They showed that those markers relate to cognition in PD (Singh *et al* 2023). Anjum *et al* have developed a specific marker based on Linear

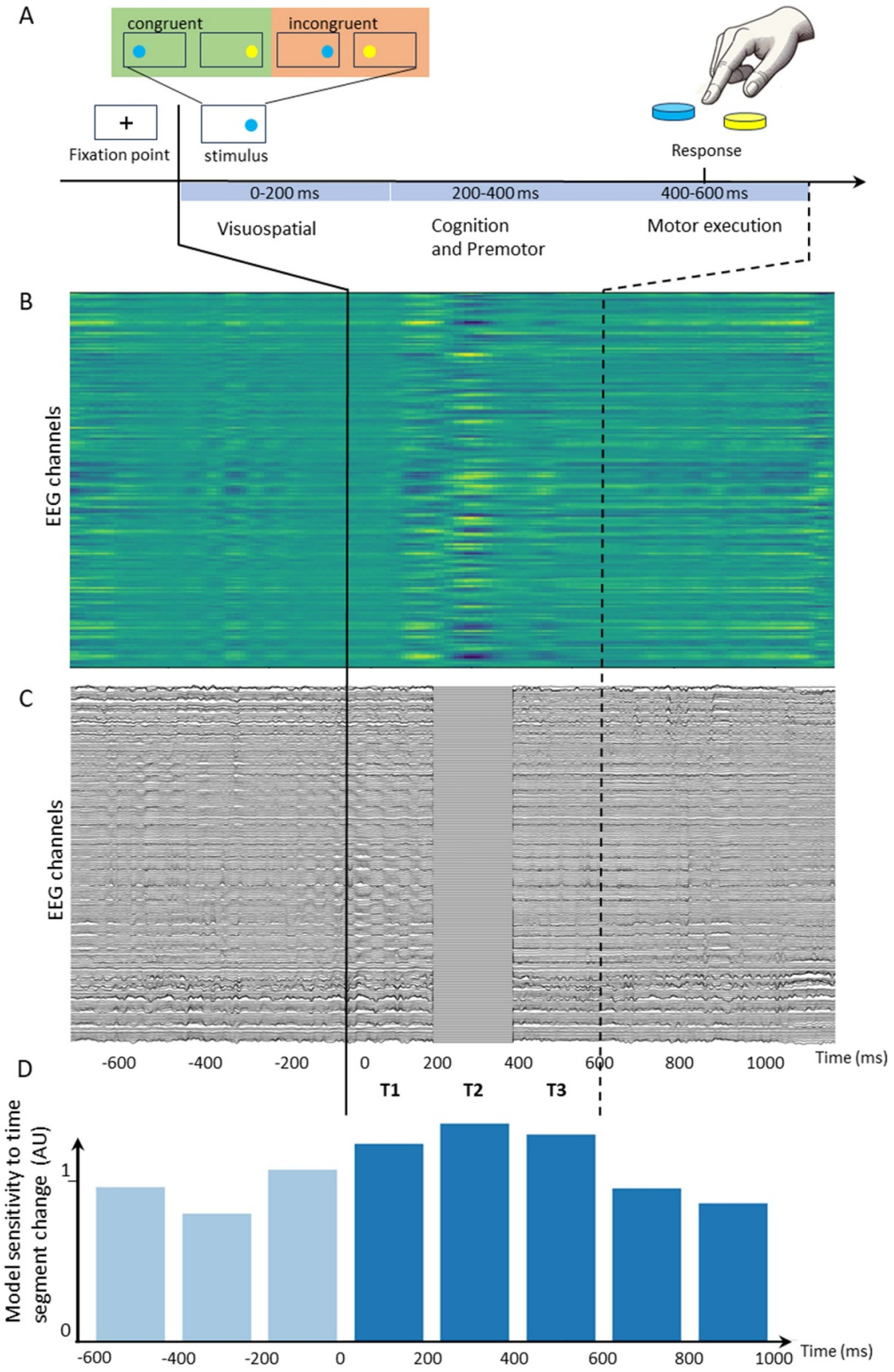

Figure 4. Simon task and sensitivity analysis in the time domain (A): Outline of the Simon task (ST): at time 0 the stimulus appears on one side of the visual field. The subject has to analyze the stimulus (T1, visuospatial task) and to decide whether it is congruent or not (T2, cognitive and premotor task) and then responds by pressing a button (T3, motor execution) (B): Integrated gradients (IG) computed and averaged across samples classified with an accuracy $> 80\%$ in the ST dataset. Time segments corresponding to the cognitive/premotor and motor execution lead to higher gradients, suggesting those segments are the most useful for the model. (**C**): Biased sample: the signal of all the electrodes is set to zero during a 200 ms time segment. 8 different biased samples are generated from every sample in the database, some of which corresponding to T1, T2 and T3. (D): the gap in the model's answer between the genuine samples and the biased ones is computed and averaged across all samples for each time segment. It is maximal during the premotor and motor components of the task, confirming the IG results with regard to the contribution of the EEG signal to the model's response in the electrode domain, the anterior regions, and to lesser extent the parietal regions, play a more important role in the ST condition than in the RS condition, as shown in figure 5.

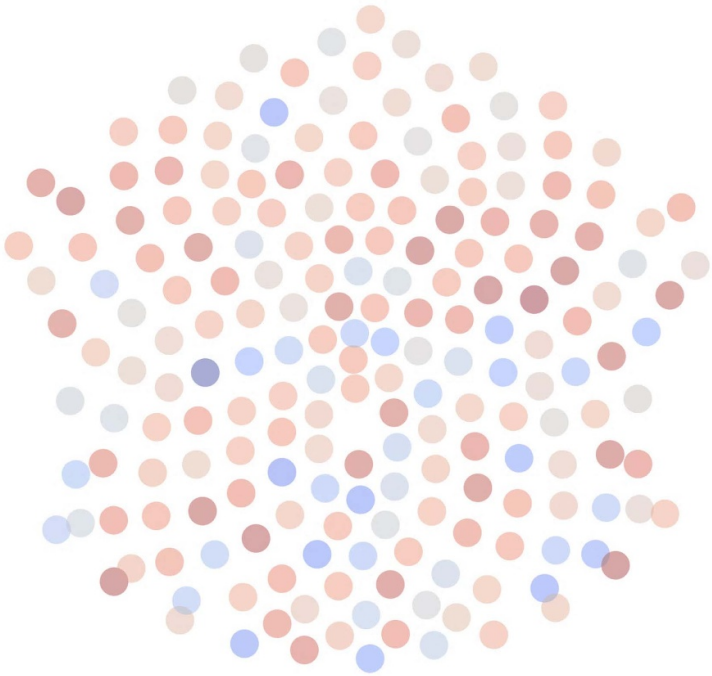

Figure 5. Comparison of integrated gradients (IG) between the Simon task (ST) and the resting state (RS) datasets in the electrode domain. IG are computed across samples classified with an accuracy $> 80\%$, then averaged by electrode along the time axis. The results in ST and RS dataset are subtracted and normalized. The electrodes corresponding to the EEG signals that the model uses predominantly to classify subjects from the ST dataset are indicated in red. Electrodes predominantly used to classify subjects from the RS dataset are shown in blue. EEG signal generated by frontal, and to a lesser extent parietal convexity cortical regions, involved in ST execution, are more discriminating in the ST dataset. (for clarity electrode names are specified in supplementary figure 1).

Predictive Coding on RS-EEG, which enables them to detect cognitive disorders in PD patients (Anjum *et al* 2020, 2024). We cannot directly compare our results with these studies, as our aim was to find a performance gain related to the ST using a subject-wise dataset. However, we have tested and combined in a new way different methods recently proven useful in this field, illustrating a new generation of biomarkers.

4.2. Preprocessing, network's architecture and hyperparameters

After preprocessing of EEG signals (Xie and Oniga 2020), feature extraction can be used to reduce the size of the data or increase its relevance. Four of the studies cited above that used DL on EEG to detect PD included feature extraction methods, three of which with 2D maps, including spectrograms. Some of them were based on established PD biomarkers such as PLV (Qiu *et al* 2022), a biomarker known to be modified in PD (de Hemptinne *et al* 2013).

Since CNN was primarily developed for 2D images, many EEG studies use 2D spectrograms as input (Craik *et al* 2019). However, such a transformation emphasizes a selected feature of the information that may not be the most pertinent for the model, unless one is defining dedicated features. For instance, Chu *et al* proposed identifying the optimal frequencies for detecting PD, which they found to be the high delta and low alpha bands. They then constructed a quantitative index of spectral characteristics incorporating spatial distribution (Chu *et al* 2021). Although we did not investigate this comparison in our study, it is conceivable that preprocessing techniques that significantly increase the input size, such as the 2D spectrogram, should be used with caution, as the desired enhancement of time–frequency information may be outweighed by the increase in model size, increasing the risk of overfitting. On the other hand, methods that reduce the size of the information may lose components that would have been relevant for classification. In the end, we found it preferable to use the raw data directly.

Most of the studies cited above have used CNN or CNN-derived architectures, sometimes in combination with recurrent networks. For example, Lee *et al* (2021) explored light and fast NNs with an innovative architecture that uses convolutional layers whose kernels are 'orthogonal' (i.e. the first convolution is in the time domain, the second in the electrode domain) and feed recurrent layers. The authors tested a range of hyperparameters and found that the best kernel number for the first convolutional layer was 35, which was also the highest tested. In addition, they experimented with reversing the time in the input, which did not significantly affect performance. This suggests that the use of recurrent layers may not be necessary for fixed-size samples. Despite different model architectures, all the above studies achieve similar

performance, with the exception of Oh *et al* (2020), which performed slightly worse. The NN in this study is characterized by large kernels in the convolutional layers, contrasting with a small number of kernels. This suggests that a small number of highly complex features may be less discriminatory than a larger number of simpler ones.

In our study, we tested several combinations of hyperparameters and architectural variations. Most of those did not improve performance compared to a simple CNN alternating pooling and convolutional layers feeding dense layers. Nevertheless, a two-scale network, with one branch capturing frequencies up to gamma and the other up to beta before merging, slightly improved performance, potentially due to the PAC between the beta and gamma frequency bands, altered in PD.

4.3. Accuracy, overfitting and 'disease-related' versus 'subject-related' features

When considering the speed of convergence of the model parameters, it is important to note that the training dataset contains considerable redundancy of subject-related information, as each subject is represented by an average of 546 samples. The learning curves reveal an early occurrence of overfitting when using subject-wise datasets. The accuracy on the validation set stabilized at 0.7–0.8 after only 2 or 3 epochs, depending on the spatial resolution and architecture of the model (figure 3(A)). However, when using sample-wise datasets, the accuracy on the validation set increased faster, higher and for more epochs, reaching above 0.98 after 5 epochs (figure 4(B)). Another global observation was that models with more than about 100 000 trainable parameters experienced more overfitting and loss of performance on subject-wise datasets, but not with sample-wise datasets.

We believe these observations are due to the greater weight of subject-specific information in the EEG data, due to the individual peculiarities of anatomy and recordings. The model focuses more on subject-specific features rather than disease-specific ones. To confirm this, an additional learning process with a sample-wise split and a number of subjects varying from 40 to 8, was conducted. As the number of subjects decreased, performance improved, supporting the idea that non-relevant subject-specific information (stronger with fewer subjects) contributes significantly to overfitting. This validates the use of a subject-wise splitting schema.

4.4. ST-EEG versus RS-EEG and sensitivity analysis

In HC, cortical activity is modified by visuospatial (Womelsdorf *et al* 2006, Muthukumaraswamy and Singh 2013) and cognitive tasks in large cortical networks and in frequency bands ranging from theta to gamma. Studies have shown that gamma rhythms in the lateral occipital cortex are specifically enhanced by stimulus-response (Simon) interference (Wiesman and Wilson 2020). Specifically, ST causes a decrease in the alpha band in posterior regions and the cuneus, and an increase in the theta band in frontal and parietal regions of the attention network (Son *et al* 2023). Functional MRI and EEG studies have shown that the midcingulate, prefrontal, and parietal cortical regions (Wittfoth *et al* 2006, Cohen and Ridderinkhof 2013, Wang and Weekes 2014) are involved in ST-related cortical activity. These regions differ from the DMN, which is active during rest and includes deep cortical areas, whose activity is more difficult to capture with surface EEG.

Regarding PD, The concept of leveraging disease-affected networks to enhance available information is well-established. While motor and cognitive functions are most studied, patients also have visual perceptual disorders. Vanegas *et al* found that ML methods can achieve high accuracy in distinguishing PD from HC based on EEG spectral data in posterior regions during visual processing (Vanegas *et al* 2018).

In our case, the model was also supposed to leverage the cortical networks activated by the conflict task, as demonstrated by the sensitivity analysis. The calculation of IG showed that EEG components corresponding to specific regions of the cerebral cortex and time segments related to stimulus processing, premotor, and motor sequences of the ST, played a more significant role in the ST dataset than in the RS dataset. We also found that setting to zero these segments led to higher changes in prediction accuracy.

This indicates that the model relies heavily on these components of the EEG, highlighting differences between PD and HC in cognitive and premotor activation. Moreover, maximum accuracy was slightly superior with ST-EEG than with RS-EEG, supporting our hypothesis regarding the gain achieved by activating cortical networks via the ST.

The lack of explainability is a constant drawback of deep NN. To get more clues about the patterns used by the model, especially in terms of frequency bands, we could use another type of sensitivity analysis, e.g. using band-stop filters. However, we considered that the size of our database and the overall performance of the models on the whole signal were not sufficient for this type of analysis. A number of algorithms have been developed by teams working on DL to enhance the explicability of models. These algorithms are based on general-purpose (Islam *et al* 2022) or more specialized (Khare and Acharya 2023) models. This is exemplified by the adaptive gated graph convolutional network proposed by Klepl *et al* which exhibits a level of explicability that surpasses that of a standard CNN (2023).

4.5. The role of spatial resolution

Increasing the number of electrodes improved accuracy. Higher spatial resolution increases the amount of information conveyed by the EEG samples in source

reconstruction (Lantz *et al* 2003, Vorderwülbecke *et al* 2020) or when using ICA (Janani *et al* 2018). Fast frequencies, being more localized and less powerful than slow rhythms, may fall between the electrodes, explaining the superiority of HD-EEG in studying conditions like PD that affect beta and gamma rhythms. However, HD-EEG is susceptible to errors in electrode positioning. To mitigate this, a network architecture could use a rectangular 2D grid mapped to the scalp topology to minimize deformation, with an initial 2D convolution layer in the spatial domain feeding into 1D convolution layers in the temporal domain.

4.6. The role of time resolution

The initial pooling rate that gave optimal results for our data in the time domain was 1/12 for 1000 Hz data, corresponding to a high cutoff frequency in the low gamma band. Of note, using a small or no initial pooling rate resulted in higher accuracy for sample-wise datasets, while subject-wise datasets exhibited lower accuracy. This may be because information contained in the high frequencies (>40 Hz) tends to be highly specific to individual recordings, possibly due to the presence of power line or muscle artifacts in some channels. These artifacts constitute part of the subject-specific information that dilutes the disease-specific information, as discussed in 4.6 section. Alternatively, increasing the sampling rate in the first convolution layer requires more parameters, raising the risk of overfitting due to higher variance error.

4.7. Results on another database

The results were not satisfactory as there were biases, in particular evidence of misclassification of several control subjects as PD patients. The origin of this bias remains unclear. However, automatic noise suppression techniques, may have contributed. We were unable to find an appropriate compromise for the threshold defined in EEGlab for rejecting independent components based on their identification as predominantly muscle, eye movement, or DC artifacts. Although additional manual processing could have improved accuracy, our aim was to evaluate the automation of the entire process on any EEG datasets after training the model with our data.

5. Limitations of this study

This study has several limitations. First, the performance of the model on subject-wise datasets is significantly lower than that reported in previous studies, although it is comparable for sample-wise datasets. While part of this discrepancy may be due to differences in the making of the datasets, additional factors related to processing techniques, may also contribute.

Our database is limited in size, which renders it susceptible to sampling bias. Consequently, a decision was taken against conducting a subgroup analysis, which would have examined factors such as time or accuracy of response to the task, congruence or incongruence of the stimulus, lateralization of patients, or other criteria. This would have opened up interesting perspectives, but ML models require large datasets, and we considered that splitting the data to perform finer analyses would have made the paradigm too vulnerable to sampling bias.

It is obvious that the ST paradigm is more complex to implement and may be more difficult to reproduce than an EEG SR. This complexity could potentially offset any enhancements in classification performance, although it can be argued that a simple paradigm in which the subject is instructed to optimize his performance, could activate NNs in a more reproducible way than a RS condition. Finally, the fusion of model responses in RS and ST conditions has not been studied, as the aim was a comparison between the RS and a cognitive task.

In addition to NN, several other ML techniques have been used for EEG classification in PD. Although our main focus is not an exhaustive review of these techniques, the comparative analysis between NN and other ML techniques remains valuable, as shown by Aljalal *et al* (2022). DL techniques are time-consuming and computationally intensive, which is questionable at a time when energy savings are crucial. It is recommended to use less expensive statistical learning tools instead of NN when their performances are comparable. In addition, downsampling of the data can be considered if it does not degrade performance, as is the case in the time domain.

The opacity of the model is inherent to NN, although more modern and sophisticated algorithms than a CNN can provide clues to the features exploited by the model, in terms of frequency (Khare and Acharya 2023) or connectivity (Klepl *et al* 2023). Sensitivity analysis is an indirect technique that may be susceptible to biases, as it only gives an indication of the information that the model has found to be most relevant.

6. Conclusion

EEG is a potential biomarker for PD in DL models. Our results indicate that EEG contains more information during the activation of cortical networks known to be impaired by the disease, compared to RS-EEG. Furthermore, the sensitivity analysis suggests that the model predominantly uses EEG components that are specifically involved in the task, in both the time domain and the spatial domain. Spatial resolution is a crucial factor, as more electrodes can improve the results, although the gain is not proportional. Conversely, temporal resolution has limitations, typically at a sampling frequency of 100 Hz, beyond which increased precision is unnecessary and could lead to overfitting.

Prudent construction of training and validation datasets is imperative, particularly when they are derived from subsets that may share the same label and non-relevant but powerful information, typically one EEG recording per subject split into hundreds of samples. Disease-related information can be weaker than subject-related information, potentially predisposing sample-wise datasets to yield optimistic performances. Overfitting can also quickly occur with such data. The limited number of subjects in existing studies may explain why smaller models perform as well or even better than larger ones. Our results suggest that simple NN architectures perform fairly well in classification on these databases, although more sophisticated algorithms could improve explainability. Increasing the number of subjects may dilute subject-related information in favor of disease-related information, making the data more exploitable by larger models that could extract more subtle features. However, the accessibility of databases with large patient cohorts while maintaining homogeneity of acquisition devices and parameters remains a challenge.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Funding

This study was not supported by any sponsor or funder.

References

Abadi M *et al* 2016 {TensorFlow}: a system for {Large-Scale} machine learning *12th USENIX Symp. on Operating Systems Design and Implementation (OSDI 16)* pp 265–83

Aljalal M, Aldosari S A, AlSharabi K, Abdurraqeeb A M and Alturki F A 2022 Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques *Diagnostics* **12** 1033

Anjum M F, Dasgupta S, Mudumbai R, Singh A, Cavanagh J F and Narayanan N S 2020 Linear predictive coding distinguishes spectral EEG features of Parkinson's disease *Parkinsonism Relat. Disorders* **79** 79–85

Anjum M F, Espinoza A I, Cole R C, Singh A, May P, Uc E Y, Dasgupta S and Narayanan N S 2024 Resting-state EEG measures cognitive impairment in Parkinson's disease *npj Parkinsons Dis.* **10** 6

Babiloni C *et al* 2016 Alpha, beta and gamma electrocorticographic rhythms in somatosensory, motor, premotor and prefrontal cortical areas differ in movement execution and observation in humans *Clin. Neurophysiol.* **127** 641–54

Brown P and Marsden C D 1999 Bradykinesia and impairment of EEG desynchronization in Parkinson's disease *Mov. Disorders* **14** 423–9

Cavanagh J F, Ryman S and Richardson S P 2022 Cognitive control in Parkinson's disease *Prog. Brain Res.* **269** 137–52

Chu C, Zhang Z, Wang J, Liu S, Wang F, Sun Y, Han X, Li Z, Zhu X and Liu C 2021 Deep learning reveals personalized spatial spectral abnormalities of high delta and low alpha bands in EEG of patients with early Parkinson's disease *J. Neural Eng.* **18** 066036

Cohen M X and Ridderinkhof K R 2013 EEG source reconstruction reveals frontal-parietal dynamics of spatial conflict processing *PLoS One* **8** e57293

Craik A, He Y and Contreras-Vidal J L 2019 Deep learning for electroencephalogram (EEG) classification tasks: a review *J. Neural Eng.* **16** 031001

Das A, de Los Angeles C and Menon V 2022 Electrophysiological foundations of the human default-mode network revealed by intracranial-EEG recordings during resting-state and cognition *Neuroimage* **250** 118927

de Hemptinne C, Ryapolova-Webb E S, Air E L, Garcia P A, Miller K J, Ojemann J G, Ostrem J L, Galifianakis N B and Starr P A 2013 Exaggerated phase-amplitude coupling in the primary motor cortex in Parkinson disease *Proc. Natl Acad. Sci. USA* **110** 4780–5

de Hemptinne C, Swann N C, Ostrem J L, Ryapolova-Webb E S, San Luciano M, Galifianakis N B and Starr P A 2015 Therapeutic deep brain stimulation reduces cortical phase-amplitude coupling in Parkinson's disease *Nat. Neurosci.* **18** 779–86

Defebvre L, Derambure P, Bourriez J L, Destée A and Guieu J D 1999 Event-related desynchronization and Parkinson disease. Importance in the analysis of the phase of preparation for movement *Neurophysiol. Clin.* **29** 71–89

Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21

Devos D 2004 Subthalamic nucleus stimulation modulates motor cortex oscillatory activity in Parkinson's disease *Brain* **127** 408–19

Dirnberger G and Jahanshahi M 2013 Executive dysfunction in Parkinson's disease: a review *J. Neuropsychol.* **7** 193–224

Dujardin K, Moonen A J H, Behal H, Defebvre L, Duhamel A, Duits A A, Plomhause L, Tard C and Leentjens A F G 2015 Cognitive disorders in Parkinson's disease: confirmation of a spectrum of severity *Parkinsonism Relat. Disorders* **21** 1299–305

Duprez J *et al* 2022 Spatio-temporal dynamics of large-scale electrophysiological networks during cognitive action control in healthy controls and Parkinson's disease patients *Neuroimage* **258** 119331

Duprez J, Houvenaghel J-F, Argaud S, Naudet F, Robert G, Drapier D, Vérin M and Sauleau P 2017 Impulsive oculomotor action selection in Parkinson's disease *Neuropsychologia* **95** 250–8

Fereshtehnejad S-M, Zeighami Y, Dagher A and Postuma R B 2017 Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression *Brain* **140** 1959–76

Galvan A and Wichmann T 2008 Pathophysiology of parkinsonism *Clin. Neurophysiol.* **119** 1459–74

George J S, Strunk J, Mak-mccully R, Houser M, Poizner H and Aron A R 2013 Dopaminergic therapy in Parkinson's disease decreases cortical beta band coherence in the resting state and increases cortical beta band power during executive control *Neuroimage Clin.* **3** 261–70

Geraedts V J, Koch M, Contarino M F, Middelkoop H A M, Wang H, van Hilten J J, Bäck T H W and Tannemaat M R 2021 Machine learning for automated EEG-based biomarkers of cognitive impairment during deep brain stimulation screening in patients with Parkinson's disease *Clin. Neurophysiol.* **132** 1041–8

Hassan M, Benquet P, Biraben A, Berrou C, Dufor O and Wendling F 2015 Dynamic reorganization of functional brain networks during picture naming *Cortex* **73** 276–88

Heida T, Poppe N R, de Vos C C, van Putten M J A M and van Vugt J P P 2014 Event-related mu-rhythm desynchronization during movement observation is impaired in Parkinson's disease *Clin. Neurophysiol.* **125** 1819–25

Islam M S, Hussain I, Rahman M M, Park S J and Hossain M A 2022 Explainable artificial intelligence model for stroke prediction using EEG signal *Sensors* **22** 9859

Janani A S, Grummett T S, Bakhshayesh H, Lewis T W, Willoughby J O and Pope K J 2018 How many channels are enough? evaluation of tonic cranial muscle artefact reduction using ICA with different numbers of EEG channels *2018 26th European Signal Processing Conf. (EUSIPCO)* pp 101–5

Kawai Y, Tachikawa K, Park J and Asada M 2022 Compensated integrated gradients for reliable explanation of electroencephalogram signal classification *Brain Sci.* **12** 849

Khare S K and Acharya U R 2023 Adazd-Net: automated adaptive and explainable Alzheimer's disease detection system using EEG signals *Knowl.-Based Syst.* **278** 110858

Khare S K, Bajaj V and Acharya U R 2021 PDCNNet: an automatic framework for the detection of Parkinson's disease using EEG signals *IEEE Sens. J.* **21** 17017–24

Klepl D, He F, Wu M, Blackburn D J and Sarrigiannis P 2023 Adaptive gated graph convolutional network for explainable diagnosis of Alzheimer's disease using EEG data *IEEE Trans. Neural Syst. Rehabil. Eng.* **31** 3978–87

Lantz G, Grave de Peralta R, Spinelli L, Seeck M and Michel C M 2003 Epileptic source localization with high density EEG: how many electrodes are needed? *Clin. Neurophysiol.* **114** 63–69

Lee S, Hussein R, Ward R, Jane Wang Z and McKeown M J 2021 A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease *J. Neurosci. Methods* **361** 109282

Li K, Ao B, Wu X, Wen Q, Ul Haq E and Yin J 2023 Parkinson's disease detection and classification using EEG based on deep CNN-LSTM model *Biotechnol. Genet. Eng. Rev.* **40** 2577–96

Loh H W, Ooi C P, Palmer E, Barua P D, Dogan S, Tuncer T, Baygin M and Acharya U R 2021 GaborPDNet: Gabor transformation and deep neural network for Parkinson's disease detection using EEG signals *Electronics* **10** 1740

Muthukumaraswamy S D and Singh K D 2013 Visual gamma oscillations: the effects of stimulus type, visual field coverage and stimulus motion on MEG and EEG recordings *Neuroimage* **69** 223–30

Nour M, Senturk U and Polat K 2023 Diagnosis and classification of Parkinson's disease using ensemble learning and 1D-PDCovNN *Comput. Biol. Med.* **161** 107031

Nwogo R O, Kammermeier S and Singh A 2022 Abnormal neural oscillations during gait and dual-task in Parkinson's disease *Front. Syst. Neurosci.* **16** 995375

Oh S L, Hagiwara Y, Raghavendra U, Yuvaraj R, Arunkumar N, Murugappan M and Acharya U R 2020 A deep learning approach for Parkinson's disease diagnosis from EEG signals *Neural Comput. Appl.* **32** 10927–33

Ott T, Stein A M and Nieder A 2023 Dopamine receptor activation regulates reward expectancy signals during cognitive control in primate prefrontal neurons *Nat. Commun.* **14** 7537

Pfurtscheller G and Lopes da Silva F H 1999 Event-related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842–57

Postuma R B *et al* 2015 MDS clinical diagnostic criteria for Parkinson's disease *Mov. Disorders* **30** 1591–601

Qiu L, Li J and Pan J 2022 Parkinson's disease detection based on multi-pattern analysis and multi-scale convolutional neural networks *Front. Neurosci.* **16** 957181

Rosenblum Y, Shiner T, Bregman N, Fahoum F, Giladi N, Maidan I and Mirelman A 2022 Event-related oscillations differentiate between cognitive, motor and visual impairments *J. Neurol.* **269** 3529–40

Schmiedt-Fehr C, Schwendemann G, Herrmann M and Basar-Eroglu C 2007 Parkinson's disease and age-related alterations in brain oscillations during a Simon task *Neuroreport* **18** 277–81

Shaban M and Amara A W 2022 Resting-state electroencephalography based deep-learning for the detection of Parkinson's disease *PLoS One* **17** e0263159

Shi X, Wang T, Wang L, Liu H and Yan N 2019 Hybrid convolutional recurrent neural networks outperform CNN and RNN in task-state EEG detection for Parkinson's disease *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC)* (IEEE) pp 939–44

Simon J R and Rudell A P 1967 Auditory S-R compatibility: the effect of an irrelevant cue on information processing *J. Appl. Psychol.* **51** 300–4

Singh A, Cole R C, Espinoza A I, Wessel J R, Cavanagh J F and Narayanan N S 2023 Evoked mid-frontal activity predicts cognitive dysfunction in Parkinson's disease *J. Neurol. Neurosurg. Psychiatry* **94** 945–53

Singh A, Richardson S P, Narayanan N and Cavanagh J F 2018 Mid-frontal theta activity is diminished during cognitive control in Parkinson's disease *Neuropsychologia* **117** 113–22

Son J J, Arif Y, Schantell M, Willett M P, Johnson H J, Okelberry H J, Embury C M and Wilson T W 2023 Oscillatory dynamics serving visual selective attention during a Simon task *Brain Commun.* **5** fcad131

Stuart S, Wagner J, Makeig S and Mancini M 2021 Brain activity response to visual cues for gait impairment in Parkinson's disease: an EEG study *Neurorehabilit. Neural Repair* **35** 996–1009

Sundararajan M, Taly A and Yan Q 2017 Axiomatic attribution for deep networks *Proc. 34th Int. Conf. on Machine Learning (PMLR)* pp 3319–28

Swann N C, de Hemptinne C, Aron A R, Ostrem J L, Knight R T and Starr P A 2015 Elevated synchrony in Parkinson disease detected with electroencephalography: elevated synchrony in PD *Ann. Neurol* **78** 742–50

Tanveer M, Rashid A H, Kumar R and Balasubramanian R 2022 Parkinson's disease diagnosis using neural networks: survey and comprehensive evaluation *Inf. Process. Manag.* **59** 102909

Tomlinson C L, Stowe R, Patel S, Rick C, Gray R and Clarke C E 2010 Systematic review of levodopa dose equivalency reporting in Parkinson's disease *Mov. Disorders* **25** 2649–53

Vanegas M I, Ghilardi M F, Kelly S P and Blangero A 2018. Machine learning for EEG-based biomarkers in Parkinson's disease *2018 IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)* (IEEE) pp 2661–5

Vorderwülbecke B J, Carboni M, Tourbier S, Brunet D, Seeber M, Spinelli L, Seeck M and Vulliemoz S 2020 High-density electric source imaging of interictal epileptic discharges: how many electrodes and which time point? *Clin. Neurophysiol.* **131** 2795–803

Wang L and Weekes B 2014 Neural correlates of the simon effect modulated by practice with spatial mapping *Neuropsychologia* **63** 72–84

Waninger S, Berka C, Stevanovic Karic M, Korszen S, Mozley P D, Henchcliffe C, Kang Y, Hesterman J, Mangoubi T and Verma A 2020 Neurophysiological biomarkers of Parkinson's disease *J. Parkinsons. Dis.* **10** 471–80

Wiesman A I and Wilson T W 2020 Posterior alpha and gamma oscillations index divergent and superadditive effects of cognitive interference *Cereb. Cortex* **30** 1931–45

Wittfoth M, Buck D, Fahle M and Herrmann M 2006 Comparison of two simon tasks: neuronal correlates of conflict resolution based on coherent motion perception *NeuroImage* **32** 921–9

Womelsdorf T, Fries P, Mitra P P and Desimone R 2006 Gamma-band synchronization in visual cortex predicts speed of change detection *Nature* **439** 733–6
Wylie S A, Ridderinkhof K R, Bashore T R and van den Wildenberg W P M 2010 The effect of Parkinson's disease on the dynamics of on-line and proactive cognitive control during action selection *J. Cogn. Neurosci.* **22** 2058–73
Xie Y and Oniga S 2020 A review of processing methods and classification algorithm for EEG signal *Carpath. J. Electr. Comput. Eng.* **13** 23–29