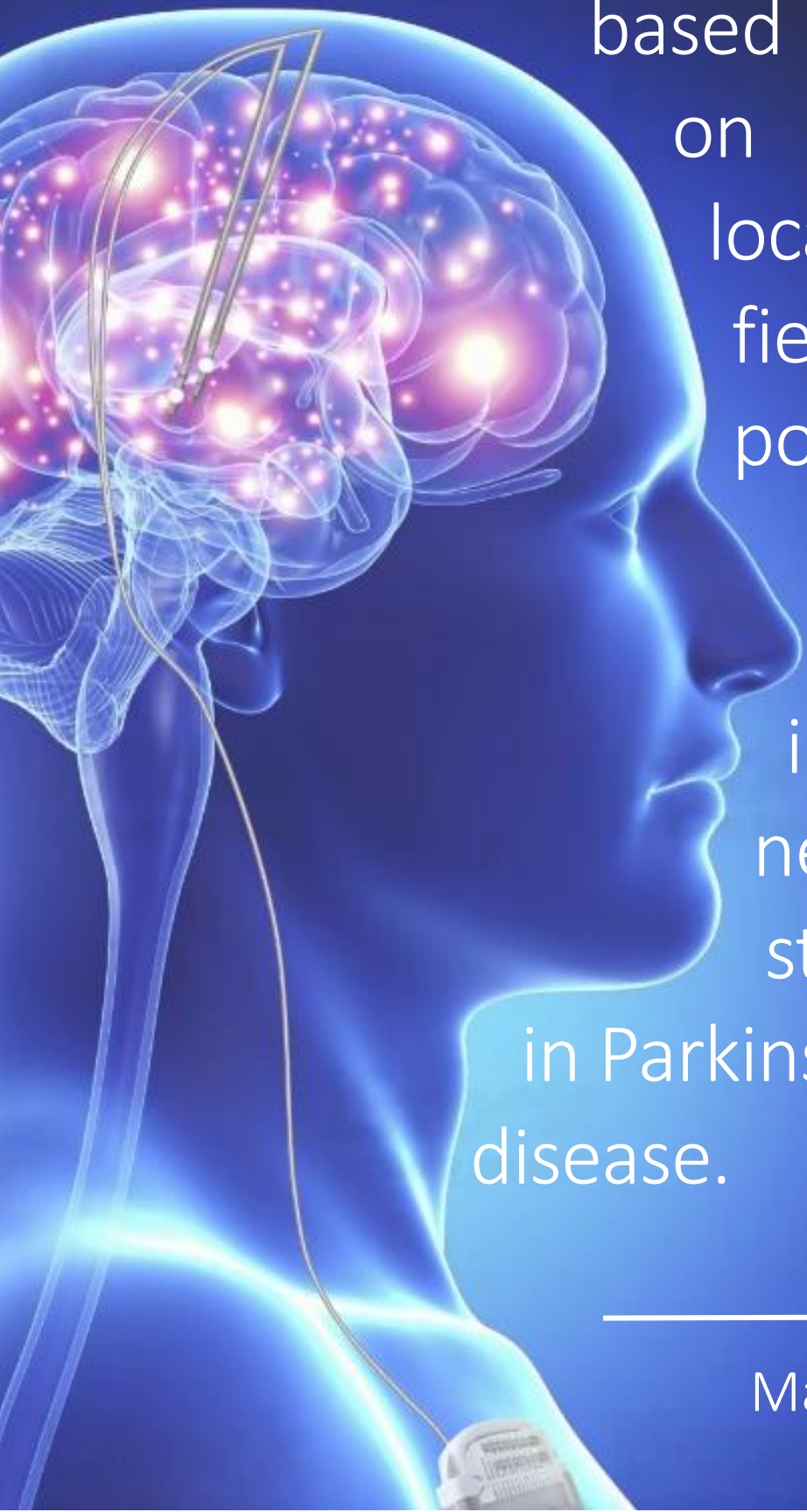
Master's Thesis - MSc Technical Medicine - August 2023

Prediction of the optimal deep brain stimulation contact based on local field potentials measured by the implanted neuro-stimulator in Parkinson's disease.



Marjolein Muller

This page was intentionally left blank.

PREDICTION OF THE OPTIMAL DEEP BRAIN STIMULATION CONTACTS BASED ON LOCAL FIELD POTENTIALS MEASURED BY THE IMPLANTED NEUROSTIMULATOR IN PARKINSON'S DISEASE.

Marjolein, Muller

Student number : 4649095

4 August 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)
Department of Neurology, HagaZiekenhuis
Biomechanical engineering, TU Delft
September 2022 – August 2023
Feasibility study (15 ECTS): April 2023 – June 2023

Supervisors

Prof. dr. ir. A.C. (Alfred) Schouten, TU Delft (Technical Supervisor)
Dr. M.F. (Fiorella) Contarino, HagaZiekenhuis (Medical Supervisor)

Thesis committee members

Prof. dr. ir. A.C. (Alfred) Schouten, TU Delft (chair)
Dr. M.F. (Fiorella) Contarino, HagaZiekenhuis
Dr. M.R (Martijn) Tannemaat, LUMC
Prof. dr. ir. W.A. (Wouter) Serdijn, TU Delft

This page was intentionally left blank.

Preface

This master thesis report represents the final element before obtaining my masters degree in Technical Medicine. It concludes three years of studying the master Technical Medicine at the Technical University Delft, Erasmus University Rotterdam and University Leiden. The focus of this master thesis was within the field of neurology. Specifically where technical innovations meet the clinical practice, an ideal combination for concluding the master Technical Medicine. Furthermore, the neurological field already gained my interest at the start of my studies. The known working mechanisms of the human mind, how we think, feel and communicate but also the remaining mysteries of the brain instantly intrigued me. During my master thesis I had the opportunity to explore this interest and increase my knowledge on this subject as well as on technological innovations within this field. Furthermore, I had the chance to contribute to local as well as international collaborations. I truly enjoyed working on this project and am looking forward to performing further research within this field.

First, I would like to thank my medical supervisor, Fiorella Contarino, who assisted me with finding my way through the complex matters of this field for the entirety of my master thesis project. Her enthusiasm for research is amazing and contagious. I always enjoyed our discussion and considered her advices to be extremely valuable. Next, I would like to thank my technical supervisor, Alfred Schouten. He was always able to provide insightful advice on the technical elements that formed the backbone of my analyses. Furthermore, I would like to thank the NMC lab team and attending students for their valuable responses to my complicated questions. I would also like to thank the Haga Teaching Hospital employees who contributed to my professional and clinical development. Finally, many thanks go out to my friends and family, my parents and Peter in particular. Thank you for your unconditional support throughout the entire journey of my master thesis.

Marjolein Muller
Delft, August 2023

Summary

Introduction: Over the past two decades deep brain stimulation (DBS) has emerged as an important therapeutic option for Parkinson's disease (PD). However, the current DBS programming method, monopolar review (MPR), is time-consuming, requires highly trained personnel and causes discomfort for patients. This study aimed to predict the optimal stimulation contact(s) based on local field potential (LFP) recordings by the implanted leads and a sensing enabled DBS system, and as such, improve the efficiency of DBS programming in PD patients.

Methods: Level-based LFP recordings (OFF-medication) within the first two post-operative weeks in PD patients implanted with directional Sensight leads® and the Percept PC® neurostimulator in the Haga Teaching Hospital were retrospectively analysed. Time and frequency domain data were inspected for artefacts. From the individual theta (4-7 Hz), alpha (8-12 Hz), beta (13-35Hz) and gamma (≥36 Hz) bands the maximum power (Max.) and area under the curve above 1/frequency (AUC_flat) were extracted. The clinically chosen contact during MPR served as reference for all predictions. Machine learning models using AUC_flat features from frequency band combinations were evaluated using nested cross-validation. Two custom ranking methods, pattern based and decision tree, were developed for both beta band features individually. The predictive accuracy (Acc.) of the 1st and 2nd prediction combined was evaluated on a training and unseen test set, considering all data and subgroups based on amount of symptoms during MPR and amount of beta activity above 1/frequency. The ranking methods were additionally compared to an existing algorithm (DETEC). Sub-analyses were conducted to evaluate the impact of time, disease, and recording-related factors on the Acc., as well as the Acc. of segment-based LFP recording predictions.

Results: Recordings from 34 patients (68 subthalamic nuclei) were analysed. Artefacts did not overlap with frequencies of interest or were sporadic of nature. The machine learning model with the highest performance was a linear discriminant analysis combining raw beta and alpha features (AUC: Design: 0.86, Test: 0.69). For the 1st and 2nd predicted contacts combined, the two best performing ranking models were pattern based using AUC_flat (Acc. training set: 86.2%; test set: 100%) and decision tree using Max. (Acc. training set: 87.9%; test set: 100%). Correct pattern based (AUC_flat) predictions were more often 1st opposed to 2nd predictions than correct decision tree (Max.) predictions (Tr = 55.2%, T = 90% vs. Tr = 10.3%, T = 10%). Acc. obtained for subgroups based on amount of symptoms during MPR and amount of beta activity above 1/frequency were similar across all subgroups for the custom ranking methods. The Acc. of the DETEC algorithm was inferior to all custom ranking methods. No significant differences were found for the sub-analyses, with the exception that segment-based LFP predictions were significantly inferior to level-based LFP predictions.

Conclusions: This study demonstrates the feasibility of using level-based LFP recordings to predict the optimal stimulation contact in patients with PD. The best results were obtained using the pattern based (AUC_flat) and decision tree (Max.) custom ranking methods. For clinical implementation the decision tree (Max.) ranking method is expected to be favoured. Although prospective research is required to identify the true Acc. of the models in clinical practice, these results show potential to halve the required DBS programming time (only two out of four contacts require evaluation), and can thus improve DBS programming efficiency.

Table of contents

List of Abbreviations

PD	Parkinson's Disease
DBS	Deep Brain Stimulation
STN	Subthalamic Nucleus
IPG	Implantable Pulse Generator
MPR	Monopolar Review
LFP(s)	Local Field Potential(s)
METC-LDD	Medical Ethics Committee-Leiden, The Hague, Delft
1/f	One divided by the frequency (= aperiodic signal)
PCA	Principle Component Analysis
PLSR	Partial Least Squares Regression
L1	Lasso/L1 Regularisation
LDA	Linear Discriminant Analysis
K-NN	K-Nearest Neighbours
RF	Random Forest
AUC_flat	Area Under the Curve after removal of 1/f
Max.	Maximum
FOOOF	Fitting Oscillations and One-Over-F
PSD	Power Spectral Density
n/#	Number (of)
(ROC-)AUC	Area Under the Receiver Operator Curve
SD	Standard Deviation
T.	Test set
D.	Design set
Tr.	Training set
CI	Confidence Interval

1. Introduction

1.1 Parkinson's disease

Parkinson's disease (PD) is the second most common age-related neurodegenerative disorder worldwide (1). The frequency of PD increases with age, however, specific gene mutations can lead to young onset PD at ages between 21 and 45 (2). Nevertheless, most cases of PD are sporadic and not hereditary. The precise cause of PD currently remains unknown, however, a combination of genetics and exposure to one or more unknown environmental factors are expected to trigger the disease (3).

In PD, the protein α-synuclein accumulates in Lewy bodies and Lewy neurites leading to cell death (4). This degeneration mainly takes place in the nigrostriatal dopaminergic system of the basal ganglia (4). This system is responsible for the production of dopamine, the neurotransmitter for transmitting signals which normally lead to the initiation of smooth and purposeful movements. An impaired nigrostriatal dopaminergic system causes a loss of dopamine production, which causes abnormal nerve firing patterns resulting in impaired movement (3).

However, the neurodegeneration caused by PD is not limited to the nigrostriatal pathway. The disease process of PD is multifocal and heterogenic, therefore specific motor and non-motor symptoms vary between individuals. Bradykinesia (slowness), rigidity (stiffness) and tremor are the most common motor symptoms of PD. Next to these key motor symptoms, other motor symptoms such as gait dysfunction, freezing of gait, postural instability, speech difficulties, swallowing impairments and autonomic disturbances can been seen in PD. Non-motor symptoms can include sensory alterations, mood disorders, sleep dysfunction, cognitive impairment, dementia and others (5).

Diagnosing PD can be a difficult process as the disease has a gradual onset which is similar to the onset of several other forms of parkinsonism. According to the diagnostic criteria of the International Parkinson and Movement Disorder Society the diagnosis of PD can only be made when bradykinesia is present in combination with either rigidity, tremor or both (6). This diagnosis can be strengthened by the presence of supportive criteria such as responsiveness to dopaminergic therapy. However, absolute exclusion criteria and red flags, which eliminate or decrease the likelihood of PD, are also important to investigate (6).

1.2 PD therapy options

Over the last 25 years the prevalence of PD has doubled. In 2019 global estimates indicated that PD affected over 8.5 million people (7). Additionally, disability and death due to PD are increasing faster than for any other neurological disorder worldwide. Nonetheless, currently, only symptomatic treatments exist.

The primary, and traditionally most effective, treatment of PD consists of dopaminergic therapy often including a combination of Levodopa and Carbidopa (8, 9). Levodopa is a natural precursor of dopamine that can pass through the blood brain barrier. By combining Levodopa with Carbidopa early conversion to dopamine outside the brain is prevented, leading to more efficient dopaminergic therapy (8, 9).

An alternative to dopaminergic therapy is the use of dopamine agonists. These substances mimic dopamine effects. However, dopamine agonists have shown to be less effective than Levodopa (9, 10). As the effect of dopamine agonists lasts longer than that of Levodopa it is often used as an additional therapy to smoothen the possible fluctuations in the effect of Levodopa (9). Another medicinal therapy option includes MAO-B inhibitors and/or COMT inhibitors which reduce the activity of specific enzymes that break down dopamine in the brain, prolonging the neuronal dopamine exposure time (9, 10). Further medicinal therapy options include amantadine (to limit levodopa-induced dyskinesia

and reduce PD symptoms), anticholinergics (to help reduce tremor) and adenosine (A2A) receptor antagonists (to increase response to dopamine and allow more natural dopamine release) (9, 10).

However, as mentioned before, medicinal therapies often lead to fluctuations during the day, referred to as on-and-off or wearing-off effects. Next to this, all of these medicinal treatments have side-effects. Due to progression of PD, increased dosages of pharmaceutical treatments are required over time, which can cause an increase of these side-effects (11, 12). Furthermore, for a proportion of PD patients, symptoms cannot be adequately controlled by medication (13).

To overcome these drawbacks, over the last two decades, deep brain stimulation (DBS) in the subthalamic nucleus (STN) has emerged as an effective additional treatment for patients with PD. It has the potential to reduce therapy fluctuations, improve motor outcomes and/or allow a decrease in medication dosages (13, 14). DBS can be used to stimulate the STN (one of the basal ganglia), which in turn, through a certain pathway, leads to an increase of stimulation towards the (motor-)cortex. An overview of the effect of PD and DBS on the basal ganglia pathways is provided in Figure 1.

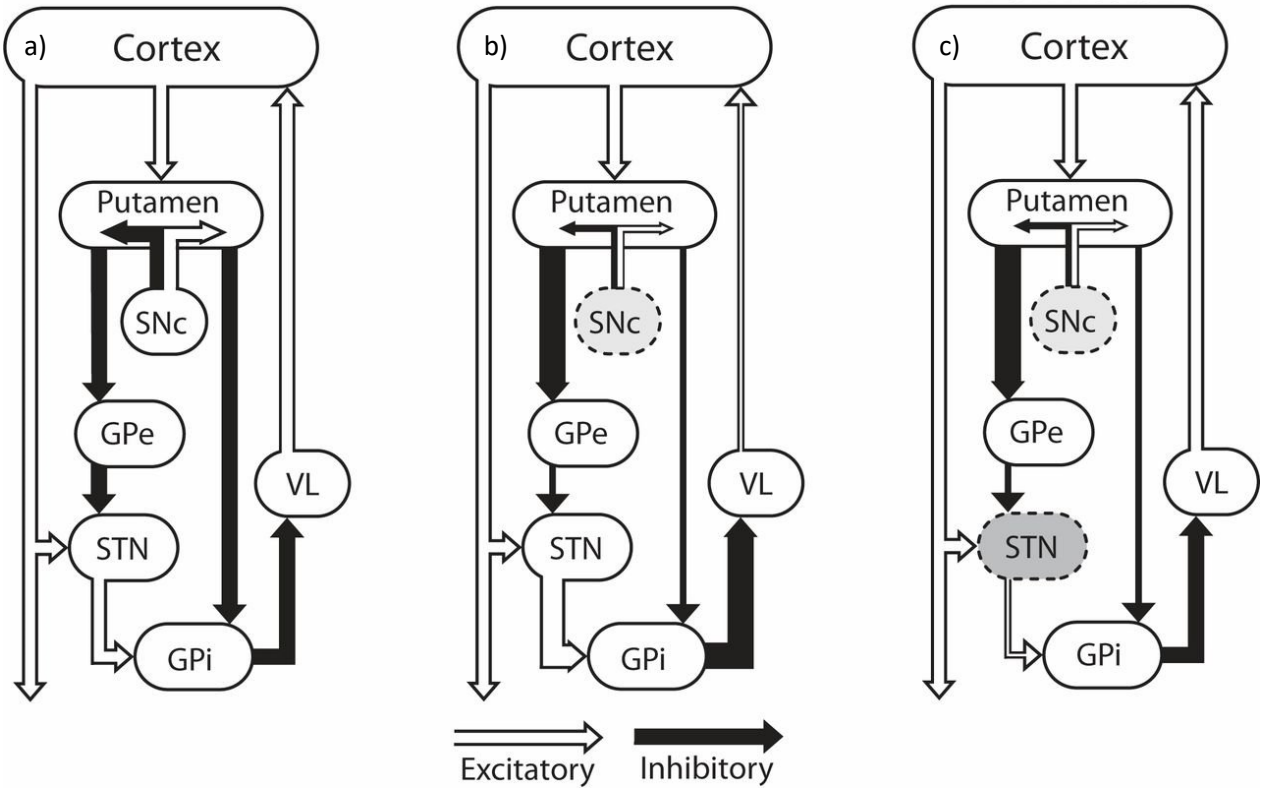


Figure 1. Overview of basal ganglia pathways in **a)** healthy subject, **b)** subject with Parkinson's disease, and **c)** subject with Parkinson's disease and deep brain stimulation within the subthalamic nucleus (15).

1.3 Deep brain stimulation

The origin of DBS lies within the ablative surgeries such as thalamotomies and pallidotomies (16). During these surgeries it was identified that high frequency stimulation could achieve similar effects as the ablative treatments for patients with PD (17). However, as these surgical interventions became less popular with the introduction of Levodopa in 1969, the possibilities of this high frequency stimulation were never thoroughly researched (18). Only when the drawbacks of Levodopa, such as motor fluctuations and Levodopa-induced dyskinesias, became more prominent and advances took place in the field of stereotactic surgery the possibility of applying DBS revived. Opposed to ablative surgery the effects of DBS are generally reversible by turning off the stimulation, making it a safer and potentially more effective alternative (19).

In DBS chronic stimulation is applied within the brain using implanted electrodes. The effects of the stimulation depend on the anatomical location of the electrodes. Therefore, DBS can be used for a variety of neural disorders such as PD, essential tremor, dystonia and possibly also epilepsy, obsessive compulsive disorder and major depression (20). As mentioned earlier, for PD the electrodes are most often implanted within the STN but the internal globus pallidus is a possible alternative (21).

Traditionally, DBS systems use four-contact stimulating electrodes (i.e. leads) that are stereotactically implanted in a certain target area within the brain. These leads are most often placed bilaterally and connected to an implantable pulse generator (IPG), a pacemaker-like device implanted within the chest wall, via a subcutaneous wire. Recently, directional leads, such as the Medtronic Sensight lead® or Boston Scientific Vercise Cartesia lead®, have increased in popularity. These leads consist of 2 ring electrodes, one above and one below two segmented electrodes. Each segmented electrode consists of 3 individual 120 degrees electrode segments, see Figure 2 (22). Segmented electrodes allow current steering, which can be useful in overcoming side-effects caused by DBS (23).

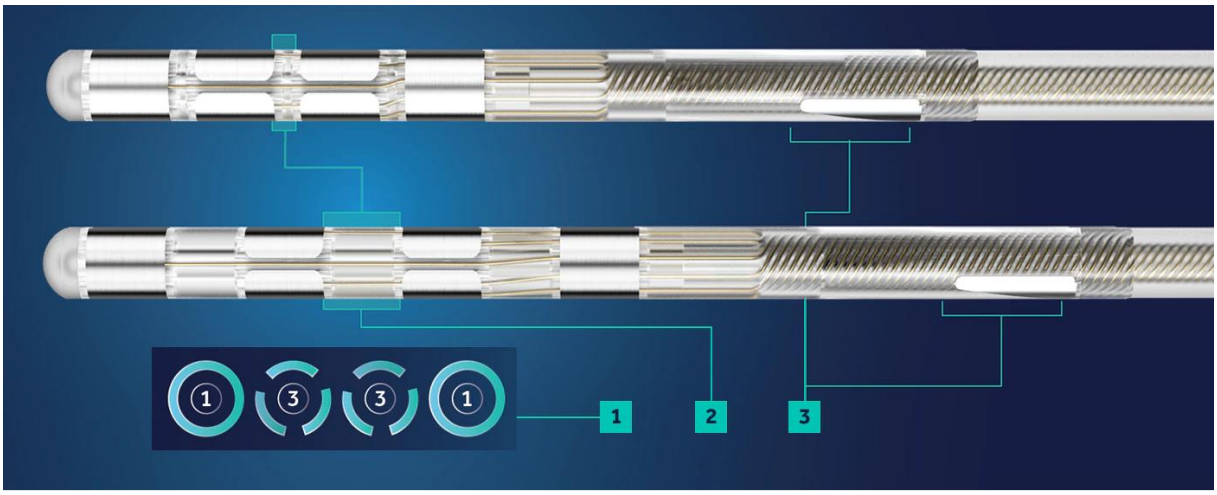


Figure 2. Medtronic Sensight directional lead®, including 4 stimulation levels, **1)** the middle two levels contain 3 individual 120 degrees segments, **2)** with either 1.5mm or 0.5mm spacing between levels, **3)** including completely insulated orientation markers to guide directional programming (24).

After lead and IPG implantation the neurologist uses a handheld device to wirelessly communicate with the IPG, allowing adjustment of the stimulation parameters (21). An important element of effective DBS, next to a precise implantation, is appropriate DBS programming. The programming of DBS starts by identifying the single most ideal stimulation contact. Currently, this contact selection is often achieved by performing a monopolar review (MPR) (25). During MPR the effects and side-effects are tested for each individual contact by gradually increasing the stimulation current. After this evaluation, the contact with the largest therapeutic window (window between the stimulation current that produces adverse effects and the stimulation current required to produce a beneficial effect) is initially chosen for chronic stimulation. Next, other parameters such as stimulation frequency, stimulation pulse width and stimulation amplitude are set (26).

This contact selection process is, however, time-consuming (20-30 minutes per hemisphere), requires highly trained personnel and can be exhausting and uncomfortable for patients (27). This is especially the case for directional DBS leads, which often contain at least 8 (instead of 4) individual contacts, doubling the required programming time (23). Furthermore, MPR can be influenced by several confounders such as patient fatigue, the 'stun' effect caused by lead implantation and the fact that several therapeutic effects of DBS have a latency period (28).

1.4 Local field potentials

A possible solution to these problems may lie within the measurement of local field potentials (LFPs). LFPs are transient electrical signals generated in nervous tissues by the summed and synchronous electrical activity of individual neurons (29). These signals can be recorded using an extracellular electrode at any specific location within the brain and can provide direct insight into, for instance, the functioning of basal ganglia (29, 30).

LFPs are generally analysed and interpreted within the frequency domain. LFP activity within the beta-frequency band (13-35 Hz) has been shown to correlate well with PD symptoms such as rigidity and bradykinesia and can be modulated by several PD therapies (e.g. medication/DBS) (31). Suppression of the beta-frequency band has been shown to correspond with motor improvement in PD patients (32). LFP activity within other frequency bands such as the theta (4-7 Hz), alpha (8-12 Hz) and gamma (36-200 Hz) bands as well as high frequency oscillations (>200 Hz) have also shown potential connections to certain PD symptoms and/or disease states (33).

Figure 3. Percept PC® internal pulse generator in combination with the patient remote and clinical programming interface which allow the recording of various local field potential measurements (34).

In November 2020 the Medtronic Percept PC® IPG was introduced, see Figure 3. This device has the ability to record LFPs within a frequency range of 0 to 96.68 Hz via the macroelectrodes of the implanted DBS leads (34). The Percept PC® has multiple recording modes enabling in clinic as well as at home LFP recordings. A summary of all Percept PC® recording modes has been provided by Thenaisie et al. (2021) (35).

In the past in clinic LFP recordings have shown the potential to aid the optimal stimulation contact selection during DBS programming (13, 35-39). However, this potential has not yet been thoroughly

researched for LFP recordings performed using an implanted IPG. Furthermore, to date, no method for using LFP recordings during the initial contact selection has been implemented in clinical practice.

1.5 Thesis objective

The main objective of this study was to predict the optimal stimulation contact(s) based on LFP recordings performed using an implanted lead and sensing enabled IPG, and as such, improve the efficiency of DBS programming in PD patients. To do so, LFP recordings were retrospectively used to predict a single, multiple or a ranking of the optimal stimulation contact(s). Predictions were made using several machine learning models, custom ranking methods as well as by means of one existing ranking method. The contact chosen for stimulation by the clinician during MPR served as a reference for all predictions. As a secondary objective it was investigated whether the lack of symptoms during MPR or the lack of beta activity above 1/frequency (1/f) affects the predictive accuracy of the developed methods. Finally, several sub-analyses were conducted to evaluate the impact of time after surgery, tremor as main symptom, recording impedances, on the predictive accuracy, as well as the predictive accuracy of segment-based LFP recording predictions.

2. Methods

2.1 Study design

This concerns a retrospective study of data gathered as part of the clinical routine amongst patients implanted with the Percept PC® neurostimulator in combination with directional Sensight leads®, for the treatment of PD. Patients were implanted in the Haga Teaching Hospital between November 1st, 2021 and June 5th, 2023. Only those patients that did not object to the use of their data for research purposes and for which the required measurements were available were included in the study. Ethics approval was obtained from the regional Dutch medical ethical committee (METC-LDD).

2.2 Data acquisition

2.2.1 LFP measurements

Predictions for the optimal stimulation contact were made using LFP recordings obtained through a BrainSense Survey conducted in the OFF medication state within the first two weeks after surgery.

A Survey measures the differential power signal between all possible contact combinations available on a single directional Sensight lead®, including rings as well as segments. The results are measures of how similar or different two regions within a single hemisphere of the brain are. This measurement is always performed in the stimulation OFF mode with a sample frequency of 250 Hz and a total duration of approximately 90 seconds. For each channel the time domain data (approximately 21 seconds per channel) is converted to the frequency domain using a Fast Fourier Transform and 0.98 Hz bins. The resulting frequency signal ranges from 0 to 96.68 Hz. The Survey measurement is always conducted in two individual passes. The first pass includes all stimulation compatible LFP recording pairs (e.g. channels: 0-2, 0-3, 1-3), while the second pass includes all immediately adjacent LFP recording pairs (e.g. channels: 0-1, 1-2, 2-3) (40). Survey measurements need to be performed for each hemisphere separately. Furthermore, to collect data for the segments, an additional Survey tailored to the segments has to be performed.

The user interface (tablet) only displays LFP magnitude (micro volts peak (µvp)) compared to frequency (Hz). However, offline, both frequency domain and time domain data are available and can be downloaded as part of a .json file.

2.2.2 Reference contact

The contact chosen for chronic stimulation by the clinician during MPR in the OFF medication state at approximately 10 days post-operatively was used as a reference for all the LFP based predictions.

2.3 Data analysis

Matlab (version R2023a) was used to perform all data preprocessing and data analysis steps.

2.3.1 Data preprocessing

Available Survey measurements were visually inspected for potential artefacts in the frequency as well as time domain. For the frequency domain both the readily available Percept PC® frequency domain data as well as a custom conversion of the Percept PC® time domain data to the frequency domain were used. The custom time to frequency conversion was performed using Welch's method, employing a Hanning window with 256 samples and 50% overlap. The frequency domain data obtained through this custom conversion had a higher resolution (5288 samples) than the readily available Percept PC® frequency domain data (100 samples). However, it should be noted that the frequency domain data provided by the Percept PC® system generally contains fewer artefacts, as the system already applies a form of automatic artefact removal, see Appendix A, Figure 1. The readily available Percept PC® frequency domain data was used for all predictions.

The frequency bands considered here were the theta (4-7 Hz), alpha (8-12 Hz), beta (13-35 Hz) and gamma (≥36 Hz) frequency bands. From each frequency band two features were extracted per measurement channel. The first feature was the maximum power (Max.) within the respective frequency band, which is commonly used in clinical practice. The second feature was the area under the curve (AUC_flat) within the respective frequency band after removing the aperiodic 1/f component of the data by means of the existing FOOOF algorithm (41), see Figure 4.

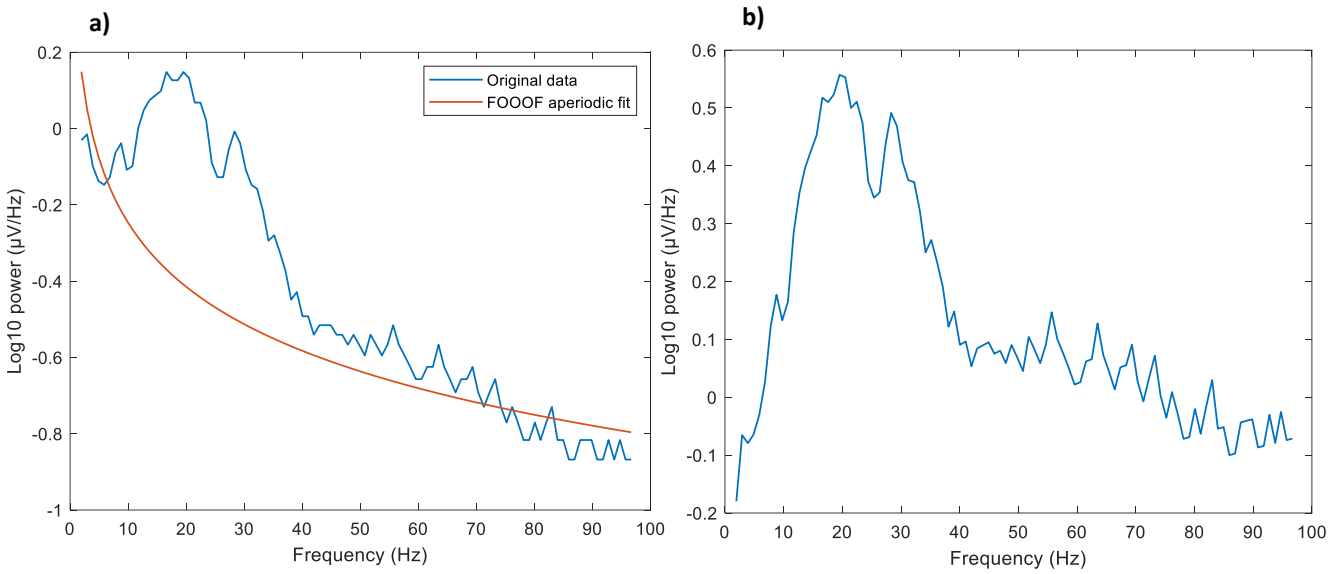


Figure 4. Example of removing aperiodic (1/frequency) information from data. **a)** Original data in logarithmic scale with corresponding aperiodic fit by the existing FOOOF algorithm. **b)** Resulting data in logarithmic scale after subtracting FOOOF-based aperiodic fit.

2.3.2 Machine learning models

To obtain the prediction of a single optimal stimulation contact several machine learning models in combination with several preprocessing methods were evaluated. These models were developed using either beta frequency features alone, a combination of beta with either theta, alpha or gamma band features and for a combination of features from all four frequency bands.

All machine learning models were evaluated using nested k-fold cross-validation. K-fold cross-validation splits the group into k parts, all of these parts are used as the test set once amongst the k repetitions of the analysis. This allows all the data to be used for training as well as testing whilst ensuring that testing is only performed using unseen data. In the nested form two k-fold cross-validations are applied within each other. This allows hyperparameter optimisation and thus model optimisation within the inner-loop. In general, k is set to 3, 5 or 10, taking into account the bias-variance trade-off (42). Here k was set to 10 for the outer-loop as this prevents excessive variance within the testing procedure whilst still minimising bias (42). As the complexity of one of the used classifiers is high, a high running time was expected. To minimise this k was set to 3 for the inner-loop.

The outer-loop was used to randomly split the data into a design and test set. Stratification was applied to ensure even distribution of the outcome classes amongst the design and test set. In case of the application of a preprocessing method the design set was used to fit this model. The preprocessing model was thereafter applied on the unseen test set. Within the inner-loop the design set was randomly split into a training and validation group whilst, once again, applying stratification. The training group was used to fit the machine learning models with different hyperparameter settings. The performance of each hyperparameter setting was evaluated on the unseen validation set. The median of the optimal hyperparameters for each model (i.e. the hyperparameter that resulted in the

smallest error for the validation sets), was thereafter chosen and used to train the model using the full design dataset within the outer-loop. The eventual model performance was evaluated by means of application of the model on the unseen test data and comparing the predicted outcome to the true outcome. An overview of the nested cross-validation method can be found in Figure 5.

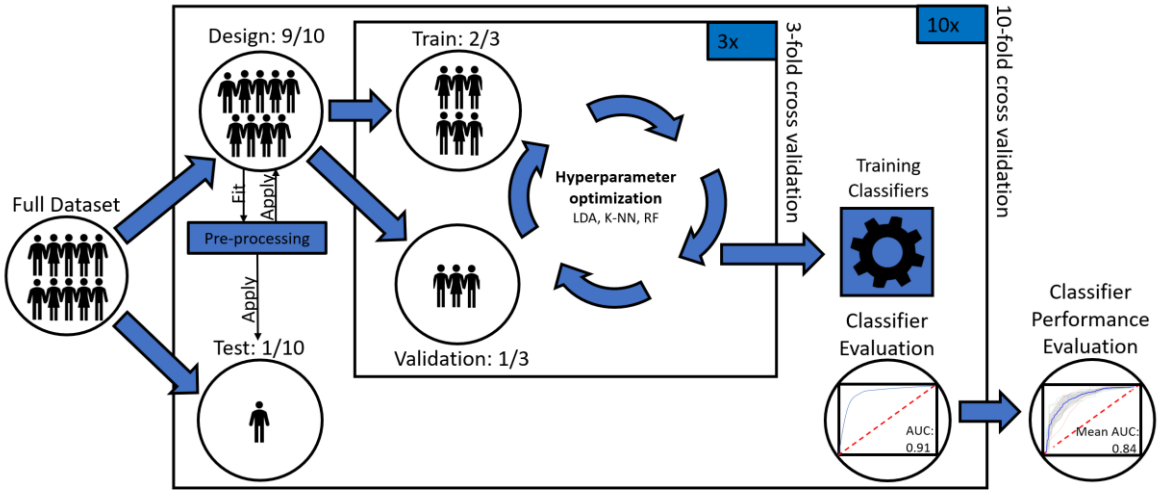


Figure 5. Overview of the nested cross-validation method, where the 10-fold outer-loop splits all available data into a design and test set and the 3-fold inner-loop split the design set into a train and validation set. The train and validation sets served for hyperparameter optimisation whilst the design and test set were used for fitting and evaluating the preprocessing and machine learning models.

The preprocessing methods evaluated here were: principal component analysis (PCA), partial least squares regression (PLSR) and lasso regularisation (L1). An in depth explanation of the characteristics and applied cut-offs for each method can be found in Appendix A, Table 1. Three machine learning models were evaluated in combination with each of the three preprocessing methods as well as without applying preprocessing. The machine learning models evaluated were: linear discriminant analysis (LDA), k-nearest neighbours (K-NN) and random forest (RF). An in depth explanation of the characteristics and hyperparameter ranges used for optimisation are provided in Appendix A, Table 2.

The model performance of the method with the best performance for all data was also trained and evaluated for three subgroups: a subgroup of data from hemispheres where PD symptoms were present during MPR, data of hemispheres showing beta activity in AUC_flat for at least one channel, and hemispheres with both of these aspects (symptoms during MPR as well as beta activity in AUC_flat for at least one channel).

2.3.3 Ranking models

To allow prediction of an optimal ranking of contacts instead of a single optimal contact several custom ranking methods were developed and evaluated together with an existing ranking model. For the design all data available up to March 13th, 2023 was used, this was considered as training data. Data that became available after this period and before June 5th, 2023 was used for evaluation of the predictive accuracy, and was considered as unseen test data. All ranking models were evaluated for all available data (training and test set separately) as well as for subgroups of hemispheres of patients with/without symptoms during MPR and hemispheres of patients with clear/little/no beta activity above 1/f. The amount of beta activity was evaluated by means of the AUC_flat feature. Hemispheres were considered as “clear beta above 1/f” when at least one channel showed an AUC_flat value above 0.6 µV. The label “little beta above 1/f” was given when one or more channels showed an AUC_flat value above 0.0 µV but below 0.6 µV. Finally, the label “no beta above 1/f” was appointed when all

channels had AUC_flat values below 0.0 µV. In the case of little or no beta above 1/f there remained an order in the power of the recording channels, therefore, ranking was still possible in these cases.

2.3.3.a Pattern based (custom)

For the pattern based ranking method the beta power for either the Max. or AUC_flat feature of all the bipolar measurements was captured within a single greyscale grid. This grid was used to visualise the distribution of the measured bipolar powers across the possible monopolar stimulation contacts. Furthermore, it was used as a method to predict the optimal stimulation contact, see Figure 6. For the latter, two methods were combined. The first method was calculating the average power per contact for all rows and columns that include this particular contact (e.g. contact 1 = average of 0-1, 1-2, 1-3).

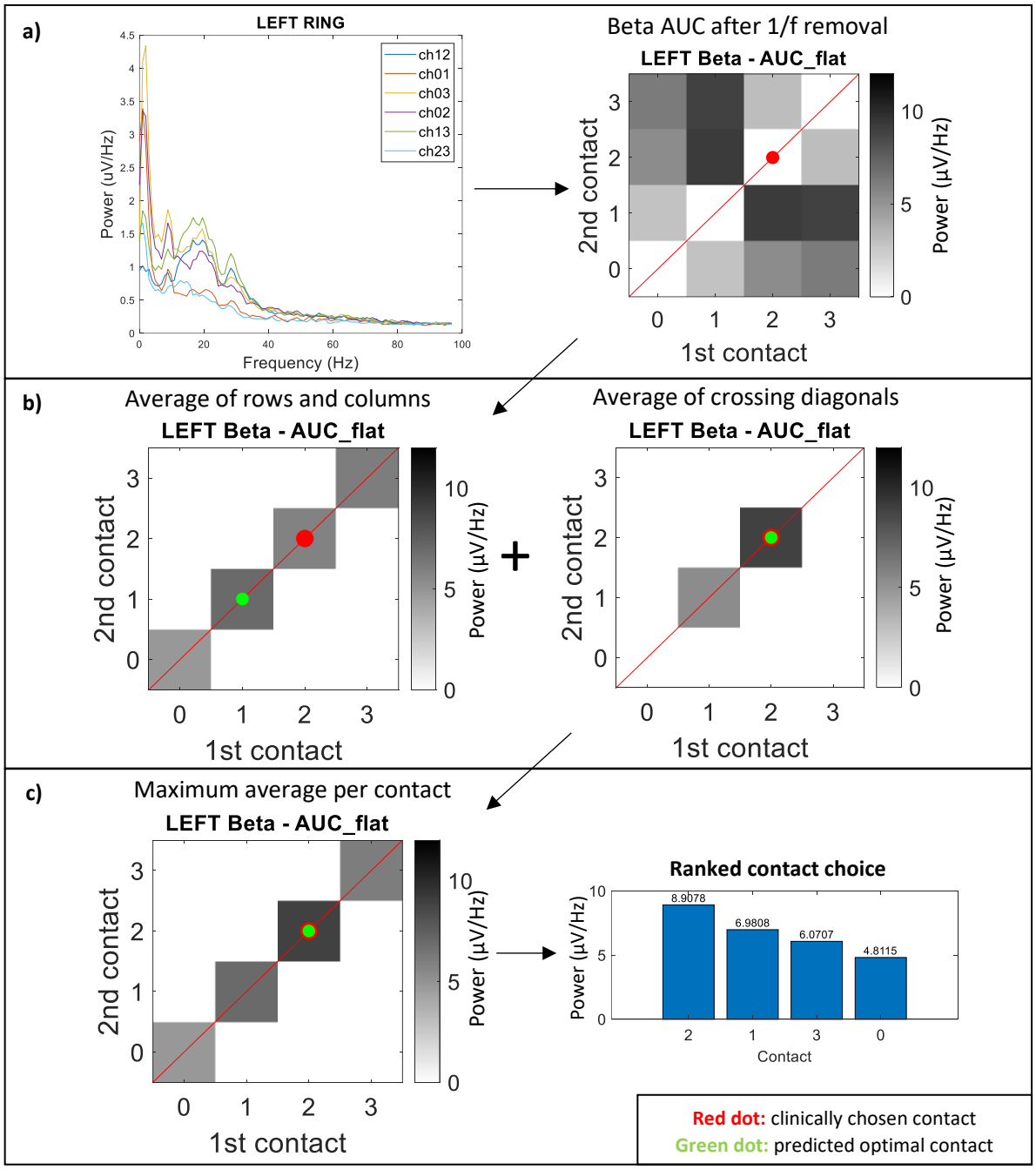


Figure 6. Working mechanism of pattern based ranking method using the area under the curve beta feature after removing the aperiodic component. **a)** greyscale grid is constructed using level-based LFP recordings. **b)** Two methods (averaging over rows/columns and averaging of crossing diagonals) are applied to obtain power per contact choice. **c)** Maximum of two methods is used to obtain and rank power value per contact choice.

The second method was to calculate the average power for the diagonals that cross each contact (e.g. contact 1 = 0-2), this is only available for contact 1 and 2. These methods were combined by taking the maximum possible result for each contact, resulting in a single value per contact choice. The contact with the highest value was thereafter considered as the most optimal stimulation contact. Second, third and fourth most optimal stimulation contacts were deducted by means of the same method.

2.3.3.b Decision tree (custom)

For the decision tree based ranking method two sets of decision trees were applied. In the first set the two or three bipolar channels with the lowest beta power were used to eliminate one or two stimulation contacts (see Appendix B, Part I for elimination decision trees). In the second set the two or three bipolar channels with the highest beta power were used to select two eligible stimulation contacts (see Appendix B, Part II for selection decision trees). By combining the results of both methods the single or two most eligible as well as the two or three least eligible stimulation contacts could be selected, see the example in Figure 7. This method was incorporated in a standalone MATLAB GUI using the App Designer functionality to allow easy implementation in clinical practice (Appendix B, Part III) and was evaluated for both the Max. and AUC_flat beta-band features individually.

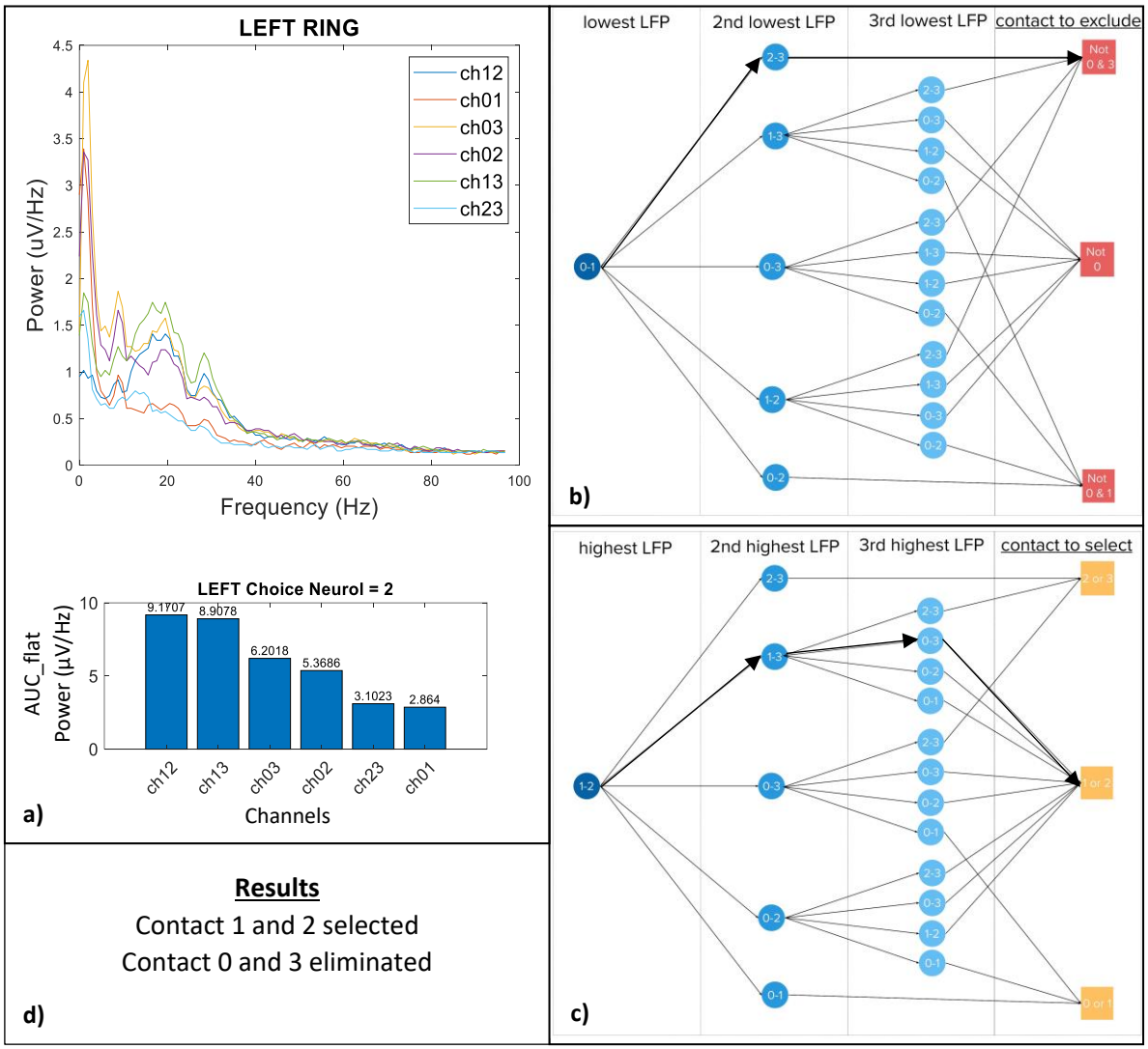


Figure 7. Working mechanism of decision tree ranking method using the area under the curve beta feature after removing the aperiodic (1/frequency) component. **a)** level-based LFP recordings and corresponding ranking in bar graph. **b)** route for contact elimination (not contact 0 and 3) using lowest (0-1) and 2nd lowest (2-3) ranked channels. **c)** route for contact selection (contact 1 or 2) using highest (1-2), 2nd highest (1-3) and 3rd highest (0-3) ranked channels. **d)** decision tree ranking results: possibly contact 1 or 2, not contact 0 and 3.

2.3.3.c DETEC algorithm (Strelow et al. (2022))

The DETEC algorithm is an existing algorithm published by J. Strelow et al. (2022) (38). This method calculates the weighted average beta (13-35Hz) power per monopolar contact based on bipolar recordings whilst correcting for the distance between the contacts using Equation(1). Before doing so the aperiodic 1/f component is removed by the FOOOF algorithm (41). Furthermore, an automated function determining the local maxima is implemented. If the algorithm detects a peak in at least one recording in a certain hemisphere, normalised activity at this peak is used for the analysis of this hemisphere. If no peak is detected, the algorithm uses normalised activity within the low beta range. Eventually, the result of the algorithm is a single "monopolar" power value per contact. The contact with the highest power is considered most eligible and the one with the lowest power is considered least eligible, this allows the ranking of all contacts. The DETEC algorithm was separately applied on the training and testing datasets to allow a better comparison with the pattern based and decision tree methods, however, for this existing method no additional training was required.

$$\mathrm{PSD}_{\mathbf{weighted}} = \frac{\sum_{i=1}^{n} PSD_i * \frac{1}{d_i}}{\sum_{i=1}^{n} \frac{1}{d_i}} \quad (1)$$

In Equation(1), PSD_i is the power spectral density (PSD) from bipolar recording ί of the n bipolar recordings involving the investigated contact. d_i is the distance between the centre of the investigated contact and its bipolar recording partner for bipolar recording ί (mm).

2.3.3.d Fixed ranking (reference)

The final ranking method that was evaluated was a fixed ranking. This was constructed based on the frequency of clinically choosing each contact in the Haga Teaching Hospital. This resulted in the following ranking method: contact #2 → contact #1 → contact #3 → contact #0. This method was not considered as a prediction method but was used to gain additional insight in the choices that are currently made for the chronic stimulation contact in the Haga Teaching Hospital.

2.4 Sub-analyses

Factors that may be of influence to the predictive accuracy of the evaluated methods were investigated in several sub-analyses. One sub-analysis was used to investigate whether there is a difference in predictive accuracy between groups where LFP recordings are performed within 5 days or at more than 5 days post-operatively. This analysis also allowed the comparison of recording LFPs at a time difference of either <5 days or ≥5 days from the moment of the clinical contact choice during MPR (10 days post-operatively). Another factor of which the effect on the predictive accuracy was evaluated is the effect of (one of) the patients dominant symptoms being tremor. To this end predictive accuracies were compared between subgroups with and without tremor as main symptom. A third sub-analysis was performed to evaluate whether differences in impedances affect the predictive accuracy of the developed ranking methods. This was done by comparing the results of STN for which all channels had bipolar impedances <5000 Ω to STN with one or more channels with bipolar impedances ≥5000 Ω.

Predictions based on segment LFP recordings were evaluated in a final sub-analysis. This was possible for cases where contact 1 or 2 were clinically chosen for chronic stimulation. These cases could only be included in the analysis if segment LFP Survey recordings were available for this specific case. Two techniques were evaluated to predict whether contact 1 or 2 should be preferred according to the segment LFP recordings. The first technique used the average power of all segment LFPs across one level to determine the optimal level (e.g. if average of 1a-1b, 1a-1c and 1b-1c is higher than average of

2a-2b, 2a-2c and 2b-2c then level 1 is chosen). The second used the level location of the segment LFP with the highest power (e.g. if highest among 1a-1b, 1a-1c, 1b-1c, 2a-2b, 2a-2c and 2b-2c is 1a-1b then level 1 is chosen). Both methods were tested using the Max. as well as the AUC_flat feature of the beta-frequency band. Results were additionally compared to the performance of the best performing level-based LFP method when considering the prediction of a single contact.

2.5 Statistical analysis

2.5.1 Machine learning models

The performance of machine learning models was evaluated across the design set as well as the unseen test set. In both cases the total area under the receiver operator curve (ROC-)AUC across all 10 outer cross-validation loops was identified in combination with the corresponding confusion matrix using the "perfcurve()"and "confusionchart()" Matlab functions, respectively. The lower limit for considering the model performance as promising, was an ROC-AUC of 0.7. This boundary was chosen as it can provide a first indication of whether predicting the stimulation contact based on a certain machine learning method is feasible (43). If this result is achieved it is expected that additional data and tuning can improve the model performance. False predictions are not expected to negatively influence the patient outcomes. If the required effect is not achieved by means of the predicted contact alternative contacts will be tested, similar to the current clinical practice. By means of a confusion matrix additional insight will be gained in which classes are being predicted correctly and what type of errors are being made.

2.5.2 Ranking models

To evaluate the performance of the ranking models the frequency of the 1st, 2nd, 3rd and 4th predicted contact being the clinically chosen contact was documented. In addition the combined frequency of the clinically chosen stimulation contact being either the 1st or 2nd predicted contact was calculated. This allows a more equal comparison between the decision tree method, where often two optimal contacts are selected simultaneously, and the other ranking models, where a single optimal contact can be selected. Furthermore, the 1st and 2nd prediction combined represent the percentage of cases where the required programming time can be halved, as only two out of four contacts require testing, which is a clinically relevant result. The percentage of the 1st, 2nd, 3rd and 4th ranked contact being the clinically chosen stimulation contact is provided per contact option as well, as this can provide additional insight in which classes are being predicted correctly more often than others.

2.5.3 Sub-analyses

For each of the factors potentially influencing the predictive accuracy of the level-based LFP predictions, the performance of the best performing prediction methods was compared across all sets of sub-analysis subgroups. The difference in average model performance between each set was tested on statistical significance by means of a N-1 chi squared test, which was recommended by Campbell et al. (2007) (44). For obtaining 95% confidence intervals (CI) the unpaired method recommended by Altman et al. (2000) was used (45).

For evaluation of the difference in the average model performance between the best performing level-based LFP prediction method (providing a single contact prediction) and the best performing segment-based LFP prediction method the McNemar's test was applied, as this required comparison of the paired proportions of correct predictions (46). Here, for obtaining 95% CI the paired method recommended by Altman et al. (2000) was applied (45).

In all statistical evaluations a p-value smaller than 0.05 was considered significant.

3. Results

3.1 Subject characteristics

Between November 1st 2021 and June 5th 2023 a total of 62 (124 STN) patients were implanted with the Percept PC® neurostimulator in combination with directional Sensight leads®. For 34 (68 STN) of these 62 patients an OFF medication BrainSense Survey measurement was available within the first two post-operative weeks. An overview of the patient characteristics is provided in Table 1.

Table 1. Patient characteristics (n=34)

Age (years)	63.1 (SD 7.9)
Gender (male)	23 (67.6%)
Disease duration (years)	9.8 (SD 4.8)
Time since DBS placement (months)	23.9 (SD 13.7)

Abbreviations: n: number of patients; SD: standard deviation; DBS: deep brain stimulation

For each patient the UPDRS-III scores per contact as well as the corresponding stimulation amplitude were documented during MPR (Appendix C, Table 1). The contact for each hemisphere chosen during MPR and the contact with the best performance based on the UPDRS-III scores are provided in Table 2. This table also shows the amount of time between surgery and the LFP measurement as well as the post-operative timepoint at which MPR took place.

Table 2. Measurement and stimulation information per patient and hemisphere.

Patient	Post-op time LFPs (days)	Post-op time MPR (days)	UPDRS-III best contact		Clinically chosen contact	
NL_007	10	10	LH:2	RH:NA	LH:2	RH:2
NL_008	10	10	LH:NA	RH:NA	LH:1	RH:1
NL_016	10	10	LH:0	RH:1	LH:2	RH:1
NL_017	8	8	LH:1	RH:1	LH:1	RH:1
NL_018	0	10	LH:NA	RH:NA	LH:2	RH:2
NL_019	0	10	LH:1	RH:2	LH:1	RH:2
NL_020	8	8	LH:1	RH:1	LH:1	RH:2
NL_022	10	10	LH:NA	RH:0/1	LH:2	RH:2
NL_023	7	7	LH:NA	RH:2	LH:2	RH:2
NL_024	10	10	LH:NA	RH:NA	LH:2	RH:2
NL_027	10	10	LH:2	RH:3	LH:1	RH:2
NL_028	10	10	LH:NA	RH:1	LH:3	RH:2
NL_031	10	10	LH:2/3	RH:2	LH:3	RH:2
NL_037	7	7	LH:1	RH:2	LH:1	RH:1
NL_042	10	10	LH:2	RH:2	LH:2	RH:2
NL_044	8	8	LH:2	RH:1/2	LH:2	RH:2
NL_045	2	10	LH:2	RH:NA	LH:2	RH:2
NL_046	5	8	LH:0/1	RH:NA	LH:2	RH:2
NL_048	10	10	LH:1/2	RH:2/3	LH:2	RH:2
NL_049	8	8	LH:2	RH:1	LH:2	RH:1
NL_051	8	8	LH:2/3	RH:1/2	LH:2	RH:1
NL_052	10	10	LH:NA	RH:NA	LH:2	RH:2
NL_053	10	10	LH:1	RH:2	LH:2	RH:2
NL_054	10	10	LH:3	RH:2	LH:3	RH:2
NL_055	10	10	LH:3	RH:3	LH:2	RH:2

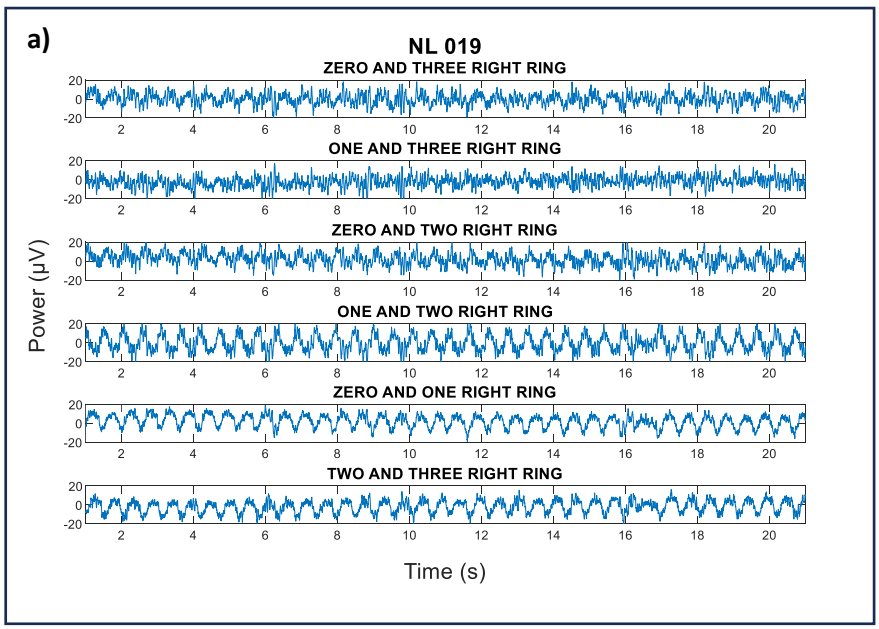
NL_056	8	8	LH:3	RH:2	LH:3	RH:2
NL_057	10	10	LH:NA	RH:1	LH:1	RH:1
NL_058	10	10	LH:NA	RH:NA	LH:2	RH:2
NL_059	10	10	LH:NA	RH:2	LH:1	RH:2
NL_065	8	8	LH:NA	RH:NA	LH:2	RH:2
NL_068	8	8	LH:2	RH:NA	LH:2	RH:1
NL_069*	10	10	LH:NA	RH:NA	LH:2	RH:2
NL_070	8	8	LH:NA	RH:NA	LH:2	RH:2
NL_071	10	10	LH:2	RH:NA	LH:1	RH:2

** For this patient no MPR UPDRS-III scores were available in the electronic patient documentation, however, MPR was performed and an initial contact choice remained available.*

Abbreviations: Post-op: post-operatively; LFPs: local field potentials; MPR: monopolar review; UPDRS: unified Parkinson's disease rating scale; LH: left hemisphere; RH: right hemisphere

3.2 Preprocessing

All data was visually inspected in the time as well as frequency domain to allow detection of potential artefacts. In 18 STN no segment recordings were available. Artefacts were present in 60 out of 68 STN for level-based LFP recordings and in 42 out of 50 STN for segment-based LFP recordings. The most common artefact found in the LFP recordings was a periodic artefact at a frequency below the frequencies of interest (<4 Hz), see Figure 8. After removing this artefact from the time domain data by means of a 4Hz high pass filter sporadic artefacts remained within other frequency bands for level-based recordings in 3 STN and for segment-based recordings in 6 STN, examples are provided in Appendix D, Figure 1. However, because of the sporadic nature of these artefacts, the fact that they did not occur across all channels within a single STN, and the expectation that these artefacts will also be present in the future clinical practice, no manual artefact removal was performed. Potential artefacts across the entire time domain recording above the 4Hz cut-off were present in 2 STN for level-based recordings and 1 STN for segment-based recordings. When looking at the frequency domain of this data a large peak in the recording power was present up to approximately 10Hz, see Appendix D, Figure 2. In signals without these potential artefacts these large power peaks were not present, see Appendix D, Figure 3. However, as it was difficult to confirm that these signals are of non-physiological origin no further artefact removal was applied.



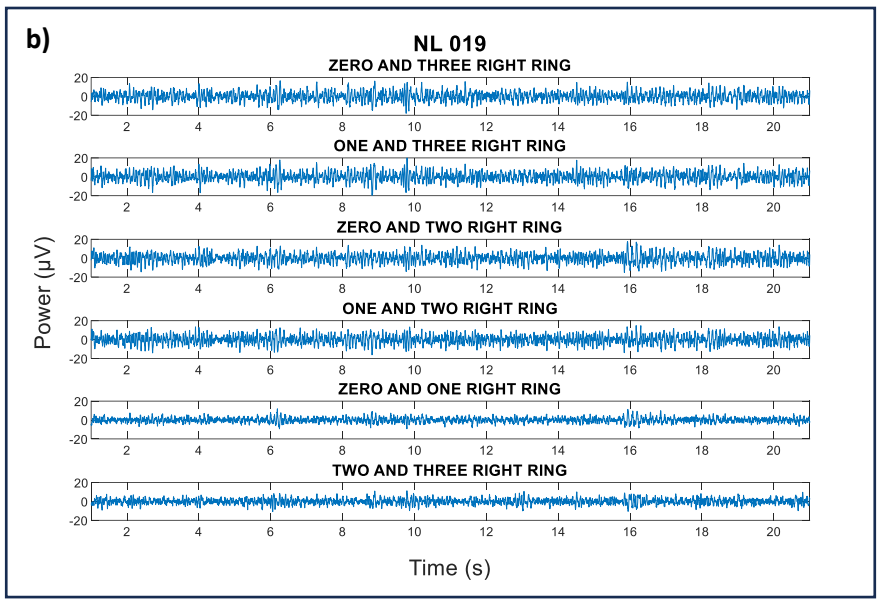


Figure 8. Example of time domain data from level-based LFP recordings across all 6 recording channels in right hemisphere of patient NL 019 **a)** before and **b)** after 4Hz high pass filter.

3.3 Machine learning models

For all machine learning results and corresponding hyperparameter settings see Appendix E, Table 1-2.

Among the different combinations of features, preprocessing techniques and machine learning methods that were evaluated here, the model with the highest overall ROC-AUC on the unseen test data was a LDA on the raw data of the beta and alpha frequency based AUC_flat features (D = 0.86, T = 0.69), see Figure 9. For this model the LDA type was set to "linear" 7 times and to "diaglinear" 3 times out of the 10 outer cross-validation loops. The second best performance was achieved by an RF model using raw data from a combination of beta and theta AUC_flat features (D = 1.00 / T = 0.64), however as the ROC-AUC for the design data was equal to 1.00 overfitting occurred. Furthermore, this model never predicted contact 3 as optimal stimulation contact, see Appendix E, Figure 1. The median number of trees applied in this model was 75 (range 25-100). The third best performance, without clear overfitting (i.e. without D = 1.00), was achieved by a LDA model in combination with PCA as preprocessing method on the AUC_flat features from all four frequency bands combined (D = 0.70, T = 0.63). Here contact 3 was, however, also never predicted as optimal stimulation contact. Additionally, only in 6 cases predictions were other than contact 2, see Appendix E, Figure 2. The applied LDA type was "linear" in all 10 outer cross-validation loops.

For the LDA method no preprocessing and preprocessing by means of PCA outperformed PLSR and L1 for all combinations of the beta and a single other frequency band when considering the ROC-AUC on the test data. For beta only and all frequency bands combined PCA outperformed the other preprocessing methods. For KNN models the combination with L1 showed the best results. However, KNN models generally performed worse than the LDA or RF models. For the RF models using combinations of the beta and a single other frequency band no preprocessing or PLSR showed the best performance. For beta only and all frequency band features combined the RF methods without preprocessing stood out. All RF based models achieved an ROC-AUC of 1.00 on the design data, indicating overfitting, this could not be resolved by decreasing the range for the hyperparameter.

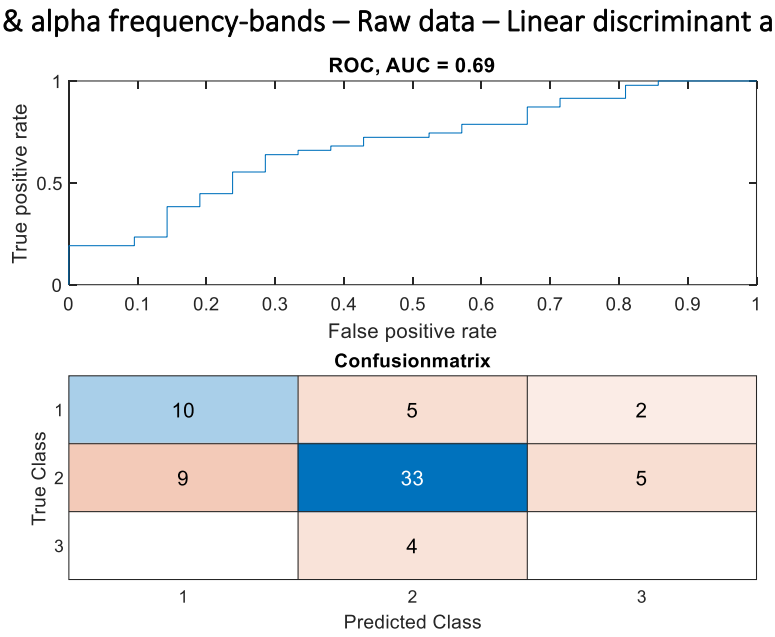


Figure 9. Receiver operator curve and confusion matrix for linear discriminant analysis model using the raw features from the beta as well as alpha band.

Training the best performing method (LDA on raw beta and alpha features) for data from hemispheres without stun effect during MPR and/or data from hemispheres showing beta activity in AUC_flat only did not improve the model performance compared to training on all data, see Table 3. The model using only data from hemispheres without stun effect during MPR did however, outperform the model that only used data from hemispheres with beta activity in AUC_flat and, especially, the models using only data from hemispheres without stun effect during MPR which also had beta activity in AUC_flat.

Table 3. Model performance of best method (linear discriminant analysis on raw feature data of beta and alpha frequency bands) when train on either data from hemispheres without stun effect during monopolar review, data from hemispheres with beta activity after 1/f removal or data from hemispheres with both of these aspects.

	No stun (n=42)		Beta in AUC_flat (n=61)		No stun & beta in AUC_flat (n=40)	
LDA	AUC	Type	AUC	Type	AUC	Type
Raw	D = 0.92 / T = 0.63	8x lin., 2x diag.	D = 0.86 / T = 0.61	7x lin., 3x diag.	D = 0.81 / T = 0.51	3x lin., 7x diag.

Abbreviations: LDA: linear discriminant analysis; n: amount of hemispheres included; AUC: area under the receiver operator curve; D: design set (90% of data); T: test set (10% of data); lin.: linear; diag.: diaglinear

3.4 Ranking models

Predictive accuracy of the 1st and 2nd prediction of the pattern based, decision tree and reference (DETEC algorithm) ranking methods categorised for either all data, with/without stun effect during MPR and clear/little/no beta activity above 1/f, is provided in Table 4. When considering all data the decision tree method using the Max. feature performed best on the training data (n=58), with 51/58 correct predictions (87.9%) for the 1st and 2nd predicted contact combined. The second best result was achieved by the pattern based method using the AUC_flat feature, with 50/58 correct predictions (86.2%). On the test set (n=10) both decision tree methods (AUC_flat and Max.) as well as the pattern based (AUC_flat) methods showed 100% predictive accuracy when considering the 1st and 2nd predicted contact. The existing ranking method (DETEC algorithm) showed the lowest predictive accuracy among the ranking methods when considering the 1st and 2nd predicted contact of all data for

the training as well as test set (Tr = 33/58 (56.9%), T = 7/10 (70.0%)). As the decision tree method often provides two contacts as prediction, correct predictions were often considered 2nd predicted contacts (AUC_flat: Tr = 43/58 , T = 9/10 ; Max.: Tr = 43/58 , T = 7/10). For the pattern based methods most correct predictions were 1st predicted contacts (AUC_flat: Tr = 32/58 , T = 9/10; Max.: Tr = 35/58 , T = 8/10) as this method provides a single optimal contact, Appendix F, Table 2-6.

Subdivisions were made for data with and without stun effect during MPR as well as for data with clear, little or no beta activity above 1/f. For both pattern based methods as well as both decision tree methods in general no decrease in performance was visible for data with stun effect during MPR or data with little or no beta activity above 1/f compared to data without stun effect during MPR and data with clear beta activity above 1/f. For the DETEC algorithm the performance on the training data did appear to decrease due to the presence of a stun effect or little to no beta activity above 1/f. This was not the case for the test set.

Table 4. Frequency of 1st or 2nd predicted contact being the clinically chosen contact per ranking model for all patients and subcategorised for with/without stun effect and clear/little/no beta above 1/f.

	AUC_flat		Max.		Reference
	Pattern based	Decision tree	Pattern based	Decision Tree	DETEC algor.
All data (Tr=58/T=10)	Tr = 50 (86.2%) / T = 10 (100%)	Tr = 49 (84.5%) / T = 10 (100%)	Tr = 43 (74.1%) / T = 9 (90.0%)	Tr = 51 (87.9%) / T = 10 (100.0%)	Tr = 33 (56.9%) / T = 7 (70.0%)
Without stun (Tr=38/T=2)	Tr = 34 (89.5%) / T = 2 (100%)	Tr = 33 (86.8%) / T = 2 (100%)	Tr = 31 (81.6%) / T = 2 (100%)	Tr = 32 (84.2%) / T = 2 (100%)	Tr = 25 (65.8%) / T = 1 (50.0%)
With stun (Tr=20/T=8)	Tr = 16 (80.0%) / T = 8 (100%)	Tr = 16 (80.0%) / T = 8 (100%)	Tr = 12 (60.0%) / T =7 (87.5%)	Tr = 19 (95.0%) / T = 8 (100%)	Tr = 8 (40.0%) / T = 6 (75.0%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 38 (88.4%) / T = 9 (100%)	Tr = 37 (85.4%) / T = 9 (100%)	Tr = 35 (81.4%) / T = 8 (88.9%)	Tr = 37 (86.1%) / T = 9 (100%)	Tr = 27 (62.8%) / T = 6 (66.7%)
Little beta above 1/f (Tr=9/T=0)	Tr = 7 (77.8%) / T = NA	Tr = 8 (88.9%) / T = NA	Tr = 3 (33.3%) / T = NA	Tr = 8 (88.9%) / T = NA	Tr = 4 (44.4%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 5 (83.3%) / T = 1 (100%)	Tr = 4 (66.7%) / T = 1 (100%)	Tr = 5 (83.3%) / T = 1 (100%)	Tr = 6 (100%) / T = 1 (100%)	Tr = 2 (33.3%) / T = 1 (100%)

Abbreviations: AUC_flat: area under the receiver operator curve after removing 1/frequency; Max.: maximum; Algor.: Algorithm; Tr: training set (size); T: test set (size); NA: not available; 1/f: 1/frequency (aperiodic signal)

Table 5 provides an overview of the predictive accuracy of the 1st and 2nd predicted contacts combined for each ranking model per clinically chosen contact. An overview of the predictive accuracy of the 1st, 2nd, 3rd and 4th predictions separately is provided in Appendix F, Table 7-11, for all models. As contact 0 was never chosen clinically there were no predictive accuracy results for this contact. Furthermore, as in none of the test data contact 3 was clinically chosen this predictive accuracy could not be evaluated for the test set either. When considering the prediction results per clinically chosen contact, clinically chosen contact 1 and 2 were predicted with similar accuracies. However, predictions for chosen contact 3 showed lower accuracies for all pattern based and decision tree methods. For the reference method (DETEC algorithm) predictive accuracy of chosen contact 3 was higher than for contacts 1 and 2 (75% vs. 59% and 46.7%, respectively). Nevertheless, contact 3 was clinically only chosen in four hemispheres whereas contact 1 and 2 were clinically chosen in 15 and 39 hemispheres, respectively.

Table 5. Frequency of 1st or 2nd predicted contact being the clinically chosen contact per ranking model for all patients and subcategorised for the clinically chosen contact.					
	AUC_flat		Max.		Reference
	Pattern based	Decision Tree	Pattern based	Decision Tree	DETEC algor.
Contact #3 (Tr=4/T=0)	Tr = 3 (75.0%) / T = NA	Tr = 3 (75.0%) / T = NA	Tr = 2 (50.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 3 (75.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 35 (89.7%) / T = 8 (100%)	Tr = 35 (89.7%) / T = 8 (100%)	Tr = 29 (74.4%) / T = 7 (87.5%)	Tr = 37 (94.9%) / T = 8 (100%)	Tr = 23 (59.0%) / T = 6 (80%)
Contact #1 (Tr=15/T = 2)	Tr = 12 (80.0%) / T =2 (100%)	Tr = 11 (73.3%) / T =2 (100%)	Tr = 12 (80.0%) / T = 2 (100%)	Tr = 14 (93.3%) / T = 2 (100%)	Tr = 7 (46.7%) / T =1 (50%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA
Abbreviations: AUC_flat: area under the receiver operator curve after removing 1/frequency; Max.: maximum; Algor.: Algorithm; Tr: training set (size); T: test set (size); NA: not available					

The results of the 1st and 2nd predicted contact combined for the fixed ranking (contact #2 → #1 → #3 → #0) are shown in Table 6. This method achieved a predictive accuracy on all data of 93.1% in the training set and 100% in the test set for the 1st and 2nd predicted contacts, illustrating that in the Haga Teaching hospital contact 2 and 1 are almost always chosen for stimulation. When comparing data with and without stun effect during MPR as well as data with clear or little beta activity above 1/f to data without beta activity above 1/f the fixed ranking technique shows better results for data with stun effect and without beta activity above 1/f, indicating that when the clinical choice is difficult contact 1 or 2 are always chosen. For the predictive accuracy of the 1st, 2nd, 3rd and 4th prediction separately for the fixed ranking model as well as the predictive accuracy of this method per clinically chosen contact, see Appendix F, Table 12 and Table 13.

Table 6. Frequency of 1st and 2nd predicted contact being the clinically chosen contact for fixed ranking #2→#1→#3→#0 subcategorised for with/without stun effect and clear/little/no beta above 1/f.					
All data (Tr=58/T=10)	Without stun (Tr=38/T=2)	With stun (Tr=20/T=8)	Clear beta above 1/f (Tr=43/T=9)	Little beta above 1/f (Tr=9/T=0)	No beta above 1/f (Tr=6/T=1)
Tr = 54 (93.1%) / T = 10 (100%)	Tr = 34 (89.5%) / T = 2 (100%)	Tr = 20 (100%) / T = 8 (100%)	Tr = 40 (93.0%) / T = 9 (100%)	Tr = 8 (88.9%) / T = NA	Tr = 6 (100%) / T = 1 (100%)
Abbreviations: Tr: training set (size); T: test set (size); 1/f: 1/frequency (aperiodic signal)					

3.5 Sub-analyses

Several sub-analyses were performed using the two best performing ranking methods, pattern based (AUC_flat) and decision tree (Max.), for all available data (train and test set combined). No significant difference in predictive accuracy of the 1st and 2nd predicted contact combined was found for either model when comparing predictive accuracy results of LFP recordings performed ≤5 days post-operatively to predictive accuracy results for LFP recordings performed > 5 days post-operatively, see Table 7. The percental difference was 13.3% (95% CI: -19.74% to 24.17%, p = 0.275) for the pattern based (AUC_flat) method and 11.7% (95% CI: -21.31% to 22.18%, p = 0.311) for the decision tree (Max.) method.

Table 7. Comparison of predictive accuracy for 1st&2nd predictions of pattern based (AUC_flat) and decision tree (Max.) methods between subgroups with LFP recordings ≤5 days / >5 days post-op.			
	LFP ≤5 days post-op (n=8)	LFP >5 days post-op (n=60)	Percental difference (95% CI, p-value*)
Pattern based (AUC_flat) 1st&2nd prediction	8 (100%)	52 (86.7%)	13.3% (95% CI: -19.74% to 24.17%, p = 0.275)
Decision tree (Max.) 1st&2nd prediction	8 (100%)	53 (88.3%)	11.7% (95% CI: -21.31% to 22.18%, p = 0.311)
**N-1 chi-square test* *Abbreviations: LFP: local field potential; post-op: post-operatively; n: number of subthalamic nuclei; CI: confidence interval; AUC_flat: area under the receiver operator curve after removing 1/frequency; Max.: maximum*			

Additionally, no significant difference in predictive accuracies was found when comparing results of LFP recordings performed in patients with tremor as main symptom to results for LFP recordings performed in patients with either bradykinesia or rigidity as main symptom, see Table 8. The percental difference was 4.2% (95% CI: -23.61% to 17.23%, p =0.687) for the pattern based (AUC_flat) method and 12.5% (95% CI: -12.56% to 23.63%, p =0.199) for the decision tree (Max.) method. In some patients without symptoms during MPR the main symptom remained unclear, in these cases the main symptom was considered to be either bradykinesia or rigidity and not tremor.

Table 8. Comparison of predictive accuracy for 1st&2nd predictions of pattern based (AUC_flat) and decision tree (Max.) methods between subgroups with tremor as main symptom or not.			
	Tremor main symptom (n=12)	Bradykinesia/ rigidity main symptom (n=56)	Percental difference (95% CI, p-value*)
Pattern based (AUC_flat) 1st&2nd prediction	11 (91.7%)	49 (87.5%)	4.2% (95% CI: -23.61% to 17.23%, p =0.687)
Decision tree (Max.) 1st&2nd prediction	12 (100%)	49 (87.5%)	12.5% (95% CI: -12.56% to 23.63%, p =0.199)
**N-1 chi-square test* *Abbreviations: n: number of subthalamic nuclei; CI: confidence interval; AUC_flat: area under the receiver operator curve after removing 1/frequency; Max.: maximum*			

Furthermore, no significant difference in predictive accuracies was found when comparing results of STN with one or more channels with bipolar impedance(s) ≥ 5000 Ω to results for STN with all impedances < 5000 Ω, see Table 9. The percental difference was 12.2% (95% CI: -5.31% to 38.85%, p = 0.211) for the pattern based (AUC_flat) method and 14.0% (95% CI: -3.15% to 40.56%, p = 0.127) for the decision tree (Max.) method. For the individual impedance measurements per channel for all STN with one or more impedance(s) ≥ 5000 Ω see Appendix G, Table 1.

Table 9. Comparison of predictive accuracy for 1st&2nd predictions of pattern based (AUC_flat) and decision tree (Max.) methods between subgroups with bipolar impedances ≤5000 Ω / >5000 Ω.			
	Impedance ≥ 5000 Ω (n=14)	Impedance < 5000 Ω (n=54)	Percental difference (95% CI, p-value*)
Pattern based (AUC_flat) 1st&2nd prediction	11 (78.6%)	49 (90.7%)	12.2% (95% CI: -5.31% to 38.85%, p = 0.211)

Decision tree (Max.) 1st&2nd prediction	11 (78.6%)	50 (92.6%)	14.0% (95% CI: -3.15% to 40.56%, p = 0.127)

**N-1 chi-square test*
Abbreviations: n: number of subthalamic nuclei; CI: confidence interval; AUC_flat: area under the receiver operator curve after removing 1/frequency; Max.: maximum

Finally, the predictive accuracy of segment-based LFP recordings for the optimal stimulation level was calculated and compared to the predictive accuracy for the optimal stimulation level of the best level-based ranking method providing a single optimal contact (pattern based using the AUC_flat feature).

For a total of 46 STN the clinically chosen contact was either contact 1 or 2 and segment LFP recordings were available. Level predictions based on the level with the highest segment LFP peak achieved higher predictive accuracies than level predictions based on the level with the highest segment LFP average, see Table 10. For prediction comparisons between all segment LFP predictions techniques as well as the best performing level-based ranking method that can provide a single optimal contact per patient hemisphere see Appendix G, Table 2.

Table 10. Predictive accuracy of 1st prediction based on either the level of the highest segment-based LFP peak or the level of the highest segment LFP average (both AUC_flat & Max.).

	Level highest segment LFP peak		Level highest segment LFP average	
	AUC_flat	Max.	AUC_flat	Max.
Correct 1st prediction # STN (%)	18 (39.1%)	18 (39.1%)	15 (32.6%)	17 (37.0%)

Abbreviations: LFP: local field potential; AUC_flat: area under the receiver operator curve after removing 1/frequency; Max.: maximum; # STN: number of subthalamic nuclei.

When comparing the best performing level prediction based on segment LFPs to the performance of the best level-based LFP method, the latter significantly outperformed the former (difference in predictive accuracy: -30.43% (95 CI: -47.92% to -9.57%, p = 0.006).

Table 11. Comparison of predictive accuracy of 1st prediction between the best segment-based LFP technique (level with highest segment LFP peak (AUC_flat)) and the best level-based LFP technique for single contact predictions (pattern based (AUC_flat)).

	Best segment[1] correct 1st prediction #STN (%)	Best segment incorrect 1st prediction #STN (%)	Total best level #STN (%) correct/incorrect
Best level[2] correct 1st prediction #STN (%)	12 (26.1%)	20 (43.5%)	32 (69.6%)
Best level incorrect 1st prediction #STN (%)	6 (13.0%)	8 (17.4%)	14 (30.4%)
Total best segment #STN (%) correct/incorrect	18 (39.1%)	28 (60.9%)	**McNemar's test:** Diff -30.43% (95% CI: -47.92% to -9.57%, p = 0.006)

[1] *Best segment: Level highest segment peak (Max.)*
[2] *Best level: Patter based (AUC_flat)*
Abbreviations: # STN: number of subthalamic nuclei; Diff: difference in predictive accuracy

4. Discussion

The aim of this research was to predict the optimal stimulation contact(s) based on LFP recordings, and as such, improve the efficiency of DBS programming in PD patients. To achieve this, several prediction methods based on LFP recordings obtained using the implanted leads and a sensing enabled DBS system were developed. Furthermore, the performance of these methods was evaluated by means of comparing the outcomes to the clinically chosen stimulation contact during MPR.

4.1 Ranking based predictions

4.1.1 Custom ranking methods

The most promising methods in this research were found to be two of the custom developed ranking methods. For the development of these custom ranking methods the level-based LFP recording data for each hemisphere was visualised using a greyscale grid. By means of this grid it was easier to visually interpret the LFP recording data in relationship to the clinically chosen contact. This allowed the development of two individual methods, one automatic (pattern based) and one manual (decision trees). The advantage of ranking based methods over methods providing a single contact prediction lies within the fact that side-effects can occur due to DBS. Side-effects may require deviation from the contact predicted based on LFP recordings and from the contact with lowest threshold for improving symptoms (theoretical best clinical contact). Ranking methods allow an informed alternative contact selection by providing 2nd, 3rd and 4th best contact options.

The decision tree (Max.) method achieved the best results among the ranking based methods for a combination of the 1st and 2nd prediction, a predictive accuracy of 87.9% and 100% on all training and testing data, respectively. The pattern based (AUC_flat) method achieved the second best result with a predictive accuracy of 86.2% and 100% on all training and testing data, respectively. However, while the pattern based (AUC_flat) method provides a single optimal contact prediction (1st prediction correct: Tr = 55.2%, T = 90.0%), the decision tree (Max.) method, due to the applied technique, often provides the two most optimal contacts, reducing the amount of correct 1st predictions (1st prediction correct: Tr = 10.3%, T = 10.0%). Furthermore, as the aperiodic (1/f) signal was fit and removed per individual channel before obtaining the AUC_flat feature, this feature is expected to be more robust against artefacts and potential differences in the amplitude of the background signal between channels than the Max. feature. Despite these advantages of the pattern based (AUC_flat) method the decision tree (Max.) method may be favoured for clinical implementation as the Max. feature is already determined in the current clinical practice whereas manual determination of the AUC_flat feature is currently challenging.

4.1.2 DETEC algorithm

The performance of the custom ranking methods was evaluated on the training data, test data and in comparison with an existing ranking method, the DETEC algorithm. In contrast to the custom developed methods, the DETEC algorithm corrects for the difference in distance between the contacts within the different bipolar channels. In the research by Strelow et al. (2022) the DETEC algorithm resulted in a predictive accuracy of 71.4% (n=14 STN) for the 1st and 2nd prediction combined (38). When applied to our data, however, this algorithm only achieved 56.9% for the combination of the 1st and 2nd prediction based on the training data (n=58 STN). The difference in these results may be due to the fact that they compared their results to the contact providing the best clinical efficacy and not the clinically chosen stimulation contact. There might be a difference between the contact providing the best efficacy and the clinically chosen stimulation contact as the possible presence of side-effects might make the examiner choose for another contact. In addition the clinically chosen contact might be examiner-dependent. Furthermore, the number of sides in which no symptoms were measured

(stun effect) could also influence the accuracy of the clinically chosen contact. In those cases the examiner might choose a contact based on previous experience, (presumed) anatomical location of the lead, or other factors. Other reasons for a dissimilarity in DETEC algorithm performance could be a difference in lead placement location during surgery, a difference in the preprocessing methods or the difference in the amount of STN that were researched.

Although the DETEC algorithm showed a performance inferior to the pattern and decision tree based ranking methods on the data from the current study, the concept of considering the difference in distance between the contacts for LFP recordings is understandable. From electroencephalography measurements it is a known concept that the distance between two electrodes influences the spatial resolution and amplitude of a bipolar measurement (47). However, as the precise influence of this difference in distance as well as the correct correction method currently remain unknown for LFP recordings this could not be taken into account in a standard manner and was therefore not included here. Furthermore, as the distance between contacts is expected to be similar for all patients, it would be expected that the effect on LFP recordings is similar between patients as well. Additionally, in this research, prediction methods were trained in a supervised manner. Therefore, the effect of distance between contacts is expected to have had minimal effect on the predictive accuracy of the evaluated methods.

4.1.3 Ranking without symptoms or beta activity above 1/frequency

One significant finding of this study is that predictions for patients with no symptoms during MPR or with little or no beta activity above 1/f achieved similar accuracies as those for patients with symptoms during MPR and patients with clear beta activity above 1/f. This suggests that, by means of the presented ranking methods, accurate contact choices can potentially be made for patients who are currently difficult to programme in the initial post-operative period. This may provide a solution to the current problems of early DBS programming in patients who exhibit a stun effect, and thus have little to no symptoms that can be used to inform MPR (48).

4.2 Multi-factor predictions

4.2.1 Machine learning models

The literature reviewed prior to this master thesis showed that a combination of multiple frequency factors within machine learning models may provide promising results when predicting the optimal stimulation contact based on LFP recording. For instance the research performed by A.T. Connolly et al. (2015) in 28 STN showed that a support vector machine based on theta and low-beta frequency features could achieve a predictive accuracy of 91% (49). Furthermore, research by S.S. Xu et al. (2022) in 92 STN showed that an RF model based on average power of evoked resonant neuronal activity, beta and anatomy obtained a mean predictive accuracy of 80% (50). Here, the best performance among the machine learning models was achieved by a LDA model using raw beta and alpha frequency data. The performance obtained by this model was an AUC of 0.69 for the unseen test data using nested 10- and 3-fold cross-validation. This result did not surpass the predefined limit for the AUC of 0.7, however, it was very close to meeting this criterion. Furthermore, considering the limited amount of available data and the high frequency of the reference contact being contact 2, this result is expected to allow improvement in the future.

In general the results obtained by the machine learning methods in this study appeared to be inferior to those obtained in the aforementioned literature. However, predictive accuracy often overestimates model performance in comparison to the ROC-AUC, as the former does not account for precision (positive predictive value) or recall (true positive rate). Because of this the performance scores obtained by the models in literature, using predictive accuracy, may have been higher than those

obtained here, using the ROC-AUC. This does however not necessarily indicate that these models achieved a more optimal, and in particular more generalisable, model performance than those developed here.

4.2.2. Frequency factors

Beta frequency features are considered to be most predictive of PD symptoms (51). Furthermore, according to literature, theta and gamma frequency features are considered more predictive of PD motor-symptoms than alpha frequency band features (51). However, for this to be true reliable recordings are expected to be essential. The fact that the best machine learning performance in this research was obtained by the combination of the beta and alpha frequency band features may, therefore, be due to the unreliability of the low (and possibly high) frequency recordings of the Percept PC®. These frequency bands are known to be susceptible to motion artefacts and may also be influenced by the relatively low sampling frequency (250 Hz) of the Percept PC® system (35). Nonetheless, the fact that a combination of frequency features outperforms predictions based on beta frequency features alone highlights that combining features from different frequency bands can improve the predictive accuracy of machine learning models when predicting the optimal stimulation contact in PD. This could indicate that more complex synchronisation patterns might explain the clinical symptoms better than a single frequency band. This is in accordance with the findings by A.T. Connolly et al. (2015), they showed that predictions based on a combination of theta and beta band features outperforms predictions based on features from the beta band alone (49).

Additionally, in literature a distinction between the low and high beta frequency band is often made (52). Several articles have shown that the low beta frequency shows better correlation with PD symptom fluctuations and/or levodopa treatment than the high beta frequency (53, 54). Nevertheless, others have shown that in some cases the high beta band shows better correlation than the low beta band (52), and in some cases only a high beta peak and no low beta peak is present (37). As it currently remains unclear which of these two bands is most useful for contact selection (55), and as the beta peak within this research was sometimes located at a frequency in between the low and high beta band, only features from the total beta band were used here.

4.3 Limitations

4.3.1 Number of subjects

Because of the retrospective nature of this research the required level-based LFP recordings were only available for a subset of 34 out of 62 patients within the first two post-operative weeks, OFF medication. Furthermore, segment-based LFP recordings were only performed in 25 of these 34 patients.

The fact that each patient had two individual leads implanted, one in each STN, made that for each patient two contact choices were made, resulting in 68 contact choices that could be used for the development of prediction methods. However, the fact that two predictions were evaluated per patient may have biased the results. Furthermore, although this is one of the largest samples so far, a sample size of 68 is still relatively small, especially when evaluating machine learning models (56). Additionally, the unbalanced subdivision among subgroups of this small dataset within the subgroup analyses of the machine learning and custom ranking methods as well as the secondary analyses may have decreased the reliability of the presented results.

4.3.2 Measurement and subject-related factors

Several LFP measurement and subject-related factors may have been of influence as well. For instance, patients with tremor as main symptom may show other patterns in LFP recordings, such as less beta and more theta activity, than those with bradykinesia or rigidity as main symptom (57).

Furthermore, tremor can potentially induce movement artefacts in LFP recordings (58). This may have led to different patterns and/or predictions in this subgroup of patients. The sub-analysis performed here did, however, not confirm a difference between these subgroups. Nevertheless, this research is not conclusive as only a small subset (12 out of 68 STN) had tremor as main symptom here.

Another factor that could potentially influence the results is the fact that a difference in impedances between recording channels may create a difference in the amount of background signal which influences the LFP recordings (i.e. LFP recordings can be biased by a difference in aperiodic 1/f signal between channels) (59). The sub-analysis performed here did not show a significant difference in predictive accuracy between subgroups with and without relatively high (≥5000Ω) bipolar impedances, for either the decision tree (Max.) or the pattern based (AUC_flat) method. Although this was not seen here, differences in impedances are expected to mainly influence the results for ranking based predictions made using the Max. feature and less for the AUC_flat feature, as for the latter the aperiodic 1/f signal is already removed.

LFP recordings and the prediction results based on these measurements can also be influenced by the amount of time between the surgery and the LFP recordings. For instance, the amplitude of the LFP recordings can be reduced in the first days and/or weeks after surgery due to the stun effect (60, 61). This effect may last for two weeks up to three months after surgery (62, 63). The occurrence of a stun effect was confirmed here as 28 out 68 STN (41.2%) showed no symptoms during MPR. In patients who did show symptoms during MPR a minimal microlesion effect could not be excluded. Furthermore, in 16 out of 68 STN (23.5%) little or no beta activity was present above 1/f within the LFP recordings. Of the 16 cases, with little or no beta activity above 1/f, 11 overlapped with those showing no symptoms during MPR. Nevertheless, the sub-analysis performed here did not show significant differences in predictive accuracy between predictions based on recordings within or more than 5 days after surgery. This may have been due to the small time difference between the two groups. However, this may also be due to the fact that prediction results are not drastically influenced by the occurrence of stun effect. The fact that prediction results obtained within subgroups without symptoms during MPR or with little or no beta activity above 1/f were similar to the results of subgroups with symptoms during MPR or clear beta activity above 1/f may provide evidence for this hypothesis.

Examples of other factors that were not evaluated in a sub-analysis but may have been of influence to the results obtained here are: the amount of movement performed by the patient during the LFP measurements; whether the patient was fully OFF medication during the LFP recordings; the type and intensity of symptoms present prior to (and after) lead implantation as well as the precise location of the lead within the STN.

4.3.3 LFP measurement reliability

The reliability of LFP recordings can be influenced by several factors. For instance, for the Percept PC® neurostimulator recordings of the lower frequency may be less reliable due to movement artefacts (35, 58). A comparison performed here showed that the readily available Percept PC® frequency domain data contains less artefacts than the manual time to frequency domain conversion. However, it remains unknown how the Percept PC® system filters or alternatively alters the frequency domain data. Furthermore, the custom time to frequency domain data has a higher signal resolution. Nevertheless, the signal resolution was thought to be of less importance than the occurrence of artefacts when obtaining the channel with highest Max. or AUC_flat. Additionally, only the readily available Percept PC® frequency domain data is available during DBS programming in the current clinical practice. Therefore, the readily available Percept PC® frequency domain data was used to perform all analyses within this study.

4.3.4 Reference method

The fact that the clinically chosen contact during MPR was used as the reference for the prediction techniques may have influenced the results. One of the reasons for this is the fact that the clinician may have been biased towards a certain contact choice. For instance, the fixed ranking method showed that in the Haga Teaching Hospital contact 2 or 1 are chosen in 93.1% of the training set and 100% of the test set. In part this is expected to be due to the planned lead implantation location during surgery. However, this may also have caused bias towards the prediction of contact 2 and 1 in all techniques, making it unclear whether the results presented here are generalisable for other centres. Furthermore, the motivation behind the clinical contact choice in patients without symptoms during MPR remains unclear and could have been influenced by the presence and pattern of side-effects.

Even in patients with symptoms, during the clinical contact choice, side-effects are taken into account, however, these are not reflected by LFP recordings. This may have led to the identification of incorrect or non-existent relationships between LFP recordings and clinical contact choices. Therefore, a reference contact which was chosen based on symptoms alone, for instance by using the contact with the best UPDRS-III score, may have shown better correlation with the LFP recordings. However, as 24 out of 68 included STN showed no symptoms during MPR this would have greatly limited the amount of available data. Additionally, in 9 STN two or more contacts were equally promising according to the UPDRS-III scores, whereas in clinical practice this may have been more nuanced (i.e. preference of patient or clinician due to particular improvements). Furthermore, when using only the UPDRS-III based contact choice we would not have gained the insight that in cases of no symptoms during MPR LFP recordings can still provide reliable predictions of the optimal stimulation contact.

As symptoms may return over time long-term contact choices may form a better representative of the true optimal contact. However, as there was a large variation in the time after surgery for the included patients and almost no LFP recordings are performed after the first two weeks post-operatively the long-term contact choice was considered to be sub-optimal as reference in this retrospective research.

4.3.5 Multi-class classification

One of the reasons for evaluating these 3 specific machine learning methods was the fact that multi-class classification with a categorical outcome variable was required. This made that several machine learning methods were not applicable or required a one vs. rest approach complicating the comparison with the currently evaluated models. As this was only a first evaluation of the feasibility of using machine learning models for the prediction of the optimal stimulation contact this was not attempted here. The three evaluated machine learning models, that vary in methodology and complexity, were chosen as a categorical multi-class prediction was required, which is not a standard option for all machine learning methods (64) and because the required model complexity remained unknown. However, other machine learning methods that were not evaluated, such as support vector machines, may result in a higher performance (49). One of the limitations of the machine learning models evaluated here, as well as of those evaluated with a similar aim in literature, is the fact that these methods only provide a single prediction for the most optimal stimulation contact. Whereas, in clinical practice, this potentially optimal stimulation contact may lead to undesired side-effects. Therefore, an optimal ranking of the stimulation contacts may be more useful in clinical practice. Although ranking methods exist in machine learning, these methods do not allow insight in the reasoning behind a certain ranking and/or are not easily tweaked or interpreted (65). To overcome these issues custom ranking methods were developed.

Another limitation of machine learning models is the risk of overfitting on the training data, which decreases the generalisability of the model. The risk for overfitting was relatively high in this research because of the limited amount of data, and especially the lack of equally distributed data among the possible classes. An attempt was made to limit overfitting and reduce training time by means of preprocessing techniques that allow dimensionality reduction (66). However, since it was unknown which features or feature combinations are of highest value for the predictions 3 different techniques as well as no preprocessing were evaluated in combination with each model. Nevertheless, all the RF models were subject to overfitting. This was proven by the fact that they achieved a predictive accuracy of 100% on the training data and a far lower predictive accuracy (≤64%) on the unseen test data. This overfitting did not decrease when decreasing the amount of features and/or number of trees. Therefore, the fact that overfitting occurred, was expected to be due to the relatively small amount of available data, indicating that the test set performance could increase by adding additional, more equally distributed, training data (67).

4.4 Future directions

4.4.1 Monopolar LFP recordings

The identification of the single optimal stimulation contact is complicated by the fact that all recordings are bipolar (38). Bipolar recordings only provide a difference in beta activity between the two included regions and not the absolute beta activity of a specific location. It is therefore expected that in the future, different recording approaches will be evaluated. For instance, monopolar LFP recordings may provide a more direct representation of the activity at an exact location among the contacts available for stimulation. Nevertheless, bipolar measurements have the advantage that common noise can be discarded. Therefore, a semi-monopolar approach, where there is still a reference but one common to all contacts within one hemisphere, may have the greatest potential. It is expected that a LFP recording option of this type will be added to the Percept PC® system in the future. However, as recording via the stimulating electrode is not possible during stimulation, a bipolar measurement configuration around the stimulating electrode is expected to remain for recordings during stimulation. This bipolar approach will also allow reduction of the stimulation artefact. As the at-home recordings, which are performed during stimulation, are expected to guide adaptive/closed-loop DBS in the future, a comprehensive understanding of the bipolar recording modalities remains important.

4.4.2 Directional leads

The directional Sensight leads® investigated here are an example of the current commercially available directional leads. Due to the segmentation of the middle two contacts current steering is enabled. The application of current steering has shown the potential to reduce side-effects caused by DBS (37). However, using a directional lead to its fullest potential requires the evaluation of additional contacts during MPR. As even leads with 32 contacts have been tested intra-operatively, (68) future use of similar devices in clinical practice would drastically increase the amount of contacts requiring evaluation. Currently, due to the limited amount of available time, segmented contacts are only tested individually in case limiting side-effects occur during level-based stimulation or when the results of DBS remain suboptimal after trying several level-based configurations over a longer period of time. The automation of the contact selection process by means of a prediction method based on LFP recordings could help reduce the amount of time required for DBS programming and thus allow the use of directional leads to their full potential. This would help decrease side-effects due to DBS and potentially even improve DBS efficacy (37). Nevertheless, the finding of this research that segment LFPs predictions are significantly inferior to level-based LFP predictions for predicting a stimulation level should be taken into account. This finding may provide evidence that future research on the

predictive accuracy of segment LFPs should focus on predicting the most optimal stimulation direction and not (only) the most optimal stimulation level.

4.4.3 Other neural disorders

In this research only patients with PD were considered. However, DBS is currently used for several other neuromuscular diseases such as dystonia and essential tremor. Furthermore, applications for epilepsy, obsessive compulsive disorder and major depression are currently being researched (20). To date, most of the research on potential physiomarkers to inform DBS programming has been performed for PD. However, DBS programming in other disorders may also benefit from the use of LFP recordings (20, 51). An example is the use of LFP recordings as predictor for the occurrence of an epileptic seizure (69). Information from LFP recordings is, however, not yet used for this purpose as there is currently an insufficient amount of clinical evidence on the predictive accuracy of LFP based physiomarkers (69). Nevertheless, as DBS amplitude allows quick activation or adjustment patients with sudden onset diseases such as epilepsy or certain subtypes of dystonia may potentially benefit more from predictions based on LFP recordings than patients with relatively slow changing diseases such as PD. Performing additional research on LFP based physiomarkers may therefore provide important insights for achieving better symptom control in patients with neural disorders other than PD.

4.4.2 Generalisability and prospective research

Future research on the topic of this study should focus on evaluation of the generalisability of the proposed techniques. This can be achieved by evaluating the performance of the methods on additional data from the Haga Teaching Hospital as well as from other clinics. Furthermore, an attempt should be made to improve the methods and models by means of prospective research and the use of the contact chosen based on UPDRS-III scores as a reference instead of the clinical contact choice (which may have been influenced by side-effects which are not captured by LFP recordings). For the machine learning methods additional training data is required to identify the true potential of the evaluated and other, non-evaluated, techniques. This additional data will also help identify the most promising combination of feature(s), preprocessing technique(s) and a machine learning method(s). Eventually, to properly assess generalisability for all methods, the performance of the most promising techniques should be evaluated on an external validation set.

For the most promising method feasibility of clinical application should be evaluated by performing prospective clinical research. This research should evaluate whether DBS efficacy achieved by using this new technique is non-inferior to the DBS efficacy achieved by means of the current clinical practice. Furthermore, the amount of time required for DBS programming should be compared. This could be achieved by means of a randomised controlled trial, preferably with a cross-over design. Within this research it should also be evaluated whether the predictive accuracy of the most promising technique improves when LFPs and the clinical contact choice are performed after the stun effect period (i.e. after LFP stabilisation). However, to do so, a more precise estimate of the moment of LFP stabilisation in time is required. In future research Survey and/or Setup measurements should, therefore, be repeated at multiple moments within the first 3 post-operative months, allowing more precise identification of the moment that LFP amplitudes stabilise (63).

5. Conclusion

By this research it was demonstrated that prediction of the optimal stimulation contact(s) in patients with PD using level-based LFP recordings is feasible and has the potential to improve DBS programming efficacy. The best results were obtained using the pattern based (AUC_flat) and decision tree (Max.) custom ranking methods. For clinical implementation the decision tree (Max.) ranking method is expected to be favoured. Prospective research is, however, required to identify the precise predictive accuracy of this technique in clinical practice as well as provide insight in the clinical limitations and potential of the developed model. Furthermore, challenges such as the limited amount of available data and bias in the reference method should be addressed in further evaluations. Despite these considerations, the advancements presented here show the potential to halve the required DBS programming time, which could play an important role in improving DBS programming efficiency as well as reducing patient discomfort.

6. References

1. Kalia LV, Lang AE. Parkinson's disease. Lancet. 2015;386(9996):896-912.
2. Kurian MA, Abela L. DNAJC6 Parkinson Disease. In: Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle; 1993.
3. Parkinson's disease : hope through research. Bethesda, Md: National Institute of Neurological Disorders and Stroke, National Institutes of Health; 2006.
4. Dickson DW. Parkinson's disease and parkinsonism: neuropathology. Cold Spring Harb Perspect Med. 2012;2(8).
5. Olanow CW, Klein C, Schapira AHV. Parkinson's Disease. In: Jameson JL, Fauci AS, Kasper DL, Hauser SL, Longo DL, Loscalzo J, editors. Harrison's Principles of Internal Medicine, 20e. New York, NY: McGraw-Hill Education; 2018.
6. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015;30(12):1591-601.
7. Parkinson disease: a public health approach. Technical brief. Geneva: World Health Organization; 2022. Contract No.: CC BY-NC-SA 3.0 IGO.
8. Fahn S. The medical treatment of Parkinson disease from James Parkinson to George Cotzias. Movement Disorders. 2015;30(1):4-18.
9. Reich SG, Savitt JM. Parkinson's Disease. Med Clin North Am. 2019;103(2):337-50.
10. Fox SH, Katzenschlager R, Lim S-Y, Barton B, de Bie RMA, Seppi K, et al. International Parkinson and movement disorder society evidence-based medicine review: Update on treatments for the motor symptoms of Parkinson's disease. Movement Disorders. 2018;33(8):1248-66.
11. Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021;397(10291):2284-303.
12. Dexter DT, Jenner P. Parkinson disease: from pathology to molecular disease mechanisms. Free Radic Biol Med. 2013;62:132-44.
13. Ince NF, Gupte A, Wichmann T, Ashe J, Henry T, Bebler M, et al. Selection of optimal programming contacts based on local field potential recordings from subthalamic nucleus in patients with Parkinson's disease. Neurosurgery. 2010;67(2):390-7.
14. Pintér D, Járdaházi E, Balás I, Harmat M, Makó T, Juhász A, et al. Antiparkinsonian Drug Reduction After Directional Versus Omnidirectional Bilateral Subthalamic Deep Brain Stimulation. Neuromodulation. 2023;26(2):374-81.
15. Vitek JL, Johnson LA. Understanding Parkinson's disease and deep brain stimulation: Role of monkey models. Proc Natl Acad Sci U S A. 2019;116(52):26259-65.
16. Ohye C, Kubota K, Hongo T, Nagao T, Narabayashi H. Ventrolateral and Subventrolateral Thalamic Stimulation: Motor Effects. Archives of Neurology. 1964;11(4):427-34.
17. Hassler R, Riechert T, Mundinger F, Umbach W, Ganglberger JA. PHYSIOLOGICAL OBSERVATIONS IN STEREOTAXIC OPERATIONS IN EXTRAPYRAMIDAL MOTOR DISTURBANCES. Brain. 1960;83(2):337-50.
18. Cotzias GC, Papavasiliou PS, Gellene R. Modification of Parkinsonism — Chronic Treatment with L-Dopa. New England Journal of Medicine. 1969;280(7):337-45.
19. Benabid AL, Pollak P, Louveau A, Henry S, de Rougemont J. Combined (Thalamotomy and Stimulation) Stereotactic Surgery of the VIM Thalamic Nucleus for Bilateral Parkinson Disease. Applied Neurophysiology. 1988;50(1-6):344-6.
20. Herrington TM, Cheng JJ, Eskandar EN. Mechanisms of deep brain stimulation. J Neurophysiol. 2016;115(1):19-38.
21. Udupa K, Chen R. The mechanisms of action of deep brain stimulation and ideas for the future development. Progress in Neurobiology. 2015;133:27-49.
22. Masuda H, Shirozu H, Ito Y, Fukuda M, Fujii Y. Surgical Strategy for Directional Deep Brain Stimulation. Neurol Med Chir (Tokyo). 2022;62(1):1-12.
23. Pollo C, Kaelin-Lang A, Oertel MF, Stieglitz L, Taub E, Fuhr P, et al. Directional deep brain stimulation: an intraoperative double-blind pilot study. Brain. 2014;137(7):2015-26.

24. Medtronic. SENSIGHT™ DIRECTIONAL LEADS: medtronic.eu; 2021 [Available from: https://europe.medtronic.com/xd-en/healthcare-professionals/products/neurological/deep-brain-stimulation-systems/sensight-lead.html.
25. Volkmann J, Moro E, Pahwa R. Basic algorithms for the programming of deep brain stimulation in Parkinson's disease. Mov Disord. 2006;21 Suppl 14:S284-9.
26. Koeglsperger T, Palleis C, Hell F, Mehrkens JH, Bötzel K. Deep Brain Stimulation Programming for Movement Disorders: Current Concepts and Evidence-Based Strategies. Frontiers in Neurology. 2019;10.
27. Thompson JA, Lanctin D, Ince NF, Abosch A. Clinical Implications of Local Field Potentials for Understanding and Treating Movement Disorders. Stereotactic and Functional Neurosurgery. 2014;92(4):251-63.
28. Mestre TA, Lang AE, Okun MS. Factors influencing the outcome of deep brain stimulation: Placebo, nocebo, lessebo, and lesion effects. Movement Disorders. 2016;31(3):290-8.
29. Lashgari R, Li X, Chen Y, Kremkow J, Bereshpolova Y, Swadlow HA, et al. Response properties of local field potentials and neighboring single neurons in awake primary visual cortex. J Neurosci. 2012;32(33):11396-413.
30. Einevoll GT, Kayser C, Logothetis NK, Panzeri S. Modelling and analysis of local field potentials for studying the function of cortical circuits. Nature Reviews Neuroscience. 2013;14(11):770-85.
31. Brittain J-S, Brown P. Oscillations and the basal ganglia: Motor control and beyond. NeuroImage. 2014;85:637-47.
32. Kühn AA, Tsui A, Aziz T, Ray N, Brücke C, Kupsch A, et al. Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity. Experimental Neurology. 2009;215(2):380-7.
33. Yin Z, Zhu G, Zhao B, Bai Y, Jiang Y, Neumann W-J, et al. Local field potentials in Parkinson's disease: A frequency-based review. Neurobiology of Disease. 2021;155:105372.
34. Medtronic. PERCEPT PC-NEUROSTIMULATOR medtronic.eu2020 [Available from: https://europe.medtronic.com/xd-en/healthcare-professionals/products/neurological/deep-brain-stimulation-systems/percept-pc.html.
35. Thenaisie Y, Palmisano C, Canessa A, Keulen BJ, Capetian P, Jiménez MC, 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).
36. Milosevic L, Scherer M, Cebi I, Guggenberger R, Machetanz K, Naros G, et al. Online Mapping With the Deep Brain Stimulation Lead: A Novel Targeting Tool in Parkinson's Disease. Mov Disord. 2020;35(9):1574-86.
37. Fernández-García C, Monje MHG, Gómez-Mayordomo V, Foffani G, Herranz R, Catalán MJ, et al. Long-term directional deep brain stimulation: Monopolar review vs. local field potential guided programming. Brain Stimul. 2022;15(3):727-36.
38. Strelow JN, Dembek TA, Baldermann JC, Andrade P, Jergas H, Visser-Vandewalle V, et al. Local Field Potential-Guided Contact Selection Using Chronically Implanted Sensing Devices for Deep Brain Stimulation in Parkinson's Disease. Brain Sci. 2022;12(12).
39. Yoshida F, Martinez-Torres I, Pogosyan A, Holl E, Petersen E, Chen CC, et al. Value of subthalamic nucleus local field potentials recordings in predicting stimulation parameters for deep brain stimulation in Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry. 2010;81(8):885-9.
40. Lopes EM, Rego R, Rito M, Chamadoira C, Dias D, Cunha JP. Estimation of ANT-DBS Electrodes on Target Positioning Based on a New PerceptTM PC LFP Signal Analysis. Sensors [Internet]. 2022; 22(17).
41. Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, et al. Parameterizing neural power spectra into periodic and aperiodic components.

42. Maleki F, Muthukrishnan N, Ovens K, Reinhold C, Forghani R. Machine Learning Algorithm Validation: From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment. Neuroimaging Clin N Am. 2020;30(4):433-45.
43. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology. 2010;5(9):1315-6.
44. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med. 2007;26(19):3661-75.
45. Altman DG, Machin D, Bryant T, Gardner M. Statistics with Confidence: Confidence Intervals and Statistical Guideline. 2nd ed: BMJ Books; 2000. 254 p.
46. Altman DG. Pratical Statistics for Medical Research. New York: Chapman and Hall/CRC; 1990. 624 p.
47. Epstein CM, Brickley GP. Interelectrode distance and amplitude of the scalp EEG. Electroencephalogr Clin Neurophysiol. 1985;60(4):287-92.
48. Deuschl G, Herzog J, Kleiner-Fisman G, Kubu C, Lozano AM, Lyons KE, et al. Deep brain stimulation: postoperative issues. Mov Disord. 2006;21 Suppl 14:S219-37.
49. Connolly A, Kaemmerer W, Dani S, Stanslaski S, Panken E, Johnson MD, et al. Guiding deep brain stimulation contact selection using local field potentials sensed by a chronically implanted device in Parkinson's disease patients. International IEEE/EMBS Conference on Neural Engineering, NER. 2015;2015:840-3.
50. Xu SS, Lee WL, Perera T, Sinclair NC, Bulluss KJ, McDermott HJ, et al. Can brain signals and anatomy refine contact choice for deep brain stimulation in Parkinson's disease? J Neurol Neurosurg Psychiatry. 2022.
51. 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):1162-77.
52. Chen PL, Chen YC, Tu PH, Liu TC, Chen MC, Wu HT, et al. Subthalamic high-beta oscillation informs the outcome of deep brain stimulation in patients with Parkinson's disease. Front Hum Neurosci. 2022;16:958521.
53. Priori A, Foffani G, Pesenti A, Tamma F, Bianchi AM, Pellegrini M, et al. Rhythm-specific pharmacological modulation of subthalamic activity in Parkinson's disease. Exp Neurol. 2004;189(2):369-79.
54. Neumann WJ, Degen K, Schneider GH, Brücke C, Huebl J, Brown P, et al. Subthalamic synchronized oscillatory activity correlates with motor impairment in patients with Parkinson's disease. Mov Disord. 2016;31(11):1748-51.
55. Shah A, Nguyen TK, Peterman K, Khawaldeh S, Debove I, Shah SA, et al. Combining Multimodal Biomarkers to Guide Deep Brain Stimulation Programming in Parkinson Disease. Neuromodulation. 2023;26(2):320-32.
56. Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. Nature Medicine. 2022;28(10):1996-9.
57. He S, Mostofi A, Syed E, Torrecillos F, Tinkhauser G, Fischer P, et al. Subthalamic beta-targeted neurofeedback speeds up movement initiation but increases tremor in Parkinsonian patients. eLife. 2020;9:e60979.
58. Del Vecchio Del Vecchio J, Hanafi I, Pozzi NG, Capetian P, Isaias IU, Haufe S, et al. Pallidal Recordings in Chronically Implanted Dystonic Patients: Mitigation of Tremor-Related Artifacts. Bioengineering (Basel). 2023;10(4).
59. Nelson MJ, Pouget P. Do Electrode Properties Create a Problem in Interpreting Local Field Potential Recordings? Journal of Neurophysiology. 2010;103(5):2315-7.
60. Meidahl AC, Tinkhauser G, Herz DM, Cagnan H, Debarros J, Brown P. Adaptive Deep Brain Stimulation for Movement Disorders: The Long Road to Clinical Therapy. Mov Disord. 2017;32(6):810-9.

61. Chen CC, Pogosyan A, Zrinzo LU, Tisch S, Limousin P, Ashkan K, et al. Intra-operative recordings of local field potentials can help localize the subthalamic nucleus in Parkinson's disease surgery. Experimental Neurology. 2006;198(1):214-21.
62. Singh A, Kammermeier S, Mehrkens JH, Bötzel K. Movement kinematic after deep brain stimulation associated microlesions. Journal of Neurology, Neurosurgery & Psychiatry. 2012;83(10):1022-6.
63. Okun MS, Gallo BV, Mandybur G, Jagid J, Foote KD, Revilla FJ, et al. Subthalamic deep brain stimulation with a constant-current device in Parkinson's disease: an open-label randomised controlled trial. Lancet Neurol. 2012;11(2):140-9.
64. Alsafy B, Mosad Z, Mutlag W. Multiclass Classification Methods: A Review2020.
65. Casalegno F. Learning to Rank: A Complete Guide to Ranking using Machine Learning: towardsdatascience.com; 2022 [Available from: https://towardsdatascience.com/learning-to-rank-a-complete-guide-to-ranking-using-machine-learning-4c9688d370d4.
66. Velliangiri S, Alagumuthukrishnan S, Thankumar joseph SI. A Review of Dimensionality Reduction Techniques for Efficient Computation. Procedia Computer Science. 2019;165:104-11.
67. Ellis C. Random forest overfitting: Crunching the Data; 2021 [Available from: https://crunchingthedata.com/random-forest-overfitting/.
68. Bour LJ, Lourens MA, Verhagen R, de Bie RM, van den Munckhof P, Schuurman PR, et al. Directional Recording of Subthalamic Spectral Power Densities in Parkinson's Disease and the Effect of Steering Deep Brain Stimulation. Brain Stimul. 2015;8(4):730-41.
69. Chua MMJ, Vissani M, Liu DD, Schaper F, Warren AEL, Caston R, et al. Initial case series of a novel sensing deep brain stimulation device in drug-resistant epilepsy and consistent identification of alpha/beta oscillatory activity: A feasibility study. Epilepsia. 2023.
70. Lavine BK, Rayens WS. 3.16 - Statistical Discriminant Analysis. In: Brown SD, Tauler R, Walczak B, editors. Comprehensive Chemometrics. Oxford: Elsevier; 2009. p. 517-40.
71. Ohri A. LASSO Regression: A Complete Understanding: u-next.com; 2021 [Available from: https://u-next.com/blogs/artificial-intelligence/lasso-regression/#:~:text=Lasso%20regression%20algorithm%20is%20defined,as%20RMSE%20and%20R%2DSquare.
72. Xiaozhou Y. Linear Discriminant Analysis, Explained towardsdatascience.com2020 [Available from: https://towardsdatascience.com/linear-discriminant-analysis-explained-f88be6c1e00b.
73. Harrison O. Machine Learning Basics with the K-Nearest Neighbors Algorithm: towardsdatascience.com; 2018 [Available from: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761.
74. Yiu T. Understanding Random Forest: towardsdatascience.com; 2019 [Available from: https://towardsdatascience.com/understanding-random-forest-58381e0602d2#:~:text=The%20random%20forest%20is%20a,that%20of%20any%20individual%20tree.

Appendices

A. Methodology for preprocessing and machine learning models

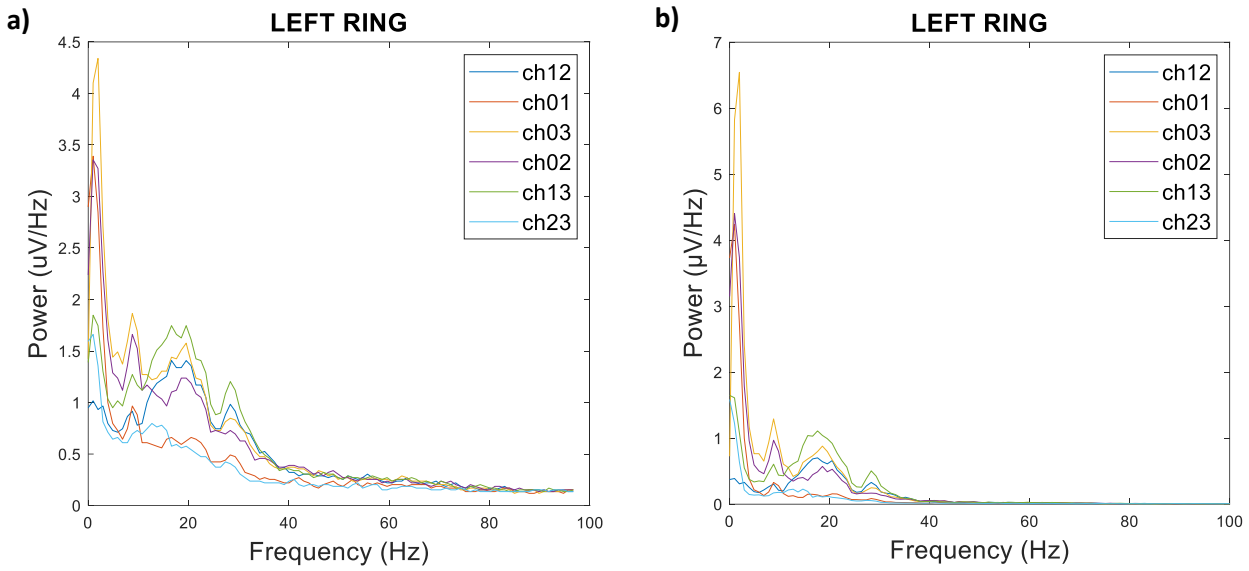


Figure 1. a) LFP level-based recordings left hemisphere NL048 as provided by the Percept PC®. **b)** LFP level-based recordings left hemisphere NL048 after manual conversion from time to frequency domain using Percept PC® time domain data.

Table 1. Overview of preprocessing methods and corresponding applied cut-off values.

Method	Mechanism	Advantages	Disadvantages	Cut-offs
Principle component analysis **(70)** Matlab function: "pca()"	Constructs new predictor variables as linear combinations of the original variables without taking the response variable into account	• Prevents overfitting • Removes correlated features • Speeds up machine learning algorithms	• Loss of information • Data standardisation • Difficult to identify most significant features • Unsupervised technique	85% variance included
Partial least squares regression **(70)** Matlab function: "plsregress()"	Constructs new predictor variables as linear combinations of the original variables whilst taking the response variable into account.	• Requires few components • Supervised technique • Removes correlated features • Speeds up machine learning algorithms	• Prone to overfitting • Loss of information • Data standardisation • Difficult to identify most significant features	75% of maximum variance included
Lasso regularisation **(71)**	Assists in the elimination of irrelevant parameters,	• Robust to outliers • Prevents overfitting	• Loss of features • Only one feature selected from a group of	75th lambda

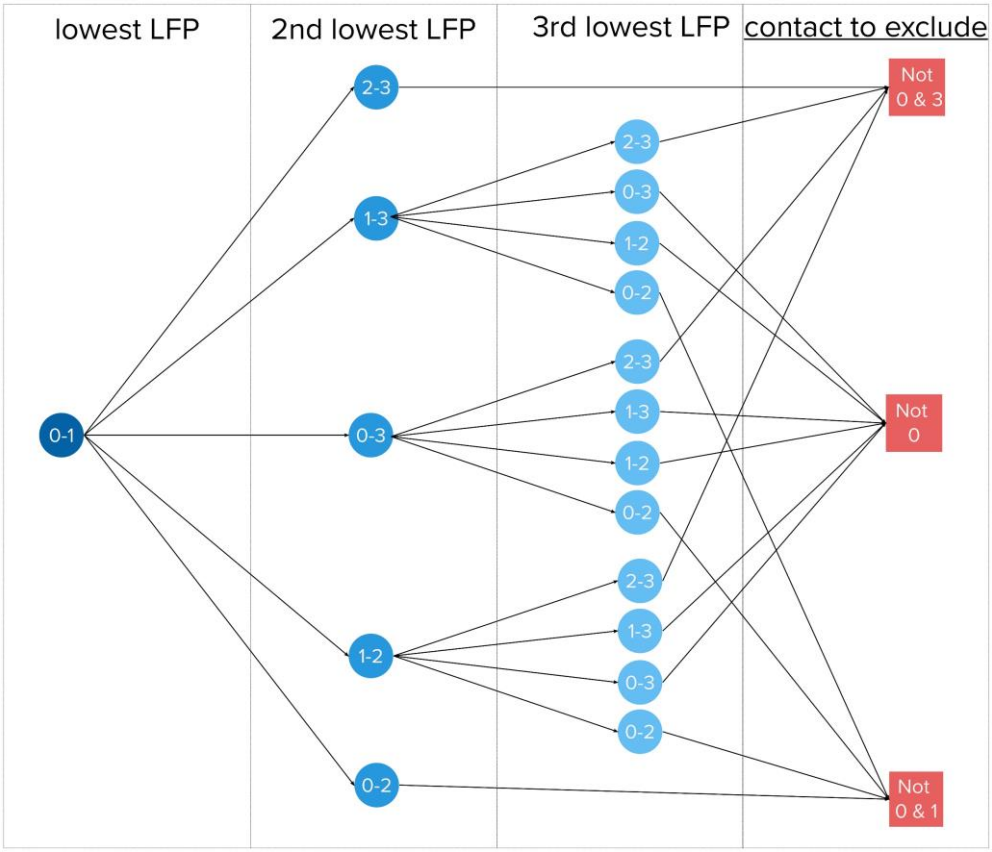
Matlab function: "lassoglm()"	thus helping in the concentration of selection and regularises the models.	• Speeds up machine learning algorithms • Most important features can be identified	correlated features (combinations not evaluated)	

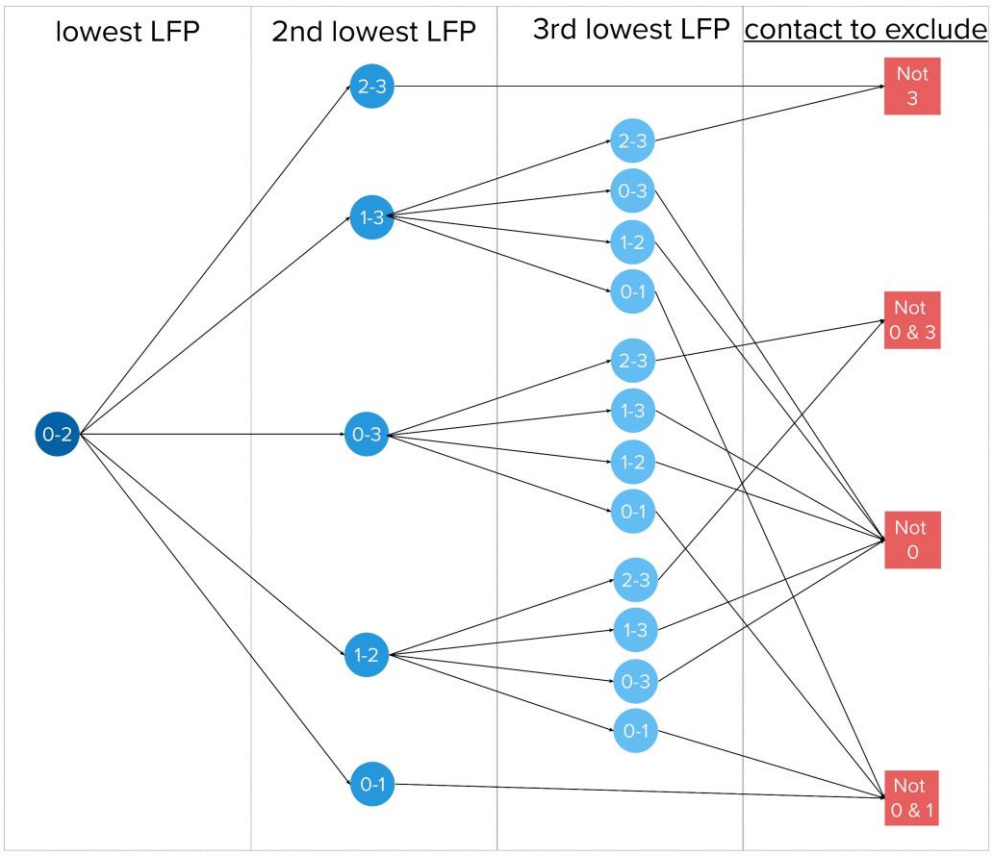
Table 2. Overview of machine learning methods and corresponding applied hyperparameter ranges.

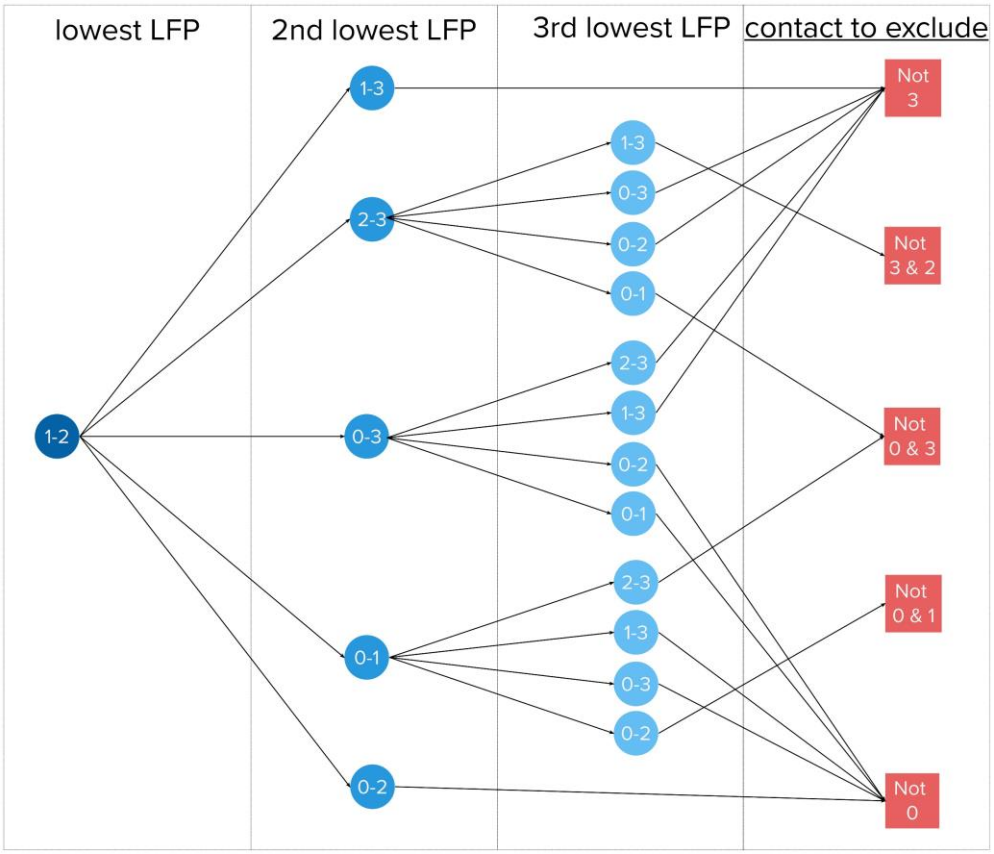
Model	Mechanism	Advantages	Disadvantages	Hyperparameter ranges
Linear discriminant analysis **(72)** Matlab function: "fitcdiscr()"	finds a linear combination of features that characterises or separates two or more classes of objects.	• Simple • Robust classification easy to implement • Low overfitting due to inherent dimensional reduction	• Linear boundaries may separate classes • Uses too many parameters in high-dimensional setting	Discriminant type: 'linear', 'pseudolinear', 'diaglinear'
K-nearest neighbours **(73)** Matlab function: "fitcknn()"	Finds k number of samples closest to point needing prediction. Evaluates which label is most frequent amongst these samples and classifies the point needing prediction as this label.	• Simple • Intuitive • New data is easy to add • Multiple distance metrics • High sensitivity • Non-linear performance	• Poor performance on imbalanced data/large datasets/high dimensions • Costly computation • Sensitive to noise/missing values/outliers	Number of neighbours: 1, 2, 3…20
Random forest **(74)** Matlab function: "TreeBagger()"	Combination of large number of decision trees. Each decision tree calculates the class prediction and the class with the highest amount of votes becomes prediction.	• Robust to outliers/missing data • Lower overfitting (vs. decision trees) • Suitable for large datasets • Non-linear performance	• Biased when dealing with categorical variables • Slow training • Parameter complexity • Overfitting risk remains	Number of trees: 25, 50, 75….500

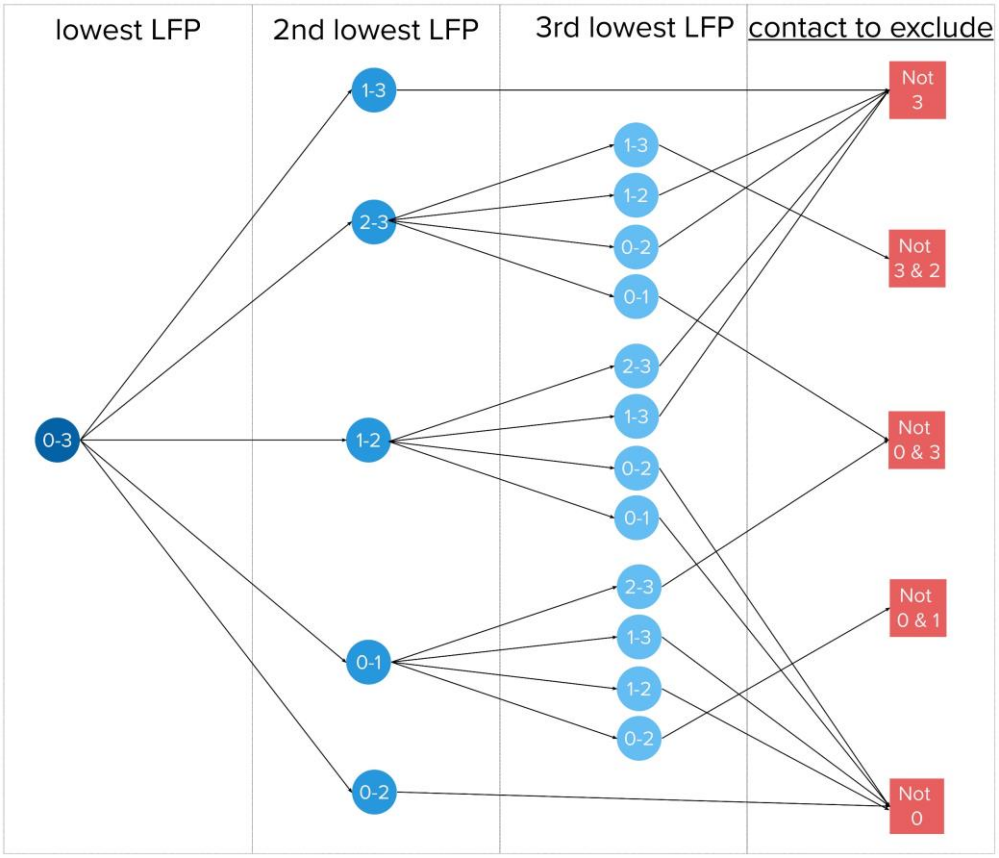
B. Decision Trees

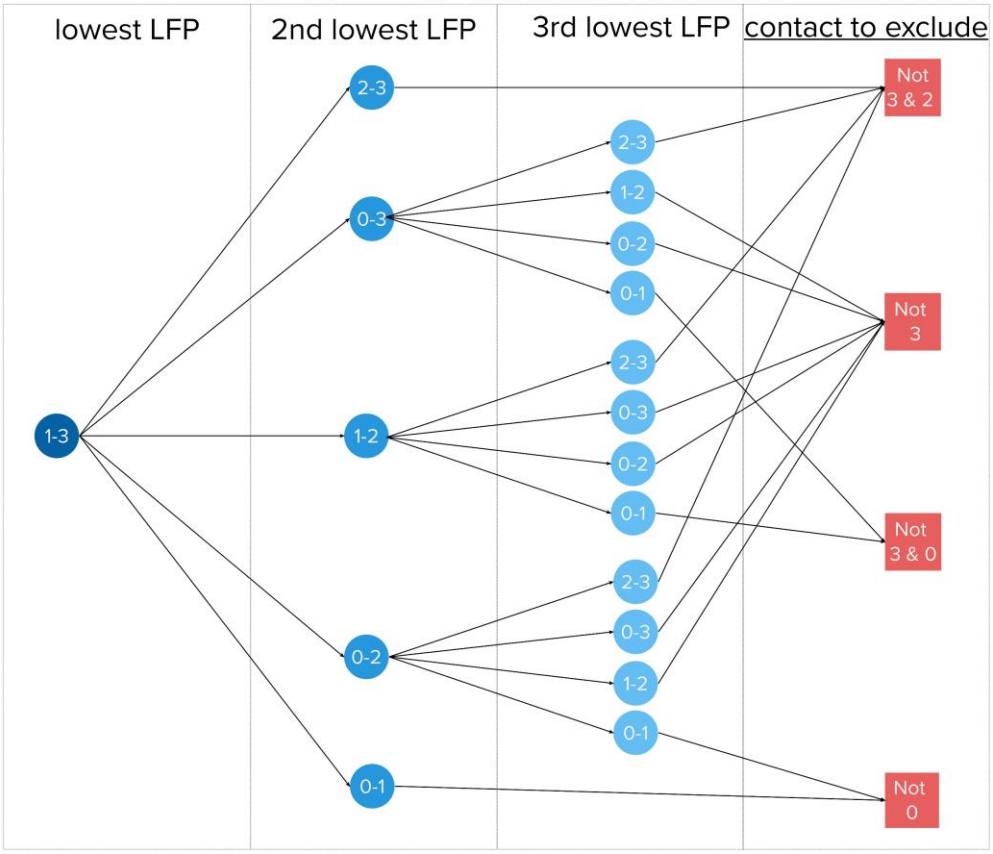
Part I. Decision Trees for contact elimination

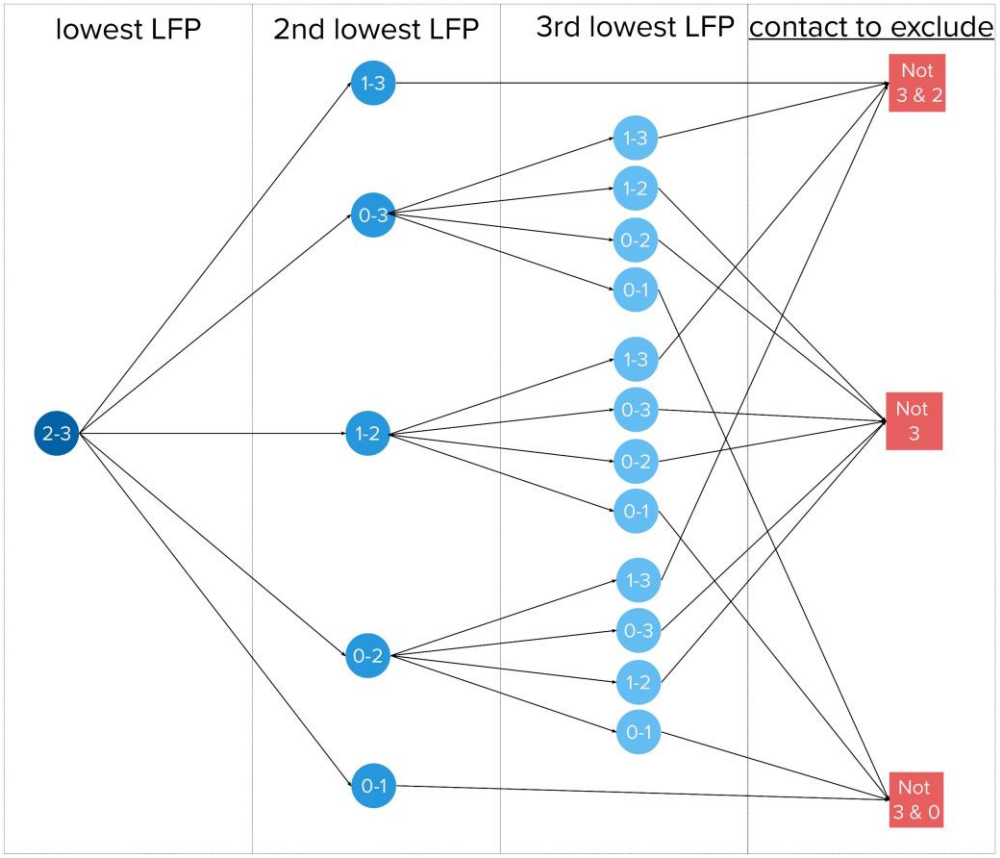




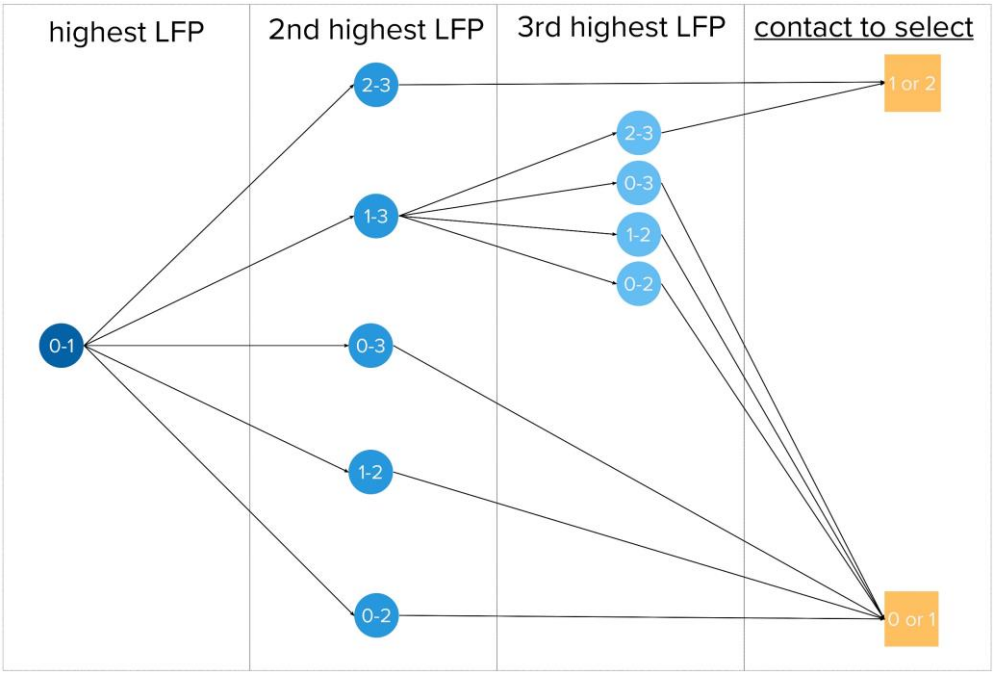

lowest LFP
2nd lowest LFP
3rd lowest LFP
contact to exclude
1-2
1-3
2-3
0-3
0-1
0-2
1-3
0-3
0-2
0-1
2-3
1-3
0-2
0-1
2-3
1-3
0-3
0-2
Not 3
Not 3 & 2
Not 0 & 3
Not 0 & 1
Not 0

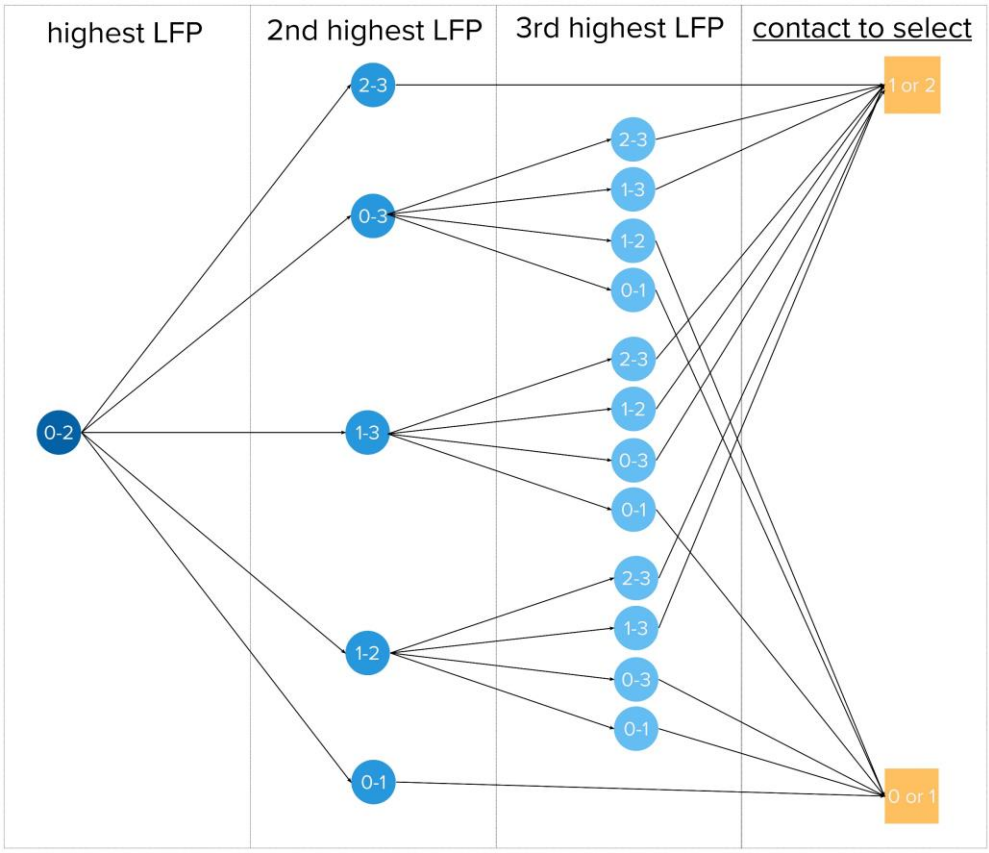

lowest LFP
2nd lowest LFP
3rd lowest LFP
contact to exclude
0-3
1-3
2-3
1-2
0-1
0-2
1-3
1-2
0-2
0-1
2-3
1-3
0-2
0-1
2-3
1-3
1-2
0-2
Not 3
Not 3 & 2
Not 0 & 3
Not 0 & 1
Not 0

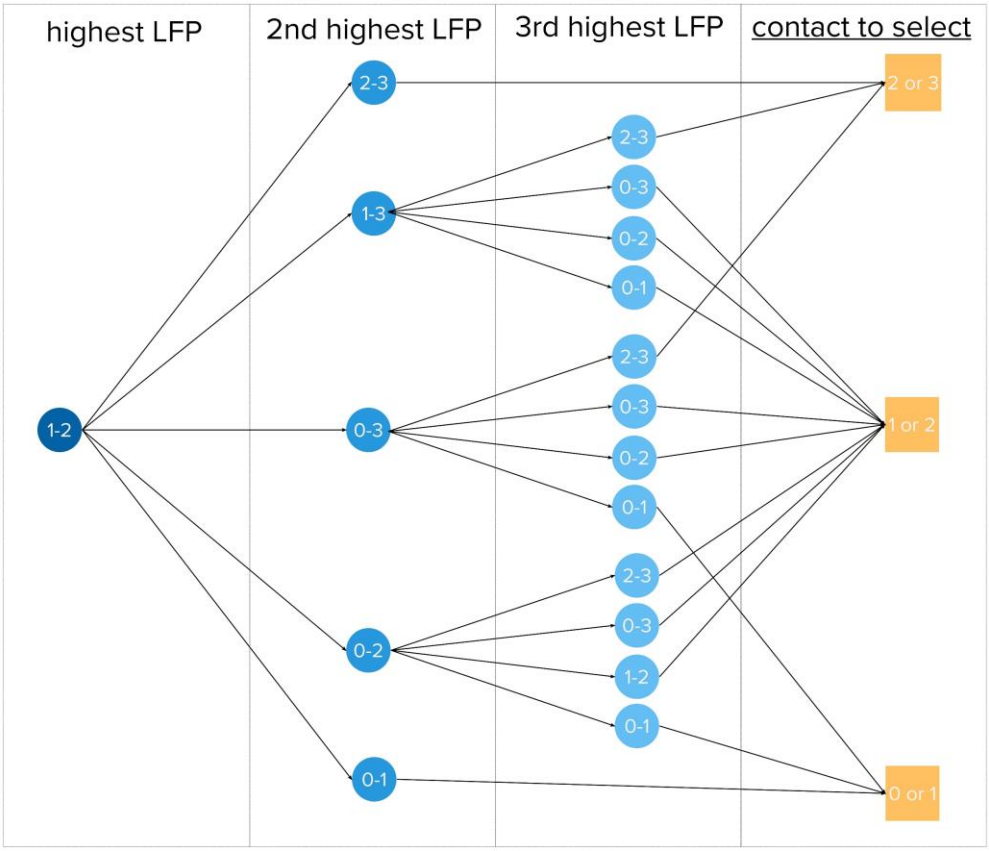

lowest LFP
2nd lowest LFP
3rd lowest LFP
contact to exclude
1-3
2-3
0-3
1-2
0-2
0-1
2-3
1-2
0-3
0-2
0-1
2-3
0-3
1-2
0-1
Not 3 & 2
Not 3
Not 3 & 0
Not 0

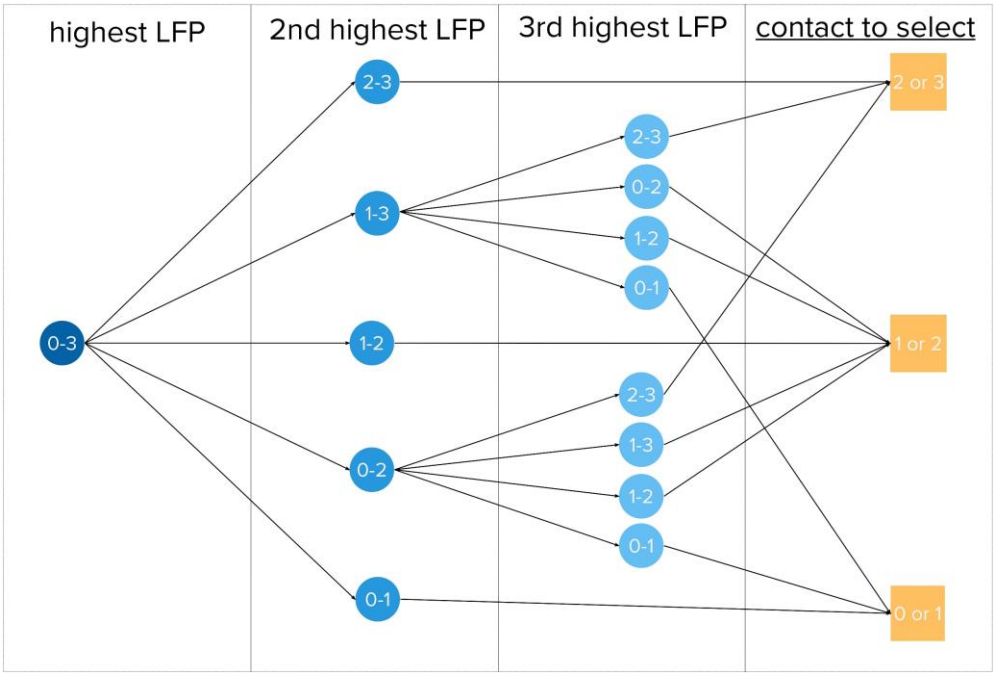

lowest LFP
2nd lowest LFP
3rd lowest LFP
contact to exclude
2-3
1-3
0-3
1-2
0-2
0-1
1-3
0-3
1-2
0-2
0-1
1-3
0-3
1-2
0-1
Not 3 & 2
Not 3
Not 3 & 0

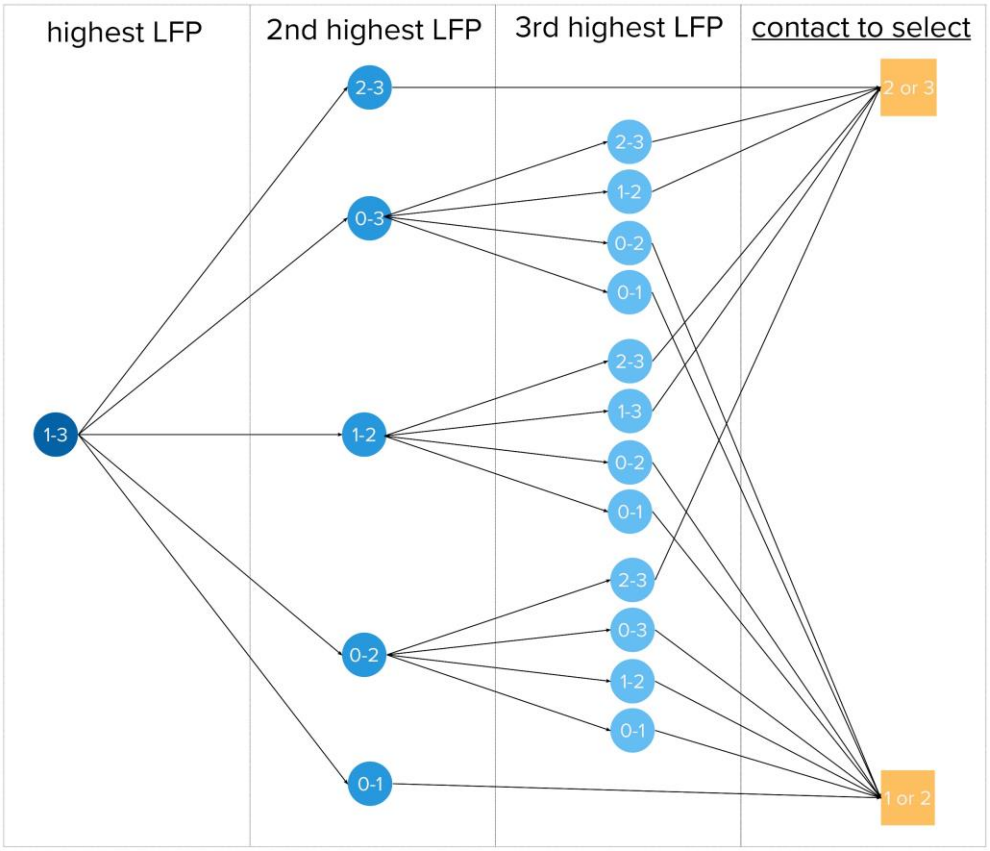
Part II. Decision Trees for contact selection

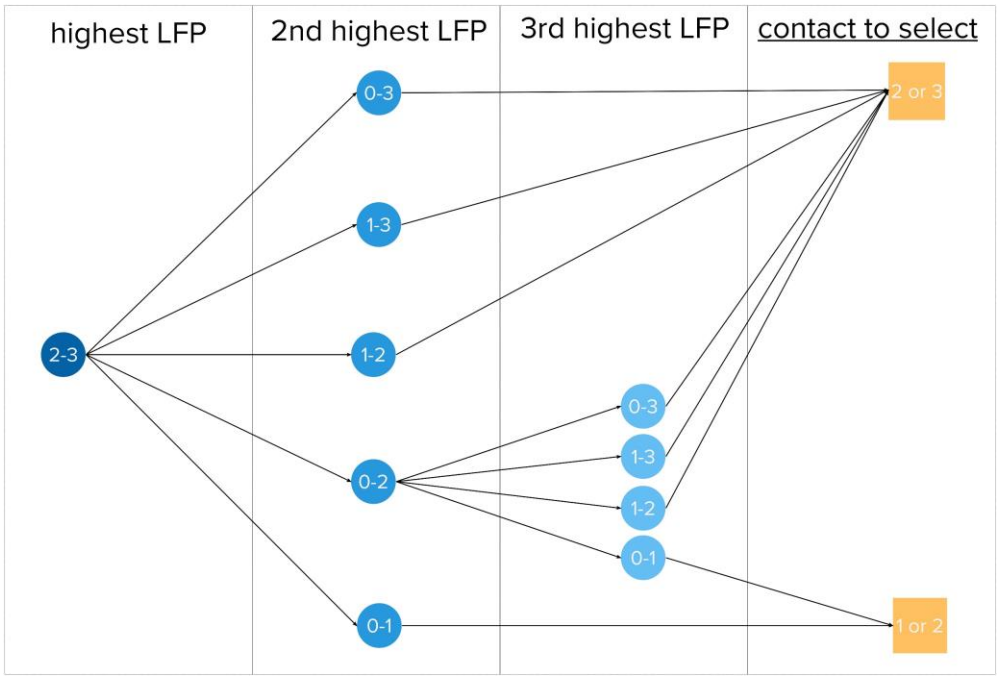





highest LFP
2nd highest LFP
3rd highest LFP
contact to select
1-2
2-3
1-3
0-3
0-2
0-1
2-3
0-3
0-2
0-1
2-3
0-3
0-2
0-1
2-3
0-3
1-2
0-1
2 or 3
1 or 2
0 or 1


highest LFP
2nd highest LFP
3rd highest LFP
contact to select
0-3
2-3
1-3
1-2
0-2
0-1
2-3
0-2
1-2
0-1
2-3
1-3
1-2
0-1
2 or 3
1 or 2
0 or 1


highest LFP
2nd highest LFP
3rd highest LFP
contact to select
1-3
2-3
0-3
1-2
0-2
0-1
2-3
1-2
0-3
1-3
0-2
0-1
2-3
1-2
0-3
1-2
0-2
0-1
2 or 3
1 or 2


highest LFP
2nd highest LFP
3rd highest LFP
contact to select
2-3
0-3
1-3
1-2
0-2
0-1
0-3
1-3
1-2
0-1
2 or 3
1 or 2

Part III. Decision Tree MATLAB GUI

For an example of the Matlab decision tree assistant interface see Figure 1.

To run with Matlab (2023a):

- Running: Open Matlab (2023a) → Type “appdesigner” in the command window → Click “Enter” button on keyboard → Open “DecisionTreeAssistant.mlapp” → Click “Run” (green arrow) → select ranking according to BrainSense Survey data → Click “Confirm ranking” → Repeat for other hemisphere by clicking “Run” (green arrow) again, select ranking according to BrainSense Survey data and click “Confirm ranking”.

To run without Matlab (2023a):

- In case Matlab is not available on the device install “Matlab Runtime” version 2023a via: https://nl.mathworks.com/products/compiler/matlab-runtime.html, this is open access software. If Matlab (2023a) is installed on your device no “Matlab Runtime” installation is required.
- Double click DecisionTreeAssistant.exe → wait for window from Figure 1. to open (this can take a while) → select ranking according to BrainSense Survey data → Click “Confirm ranking” → Repeat for other hemisphere by double clicking DecisionTreeAssistant.exe again, select ranking according to BrainSense Survey data and click “Confirm ranking”.

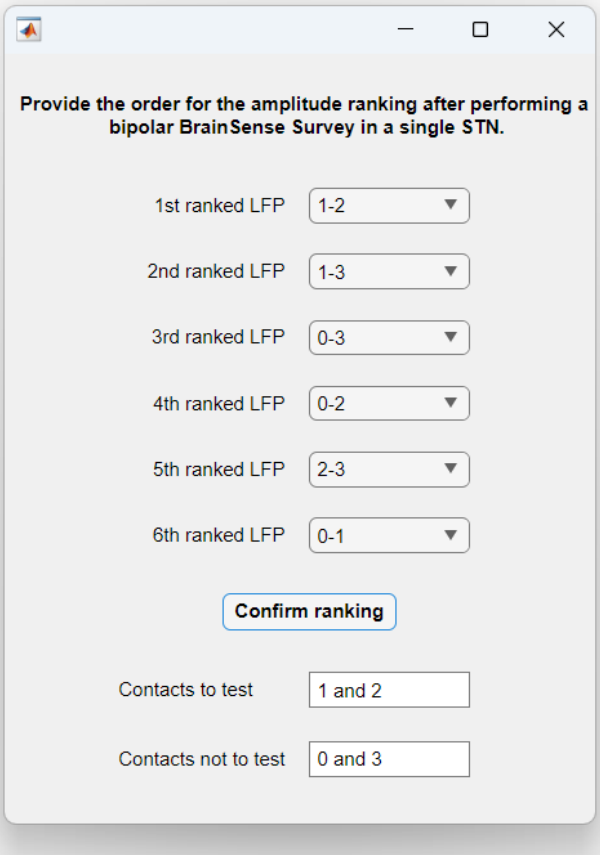


Figure 1. Matlab decision tree GUI example for case from Figure 7.

C. UPDRS-III scores

Table 1. UPDRS-III scores per patient and hemisphere scored during MPR per contact and stimulation amplitude.

		Contact 0		Contact 1		Contact 2		Contact 3	
Patient	Hemi-sphere	UPDRS (R/B/T)	mA (R/B/T)	UPDRS (R/B/T)	mA (R/B/T)	UPDRS (R/B/T)	mA (R/B/T)	UPDRS (R/B/T)	mA (R/B/T)
NL007	LH	?/?/3	?/?/4	?/?/0+	?/?/4.5	?/?/0	?/?/5	?/?/2	?/?/5
	RH								
NL008	LH								
	RH								
NL016	LH	0/0/?	1.5/1.5/?	0/1-/?	2/1/?	0/0/?	2.5/2.5/?	0/0+/?	1.5/1.5/?
	RH	1+/1+/?	2.5/2.5/?	0+/0+/?	3/3/?	0+/1/?	4/3.5/?	1/0+/?	2.5/3/?
NL017	LH	1-/?/?	2/?/?	0/?/?	1/?/?	0/?/?	1.5/?/?	0/?/?	1.5/?/?
	RH	0/0/?	3/2.5/?	0/0/?	3/2/?	0-1/0-1/?	3.5/4/?	1+/2-/?	2/2/?
NL018	LH								
	RH								
NL019	LH	1/2/?	2.5/2/?	1-/1-2/?	3.5/3.5/?	1/2-3/?	3/2/?	1-2/2-3/?	4/4/?
	RH	1/2-3/?	2/2/?	1/2+/?	4/4/?	0-1/0+/?	4/4/?	0-1/1+/?	3/4/?
NL020	LH	0/1/?	1.5/2/?	0/0+/?	1/4/?	0/1/?	2/3/?	0/1/?	2/3/?
	RH	0/1-2/?	2/4.5/?	0/1-2/?	1.5/3.5/?	0/1-2/?	1.5/5/?	0/2+/?	1.5/2.5/?
NL022	LH								
	RH	?/1/?	?/2/?	?/1/?	?/2/?	?/1/?	?/4/?	?/1/?	?/4/?
NL023	LH								
	RH	?/?/0	?/?/4	?/?/0	?/?/3	?/?/0	?/?/2.5	?/?/0	?/?/4
NL024	LH								
	RH								
NL027	LH	0/2-/1	1.5/1.5/1.5	0/1/1	1.5/1.5/1.5	0/1-/1	2/2.5/1.5	0/2--/1	1.5/2.5/1.5
	RH	3/3/?	2/2/?	3-/3-/?	2/2/?	2-/2-/?	4.5/4.5/?	1/1+/?	5/5/?
NL028	LH								
	RH	1/2/?	2.5/2.5/?	1-/1/?	2/2/?	1/1/?	2/2/?	1-/1-2/?	2/2/?
NL031	LH	1/3/2	2.5/2.5/2.5	1/1-2/2-	2.5/4/3	1/1-2/1+	3/3/4.	1/1-2/1+	3/3/4.
	RH	0/0-1/0-1	4/4/4	0/1/0	2.5/2.5/2.5	0/0-1/0	2.5/3.5/2.5	0/1+/0	2.5/3/2.5
NL037	LH	0+/0+/?	1.5/2.5/?	0/0/?	2/3.5/?	0/0+/?	1.5/3.5/?	0+/0+/?	1.5/4.5/?
	RH	0/1/?	1.5/2/?	0/0/?	2.5/2.5/?	0/0/?	2/2/?	0/2/?	2/3.5/?
NL042	LH	?/0+/?	?/2/?	?/0+/?	?/2/?	?/0+/?	?/1.5/?	?/1/?	?/1.5/?
	RH	0/1-/?	2.5/2.5/?	0+/1-/?	2/2.5/?	0/1-/?	2/2/?	0/1+/?	4.5/4.5/?
NL044	LH	?/0-1/0-1	?/1.5/2.5	?/0/0	?/4/4	?/?/0	?/?/1	?/?/0-1	?/?/3.5
	RH	?/0/?	?/3.5/?	?/0/?	?/3/?	?/0/?	?/3/?	?/2/?	?/4/?
NL045	LH	0/2/?	3/3/?	0/1-/?	1.5/4/?	0/0/?	2/2/?	0/0-1/?	1.5/1.5/?
	RH								
NL046	LH	?/3/?	?/1.5/?	?/3/?	?/1.5/?	?/3/?	?/3/?	?/3/?	?/3/?
	RH								
NL048	LH	?/2--/?	?/3.5/?	?/1-/?	?/1.5/?	?/1-/?	?/1.5/?	?/2-3/?	?/2/?
	RH	2/1/?	2/2/?	0/1/?	1.5/2/?	0/0+/?	1.5/2/?	0/0+/?	1.5/2/?
NL049	LH	1/1+/?	2.5/2.5/?	1/0/?	2/3/?	0/0/?	2/2.5/?	0+/0+/?	3/3/?
	RH	0+/0/?	3.5/3.5/?	0+/0/?	2.5/2.5/?	0+/1-/?	3.5/3.5/?	0+/1-/?	3/4/?
NL051	LH	0/2+/?	0/2.5/?	0/1-/?	0/3/?	0/0+/?	0/2.5/?	0/0+/?	0/2.5/?
	RH	0/0/?	0/2.5/?	0/0/?	0/2/?	0/0/?	0/2/?	0/2/?	0/2/?

NL052	LH								
	RH								
NL053	LH	1/1-2/?	3.5/3.5/?	0+/0-1/?	3/3/?	0-1/1/?	4/3/?	1/2/?	3/4/?
	RH	1/1/?	3/4/?	0-1/0-1/?	3/3/?	0+/0/?	2.5/4/?	1/1/?	2/2/?
NL054	LH	???	3.5	1-/1-/0	3.5/3.5/3	0-1/0-1/?	3/3/?	0/0/?	2.5/2/?
	RH	1-0/2-/1	3/3/2.5	0/0/?	1.5/2/?	0/0/?	1.5/1/?	0/0/?	2/1/?
NL055	LH	1/0/?	2/2.5/?	0+/0/?	2/2/?	0+/0/?	2/2/?	0+/0/?	1.5/2/?
	RH	3-/2/?	1.5/1.5/?	2/1/?	2/2/?	1/0/?	2.5/2.5/?	1-/0/?	4/2.5/?
NL056	LH	0/3-/?	1/2/?	0/3-/?	1/2/?	0/1/?	1/3/?	0/0/?	1/4.5/?
	RH	0/2/?	1/2/?	0/1-/?	1/3.5/?	0/0/?	1/3/?	0/1/?	1/2.5/?
NL057	LH								
	RH	?/?/3	?/?/3	?/?/0	?/?/4	?/?/0+	?/?/3.5	?/?/0+	?/?/4
NL058	LH								
	RH								
NL059	LH								
	RH	0/1-/2	1.5/0/0	0/1/?	1.5/0/?	0/0+/2	1/1/2.5	0/0+/?	1.5/2/?
NL065	LH								
	RH								
NL068	LH	0/0+/?	1/2/?	0/0+/?	1/1/?	0/0/?	1/1.5/?	0/0+/?	1.5/1.5/?
	RH								
NL069	LH	NA	NA	NA	NA	NA	NA	NA	NA
	RH	NA	NA	NA	NA	NA	NA	NA	NA
NL070	LH								
	RH								
NL071	LH	?/0+/0	?/2/2	?/0+/0	?/2/2.5	?/?/0	?/?/2.5	?/?/0	?/?/2.5
	RH								

Abbreviations: UPDRS: unified parkinson's disease rating scale; R/B/T: rigidity/bradykinesia/tremor; mA: milli Ampère; LH: left hemisphere; RH: right hemisphere; NA: not available

Legend	
	Best UPDRS & Chosen contact
	Best UPDRS (no chosen contact)
	Chosen contact but not best UPDRS
	No symptoms
	Chosen contact when no symptoms

D. Non-artefact recording

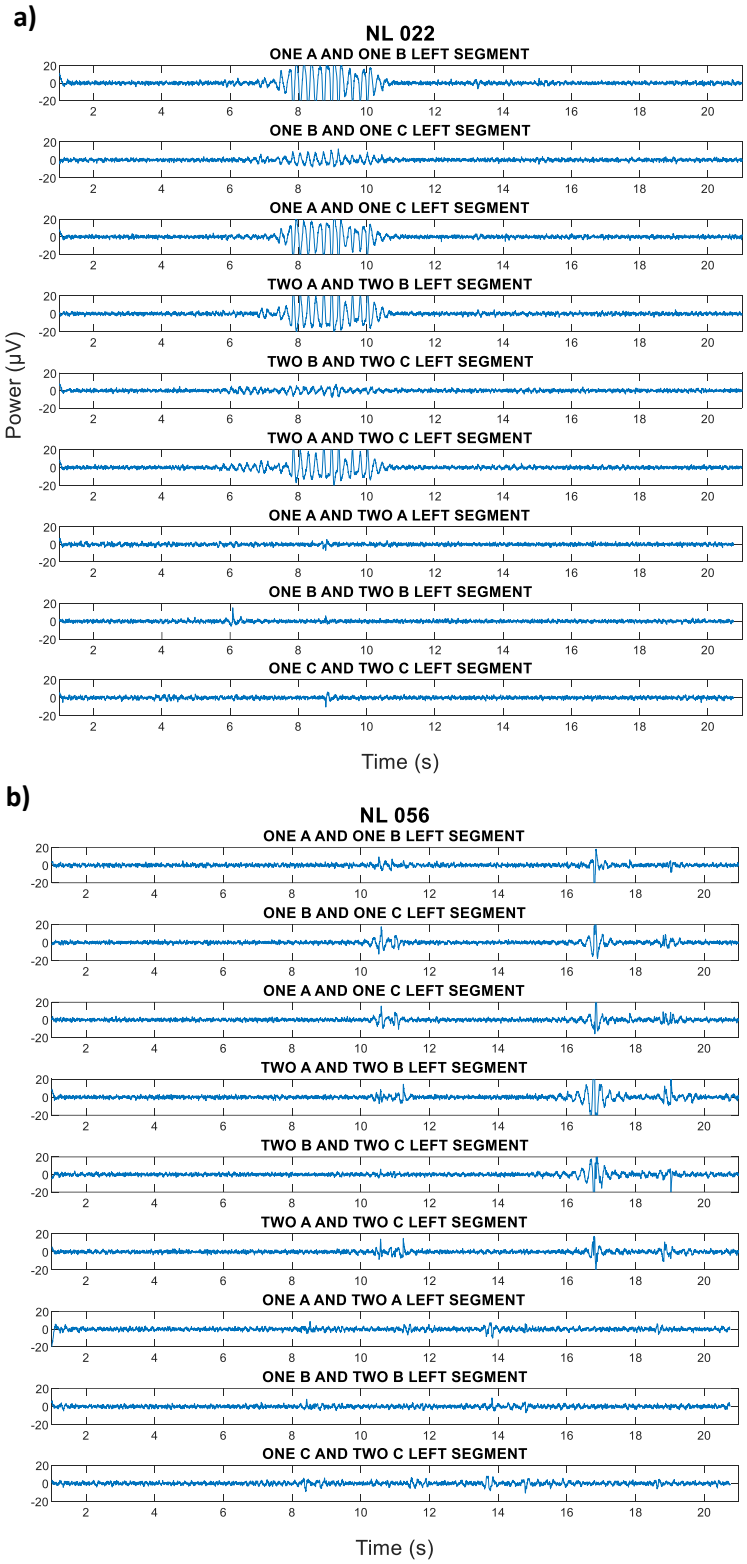


Figure 1. Time domain data after 4Hz high pass filtering from segment recordings across all 9 channels in the left hemisphere of two patients, **a)** NL022 and **b)** NL056, with sporadic artefacts remaining across several channels.

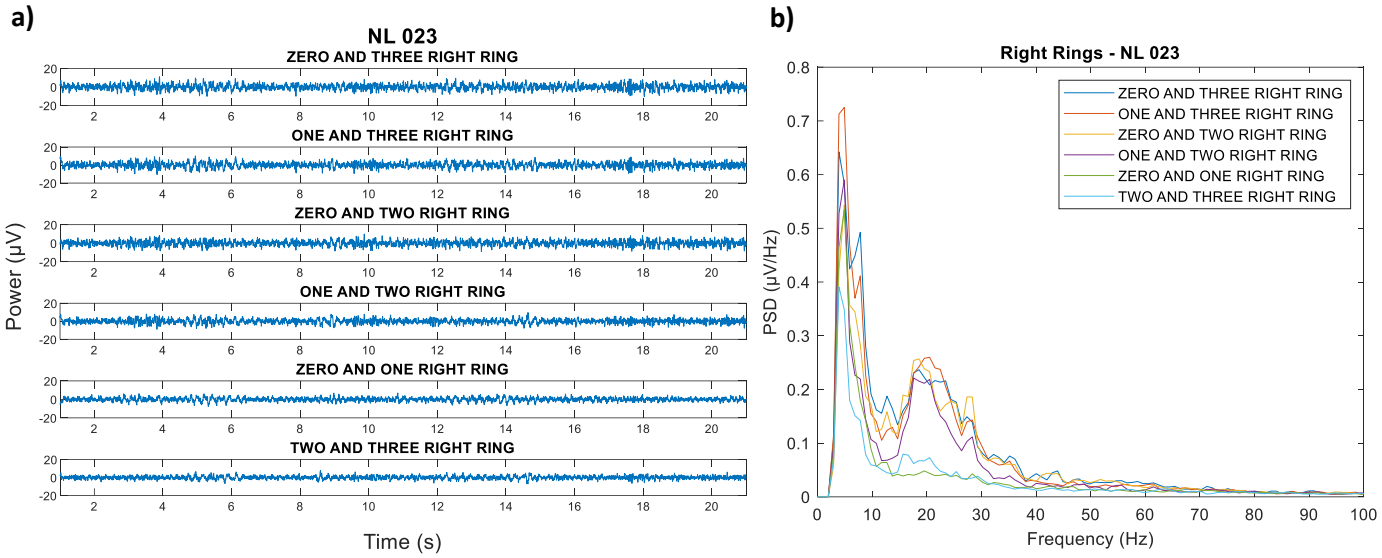


Figure 2. a) Time domain and **b)** frequency domain data after 4Hz high pass filtering from level-based recordings across all 6 channels for the right hemisphere of patient NL023, with a remaining artefact throughout all the time domain data and below 10Hz for all channels in the frequency domain data.

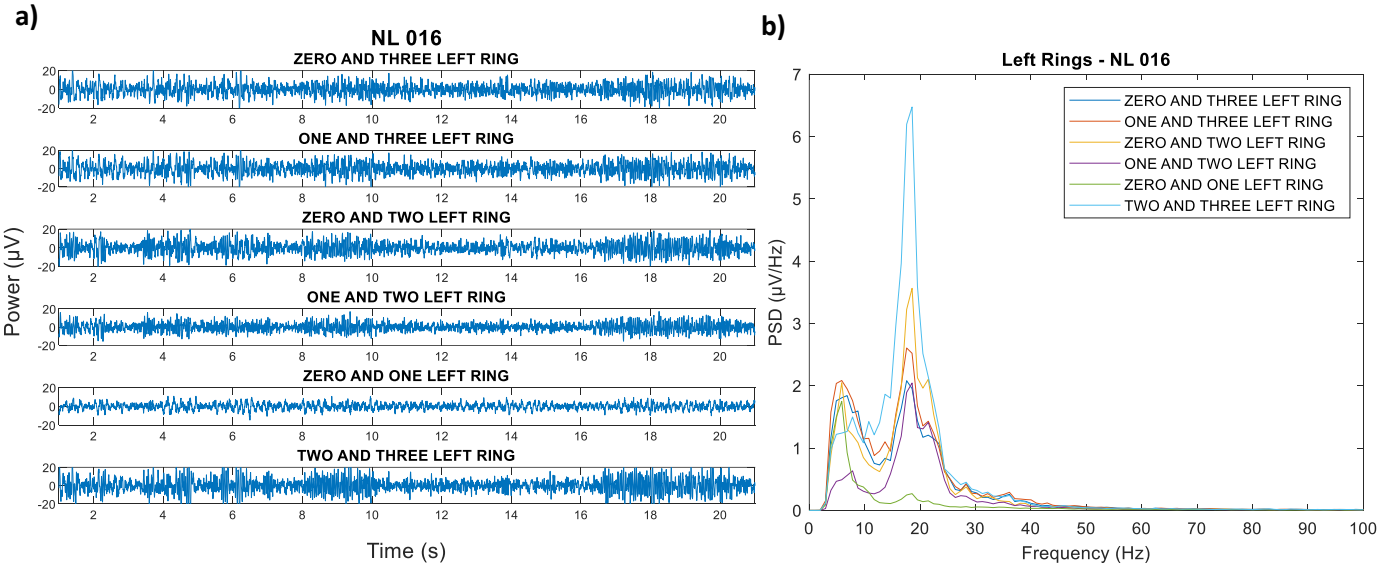


Figure 3. a) Time domain and **b)** frequency domain data after 4Hz high pass filtering from level-based recordings across all 6 channels for the left hemisphere of patient NL016, without a remaining artefact throughout the time domain data and above 4Hz for channels in the frequency domain data.

E. Machine learning results

Table 1. Machine learning model performance for models with beta frequency only and all frequencies combined.

	Beta only		Theta + Alpha + Beta + Gamma	
LDA	ROC-AUC	Type	ROC-AUC	Type
Raw	D = 0.72/T = 0.53	7x lin.,3x diag.	D = 0.95/T = 0.59	8x lin.,2x diag.
PCA	D = 0.67/T = 0.61	10x lin.	D = 0.70/T = 0.63	10x lin.
PLSR	D = 0.72/T = 0.55	10x lin.	D = 0.81/T = 0.60	8x lin., 2x diag.
L1	D = 0.72/T = 0.59	8x lin.,2x diag.	D = 0.82/T = 0.54	6x lin.,4x diag.
KNN	ROC-AUC	Median type (range)	ROC-AUC	Median type (range)
Raw	D = 0.72/T = 0.46	6 (4-12)	D = 0.70/T = 0.48	6.5 (4-9)
PCA	D = 0.80/T = 0.52	4.5 (3-7)	D = 0.77/T = 0.47	6 (3-7)
PLSR	D = 0.73/T = 0.45	7.5 (4-9)	D = 0.78/T = 0.50	5.5 (4-11)
L1	D = 0.74/T = 0.52	6.5 (4-9)	D = 0.76/T = 0.55	5 (3-7)
RF	ROC-AUC	Median type (range)	ROC-AUC	Median type (range)
Raw	D = 1.00/T = 0.55	25 (25-200)	D = 1.00/T = 0.55	62.5 (25-325)
PCA	D = 1.00/T = 0.58	25 (25-150)	D = 1.00/T = 0.54	50 (25-250)
PLSR	D = 1.00/T= 0.59	50 (25-75)	D = 1.00/T = 0.55	62.5 (25-125)
L1	D = 1.00/T = 0.53	50 (25-100)	D = 1.00/T = 0.57	50 (25-425)

Abbreviations: ROC-AUC: area under the receiver operator curve; PCA: principle component analysis; PLSR: partial least square regression; L1: lasso regression; D: design set (90% of data); T: test set (10% of data); lin.: linear; diag.: diaglinear

Table 2. Machine learning model performance for models with the beta frequency combined with either the theta, alpha or gamma frequency.

	Beta + Theta		Beta + Alpha		Beta + Gamma	
LDA	ROC-AUC	Type	ROC-AUC	Type	ROC-AUC	Type
Raw	D = 0.80/T = 0.59	5x lin.,5x diag.	D = 0.86/T = 0.69	7x lin.,3x diag.	D = 0.73/T = 0.56	5x lin.,5x diag.
PCA	D = 0.67/T = 0.62	10x lin.	D = 0.68/T = 0.62	9x lin.,1x diag.	D = 0.67/T = 0.60	8x lin.,2x diag.
PLSR	D = 0.75/T = 0.57	8x lin.,2x diag.	D = 0.78/T = 0.52	5x lin.,5x diag.	D = 0.78/T = 0.56	7x lin.,3x diag.
L1	D = 0.77/T = 0.56	6x lin.,4x diag.	D = 0.84/T = 0.59	8x lin.,2x diag.	D = 0.76/T = 0.49	6x lin.,4x diag.
KNN	ROC-AUC	Median type (range)	ROC-AUC	Median type (range)	ROC-AUC	Median type (range)
Raw	D = 0.75/T = 0.53	5 (3-8)	D = 0.75/T = 0.54	7 (3-11)	D = 0.68/T = 0.40	6 (3-10)
PCA	D = 0.80/T = 0.53	5 (3-7)	D = 0.75/T = 0.53	6 (4-7)	D = 0.74/T = 0.48	6 (3-9)
PLSR	D = 0.76/T = 0.46	6.5 (4-8)	D = 0.75/T = 0.53	5.5 (3-10)	D = 0.92/T = 0.48	5 (2-8)
L1	D = 0.74/T = 0.61	7 (4-9)	D = 0.76/T = 0.53	6.5 (5-8)	D = 0.77/T = 0.56	6 (4-8)
RF	ROC-AUC	Median type (range)	ROC-AUC	Median type (range)	ROC-AUC	Median type (range)
Raw	D = 1.00/T = 0.64	75 (25-100)	D = 1.00/T = 0.60	75 (25-125)	D = 1.00/T = 0.51	50 (25-275)
PCA	D = 1.00/T = 0.55	50 (25-150)	D = 1.00/T = 0.53	50 (25-75)	D = 1.00/T = 0.50	75 (25-175)
PLSR	D = 1.00/T = 0.63	50 (25-250)	D = 1.00/T = 0.58	50 (25-75)	D = 1.00/T = 0.55	50 (25-125)
L1	D = 1.00/T = 0.58	75 (50-125)	D = 1.00/T = 0.57	50 (25-225)	D = 1.00/T = 0.41	50 (25-200)

Abbreviations: ROC-AUC: area under the receiver operator curve; PCA: principle component analysis; PLSR: partial least square regression; L1: lasso regression; D: design set (90% of data); T: test set (10% of data); lin.: linear; diag.: diaglinear

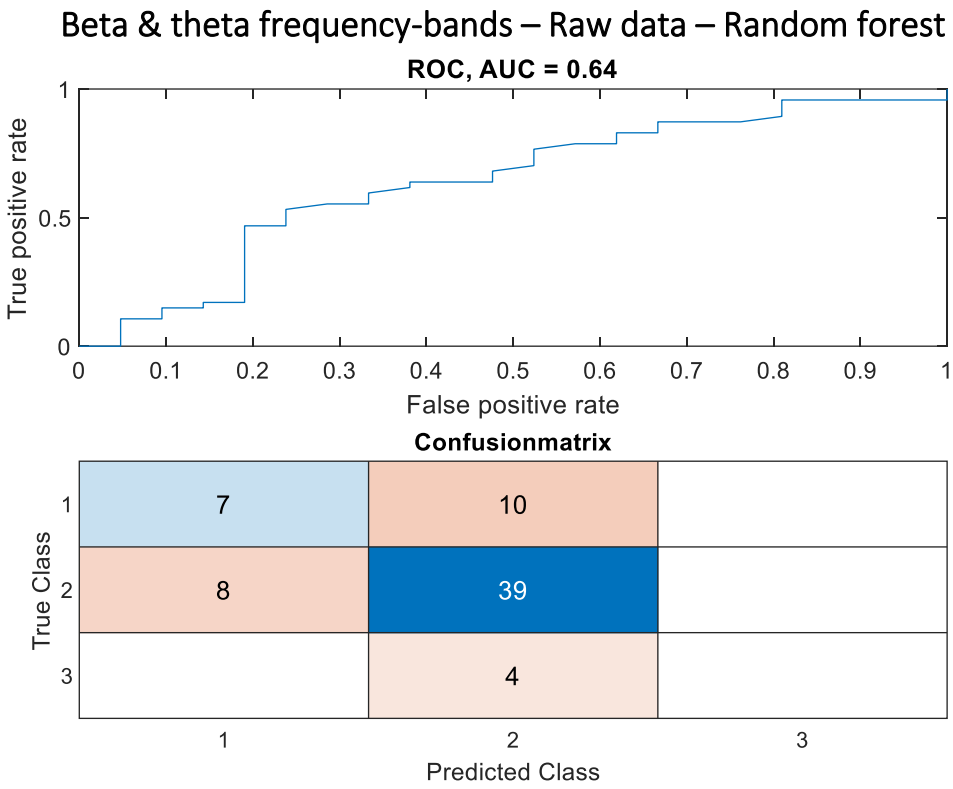


Figure 1. Receiver operator curve and confusion matrix for random forest model using the raw features from the beta as well theta band.

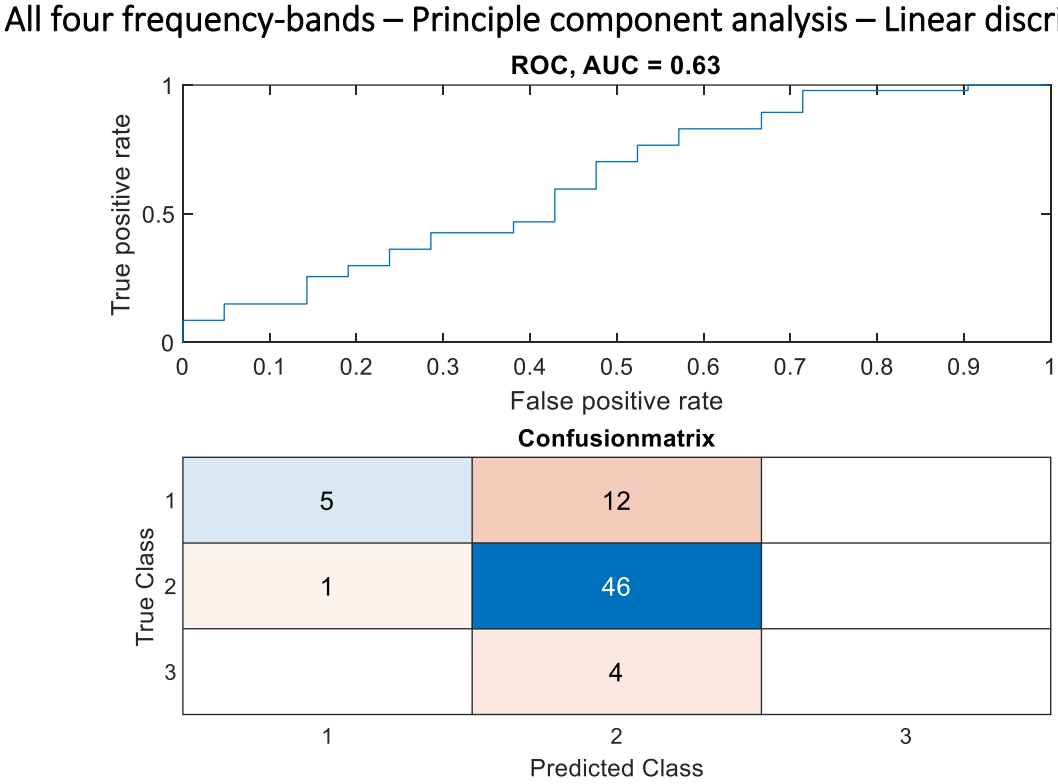


Figure 2. Receiver operator curve and confusion matrix for linear discriminant analysis model in combination with principle component analysis on the features from the theta, alpha, beta as well as gamma frequency bands.

F. Ranking method results

Table 1. Overview of ranking method results as well as the presence of stun effect and beta peak activity after removal of aperiodic signal per hemisphere for all patients.

					Pattern based		Decision trees		DETEC algorithm	
	Patient	Hemi-sphere	Stun effect?	Beta peak?	AUC_flat	Max.	AUC_flat	Max.	Norm-low	Norm-peak
Train data	NL_007	LH	Yes	Little	2-3-0-1	0-1-2-3	1 or 2 (or 0)	1 or 2	2-0-1-3	0-1-3-2
		RH	Yes	Little	3-2-0-1	1-0-3-2	2 or 3	1 or 2	0-2-3-1	0-1-2-3
	NL_008	LH	Yes	No	2-1-3-0	2-1-3-0	1 or 2 (or 3)	1 or 2	3-2-0-1	3-2-0-1
		RH	Yes	No	2-1-3-0	2-3-0-1	1 or 2 (or 3)	1 or 2	3-0-1-2	0-1-2-3
	NL_016	LH	No	Yes	2-3-1-0	2-3-1-0	2 or 3 (or 1)	2 or 3 (or 1)	3-2-1-0	3-2-1-0
		RH	No	Yes	3-2-1-0	1-3-2-0	2 or 3 (or 1)	2 or 3 (or 1)	3-2-0-1	3-2-0-1
	NL_017	LH	No	Yes	2-1-3-0	2-1-3-0	1 or 2	1 or 2	2-1-3-0	1-2-3-0
		RH	No	Yes	1-0-3-2	1-0-3-2	0 or 1	0 or 1	0-1-2-3	0-1-2-3
	NL_018	LH	Yes	Little	2-3-0-1	3-1-2-0	2 or 3 (or 1)	2 or 3 (or 1)	3-0-1-2	3-2-0-1
		RH	Yes	Little	3-2-0-1	1-0-3-2	2 or 3 (or 1)	1 or 2	2-0-3-1	0-3-1-2
	NL_019	LH	No	Yes	1-2-0-3	2-1-3-0	1 or 2	1 or 2	1-2-3-0	2-1-3-0
		RH	No	Yes	2-1-3-0	2-1-3-0	1 or 2	1 or 2	3-1-0-2	3-1-2-0
	NL_020	LH	No	Yes	2-1-3-0	2-3-1-0	1 or 2	1 or 2	0-3-2-1	3-2-1-0
		RH	No	Yes	3-2-0-1	2-3-0-1	2 or 3	1 or 2	3-2-0-1	3-2-0-1
	NL_022	LH	Yes	No	2-3-1-0	2-3-0-1	1 or 2	1 or 2	3-0-2-1	1-2-0-3
		RH	No	Yes	2-3-1-0	2-1-3-0	1 or 2	1 or 2	2-1-3-0	3-2-0-1
	NL_023	LH	Yes	Little	0-1-2-3	3-2-0-1	1 or 2	1 or 2	3-0-1-2	1-2-0-3
		RH	No	Yes	2-1-3-0	2-1-3-0	1 or 2	1 or 2	2-3-1-0	0-2-1-3
	NL_024	LH	Yes	Yes	2-3-1-0	2-1-3-0	2 or 3	2 or 3 (or 1)	1-0-2-3	1-0-2-3
		RH	Yes	Yes	1-2-0-3	1-0-2-3	0 or 1 (or 2 or 3)	1 or 2	3-1-2-0	2-0-1-3
	NL_027	LH	No	No	3-2-1-0	2-1-3-0	2 or 3 (or 1)	1 or 2	3-0-2-1	2-1-3-0
		RH	No	Yes	1-2-3-0	1-0-2-3	1 or 2	1 or 2	0-3-1-2	0-1-3-2
	NL_028	LH	Yes	No	1-2-0-3	2-3-0-1	0 or 1 (or 2)	1 or 2	3-0-2-1	3-2-0-1
		RH	No	Yes	2-3-1-0	2-3-0-1	2 or 3	1 or 2	3-0-2-1	3-2-1-0
	NL_031	LH	No	Yes	2-3-0-1	2-3-0-1	2 or 3	2 or 3	3-2-0-1	3-2-1-0
		RH	No	Yes	2-1-3-0	2-3-1-0	2 *or 3* (or 1)	2 *or 3* (or 1)	3-1-2-0	3-1-2-0
	NL_037	LH	No	Yes	0-1-2-3	1-0-3-2	0 or 1	0 or 1	1-2-0-3	1-2-0-3
		RH	No	Yes	2-1-0-3	1-2-0-3	1 or 2	*0 or* 1 (or 2)	1-2-3-0	1-2-3-0
	NL_042	LH	No	Yes	2-3-1-0	2-3-1-0	2 *or 3* (or 1)	2 *or 3* (or 1)	1-3-2-0	0-1-2-3
		RH	No	Yes	2-3-1-0	2-3-1-0	2 or 3	2 *or 3* (or 1)	3-1-2-0	1-0-3-2
	NL_044	LH	No	Yes	2-1-3-0	2-1-0-3	2 *or 3* (or 1)	1 or 2 (or 0)	0-3-2-1	3-2-0-1
		RH	No	Yes	2-1-3-0	1-0-2-3	1 or 2 (or 0)	1 or 2	3-2-1-0	3-2-1-0
	NL_045	LH	No	Yes	2-3-1-0	2-3-1-0	2 or 3	2 or 3	3-2-1-0	3-2-1-0
		RH	Yes	Yes	2-1-3-0	2-3-0-1	1 or 2	1 or 2	3-2-1-0	3-2-1-0
	NL_046	LH	No	No	1-2-3-0	2-3-0-1	1 or 2	1 or 2	3-2-0-1	2-0-1-3
		RH	Yes	Yes	1-2-3-0	1-2-3-0	1 or 2	1 or 2	0-3-1-2	2-3-1-0
	NL_048	LH	No	Yes	2-1-3-0	2-3-1-0	1 or 2	2 *or 3* (or 1)	3-2-1-0	0-1-2-3
		RH	No	Yes	2-3-1-0	2-3-0-1	2 or 3	2 or 3	3-2-0-1	3-2-0-1
	NL_049	LH	No	Yes	2-3-1-0	2-3-1-0	2 or 3	2 or 3 (or 1)	3-2-0-1	3-2-1-0
		RH	No	Yes	2-1-3-0	2-3-1-0	2 or 3	2 *or 3* (or 1)	3-1-2-0	1-3-2-0

	NL_051	LH	No	Yes	2-1-3-0	2-3-0-1	2 or 3	1 or 2	3-0-2-1	1-2-3-0
		RH	No	Yes	1-2-0-3	1-2-0-3	*0 or* 1 (or 2)	1 or 2	3-2-0-1	3-2-1-0
	NL_052	LH	Yes	Yes	2-3-1-0	2-3-1-0	2 or 3 (or 1)	2 or 3 (or 1)	3-0-1-2	3-0-2-1
		RH	Yes	Yes	3-2-1-0	2-3-1-0	2 or 3 (or 1)	2 or 3	3-2-1-0	3-0-2-1
	NL_053	LH	No	Yes	1-0-3-2	1-0-2-3	0 or 1	0 or 1	0-2-1-3	0-2-1-3
		RH	No	Yes	2-3-1-0	2-1-3-0	2 or 3	2 *or 3* (or 1)	3-2-1-0	3-2-1-0
	NL_054	LH	No	Yes	2-3-1-0	2-3-1-0	2 or 3	1 or 2	0-1-2-3	3-1-2-0
		RH	No	Yes	2-3-1-0	2-3-1-0	2 *or 3* (or 1)	2 *or 3* (or 1)	0-2-3-1	3-1-2-0
	NL_055	LH	No	Yes	2-3-1-0	2-3-0-1	2 or 3	1 or 2	3-2-1-0	3-2-1-0
		RH	No	Little	3-2-1-0	2-3-0-1	2 or 3	1 or 2	3-2-1-0	0-1-2-3
	NL_056	LH	No	Little	3-0-1-2	2-1-0-3	1 or 2 (or 3)	1 or 2	3-1-0-2	1-2-3-0
		RH	No	Little	2-3-1-0	1-0-3-2	2 or 3 (or 1)	1 or 2	0-2-3-1	3-1-2-0
	NL_057	LH	Yes	Yes	1-0-2-3	1-0-3-2	0 or 1 (or 2)	0 or 1	3-2-1-0	2-1-0-3
		RH	Yes	Yes	1-2-0-3	1-0-3-2	1 or 2	1 or 2	0-2-1-3	3-1-0-2
	NL_058	LH	Yes	Yes	1-0-2-3	1-0-3-2	0 or 1	1 or 2	3-2-0-1	1-0-3-2
		RH	Yes	Yes	1-0-2-3	1-0-2-3	1 or 2	0 or 1	1-0-2-3	3-2-1-0
	NL_059	LH	Yes	Little	2-3-1-0	1-0-3-2	2 or 3	0 or 1	2-1-3-0	2-1-3-0
		RH	No	Yes	2-3-1-0	2-3-1-0	2 or 3 (or 1)	2 or 3 (or 1)	3-2-1-0	3-2-1-0
Test data	NL_065	LH	Yes	Yes	2-3-1-0	2-3-1-0	2 or 3	2 *or 3*	1-2-0-3	3-0-2-1
		RH	Yes	Yes	1-2-3-0	1-0-2-3	1 or 2	1 or 2	3-2-1-0	1-2-3-0
	NL_068	LH	No	Yes	2-3-1-0	1-2-3-0	1 or 2	1 or 2	3-2-0-1	3-2-0-1
		RH	Yes	Yes	1-2-0-3	1-0-2-3	*0 or* 1 (or 2 or 3)	0 or 1	2-1-0-3	0-1-2-3
	NL_069	LH	NA	Yes	2-3-1-0	2-3-1-0	2 or 3	2 or 3	3-2-1-0	3-2-1-0
		RH	NA	Yes	2-3-1-0	2-3-0-1	2 or 3	2 or 3	3-2-1-0	3-2-1-0
	NL_070	LH	Yes	Yes	2-1-3-0	2-1-3-0	1 or 2 (or 0)	2 *or 3* (or 1)	3-2-1-0	3-2-1-0
		RH	Yes	No	2-1-3-0	2-3-1-0	1 or 2	1 or 2	3-2-1-0	2-0-1-3
	NL_071	LH	No	Yes	1-2-0-3	1-2-0-3	1 or 2	*0 or* 1 (or 2 or 3)	2-3-1-0	2-3-0-1
		RH	Yes	Yes	2-3-1-0	2-3-0-1	2 or 3	1 or 2	3-2-1-0	0-1-2-3

Abbreviations: LH: left hemisphere; RH: right hemisphere; NA: not available

Legend	
	1st prediction = clinically chosen contact
	2nd prediction = clinically chosen contact / no stun effect / beta peak present
	3rd prediction = clinically chosen contact / little beta peak present
	4th prediction = clinically chosen contact / stun effect / no beta peak present
2 or 1 or 3	Clinically chosen contact number
Red number(s)	Contact eliminated / Not used for DETEC

Table 2. Pattern based ranking method results per ranked prediction using AUC_flat for all data as well as the with/without stun effect and clear/little/no beta above 1/f subgroups.

	1st	2nd	3rd	4th	1st or 2nd
All data (Tr=58/T=10)	Tr = 32 (55.2%) / T = 9 (90.0%)	Tr = 18 (31.0%) / T = 1 (10.0%)	Tr = 7 (12.1%) / T = 0 (0.0%)	Tr = 1 (1.7%) / T = 0 (0.0%)	Tr = 50 (86.2%) / T = 10 (100%)
Without stun (Tr=38/T=2)	Tr = 24 (63.2%) / T = 2 (100%)	Tr = 10 (26.3%) / T = 0 (0.0%)	Tr = 3 (7.9%) / T = 0 (0.0%)	Tr = 1 (2.6%) / T = 0 (0.0%)	Tr = 34 (89.5%) / T = 2 (100%)
With stun (Tr=20/T=8)	Tr = 8 (40.0%) / T = 7 (87.5%)	Tr = 8 (40.0%) / T = 1 (12.5%)	Tr = 4 (20.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 16 (80.0%) / T = 8 (100%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 27 (62.8%) / T = 8 (88.9%)	Tr = 11 (25.6%) / T = 1 (11.1%)	Tr = 4 (9.3%) / T = 0 (0.0%)	Tr = 1 (2.3%) / T = 0 (0.0%)	Tr = 38 (88.4%) / T = 9 (100%)
Little beta above 1/f (Tr=9/T=0)	Tr = 4 (44.4%) / T = NA	Tr = 3 (33.3%) / T = NA	Tr = 2 (22.2%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 7 (77.8%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 1 (16.7%) / T = 1 (100%)	Tr = 4 (66.7%) / T = 0 (0.0%)	Tr = 1 (16.7%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 5 (83.3%) / T = 1 (100%)

Abbreviations: AUC_flat: area under the receiver operator curve after removal of 1/frequency; 1st/2nd/3rd/4th: $1^{st}/2^{nd}/3^{rd}/4^{th}$ prediction; Tr: training set (size); T: test set (size); NA: not available; 1/f: 1/frequency (aperiodic signal)

Table 3. Pattern based ranking method results per ranked prediction using Max. for all data as well as the with/without stun effect and clear/little/no beta above 1/f subgroups.

	1st	2nd	3rd	4th	1st or 2nd
All data (Tr=58/T=10)	Tr = 35 (60.3%) / T = 8 (80.0%)	Tr = 8 (13.8%) / T = 1 (10.0%)	Tr = 8 (13.8%) / T = 1 (10.0%)	Tr = 7 (12.1%) / T = 0 (0.0%)	Tr = 43 (74.1%) / T = 9 (90.0%)
Without stun (Tr=38/T=2)	Tr = 26 (68.4%) / T = 1 (50%)	Tr = 5 (13.2%) / T = 1 (50.0%)	Tr = 4 (10.5%) / T = 0 (0.0%)	Tr = 3 (7.9%) / T = 0 (0.0%)	Tr = 31 (81.6%) / T = 2 (100%)
With stun (Tr=20/T=8)	Tr = 9 (45.0%) / T = 7 (87.5%)	Tr = 3 (15.0%) / T = 0 (0.0%)	Tr = 3 (15.0%) / T = 1 (12.5%)	Tr = 4 (20.0%) / T = 0 (0.0%)	Tr = 12 (60.0%) / T =7 (87.5%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 30 (69.8%) / T = 7 (77.8%)	Tr = 5 (11.6%) / T = 1 (11.1%)	Tr = 6 (14.0%) / T = 1 (11.1%)	Tr = 2 (4.65%) / T = 0 (0.0%)	Tr = 35 (81.4%) / T = 8 (88.9%)
Little beta above 1/f (Tr=9/T=0)	Tr = 2 (22.2%) / T = NA	Tr = 1 (11.1%) / T = NA	Tr = 2 (22.2%) / T = NA	Tr = 4 (44.4%) / T = NA	Tr = 3 (33.3%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 3 (50.0%) / T = 1 (100%)	Tr = 2 (33.3%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 1 (16.7%) / T = 0 (0.0%)	Tr = 5 (83.3%) / T = 1 (100%)

Abbreviations: Max.: maximum; 1st/2nd/3rd/4th: $1^{st}/2^{nd}/3^{rd}/4^{th}$ prediction; Tr: training set (size); T: test set (size); not available; 1/f: 1/frequency (aperiodic signal)

Table 4. Decision tree ranking method results per ranked prediction using AUC_flat for all data as well as the with/without stun effect and clear/little/no beta above 1/f subgroups.

	1st*	2nd	3rd	4th	1st or 2nd
All data (Tr=58/T=10)	Tr = 6 (10.3%) / T = 1 (10.0%)	Tr = 43 (74.1%) / T = 9 (90.0%)	Tr = 4 (6.9%) / T = 0 (0.0%)	Tr = 5 (8.6%) / T = 0 (0.0%)	Tr = 49 (84.5%) / T = 10 (100%)
Without stun (Tr=38/T=2)	Tr = 5 (13.2%) / T = 0 (0.0%)	Tr = 28 (73.7%) / T = 2 (100%)	Tr = 2 (5.3%) / T = 0 (0.0%)	Tr = 3 (7.9%) / T = 0 (0.0%)	Tr = 33 (86.8%) / T = 2 (100%)
With stun (Tr=20/T=8)	Tr = 1 (5.0%) / T = 1 (12.5%)	Tr = 15 (75.0%) / T = 7 (87.5%)	Tr = 2 (10.0%) / T = 0 (0.0%)	Tr = 2 (10.0%) / T = 0 (0.0%)	Tr = 16 (80.0%) / T = 8 (100%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 6 (14.0%) / T = 1 (11.1%)	Tr = 31 (81.6%) / T = 9 (88.9%)	Tr = 2 (4.7%) / T = 0 (0.0%)	Tr = 4 (9.3%) / T = 0 (0.0%)	Tr = 37 (85.4%) / T = 9 (100%)
Little beta above 1/f (Tr=9/T=0)	Tr = 0 (0.0%) / T = NA	Tr = 8 (88.9%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 1 (11.1%) / T = NA	Tr = 8 (88.9%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 4 (66.7%) / T = 1 (100%)	Tr = 2 (33.3%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 4 (66.7%) / T = 1 (100%)

** Decision tree techniques often leads to prediction of the two most optimal contacts, which were in this case places under "2nd" predictions and not "1st" predictions, leading to less "1st" predictions and more "2nd" predictions. Abbreviations: AUC_flat: area under the receiver operator curve after removal of 1/frequency; 1st/2nd/3rd/4th: 1st/2nd/3rd/4th prediction; Tr: training set (size); T: test set (size); NA: not available; 1/f: 1/frequency (aperiodic signal)*

Table 5. Decision tree ranking method results per ranked prediction using Max. for all data as well as the with/without stun effect and clear/little/no beta above 1/f subgroups.

	1st*	2nd	3rd	4th	1st or 2nd
All data (Tr=58/T=10)	Tr = 8 (13.8%) / T = 3 (30.0%)	Tr = 43 (74.1%) / T = 7 (70.0%)	Tr = 1 (1.7%) / T = 0 (0.0%)	Tr = 6 (10.3%) / T = 0 (0.0%)	Tr = 51 (87.9%) / T = 10 (100%)
Without stun (Tr=38/T=2)	Tr = 7 (18.4%) / T = 1 (50%)	Tr = 25 (65.8%) / T = 1 (50.0%)	Tr = 1 (2.6%) / T = 0 (0.0%)	Tr = 5 (13.2%) / T = 0 (0.0%)	Tr = 32 (84.2%) / T = 2 (100%)
With stun (Tr=20/T=8)	Tr = 1 (5.0%) / T = 2 (25.0%)	Tr = 18 (90.0%) / T = 6 (75.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 1 (5.0%) / T = 0 (0.0%)	Tr = 19 (95.0%) / T = 8 (100%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 8 (18.6%) / T = 3 (33.3%)	Tr = 29 (67.4%) / T = 6 (66.7%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 5 (11.6%) / T = 0 (0.0%)	Tr = 37 (86.1%) / T = 9 (100%)
Little beta above 1/f (Tr=9/T=0)	Tr = 0 (0.0%) / T = NA	Tr = 8 (88.9%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 1 (11.1%) / T = NA	Tr = 8 (88.9%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 6 (100%) / T = 1 (100%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 6 (100%) / T = 1 (100%)

** Decision tree techniques often leads to prediction of the two most optimal contacts, which were in this case places under "2nd" predictions and not "1st" predictions, leading to less "1st" predictions and more "2nd" predictions. Abbreviations: Max.: maximum; 1st/2nd/3rd/4th: 1st/2nd/3rd/4th prediction; Tr: training set (size); T: test set (size); NA: not available; 1/f: 1/frequency (aperiodic signal)*

Table 6. DETEC algorithm reference method results per ranked prediction for all data as well as the with/without stun effect and clear/little/no beta above 1/f subgroups.

	1st	2nd	3rd	4th	1st or 2nd
All data (Tr=58/T=10)	Tr = 10 (17.2%) / T = 0 (0.0%)	Tr = 23 (39.7%) / T = 7 (70.0%)	Tr = 17 (29.3%) / T = 2 (20.0%)	Tr = 8 (13.8%) / T = 1 (10.0%)	Tr = 33 (56.9%) / T = 7 (70.0%)
Without stun (Tr=38/T=2)	Tr = 7 (18.4%) / T = 0 (0.0%)	Tr = 18 (47.4%) / T = 1 (50.0%)	Tr = 9 (23.7%) / T = 1 (50.0%)	Tr = 4 (10.5%) / T = 0 (0.0%)	Tr = 25 (65.8%) / T = 1 (50.0%)
With stun (Tr=20/T=8)	Tr = 3 (15.0%) / T = 0 (0.0%)	Tr = 5 (25.0%) / T = 6 (75.0%)	Tr = 8 (40.0%) / T = 2 (25.0%)	Tr = 4 (20.0%) / T = 0 (0.0%)	Tr = 8 (40.0%) / T = 6 (75.0%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 8 (18.6%) / T = 0 (0.0%)	Tr = 19 (44.2%) / T = 6 (66.7%)	Tr = 12 (27.9%) / T = 2 (22.2%)	Tr = 4 (9.3%) / T = 1 (11.1%)	Tr = 27 (62.8%) / T = 6 (66.7%)
Little beta above 1/f (Tr=9/T=0)	Tr = 1 (11.1%) / T = NA	Tr = 3 (33.3%) / T = NA	Tr = 3 (33.3%) / T = NA	Tr = 2 (22.2%) / T = NA	Tr = 4 (44.4%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 1 (16.7%) / T = 0 (0.0%)	Tr = 1 (16.7%) / T = 1 (100%)	Tr = 2 (33.3%) / T = 0 (0.0%)	Tr = 2 (33.3%) / T = 0 (0.0%)	Tr = 2 (33.3%) / T = 1 (100%)

Abbreviations: 1st/2nd/3rd/4th: $1^{st}/2^{nd}/3^{rd}/4^{th}$ prediction; Tr: training set (size); T: test set (size); NA: not available; 1/f: 1/frequency (aperiodic signal)

Table 7. Pattern based ranking method results for each clinically chosen contact per ranked prediction using AUC_flat for all data.

	1st	2nd	3rd	4th	1st or 2nd
Contact #3 (Tr=4/T=0)	Tr = 1 (25.0%) / T = NA	Tr = 2 (50.0%) / T = NA	Tr = 1 (25.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 3 (75.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 26 (66.7%) / T = 7 (87.5%)	Tr = 9 (23.1%) / T = 1 (12.5%)	Tr = 3 (7.7%) / T = 0 (0.0%)	Tr = 1 (2.6%) / T = 0 (0.0%)	Tr = 35 (89.7%) / T = 8 (100%)
Contact #1 (Tr=15/T = 2)	Tr = 5 (33.3%) / T = 2 (100%)	Tr = 7 (46.7%) / T = 0 (0.0%)	Tr = 3 (20.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 12 (80.0%) / T =2 (100%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA

Abbreviations: AUC_flat: area under the receiver operator curve after removal of 1/frequency; 1st/2nd/3rd/4th: $1^{st}/2^{nd}/3^{rd}/4^{th}$ prediction; Tr: training set (size); T: test set (size); NA: not available

Table 8. Pattern based ranking method results for each clinically chosen contact per ranked prediction using Max. for all data.					
	1st	2nd	3rd	4th	1st or 2nd
Contact #3 (Tr=4/T=0)	Tr = 0 (0.0%) / T = NA	Tr = 2 (50.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 2 (50.0%) / T = NA	Tr = 2 (50.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 27 (69.2%) / T = 6 (75%)	Tr = 2 (5.1%) / T = 1 (12.5%)	Tr = 6 (15.4%) / T = 1 (12.5%)	Tr = 4 (10.3%) / T = 0 (0.0%)	Tr = 29 (74.4%) / T = 7 (87.5%)
Contact #1 (Tr=15/T = 2)	Tr = 8 (53.3%) / T = 2 (100%)	Tr = 4 (26.7%) / T = 0 (0.0%)	Tr = 2 (13.3%) / T = 0 (0.0%)	Tr = 1 (6.7%) / T = 0 (0.0%)	Tr = 12 (80.0%) / T = 2 (100%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA
Abbreviations: Max.: maximum; 1st/2nd/3rd/4th: 1st/2nd/3rd/4th prediction; Tr: training set (size); T: test set (size); NA: not available					

Table 9. Decision tree ranking method results for each clinically chosen contact per ranked prediction using AUC_flat for all data.					
	1st*	2nd	3rd	4th	1st or 2nd
Contact #3 (Tr=4/T=0)	Tr = 0 (0.0%) / T = NA	Tr = 3 (75.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 1 (25.0%) / T = NA	Tr = 3 (75.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 4 (10.3%) / T = 8 (100%)	Tr =31 (79.5%) / T = 0 (0.0%)	Tr = 2 (5.1%) / T = 0 (0.0%)	Tr = 2 (5.1%) / T = 0 (0.0%)	Tr = 35 (89.7%) / T = 8 (100%)
Contact #1 (Tr=15/T = 2)	Tr = 1 (6.7%) / T = 1(50.0%)	Tr = 10 (66.7%) / T = 1 (50.0%)	Tr = 2 (13.3%) / T = 0 (0.0%)	Tr = 2 (13.3%) / T = 0 (0.0%)	Tr = 11 (73.3%) / T =2 (100%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA
** Decision tree techniques often leads to prediction of the two most optimal contacts, which were in this case places under "2nd" predictions and not "1st" predictions, leading to less "1st" predictions and more "2nd" Abbreviations: AUC_flat: area under the receiver operator curve after removal of 1/frequency; 1st/2nd/3rd/4th: 1st/2nd/3rd/4th prediction; Tr: training set (size); T: test set (size); NA: not available*					

Table 10. Decision tree ranking method results for each clinically chosen contact per ranked prediction using Max. for all data.

	1st*	2nd	3rd	4th	1st or 2nd
Contact #3 (Tr=4/T=0)	Tr = 0 (0.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 4 (100.0%) / T = NA	Tr = 0 (0.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 6 (15.4%) / T = 2 (20.0%)	Tr = 31 (79.5%) / T = 6 (80.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 2 (5.1%) / T = 0 (0.0%)	Tr = 37 (94.9%) / T = 8 (100%)
Contact #1 (Tr=15/T = 2)	Tr = 1 (6.7%) / T = 1 (50.0%)	Tr = 13 (86.7%) / T = 1 (50.0%)	Tr = 1 (6.7%) / T = 0 (0.0%)	Tr = (0.0%) / T = 0 (0.0%)	Tr = 14 (93.3%) / T = 2 (100%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA

** Decision tree techniques often leads to prediction of the two most optimal contacts, which were in this case places under "2nd" predictions and not "1st" predictions, leading to less "1st" predictions and more "2nd"*
Abbreviations: Max.: maximum; 1st/2nd/3rd/4th: $1^{st}/2^{nd}/3^{rd}/4^{th}$ prediction; Tr: training set (size); T: test set (size); NA: not available

Table 11. DETEC algorithm reference method results for each clinically chosen contact per ranked prediction for all data.

	1st	2nd	3rd	4th	1st or 2nd
Contact #3 (Tr=4/T=0)	Tr = 2 (50.0%) / T = NA	Tr = 1 (25.0%) / T = NA	Tr = 1 (25.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 3 (75.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 5 (12.8%) / T = 0 (0.0%)	Tr =18 (46.2%) / T = 6 (80.0%)	Tr = 13 (33.3%) / T = 2 (20.0%)	Tr = 3 (7.7%) / T = 0 (0.0%)	Tr = 23 (59.0%) / T = 6 (80%)
Contact #1 (Tr=15/T = 2)	Tr = 3 (20.0%) / T = 0 (0.0%)	Tr = 4 (26.7%) / T = 1 (50.0%)	Tr = 3 (20.0%) / T = 0 (0.0%)	Tr = 5 (33.3%) / T = 1 (50.0%)	Tr = 7 (46.7%) / T =1 (50%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA

Abbreviations: 1st/2nd/3rd/4th: $1^{st}/2^{nd}/3^{rd}/4^{th}$ prediction; Tr: training set (size); T: test set (size); NA: not available

Table 12. Fixed ranking #2→#1→#3→#0 method results per ranked prediction for all data as well as the with/without stun effect and clear/little/no beta above 1/f subgroups.

	1st	2nd	3rd	4th	1st or 2nd
All data (Tr=58/T=10)	Tr = 39 (67.2%) / T = 8 (80.0%)	Tr = 15 (25.9%) / T = 2 (20.0%)	Tr = 4 (6.9%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 54 (93.1%) / T = 10 (100%)
Without stun (Tr=38/T=2)	Tr = 24 (63.2%) / T = 1 (50%)	Tr = 10 (26.3%) / T = 1 (50.0%)	Tr = 4 (10.5%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 34 (89.5%) / T = 2 (100%)
With stun (Tr=20/T=8)	Tr = 15 (75.0%) / T = 7 (87.5%)	Tr = 5 (25.0%) / T = 1 (12.5%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 20 (100%) / T = 8 (100%)
Clear beta above 1/f (Tr=43/T=9)	Tr = 29 (67.4%) / T = 7 (77.8%)	Tr = 11 (25.6%) / T = 2 (22.2%)	Tr = 3 (7.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 40 (93.0%) / T = 9 (100%)
Little beta above 1/f (Tr=9/T=0)	Tr = 7 (77.8%) / T = NA	Tr = 1 (11.1%) / T = NA	Tr = 1 (11.1%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 8 (88.9%) / T = NA
No beta above 1/f (Tr=6/T=1)	Tr = 3 (50.0%) / T = 1 (100%)	Tr = 3 (50.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 6 (100%) / T = 1 (100%)

Abbreviations: 1st/2nd/3rd/4th: 1st/2nd/3rd/4th prediction; Tr: training set (size); T: test set (size); NA: not available; 1/f: 1/frequency (aperiodic signal)

Table 13. Fixed ranking #2→#1→#3→#0 method results for each clinically chosen contact per ranked prediction for all data.

	1st	2nd	3rd	4th	1st or 2nd
Contact #3 (Tr=4/T=0)	Tr = 0 (0.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 4 (100.0%) / T = NA	Tr = 0 (0.0%) / T = NA	Tr = 0 (0.0%) / T = NA
Contact #2 (Tr=39/T = 8)	Tr = 39 (100%) / T = 8 (100%)	Tr = (0.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 39 (100%) / T = 8 (100%)
Contact #1 (Tr=15/T = 2)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 15 (100%) / T = 2 (100%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 0 (0.0%) / T = 0 (0.0%)	Tr = 15 (100%) / T =2 (100%)
Contact #0 (Tr = 0/T = 0)	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA	Tr = NA / T = NA

Abbreviations: 1st/2nd/3rd/4th: 1st/2nd/3rd/4th prediction; Tr: training set (size); T: test set (size); NA: not available

G. Sub-analyses results

Table 1. Individual bipolar impedance measurements per channel for all subthalamic nuclei with one or more channels with bipolar impedance(s) ≥ 5000 Ω.

Channel		NL022_R	NL024_L	NL024_R	NL027_R	NL028_L	NL045_L	NL049_L	NL053_L	NL053_R	NL054_R	NL055_L	NL058_R	NL058_R	NL071_L
0	1a	4784	3839	3090	2661	4989	3829	3461	4593	3705	3667	3802	2639	3846	3988
0	1b	4545	3322	2301	2729	5516	4436	4447	3462	3312	4241	3816	3247	4096	3073
0	1c	4440	3403	3594	3847	5146	3785	3529	3612	4245	4169	2852	2798	5404	3333
0	2a	5217	3703	2600	2606	4951	4002	3488	4514	3692	4172	4388	3251	5180	3729
0	2b	4389	4406	2396	2988	5612	4324	4041	3667	3230	4323	3981	3881	5081	3004
0	2c	4055	4600	3050	3851	5081	3580	3051	3913	3852	4446	2990	3927	5810	3210
0	3	2699	2287	1647	1666	3143	2289	2327	2383	2110	2444	1998	1994	2823	2061
1a	1b	6563	5002	3662	3928	7535	6017	5113	5319	4538	5634	5505	3968	5367	4541
1a	1c	6280	5117	5019	4788	7408	5172	3983	5319	5678	5497	4642	3845	6471	4865
1a	2a	7213	5440	3934	3833	7121	5742	4017	6398	5084	5419	6164	3964	6522	5299
1a	2b	6471	6124	3852	4297	8083	6169	4731	5483	4572	5736	5843	4696	6508	4667
1a	2c	6160	6281	4563	5095	7487	5213	3789	5705	5330	5781	4891	4965	7221	4944
1a	3	4808	4086	3199	3001	5649	4092	3065	4343	3602	3868	4003	3110	4409	3823
1b	1c	6290	4641	4127	4839	7918	6142	5103	4012	5261	6047	4571	4370	6566	4001
1b	2a	7220	5049	3267	3983	8007	6526	5059	5292	4712	6105	6372	4754	6665	4660
1b	2b	6266	5498	2981	4260	8421	6746	5509	4197	4077	6122	5645	5224	6372	3948
1b	2c	6051	5788	3675	5146	8110	5851	4621	4490	4859	6320	4871	5565	7238	4165
1b	3	4716	3594	2355	3070	6241	4739	3877	3243	3229	4415	4081	3777	4420	3141
1c	2a	7098	5151	4726	4778	7641	5817	4195	5384	5712	6030	5467	4492	8239	4978
1c	2b	6256	5716	4508	5149	8271	6129	4802	4357	5104	6160	4887	5054	8035	4206
1c	2c	5816	5815	5053	5871	7408	4992	3707	4456	5732	6075	3838	5132	8710	4377
1c	3	4593	3730	3947	3796	5858	3990	3133	3284	4073	4316	3133	3407	5994	3483
2a	2b	6631	5716	2913	3956	7374	5895	4437	5241	4214	5795	6130	4894	7309	4343
2a	2c	6235	5716	3662	4737	6724	4680	3325	5357	5077	5808	5122	5313	7841	4561
2a	3	5067	3703	2287	2647	5092	3932	2792	4053	3734	3923	4485	3527	5329	3701
2b	2c	5527	6220	3430	5190	7634	5046	4044	4384	4487	6160	4473	5952	7831	3988
2b	3	4324	4273	2150	3206	5873	4216	3461	3283	2833	4200	3913	4020	5227	2908
2c	3	4123	4545	2978	3922	5375	3464	2369	3651	3598	4343	2922	4418	6263	3249

Legend	
	Bipolar impedance < 3000 Ω
	Bipolar impedance ≥ 3000 Ω & < 5000 Ω
	Bipolar impedance ≥ 5000 Ω & < 7000 Ω
	Bipolar impedance ≥ 7000 Ω
Red text	Incorrect prediction

Table 2. Segment LFP recordings predictions based on either the level with the highest segment LFP peak of highest segment LFP average, as well as clinically chosen contact and best level-based LFP prediction method with a single optimal prediction per hemisphere for a total 46 STN, for which the clinically chosen contact was either contact 1 or 2 and segment LFP recordings were available.

		Pattern based (level)	Level with highest segment LFP peak		Level with highest segment LFP average	
STN	MPR contact	AUC_flat	AUC_flat	Max.	AUC_flat	Max.
NL_007_L	2	2	2	1	2	1
NL_007_R	2	3	1	1	1	1
NL_008_L	1	1	1	2	1	2
NL_008_R	1	1	2	2	2	2
NL_016_L	2	2	2	2	2	2
NL_016_R	1	3	2	2	1	2
NL_017_L	1	2	1	1	1	1
NL_017_R	1	1	1	1	1	1
NL_018_L	2	2	2	1	1	1
NL_018_R	2	3	1	1	1	1
NL_019_L	1	1	2	1	2	1
NL_019_R	2	2	2	2	2	2
NL_020_L	1	2	1	1	1	1
NL_020_R	2	3	1	1	1	1
NL_022_L	2	2	2	2	1	2
NL_022_R	2	2	2	2	2	2
NL_023_L	2	0	2	1	1	2
NL_023_R	2	2	1	2	1	2
NL_028_L	2	1	1	1	1	1
NL_031_L	2	2	2	2	2	2
NL_031_R	2	2	1	1	1	1
NL_044_L	2	2	1	1	1	1
NL_044_R	2	2	1	2	1	2
NL_045_L	2	2	1	1	1	1
NL_045_R	2	2	1	2	1	1
NL_046_L	2	1	2	1	1	1
NL_046_R	2	1	2	2	2	2
NL_051_L	2	2	1	1	1	1
NL_053_L	2	1	1	2	1	1
NL_053_R	2	2	1	1	1	1
NL_054_R	2	2	1	1	1	1
NL_056_R	2	2	1	1	2	1

NL_057_L	1	1	2	2	2	2
NL_057_R	1	1	2	2	2	1
NL_058_L	2	1	2	2	1	2
NL_058_R	2	1	2	2	2	1
NL_065_L	2	2	1	1	1	1
NL_065_R	2	1	1	1	1	1
NL_068_L	2	2	1	2	1	2
NL_068_R	1	1	2	2	2	2
NL_069_L	2	2	1	1	1	1
NL_069_R	2	2	2	1	1	1
NL_070_L	2	2	1	1	1	1
NL_070_R	2	2	1	2	1	2
NL_071_L	1	1	1	2	1	2
NL_071_R	2	2	1	1	2	1

Abbreviations: STN: subthalamic nucleus; MPR: monopolar review; AUC_flat: area under the receiver operator curve after removal of 1/frequency; LFP: local field potential; Max.: maximum; _L: left hemisphere; _R: right hemisphere