Deep Learning-Driven EEG Analysis for Personalized Deep Brain Stimulation Programming in Parkinson's Disease

Nicolas Calvo Peiro[a,b,c], Mathias Ramm Haugland[a], Alena Kutuzova[a,d], Cosima Graef[a,b,c], Aminata Bocum [e], Yen Foung Tai[a,d,e], Anastasia Borovykh[f], Shlomi Haar[a,d,*]

[a] Care Research and Technology Centre, UK Dementia Research Institute, Imperial College London, London, United Kingdom
[b] Department of Computing, Imperial College London, London, United Kingdom
[c] UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London, UK
[d] Department of Brain Sciences, Imperial College London, London, United Kingdom
[e] Department of Neurology, Charing Cross Hospital, London, United Kingdom
[f] Department of Mathematics, Imperial College London, London, United Kingdom

Correspondence to Shlomi Haar (s.haar@imperial.ac.uk)

Abstract

Deep Brain Stimulation (DBS) is an invasive procedure used to alleviate motor symptoms in Parkinson's Disease (PD) patients. While brain activity can be used to optimise DBS parameters, the impact of DBS parameters on brain activity remains unclear. We aimed to identify the cortical neural response to changes in DBS parameters, which are sensitive to the effect of small changes in the stimulation parameters and could be used as neural biomarkers. We recorded in-clinic EEG data from seven hemispheres of PD patients during DBS programming sessions.

Here we developed a siamese adaptation of the EEGNet deep learning architecture and trained it to distinguish whether two short (1-sec-long) segments of brain activity were taken with the same stimulation parameters, or if either the strength or location of the stimulation had changed. 13 independent models were trained independently in each hemisphere for stimulation amplitude or contact, and all achieved high accuracy with an average of 78%. Our models are sensitive to changes in brain activity recorded at the scalp of the patients following changes as small as 0.3mA in the DBS parameters.

Next, we interpreted what our black-box AI models learned with an ablation-based explainability method, that extracts frequency bands learned by the models through a perturbation of the input’s frequency spectrum. We found that fast Narrow-Band Gamma oscillations (60-90Hz), contributed most to the models across all 7 hemispheres.

This work, using a data-driven approach, joins a recent body of evidence suggesting cortical Narrow-Band Gamma activity as the potential range for digital biomarkers for DBS optimization.

Keywords:

Deep Brain Stimulation; EEG; Explainable AI; Narrow-band gamma; Parkinson's Disease

Introduction

Deep Brain Stimulation (DBS) is an effective treatment for alleviating motor symptoms in Parkinson's Disease (PD)[1] involving the surgical insertion of electrodes into deep brain nuclei, typically the subthalamic nucleus (STN), and applying a fixed electrical current. However, programming stimulation settings relies on a sub-optimal trial-and-error approach based on in-clinic assessment by the clinicians and subjective reporting by the patients[2]. The high-dimensional stimulation parameter space (amplitude, contact point, frequency, and pulse width) is too vast for clinicians to fully explore, even in extended adjustment sessions[2]. These parameters remain unchanged until the next appointment, lacking accommodation to the brain's dynamic behaviour or the patient's varying needs throughout the day[3]. Hence, there is a need for adaptive closed-loop DBS (aDBS) systems that can read the state of the patient and adapt it accordingly[4], potentially improving long-term adverse effects from chronic stimulation[5]. In-clinic aDBS systems have shown promise[6–9] and a recent in-home clinical trial reported promising preliminary results[10]. However, the field is limited by the quality of the biomarkers used to track the brain state and to adapt the stimulation accordingly.

The neural response to DBS has been traditionally characterised via externalised recording used to localise the target during the surgery[11], and shortly after it[6,12–14]. Recently, Medtronic Percept™ DBS devices allow brain activity recordings post-recovery, but those are limited by the low sampling rate and potential artefact[15,16]. The pathological elevated activity in the beta band (13-30Hz) is the most studied marker in the context of PD and aDBS[14,17,18], primarily through Local Field Potential (LFP) recordings at the sub-cortical STN. Notably, the beta amplitude from STN LFP recordings is used as a control signal to ramp stimulation up and down in an ongoing large-scale adaptive DBS clinical trial[19]. Specific sub-bands of the beta band have shown modulation with the medication and/or DBS[14,20,21], and the bursting properties of beta activity have also been explore [17,22,23]. These works focused on the effect of DBS in an ON/OFF setting rather than the effect of DBS parameters. Cortical beta activity in PD has also been explored[18,24–27], and while findings were less consistent, it was shown to decrease during clinically effective DBS stimulation[27]. Overall, the beta activity peak and its functional correlates only appear in some patients, and the effect of DBS settings on beta activity remains unclear[4,18]. Additionally, while elevated beta oscillations may be linked to symptoms, recent studies show that this is only a modest correlation, and the pathophysiology of PD extends beyond resting-state beta oscillations[28].

Recent studies have identified Narrow-Band Gamma (NBG) activity (60-90Hz) in both STN and cortical brain activity as a new potential biomarker for DBS[10,27,29–34]. NBG has shown functional relevance relating to dyskinesia and was proposed as a control signal to reduce excessive DBS in certain patient groups[27,29,31,32]. Additionally, increased brain activity at half the stimulation frequency was observed and a 1:2 entrainment theory has been proposed to explain the induced NBG activity[27,31–34]. It has been recently shown to be amplitude and frequency-dependent as well as to have a prokinetic effect[32]. The 1:2 entrainment has also been linked to functional improvements in symptom alleviation and finger-tapping performance[10,32]. This signal has shown real-world value by being implemented in an in-home clinical trial, showing the benefits of NBG-based aDBS[10].

Despite these advancements, only a few studies have explored the effect of small parameter changes, which clinicians perform during programming sessions, on brain activity. While

existing research on biomarkers has provided insightful knowledge, it was mostly conducted in settings with clinically ineffective stimulation or non-translational paradigms, leading to a lack of robustness and real-world applicability[35–37]. To address this gap, we developed a methodology using a Siamese Neural Network (SNN) based on the EEGNet architecture[38] to discriminate brain-wide neural responses to stimulation parameter changes in a clinically relevant PD dataset. We trained the SNN to classify whether pairs of 1-second EEG segments had the same stimulation parameters or not and used perturbation-based explainability methods to uncover the features learned. We used a dataset recorded during unmodified clinical programming sessions ensuring real-world value, and advancing the field towards biomarker-guided DBS programming sessions and aDBS.

Methods

Participants

For this study, we used a dataset collected during clinical programming sessions of PD patients while the clinician tuned the DBS parameters of each patient as per established routine clinical practice, assessing the effect on motor symptoms of each parameter configuration. Patient demographics can be found in Table 1 and the range of parameters for each patient is in Supplementary Table 1. All patients gave written informed consent before the study. The study was approved by the local ethics committee (London - Harrow Research Ethics Committee) following the standards set by the Declaration of Helsinki.

	Age	Gender	Time from diagnosis (years)	Time from DBS (years)	LEDD (mg)	UPDRS Pre-programming	Device model
P01	60-65	M	18	7	375	44	Boston Vercise RC
P02	60-65	M	21	8	1325	82	Boston Vercise RC
P03	35-40	M	8	0	566	38	Boston Vercise Genus R32
P04	55-60	M	19	10	1240	26	Boston Vercise
Average	57 ± 13	N/A	17 ± 6	6 ± 4	877 ± 413	39 ± 33	N/A

Table 1. Patient characteristics. *Unified Parkinson's Disease Rating Scale (UPDRS) scores were evaluated before the start of the session. All patients apart from P03 were recorded ON medication and while DBS was ON but not optimised. P03 was recorded during his monopolar review, indicated by the time from DBS being 0, and their UPDRS was recorded OFF medication and OFF DBS. LEDD = Levodopa Equivalent Daily Dose*

Dataset

During clinical DBS programming sessions, we recorded brain activity using DSI-7 (Wearable Sensing), a wireless EEG headset with dry electrodes recording 6 channels (F3, C3, P3, F4, C4 and P4 of the 10-20 international system, with Pz as reference and Fz as ground) at 300Hz. We included 4 patients who had at least 2 different contact points and 5 or more amplitude levels in at least one hemisphere, to ensure a meaningful amount of training data. One of the patients (P04) had changes in the parameters only on the left electrode, hence, we analysed a total of 7 hemispheres. We analysed both arm and hand movements (finger tapping, hand open and close, and hand pronation-supination) and rest periods as those were most consistent across parameter sets and participants.

Post-hoc imaging analysis revealed that our patients had electrodes successfully implanted to the intended STN (Figure 1a), though for some patients, the active contacts stimulated the STN while being outside of it, and for P02, the active contacts were closer to the Globus Pallidus internus (GPi) (Figure 1b). Our cohort had a mix of monopolar and bipolar stimulation settings, with some patients having a mix of both. Only P03 was recorded during their monopolar review.

Each recording session was annotated with the stimulation parameters, which included amplitude, contact, frequency, and pulse width for the left and right electrodes. Using these annotations, we created *pairs* of 1s segments of EEG data and assigned them the label of "same" or "different". The label "same" indicates that between the two segments, there were no changes in stimulation, and "different" indicates that there was a change in one parameter between the two segments.

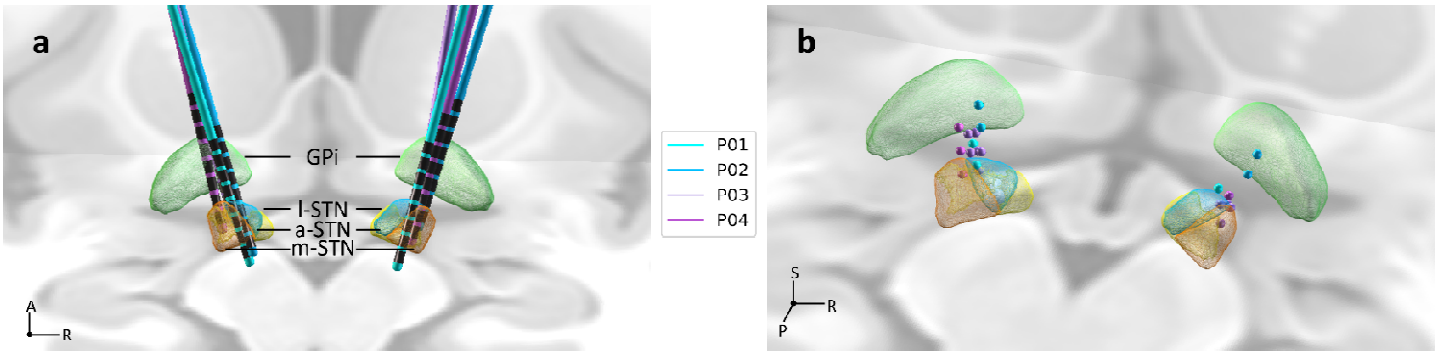


Figure 1. Electrode placement*. Visualisation of the DBS electrodes through co-registering of MRI and CT scans by means of advanced normalisation tools with subcortical refine*[39,40]*. After PaCER*[41] *algorithm-based localisation, electrode position was manually tuned for a precise estimate. Patients are colour-coded according to the legend.* ***a:*** *Localised DBS electrodes,* GPi*: Globus Pallidus Internus,* l/a/m-STN: *limbic/associative/motor subthalamic nucleus;* ***b****: Active contacts of the input segments to the model. When considering the volume of tissue activated, all patients except for P02 had STN stimulation.*

Preprocessing

EEG data was filtered at 50Hz using an Infinite Impulse Response (IIR) notch filter to remove electrical interference, bandpass filtered from 4Hz to 91Hz using a 5th-order Butterworth filter and scaled to have a mean of 1. Frames without an annotated task were discarded. To generate pairs of “same” or “different” data, an algorithm using a custom Python script was implemented. The conditions for two 1s segments to form a pair were: for all stimulation parameters to be equal except the one of interest which could vary, for all the stimulation parameters to stay constant within a segment, for a minimum time distance of 2 seconds and a maximum time distance of 15 minutes between the segments, and for the patient to be performing the same task in both segments (rest or arm movement). If two segments of a pair had the same stimulation parameters, the pair was given a label of “same”, and if the parameter of interest was changed, the pair was given the label of “different”. Additionally, pairs of different amplitude had an amplitude difference of at least 0.3mA, and in pairs of a different contact, we allowed for the amplitude to change. Since the effect of the same amplitude while stimulating with different contacts could vary significantly, so there was no justification to constrain it while comparing the same versus different contact, as it should be independent of amplitude). This was validated in the Results. A pool of segments was generated with an overlap of $o_1 = 50\%$, and possible pairs according to the previously mentioned conditions were extracted, with a maximum allowed overlap of $o_2 = 30\%$ within the pool of 2nd segments. No overlap was allowed between two segments of a pair (enforced by the minimum distance of 2 seconds). Despite the redundancy that this generated in the pairs, it increased our sample size; this can be seen as a form of data augmentation that allowed us to successfully train deep neural networks. Finally, the dataset was balanced between the classes (“same” or “different”) by removing pairs from the dominant class. Additionally, the task proportions between the two classes were balanced, to ensure the classification of the neural response, rather than the task. For every patient, we analysed differences in the amplitude and contact point of the stimulation, as those were most commonly the parameters changed by the clinicians. If a patient only had changes in some parameters, the patient was kept for analysis and only the parameters that had changes were analysed. In our case, the right amplitude in P04 and the right contact in P01 and P04 were not changed, and could therefore not be analysed. A characterisation of our patient’s distribution of monopolar/bipolar contact points can be found in Supplementary Tables 3 and 4.

The data was split into a train, validation, and test dataset, with varying split across patients – as per data availability. There was no overlap between the test dataset and the train/validation datasets, and an overlap of $o_3 = 30\%$, was allowed between the training and validation datasets. To overcome limitations from the length of the recorded data, we ran a k-fold cross-validation analysis, with the number of folds varying per patient, such that each patient had a total of 2000 test samples.

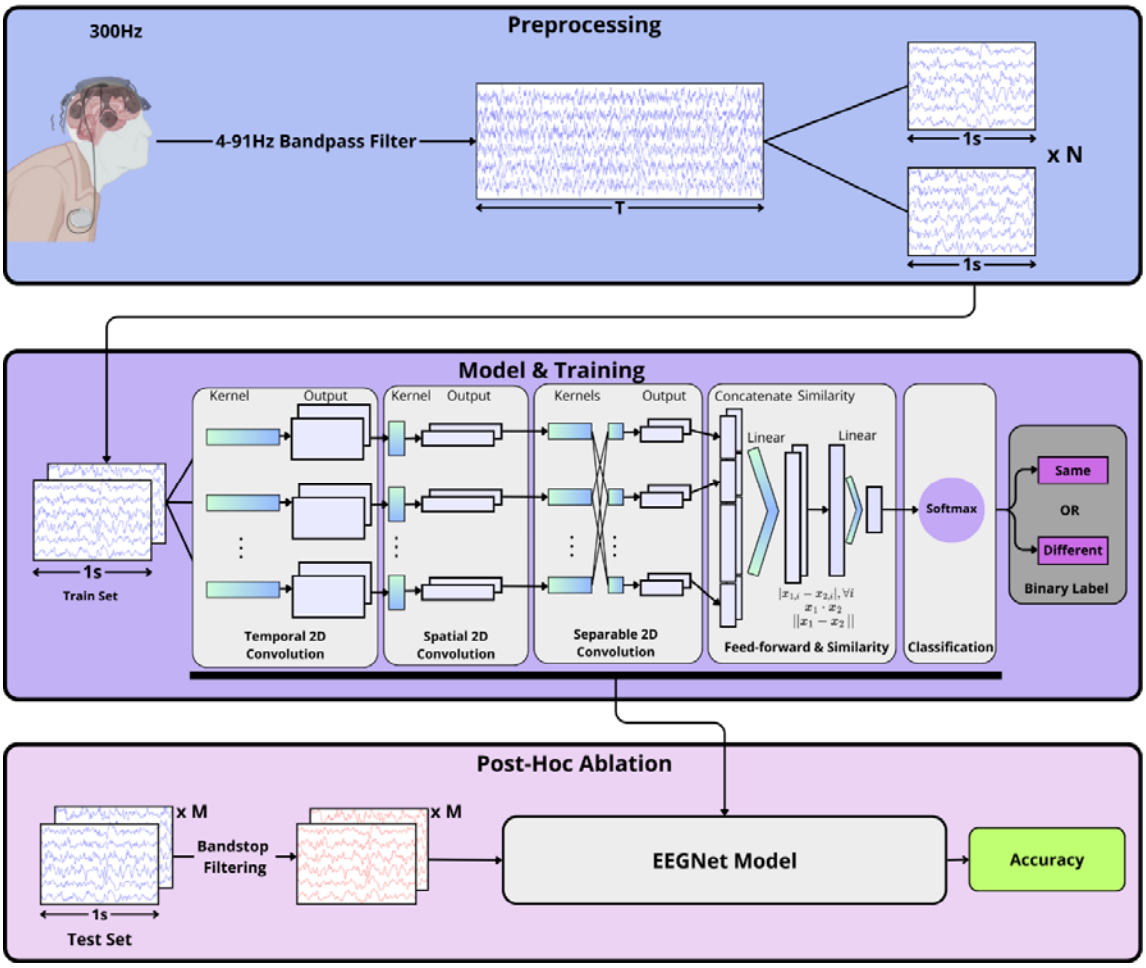


Figure 2. Data processing pipeline. *Data processing pipeline divided in three steps.* ***Preprocessing****: The EEG data is bandpass filtered between 4 to 91Hz and N pairs of 1s EEG segments are generated to train the model.* ***Model & Training****: Pairs of EEG segments are inputted as an image of 12x300 samples into the model. After convolutional feature extraction, similarity metrics are calculated and a final 2-D embedding that goes through the SoftMax layer to be classified as [0,1] or [1,0] ("same" or "different" respectively). The model dimensions were kept as in the original paper*[38]*.* ***Post-Hoc Ablation****: After the model has been trained, features are removed from the test dataset which is passed through the model to compare the accuracy to the baseline with the unfiltered data. When removing a certain feature caused large drops in accuracy, this feature was deemed to be extracted by our model and to be essential for the correct classification of the input. DBS patient schematic created in BioRender.com.*

Siamese EEGNet

For our deep learning architecture, we build upon the EEGNet; a model that has a relatively small number of parameters and has been shown to generalise across a wide range of tasks[38]. To enable the model to take a pair of 1s EEG segments as input, we adapted the EEGNet to a siamese configuration[42]; this extracts a latent-space representation from each segment in parallel using the same set of parameters and then joins them at the linear layers through similarity metrics[43]. The EEGNet architecture performs temporal and spatial convolution separately to extract features from both planes. After the convolutional feature extraction, the embedding of the input is passed through a fully connected layer that creates

a vector of dimension $d_{out} = 10$, or each of the segments. Then, three distance metrics are calculated and concatenated into a vector of dimension $d_{sim} = 12$. The first 10 elements are the element-wise absolute difference between the embedding of the pairs, the 11th element is the dot product of the vectors, and the 12th element is the norm-2 distance between the vectors. The choice of using three distance metrics resulted in a negligible increase in model parameters but led to an increased ability to infer the changes in the data. The concatenated vector (of dimension 12x1) is passed through a linear projection layer mapping into a 2-dimensional vector and finally is passed through a SoftMax and compared to the label. In this embedding, a [0,1] label represents a pair of "same" and a [1,0] label represents a pair of "different" (Figure 2). The model was implemented in PyTorch[44]. We used $F_1 = 8, D = 2$ and a dropout probability of 0.45 was added. We optimised the model using a binary cross entropy loss, Adam[45], and L2 regularisation. The L2 regularisation and learning rate were tuned on the validation dataset for every patient and parameter combination with a HyperBand scheduler[46] using 30 starting points. We used Ray[47] for the hyperparameter tuning. Our models were trained for 450 epochs, and the epoch with the lowest validation loss was extracted to be used on the test set. To control for training overfit, we set a threshold on the validation accuracy and excluded folds where the model's maximum validation accuracy across all epochs was below the threshold. We applied a variable threshold that kept the 75% top performing folds, unless the folds to be discarded reached 80% or more validation accuracy. Models were trained within a patient to discriminate changes in only one parameter at a time (e.g. amplitude of the left electrode), and accuracy was used to evaluate the model's performance.

Ablation Studies

To find the main frequency band that the model used to distinguish the datapoints, we performed an ablation test where each frequency band was filtered one by one from the test dataset using IIR bandstop filters, and the accuracy of the model trained on the full data was evaluated[43]. Frequency bands included theta (4-8Hz), alpha (9-12Hz), beta (13-30Hz), low-gamma (31-59Hz) and narrow-band gamma (60-90Hz) bands. We also separately applied moving bandstop IIR filters with a bandwidth of 12Hz to localise more precise frequency features. The filters were moved from 4Hz to 91Hz in steps of 1Hz. This was done for every patient and parameter; we present the averages across patients and parameters in Figure 3. A large drop in accuracy when removing a frequency range would indicate that features in this frequency range are essential in characterising the neural response to changes in DBS parameters.

Results

Classification Results

Our models achieved an overall 77.95% accuracy in classifying the pairs of EEG data with the binary label “same” or “different”, where “same” refers to both EEG segments in a pair having the same stimulation parameters, and “different” refers to the parameter of interest being different between the two segments. The highest accuracy was for distinguishing changes in the contact of the left electrode for P02, with an accuracy of 90.16% and the lowest was 67.88% for distinguishing changes in the contact of the right electrode for P03. The full accuracy results are shown in Table 2. To ensure the classification of contact changes, we analysed the model’s performance on pairs of “different” contacts that had no amplitude difference between the segments and obtained 75.06% accuracy. This accuracy represents the performance of models that were trained on all data when evaluated on pairs of “different” contacts that have no changes in amplitude. This accounted for 385 out of 4544 pairs of “different” contacts in our dataset. A visualisation of the accuracy in discriminating pairs of “different” contact with varying amplitude differences can be found in Supplementary Figure 1.

	P01 Left	**P01 Right**	**P02 Left**	**P02 Right**	**P03 Left**	**P03 Right**	**P04 Left**	**Average (%)**
Amplitude (%)	77.85	80.19	72.75	70.07	76.23	79.05	84.89	77.29
Contact (%)	77.20	N/A	90.16	73.11	75.62	67.88	88.33	78.72
Average (%)	77.53	80.19	81.46	71.59	75.93	73.47	86.61	**77.95**

Table 2: Test accuracies across patients and parameters. *Classification accuracy of “same” or “different” label for changes in contact and amplitude of both hemispheres. Across every patient and parameter, we achieved an accuracy of 77.95%, showing a proof-of-concept that the model can accurately distinguish between stimulation amplitude and contact points at the cortical level, with amplitude changes as small as 0.3mA.*

Frequency Ablation Study

The accuracies with the moving bandstop filters suggest some variability between patients and between parameters, but overall, the band that resulted in the largest accuracy drop was the NBG from 60-90Hz (Figure 3 and Table 3). This is evident for all hemispheres, both when evaluated for contact change (Figure 3A) and for amplitude change (Figure 3B). Yet, the subharmonic of the stimulation frequency (dashed vertical lines in all panels of Figure 3) was not generally aligned with this dip. Only in two hemispheres (P03_L and P04_L), we observed an alignment between the NBG dip in ablation and the first subharmonic of the stimulation (Figures 3A, B). P03_L and P04_L are also the only two hemispheres in which the Power Spectral Density (PSD) of the input EEG data (Figure 3C) showed an NBG peak at half of the stimulation frequency. Both the PSD subharmonic peak (Figure 3C) and the dip in accuracy (Figures 3A,B) were significantly more pronounced in P04_L than in P03_L. Additionally, for P03_L, the alignment of this dip is weaker for changes in amplitude (Figure 3A), and it is not evident in the right hemisphere of that patient (Figure 3). While performing EEG-band bandstop filtering (Table 3), we found that when averaging across hemispheres,

the band that negatively impacted accuracy the most was the NBG from 60-90Hz for every parameter. Removing the NBG frequency band (60-90Hz) caused an average decrease in accuracy of 18.41% across all hemispheres and parameters, resulting in a low average accuracy of 59.54%. For clarity purposes, the results were averaged across hemispheres (Table 3), but when inspecting each one individually, we saw that for P03's left contact and P02's right amplitude, the band that most affected the accuracy was the theta band (4-8Hz) (Supplementary Table 2). Nonetheless, even for these two cases, the NBG had a significant impact on the accuracy (Supplementary Table 2). This makes NBG the most important frequency band in 11 out of 13 cases. The full bandstop filtering results for each hemisphere and parameter can be found in Supplementary Table 2.

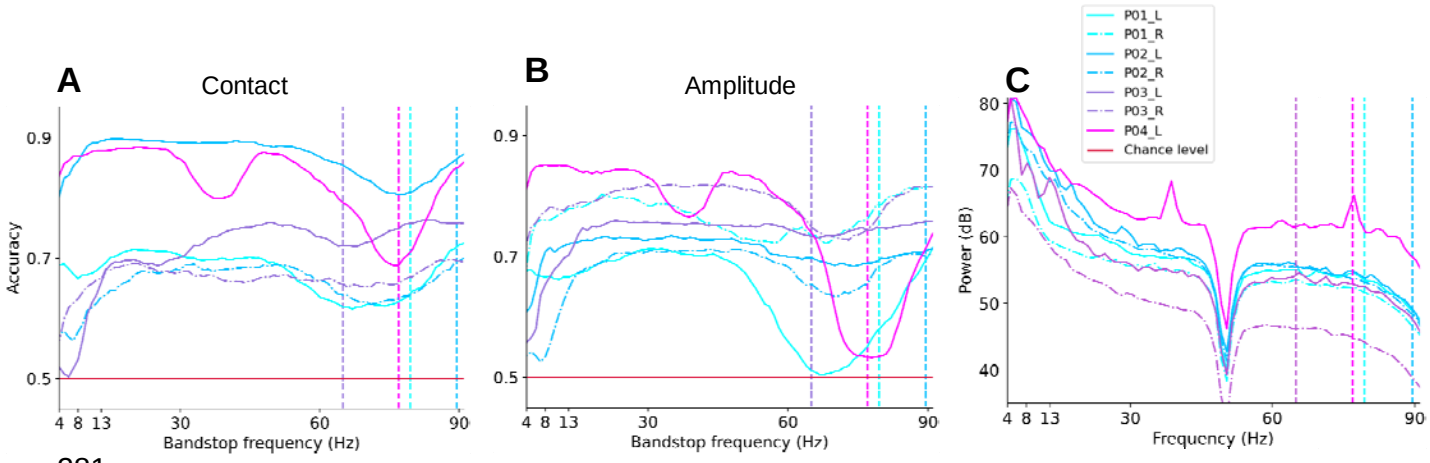


Figure 3: Accuracy results from ablation bandstop study for different patients and parameters. *A 12Hz bandstop moving filter was swept from 4Hz to 91Hz.* **A:** *Accuracy results from the ablation bandstop study of each hemisphere's changes in contact are plotted in different colours. The vertical dotted lines represent the first subharmonic of the stimulation frequency (i.e. $f_{stim}/2$) for each patient. NBG (60-90Hz) suppression was observed in all patients, but only in P04_L and P03_L this drop aligned with the first subharmonic of the stimulation frequency* **B***: Accuracy results from the ablation bandstop study for changes in amplitude. Legends as in A. Similar NBG (60-90Hz) suppression was observed across hemispheres with an even stronger dip aligned to the first subharmonic in P04_L and a weaker aligned dip in P03_L.* **C***: Power Spectral Density (PSD) of the input EEG data in dB for each patient and a single channel using Welch's periodogram during the time windows of the pairs of changes in amplitude. The channel selected for each patient represented the channel that most negatively affected the model's accuracy when removed from the test dataset. A peak in PSD aligned with the first subharmonic was observed only in P04_L and to some extent in P03_L. Similar PSDs were observed for the pairs of changes in contact.*

	Baseline	Theta (4-8Hz)	Alpha (9-12Hz)	Beta (13-30Hz)	Low-Gamma (31-59Hz)	Narrow-Band Gamma (60-90Hz)	Average (%)
Amplitude (%)	78.67	75.5	77.02	76.07	74.9	**58.65**	73.47
Contact (%)	76.79	70.29	74.53	71.14	72.5	**60.96**	71.04
Average (%)	77.95	73.5	76.06	74.17	73.98	**59.54**	72.53

Table 3. Ablation test accuracies after band-filtering. *Accuracy on the test dataset in % for each parameter averaged across hemispheres. The first column represents the baseline accuracy, taken from Table 2, and each subsequent column represents the ablation accuracies after filtering out one frequency band from the test dataset. For both parameters across hemispheres, filtering the Narrow-Band Gamma (60-90Hz) caused the largest drop in accuracy, making it the most important band for the models' performance. The total averages are not the same as the row or column averages, since the number of unique parameters analysed varied between patients (see Table 2). For a full representation of the accuracy drop per hemisphere and parameter combination, see Supplementary Table 2.*

Discussion

Our siamese CNN on 1-second EEG segments achieved an average accuracy of 77.95% across all PD patients and parameters, showing that we can accurately detect small changes in DBS parameters in a clinical setting, at the level of the scalp EEG. We achieved an accuracy of 75.06% in discriminating pairs of "different" contact with no changes in amplitude, which shows that our models are sensitive to changes in contact. Due to our limited sample size and some confounding factors between patients, – such as differences in symptoms, reaction to DBS, disease progression, or parameter ranges – we cannot conclude whether a single parameter is easier to discern. We have hereby demonstrated a novel approach achieving significantly above-chance performance using 6 dry EEG channels, 1-second time windows, and a relatively simple deep learning method. Our framework is the first to enable discrimination of the neural response to small changes in DBS parameters at the scalp level using a siamese CNN. This was done in a task-agnostic paradigm, where rest and arm movement segments were mixed in the dataset, showing the robustness of our methods. We further achieve this using EEG instead of LFP recordings; while EEG is noisier and more prone to artefacts, it captures the brain-wide activity, helping to discern the neural response to DBS. Detection of small DBS changes at the scalp level paves the way for the extraction of more robust biomarkers. Doing so with EEG recordings could potentially result in biomarker-guided programming sessions for patients with DBS devices that do not support LFP recordings. Finally, our study is among the few that explore the effects of DBS programming at the cortical level using a clinically relevant dataset. No modifications were made to the programming session protocol, which enhances the real-world value of our results.

A key goal of our study was to identify the frequency bands our models extracted to distinguish neural activity at different DBS parameters. The PSDs of our patients did not reveal any abnormalities, apart from the entrained peak of P04_L and P03_L (Figure 3C),

which has been characterised in the literature. With our full-band ablation studies that filter each frequency band, the NBG (60-90Hz) band emerged as the most significant contributor to model accuracy in 11 out of 13 cases, indicating its robustness in reflecting neural responses to DBS. Whilst each bandstop plot shows dips in accuracy in the NBG range (Figure 3), we observe higher accuracy minima than what was reported when removing the whole NBG activity at once (Table 3). This suggests that sub-bands of the NBG characterise the neural response better than others, but the remaining NBG activity should not be ignored. Despite beta power being a known marker in PD and its modulation by DBS, our results did not consistently support its significance (Table 3). It's possible that our models prioritised NBG features over beta due to easier convergence during training. While multiple factors could explain why the beta activity does not appear as a biomarker here, it is important to note that it has been strongly correlated with big changes, like DBS On vs Off but not to changes in DBS parameters[4,14,17–19]. We also observed moderately strong theta band drops in two cases (Supplementary table 2), which we attributed to the patients' tremor.

Recent studies have observed a shift from beta to gamma activity in the context of DBS, which aligns with our findings[10,27,29–34]. Initially, NBG activity recorded via LFP was used as a correlate for dyskinesia to adjust DBS levels, and this has been explored at the cortical level as well[29,30]. It is possible that our models detected NBG as a proxy for excessive stimulation leading to dyskinesia, but our current data does not allow us to confirm this. Changes in amplitude as small as 0.3mA are often subclinical and not detected using conventional clinical scales such as the UPDRS. Our ability to discriminate the neural response to these small amplitude changes suggests that the NBG oscillations could be a highly sensitive biomarker. The 1:2 entrainment of brain activity, where the brain oscillates at half the stimulation frequency, has recently gained attention[10,27,31–34,48]. Surprisingly, we only observed this in two hemispheres, despite adequate stimulation amplitude (Supplementary Table 1), and this effect being known to be stronger during movement, which was included in our dataset[32,34]. It is also surprising to observe signs of subharmonic entrainment for P03_L since the patient was in the OFF medication state[10,31]. Overall, our results suggest that the importance of cortical NBG in DBS response goes beyond the 1:2 entrainment.

A key concern in our study was the potential misclassification of stimulation artefacts as neural responses, given that our models had no a priori knowledge of which features to extract. For P04_L, we show evidence of a subharmonic entrainment, and of our models learning it (Figure 3). This supports that, in P04_L, we are detecting a brain response rather than an artefact. This is justified by the entrainment's characteristics—an upper amplitude limit, dependence on dopaminergic state, prominence in cortical areas, functional correlation, and persistence post-stimulation—indicating a neural origin[10,31,32,34]. Despite allowing amplitude to vary for changes in contact, our models achieve an accuracy of 75.06% in classifying pairs of "different" contact with no changes in amplitude, showing that they are as sensitive to contact change as pairs with contact and amplitude changes (Supplementary Figure 1). Contact points are situated as close as 0.5mm apart in deep structures of the brain, which makes contact changes unlikely to cause electrical artefacts in EEG data at the cortex level, thus pointing to the classification of neural response. We also found no bias in classifying monopolar vs. bipolar stimulation, suggesting that our models accurately identify contact points, not stimulation types (Supplementary Tables 3 and 4). Additionally, considering the common view that bipolar settings do not affect the cortex with stimulation artefacts[49,50], our results - which show similar accuracies in bipolar pairs -

represent true changes in cortical activity and not stimulation artefacts (Supplementary Tables 3 and 4). While we could not fully characterise the features our models extract, the lack of bias and ability to classify contact changes in both stimulation types support the claim that our models are detecting the neural response to changes in DBS parameters.

Despite demonstrating a clinically relevant proof-of-concept for parameter discrimination at the cortical level, our work has several limitations. We trained within-patient models, requiring a large amount of data from each patient, which restricted our sample size to 7 hemispheres, limiting the generalizability of our results. However, we identified a marker that generalised across all patients, hence providing a promising direction towards generalization. We also restricted patient tasks to rest and arm movement, as tasks like walking introduced strong EEG artefacts, though our models' success across two different tasks suggests early evidence of task robustness. Then, we lacked control over the stimulation parameters administered to patients, leading to inter-patient variability and limiting our study to a "same" vs. "different" setting, rather than a parameter prediction paradigm. As a proof-of-concept study, we used limited explainability methods, focusing on ablation studies. While we found NBG (60-90Hz) critical for model performance, we could not pinpoint the exact features extracted. Future work will apply methods like saliency maps and convolutional filter inspection to better understand the brain's neural response.

In this work, we introduced a novel method for detecting changes in DBS parameters at the cortical level, achieving significantly above chance-level performance. By testing this on 7 hemispheres, this work represents a proof of concept that we can detect changes in contact and changes in amplitude as small as 0.3mA at the cortical level with only 1s segments of 6 EEG channels. To gain preliminary insight into the neural response to changes in DBS parameters, we performed ablation studies and found that our models were extracting features from the NBG (60-90Hz) to accurately classify pairs of EEG segments as “same” or “different” depending on the DBS parameters each segment came from. Our data-driven approach produced findings which are in agreement with other recent studies linking NBG activity with DBS changes and motor symptoms in PD. Future works will focus on improving the AI methods, both in terms of accuracy and explainability, and in creating a patient-agnostic model that could be evaluated on a larger sample size with minimal training. Other steps required to bring this work to the clinic will involve using the found biomarkers to predict specific DBS parameters and their effect on symptom severity. Overall, by showing that we can reliably detect small changes in DBS parameters at the cortical level, this work brings us closer to realising biomarker-guided DBS programming, a significant step towards more effective and personalised treatments.

References

1. Lozano AM, Lipsman N, Bergman H, et al. Deep brain stimulation: current challenges and future directions. *Nature reviews Neurology*. 2019;15(3):148-160. doi:10.1038/S41582-018-0128-2

2. Degirmenci Y. Current DBS programming. *Deep Brain Stimulation*. 2024;4:29-31. doi:10.1016/j.jdbs.2023.12.002

3. Mucke D, Hermes A, Roettger TB, et al. The effects of Thalamic Deep Brain Stimulation on speech dynamics in patients with Essential Tremor: An articulographic study. *PLoS One*. 2018;13(1):e0191359. doi:10.1371/journal.pone.0191359

4. Neumann WJ, Gilron R, Little S, Tinkhauser G. Adaptive Deep Brain Stimulation: From Experimental Evidence Toward Practical Implementation. *Movement Disorders*. 2023;38(6):937-948. doi:10.1002/MDS.29415

5. Guidetti M, Marceglia S, Loh A, et al. Clinical perspectives of adaptive deep brain stimulation. *Brain Stimulation*. 2021;14(5):1238-1247. doi:10.1016/J.BRS.2021.07.063

6. Little S, Pogosyan A, Neal S, et al. Adaptive deep brain stimulation in advanced Parkinson disease. *Annals of neurology*. 2013;74(3):449-457. doi:10.1002/ANA.23951

7. Rosa M, Arlotti M, Marceglia S, et al. Adaptive deep brain stimulation controls levodopa-induced side effects in Parkinsonian patients. *Movement Disorders*. 2017;32(4):628-628. doi:10.1002/MDS.26953

8. Velisar A, Syrkin-Nikolau J, Blumenfeld Z, et al. Dual threshold neural closed loop deep brain stimulation in Parkinson disease patients. *Brain stimulation*. 2019;12(4):868-876. doi:10.1016/J.BRS.2019.02.020

9. Bocci T, Prenassi M, Arlotti M, et al. Eight-hours conventional versus adaptive deep brain stimulation of the subthalamic nucleus in Parkinson's disease. *NPJ Parkinson's disease*. 2021;7(1). doi:10.1038/S41531-021-00229-Z

10. Oehrn CR, Cernera S, Hammer LH, et al. Chronic adaptive deep brain stimulation versus conventional stimulation in Parkinson's disease: a blinded randomized feasibility trial. *Nat Med*. Published online August 19, 2024:1-12. doi:10.1038/s41591-024-03196-z

11. Moran A, Bar-Gad I, Bergman H, Israel Z. Real-time refinement of subthalamic nucleus targeting using Bayesian decision-making on the root mean square measure. *Movement Disorders*. 2006;21(9):1425-1431. doi:10.1002/mds.20995

12. Little S, Pogosyan A, Neal S, et al. Controlling Parkinson's Disease With Adaptive Deep Brain Stimulation. *JoVE (Journal of Visualized Experiments)*. 2014;(89):e51403. doi:10.3791/51403

13. Eusebio A, Thevathasan W, Gaynor LD, et al. Deep brain stimulation can suppress pathological synchronisation in parkinsonian patients. *Journal of Neurology, Neurosurgery & Psychiatry*. 2011;82(5):569-573. doi:10.1136/jnnp.2010.217489

14. Oswal A, Beudel M, Zrinzo L, et al. Deep brain stimulation modulates synchrony within spatially and spectrally distinct resting state networks in Parkinson's disease. *Brain*. 2016;139(5):1482-1496. doi:10.1093/brain/aww048

15. Sanger ZT, Henry TR, Park MC, Darrow D, McGovern RA, Netoff TI. Neural signal data collection and analysis of Percept™ PC BrainSense recordings for thalamic stimulation in epilepsy. *J Neural Eng*. 2024;21(1):012001. doi:10.1088/1741-2552/ad1dc3

16. Thenaisie Y, Palmisano C, Canessa A, et al. Towards adaptive deep brain stimulation: clinical and technical notes on a novel commercial device for chronic brain sensing. *J Neural Eng*. 2021;18(4):042002. doi:10.1088/1741-2552/ac1d5b

17. Tinkhauser G, Pogosyan A, Little S, et al. The modulatory effect of adaptive deep brain stimulation on beta bursts in Parkinson's disease. *Brain*. 2017;140(4):1053-1067. doi:10.1093/brain/awx010

18. van Wijk BCM, de Bie RMA, Beudel M. A systematic review of local field potential physiomarkers in Parkinson's disease: from clinical correlations to adaptive deep brain stimulation algorithms. *J Neurol*. 2023;270(2):1162-1177. doi:10.1007/s00415-022-11388-1

19. Stanslaski S, Summers RLS, Tonder L, et al. Sensing data and methodology from the Adaptive DBS Algorithm for Personalized Therapy in Parkinson's Disease (ADAPT-PD) clinical trial. *npj Parkinsons Dis*. 2024;10(1):1-11. doi:10.1038/s41531-024-00772-5

20. Priori A, Foffani G, Pesenti A, et al. Rhythm-specific pharmacological modulation of subthalamic activity in Parkinson's disease. *Experimental Neurology*. 2004;189(2):369-379. doi:10.1016/j.expneurol.2004.06.001

21. Tinkhauser G, Torrecillos F, Duclos Y, et al. Beta burst coupling across the motor circuit in Parkinson's disease. *Neurobiology of Disease*. 2018;117:217-225. doi:10.1016/j.nbd.2018.06.007

22. Torrecillos F, Tinkhauser G, Fischer P, et al. Modulation of Beta Bursts in the Subthalamic Nucleus Predicts Motor Performance. *J Neurosci*. 2018;38(41):8905-8917. doi:10.1523/JNEUROSCI.1314-18.2018

23. Little S, Brown P. Debugging Adaptive Deep Brain Stimulation for Parkinson's Disease. *Movement Disorders*. 2020;35(4):555-561. doi:10.1002/mds.27996

24. de Hemptinne C, Swann NC, Ostrem JL, et al. Therapeutic deep brain stimulation reduces cortical phase-amplitude coupling in Parkinson's disease. *Nat Neurosci*. 2015;18(5):779-786. doi:10.1038/nn.3997

25. Swann N, Poizner H, Houser M, et al. Deep Brain Stimulation of the Subthalamic Nucleus Alters the Cortical Profile of Response Inhibition in the Beta Frequency Band: A Scalp EEG Study in Parkinson's Disease. *J Neurosci*. 2011;31(15):5721-5729. doi:10.1523/JNEUROSCI.6135-10.2011

26. Luoma J, Pekkonen E, Airaksinen K, et al. Spontaneous sensorimotor cortical activity is suppressed by deep brain stimulation in patients with advanced Parkinson's disease. *Neuroscience Letters*. 2018;683:48-53. doi:10.1016/j.neulet.2018.06.041

27. Muthuraman M, Bange M, Koirala N, et al. Cross-frequency coupling between gamma oscillations and deep brain stimulation frequency in Parkinson's disease. *Brain: a journal of neurology*. 2020;143(11):3393-3407. doi:10.1093/BRAIN/AWAA297

28. Sumarac S, Youn J, Fearon C, et al. Clinico-physiological correlates of Parkinson's disease from multi-resolution basal ganglia recordings. *npj Parkinsons Dis*. 2024;10(1):1-11. doi:10.1038/s41531-024-00773-4

29. Swann NC, De Hemptinne C, Thompson MC, et al. Adaptive deep brain stimulation for Parkinson's disease using motor cortex sensing. *Journal of neural engineering*. 2018;15(4). doi:10.1088/1741-2552/AABC9B

30. Olaru M, Cernera S, Hahn A, et al. Motor network gamma oscillations in chronic home recordings predict dyskinesia in Parkinson's disease. *Brain*. 2024;147(6):2038-2052. doi:10.1093/brain/awae004

31. Wiest C, Tinkhauser G, Pogosyan A, et al. Subthalamic deep brain stimulation induces finely-tuned gamma oscillations in the absence of levodopa. *Neurobiology of disease*. 2021;152. doi:10.1016/J.NBD.2021.105287

32. Mathiopoulou V, Habets J, Feldmann LK, et al. DBS-induced gamma entrainment as a new biomarker for motor improvement with neuromodulation. Published online April 26, 2024:2024.04.25.24306357. doi:10.1101/2024.04.25.24306357

33. Sermon JJ, Benjaber M, Duchet B, et al. 1:2 entrainment is not a device-induced artefact, except when it is. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*. 2024;17(2):149-151. doi:10.1016/j.brs.2024.01.010

34. Sermon JJ, Olaru M, Anso J, et al. Sub-harmonic Entrainment of Cortical Gamma Oscillations to Deep Brain Stimulation in Parkinson's Disease: Predictions and Validation of a Patient-Specific Non-Linear Model. Published online March 4, 2022:2022.03.01.482549. doi:10.1101/2022.03.01.482549

35. Sand D, Arkadir D, Abu Snineh M, et al. Deep Brain Stimulation Can Differentiate Subregions of the Human Subthalamic Nucleus Area by EEG Biomarkers. *Front Syst Neurosci*. 2021;15. doi:10.3389/fnsys.2021.747681

36. Peeters J, Van Bogaert T, Boogers A, et al. EEG-based biomarkers for optimizing deep brain stimulation contact configuration in Parkinson's disease. *Front Neurosci*. 2023;17. doi:10.3389/fnins.2023.1275728

37. Spooner RK, Hizli BJ, Bahners BH, Schnitzler A, Florin E. Modulation of DBS-induced cortical responses and movement by the directionality and magnitude of current administered. *npj Parkinsons Dis*. 2024;10(1):1-10. doi:10.1038/s41531-024-00663-9

38. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. *J Neural Eng*. 2018;15(5):056013. doi:10.1088/1741-2552/aace8c

39. Avants B, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight J*. 2008;1–35. doi:10.54294/uvnhin

40. Horn A, Li N, Dembek TA, et al. Lead-DBS v2: Towards a comprehensive pipeline for deep brain stimulation imaging. *Neuroimage*. 2019;184:293-316. doi:10.1016/j.neuroimage.2018.08.068

41. Husch A, V Petersen M, Gemmar P, Goncalves J, Hertel F. PaCER - A fully automated method for electrode trajectory and contact reconstruction in deep brain stimulation. *Neuroimage Clin*. 2018;17:80-89. doi:10.1016/j.nicl.2017.10.004

42. Bromley J, Guyon I, Lecun Y, et al. Signature Verification using a “Siamese” Time Delay Neural Network. *Advances in Neural Information Processing Systems*. 1993;6.

43. Haugland MR, Borovykh A, Tai Y, Haar S. Explainable Deep Learning for Arm Classification During Deep Brain Stimulation - Towards Digital Biomarkers for Closed-Loop Stimulation. In: *2023 Conference on Cognitive Computational Neuroscience*. Cognitive Computational Neuroscience; 2023. doi:10.32470/CCN.2023.1368-0

44. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Published online December 3, 2019. doi:10.48550/arXiv.1912.01703

45. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. Published online January 29, 2017. doi:10.48550/arXiv.1412.6980

46. Li L, Jamieson K, Rostamizadeh A, et al. A System for Massively Parallel Hyperparameter Tuning. Published online March 15, 2020. doi:10.48550/arXiv.1810.05934

47. Moritz P, Nishihara R, Wang S, et al. Ray: A Distributed Framework for Emerging AI Applications. Published online September 29, 2018. doi:10.48550/arXiv.1712.05889

48. Wiest C, Torrecillos F, Tinkhauser G, et al. Finely-tuned gamma oscillations: Spectral characteristics and links to dyskinesia. *Experimental Neurology*. 2022;351:113999. doi:10.1016/j.expneurol.2022.113999

49. Frysinger RC, Quigg M, Elias WJ. Bipolar deep brain stimulation permits routine EKG, EEG, and polysomnography. *Neurology*. 2006;66(2):268-270. doi:10.1212/01.wnl.0000194272.79084.7e

50. Allen DP, Stegemöller EL, Zadikoff C, Rosenow JM, MacKinnon CD. Suppression of deep brain stimulation artifacts from the electroencephalogram by frequency-domain Hampel filtering. *Clin Neurophysiol*. 2010;121(8):1227-1232. doi:10.1016/j.clinph.2010.02.156

Statements

Data availability statement

The dataset used in this project contains sensitive patient data and will therefore not be made publicly available. The authors are keen on collaborative work with this dataset, please contact the corresponding author if interested.

The underlying code for this study will be made available in the DBS-s-EEGNet repository (https://github.com/HaarLab/DBS-s-EEGNet). The code has been adapted to be flexible and work with your own EEG data, therefore the preprocessing code is not present in the repository. See the repository description for detailed information on how to use it.

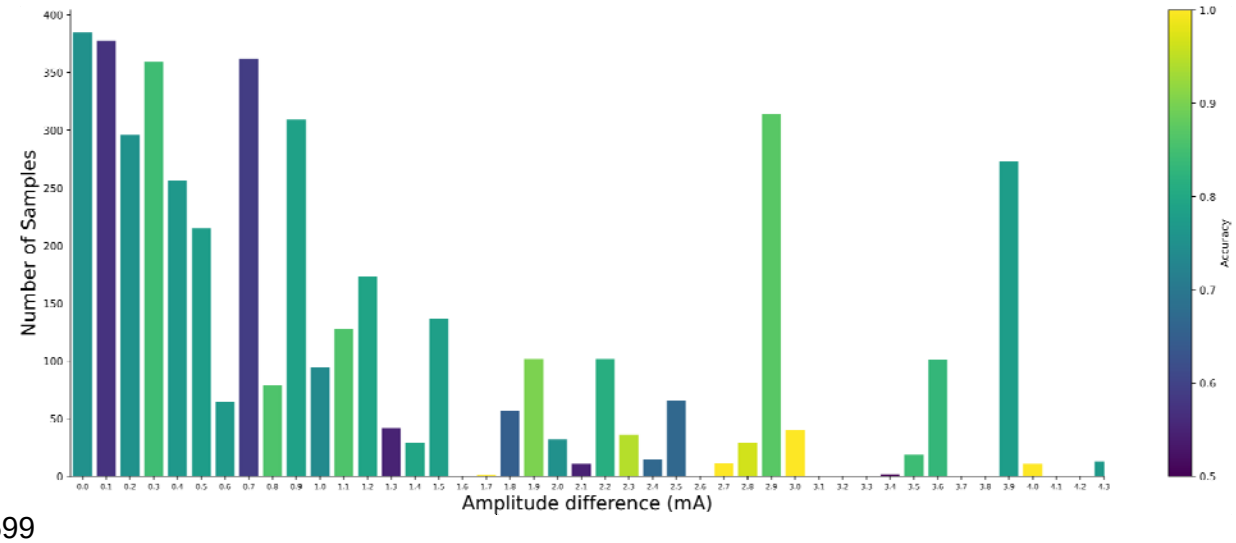
Acknowledgements

The study was supported by the Edmond and Lily Safra Fellowship awarded to S.H. and by the UK Dementia Research Institute Care Research & Technology Centre. N.C.P and C.G. are supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1];

Competing Interests

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.

Supplementary material



Supplementary Figure 1: Accuracy distribution in pairs of "different" contact with different amplitude changes*. The different sample sizes are represented as bars, where the colour represents the accuracy varying from 0 to 1. Our model achieves* 75.06% *accuracy when no changes in amplitude are present with a sample size of N = 385 pairs. We can see that our models are able to differentiate changes in contact across a wide range of amplitude differences.*

	P01	P02	P03	P04
Left Amplitude (mA)	6.3 – 6.8	3.1 – 7.0	0.5 – 2.3	0.5 – 5.3
# of Left Contact Combinations	3	2	2	6
Left Frequency (Hz)	159	179	130	154
Left Pulse-Width (µs)	80	70, 90	60	60
Right Amplitude (mA)	5.3 – 5.8	1.2 – 7.0	0.5 – 3.0	4.4
# of Right Contact Combinations	1	9	3	1
Right Frequency (Hz)	159	179	130	116
Right Pulse-Width (µs)	70, 80	60, 90	60	50

Supplementary Table 1. Parameter range across the different patients. *Parameter range for each parameter and patient. The number in the contact cells represents the number of unique contact configurations that were displayed, and in the amplitude cell, the range is represented. However, the*

changes in the amplitude were not linear or systematic, meaning that a big range does not necessarily result in more unique amplitude configurations.

	Baseline	Theta (4-8Hz)	Alpha (9-12Hz)	Beta (13-30Hz)	Low Gamma (31-59Hz)	Narrow-Band Gamma (60-90Hz)	Average
P01 Left Amplitude	77.85	73.78	74.66	64.81	68.75	46.06	67.65
P01 Left Contact	77.20	71.80	72.93	68.13	67.07	52.33	68.24
P01 Right Amplitude	80.19	80.03	79.42	78.73	76.45	60.90	75.95
P02 Left Amplitude	72.75	72.17	72.95	72.27	72.75	66.31	71.53
P02 Left Contact	90.16	86.19	90.21	89.26	88.99	71.69	86.08
P02 Right Amplitude	70.07	55.72	66.58	69.74	70.53	59.14	65.30
P02 Right Contact	73.11	67.65	72.55	66.04	67.65	55.18	67.03
P03 Left Amplitude	76.23	71.93	72.02	74.74	75.00	66.23	72.69
P03 Left Contact	75.62	63.12	70.76	67.21	73.23	66.51	69.41
P03 Right Amplitude	79.05	77.37	77.68	77.45	78.98	61.01	75.26
P03 Right Contact	67.88	62.71	66.19	65.08	65.56	59.07	64.42
P04 Left Amplitude	84.89	85.16	84.95	83.59	76.58	50.49	77.61
P04 Left Contact	88.33	87.81	87.92	87.19	80.16	59.06	81.75
Average	77.95	73.50	76.06	74.17	73.98	59.54	72.53

Supplementary Table 2. Full band-filtering results. *Accuracy in % after filtering different bands from the test set with a model trained on the full data. The results are shown for every patient and parameter that was analysed in this study. For 11 out of 13 cases, the Narrow-Band Gamma (60-90Hz) caused the most significant drop in accuracy. For P02's right amplitude and P03's left contact, the Theta band (4-8Hz) caused the most significant drop in accuracy, but the Narrow-Band Gamma still caused a significant decrease.*

		Same M-M	Same B-B	Diff M-M	Diff M-B	Diff B-B
1	**P01 Left**	N/A	88.00% N=750	N/A	N/A	66.40% N=750
2	**P01 Right**	N/A	N/A	N/A	N/A	N/A
3	**P02 Left**	N/A	96.72% N=945	N/A	N/A	83.60% N=945
4	**P02 Right**	71.24% N=445	87.83% N=269	65.24% N=607	90.00% N=100	71.43% N=7
5	**P03 Left**	77.16% N=648	N/A	74.07% N=648	N/A	N/A
6	**P03 Right**	71.24% N=540	N/A	65.37% N=540	N/A	N/A
7	**P04 Left**	8.82% N=34	88.44% N=926	94.25% N=87	91.04% N=603	90.00% N=270

Supplementary Table 3: Accuracy and number for each patient's hemisphere separated by polarity in pairs of "same" or "different" contact*. M-M means that the patients received monopolar stimulation in both segments of the pair. M-B means one segment had monopolar stimulation and the other bipolar stimulation. B-B means the patient received bipolar stimulation in both segments. This shows that our models can learn to differentiate contact even when the stimulation type is the same across both segments of a pair, and that our models can learn regardless of the stimulation type. Additionally, it shows that our data mostly had same stimulation type pairs and that our models were able to learn when presented with both stimulation types at the same time.*

		Same M-M	Same B-B	Diff M-M	Diff B-B
1	**P01 Left**	N/A	87.64% N=736	N/A	68.07% N=736
2	**P01 Right**	N/A	92.38% N=656	N/A	67.99% N=656
3	**P02 Left**	N/A	86.52% N=512	N/A	58.98% N=512
4	**P02 Right**	46.19% N=578	86.81% N=182	84.21% N=760	N/A
5	**P03 Left**	82.11% N=570	N/A	70.35% N=570	N/A
6	**P03 Right**	85.19% N=540	N/A	73.33% N=540	N/A
7	**P04 Left**	100.0% N=21	90.66% N=899	61.54% N=26	79.42% N=894

Supplementary Table 4: Accuracy and number for each patient's hemisphere separated by polarity in pairs of "same" or "different" amplitude*. M-M means that the patients received monopolar stimulation in both segments of the pair. B-B means the patient received bipolar stimulation in both segments. Since the contact was the same in two segments of a pair of amplitude change, the M-B case is not present (see Supplementary Table 3 for reference). This shows that are models can learn to differentiate amplitude regardless of the stimulation type. Additionally, it shows that our models were able to learn when presented with both stimulation types at the same time.*