Vittoria Bucciarelli*, Dorian Vogel, Teresa Nordin, Marc Stawiski, Jérôme Coste, Jean-Jacques Lemaire, Karin Wårdell, Raphael Guzman, Simone Hemm

Predicting Deep Brain Stimulation Outcomes Using Intra-Operative Stimulation Test Data

https://doi.org/10.1515/cdbme-2025-0189

Abstract: Deep Brain Stimulation (DBS) is an effective treatment for movement disorders. Optimizing stimulation parameters remains, however, a trial-and-error process. Data-driven models leveraging Probabilistic Mapping have shown promise in predicting DBS outcomes, yet current studies rely on chronic stimulation data. This study explores the feasibility of using intra-operative stimulation test data for DBS effect prediction. Probabilistic volumes of beneficial and adverse effects were computed from intra-operative stimulation test data of 65 patients (23 with Essential Tremor + 42 with Parkinson's Disease). A prediction dataset was generated including clinical, morphological, stimulation features along with features derived from probabilistic maps and simulated Volumes of Tissue Activated. Three machine learning models (Adaboost, Support Vector Classifier and Naïve Bayes) were implemented to predict stimulation effects in a classification task. The models were validated in a leave-one-out cross-validation and their performances were compared. All the developed models were able to predict DBS outcome classes. The best predictive performance was achieved by the Adaboost model with a maximum balanced accuracy of 0.71 on 3 classes. These results show that intra-operative stimulation test data can predict DBS effects with a similar approach and comparable accuracy to post-operative monopolar review data.

***Vittoria Bucciarelli (corresponding author, vittoria.bucciarelli@fhnw.ch), Dorian Vogel, Marc Stawiski, Karin Wårdell, Simone Hemm:** Institute for Medical Engineering and Medical Informatics, School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland
Teresa Nordin, Karin Wårdell, Simone Hemm: Department of Biomedical Engineering, Linköping University, Linköping, Sweden
Jérôme Coste, Jean-Jacques Lemaire: Université Clermont Auvergne, Clermont Auvergne INP, CHU Clermont-Ferrand, CNRS, Institut Pascal, F-63000 Clermont–Ferrand, France
Raphael Guzman: Department of Neurosurgery, University Hospital Basel, Basel, Switzerland

Keywords: Deep Brain Stimulation, machine learning, effect prediction.

1 Introduction

Deep Brain Stimulation (DBS) is a neurosurgical intervention primarily applied to the treatment of movement disorders such as Parkinson's Disease (PD) or Essential Tremor (ET) [1]. DBS strongly relies on the precise stimulation of target brain structures. However, for most diseases the pathological mechanisms are not completely understood, resulting in varying stimulation targets among clinical centers. While anatomical guidance can inform DBS settings selection, optimizing stimulation parameters to balance symptom relief and side effects often requires iterative adjustments over multiple sessions [2]. To alleviate the complexity of this process, recent research has focused on developing data-driven models capable of predicting the effects of specific DBS settings. These models are trained on feature sets derived from a patient population, often incorporating variables obtained through Probabilistic Mapping approaches [3],[4],[5]. Probabilistic Mapping enables the identification of average brain volumes associated with symptom improvement, known as Probabilistic Sweet Spots (PSS), or adverse effects. Activation of these regions has been demonstrated to be predictive of stimulation outcomes (tremor, rigidity, bradykinesia, axial signs) in Parkinson's Disease [6]. Nevertheless, no standardized DBS effect prediction model or established set of clinical improvement predictors currently exists. Moreover, all the available studies are based on monopolar review data or chronic stimulation parameters. This study aimed to assess the applicability of DBS effect prediction approaches to intra-operative stimulation test data. To achieve this, probabilistic volumes of beneficial and adverse effects were first derived from intra-operative stimulation tests. Subsequently, they were used to train a machine learning model designed to predict the outcomes of specific stimulation configurations.

2 Materials and methods

2.1 Patients and clinical data

The analysis was conducted on data of 65 patients (42 PD + 23 ET, Ptolemee Electrophysiologie project: IRB 5921, CE-CIC-GREN-18-03). The patients underwent bilateral DBS implantation at the Department of Neurosurgery of Clermont-Ferrand University Hospital, France, between 2008 and 2018. 28 patients (23 ET + 5 PD) were targeted in the ventrointermediate nucleus of thalamus (Vim) and 37 (all PD) in the subthalamic nucleus (STN). Pre-operative MR images were used to segment deep brain structures of interest and for the surgery planning. Intra-operative stimulation tests were performed by stimulating with a microelectrode (Alpha-Omega Engineering, Israel; frequency 130 Hz; pulse width 60 μs) in 1-mm steps over 14 mm, increasing the current amplitude between 0.2 and 3 mA in steps of 0.2 mA. Therefore, the tested positions along the trajectory ranged from -10 to +4 with 0 indicating the target designated during planning. Symptom improvement (tremor for Vim-targeted patients, rigidity for STN-targeted patients) was classified as "no improvement" (0%), "poor" (25%), "fair" (50%), "good" (75%), "excellent" (100%), adding some intermediate values if necessary. The lowest current amplitude generating improvement and occurrence of side effects (oculomotor effects, paresthesia, dysarthria) were recorded.

2.2 Probabilistic maps generation

Patients were categorized into Vim-targeted (patients=28, stimulations=878) and STN-targeted (patients=37, stimulations=1234) groups, and probabilistic maps were generated for each cohort. The Probabilistic Mapping workflow is detailed in [7]. Briefly, pre-operative MR T1 images were used to generate patient-specific brain tissue conductivity models with the software ELMA [8] and electric field (EF) spread was simulated in Comsol Multiphysics 5.5 (COMSOL AB, Sweden). The EF was thresholded at 0.2 V mm^{-1} [9] to obtain the Volume of Tissue Activated (VTA). Patients' VTAs were transferred in a group-specific anatomical template [10] and voxel-wise statistical testing was applied to compute the brain volumes associated with high symptom improvement (PSS) and side effects (SE). A Bayesian t-test was applied to the VTAs labelled with an improvement score to extract the PSS. This included voxels with a Bayes Factor (BF)≥10 in favour of the alternative hypothesis H1: "voxel improvement>median(cohort improvement)". VTAs associated with side effects of different types were grouped together to obtain a general side effect volume. Given the binary nature of the side effect information (SE/no SE) a binomial test with False Discovery Rate correction was applied. Voxels with p-value<0.05 were retained as part of the SE. Additionally, an nMap was generated to depict the stimulation frequency per voxel, along with a wMeanMap representing the weighted mean improvement score for each voxel. An example of PSS, SE and VTA is shown in **Figure 1**.

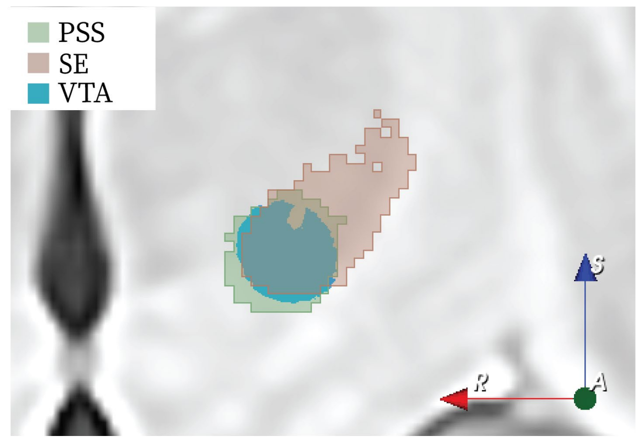

Figure 1: Probabilistic Sweet Spot (green), Side Effects map (brown) and a patient's Volume of Tissue Activated (light blue) in anatomical atlas space (coronal view, zoom on deep brain area).

2.3 DBS effects prediction

A machine learning approach was implemented to predict the effect of a specific VTA and applied separately to the Vim and STN cohort. Datasets with clinical, morphological, stimulation and VTA-related features (**Table 1**) were generated and used to train the model. The target variable was the DBS effect categorized in 3 classes:

- Side effects (class 0)
- No or low improvement (< 50%, class 1)
- High improvement (≥ 50%, class 2)

3 supervised machine learning models based on different underlying principles were implemented and evaluated: AdaBoost (Ada), Support Vector Classifier (SVC) and Naïve Bayes (NB). The Ada classification approach for the three classes was based on a one vs. one logic. This means that a separate binary classifier was trained for each pair of classes, and final predictions were made via majority voting, instead of using a multi-class approach fitting a single model to directly distinguish all classes. Due to the sensitivity of NB and SVC to multiple highly correlated features, these models were provided with the first ten principal components derived from Principal Component Analysis. Synthetic Minority Over-

sampling Technique was applied to account for class imbalance.

The models were validated using leave-one-out cross-validation. In each iteration, probabilistic maps and feature values were computed using n−1 patients for training, while the remaining patient was reserved for testing. Evaluation metrics were balanced accuracy, precision, recall, F1-score. Due to class imbalance in the dataset, balanced accuracy was used as the evaluation metric, as it provides an unbiased estimate of model performance by equally accounting for each class.

Table 1: Features used to train the DBS effect prediction models.

Category	Features list
Clinical	Age, sex, disease duration, pre-operative UPDRSIII, pre-operative levodopa dose
Morphological	Vim size, STN size
Stimulation	Hemisphere Microelectrode trajectory Position on trajectory
VTA-related	VTA volume VTA-PSS centroid distance VTA-PSS overlap (Dice coefficient) mean(BF) in VTA-PSS overlap volume VTA-SE overlap (Dice coefficient) max(wMeanMap value) in VTA mean(wMeanMap value) in VTA max(nMap value) in VTA VTA-atlas target structure overlap (Dice coefficient)

Vim and STN size were calculated from the structures segmentations on the pre-operative MRI.

3 Results

The model achieving the highest balanced accuracy was Ada with scores of 0.71 for the STN cohort and 0.61 for the Vim cohort. The classifications achieved with this model are shown in **Figure 2A** and **Figure 2B**. The diagonals of the confusion matrices show the percentages of correctly classified samples for each class. The average of such values is the balanced accuracy score. For the STN cohort the highest classification performance was achieved for class 2, likely attributable to the larger number of available samples for this class. On the contrary, the best predictive performance was shown on class 0 for the Vim cohort.

The second-best performing model was SVC with 0.69 (STN) and 0.59 (Vim), while NB reached 0.67 (STN) and 0.57 (Vim). The precision, recall and F1-scores of the 3 models are summarized in **Table 2**. Ada outperformed the other models across both cohorts, showing the highest precision, recall, and F1 scores. However, all models experienced a performance drop when moving from the STN to the Vim cohort, with recall being the most impacted. The top three most important features in the prediction were: VTA volume, VTA-PSS centroid distance and VTA-SE overlap.

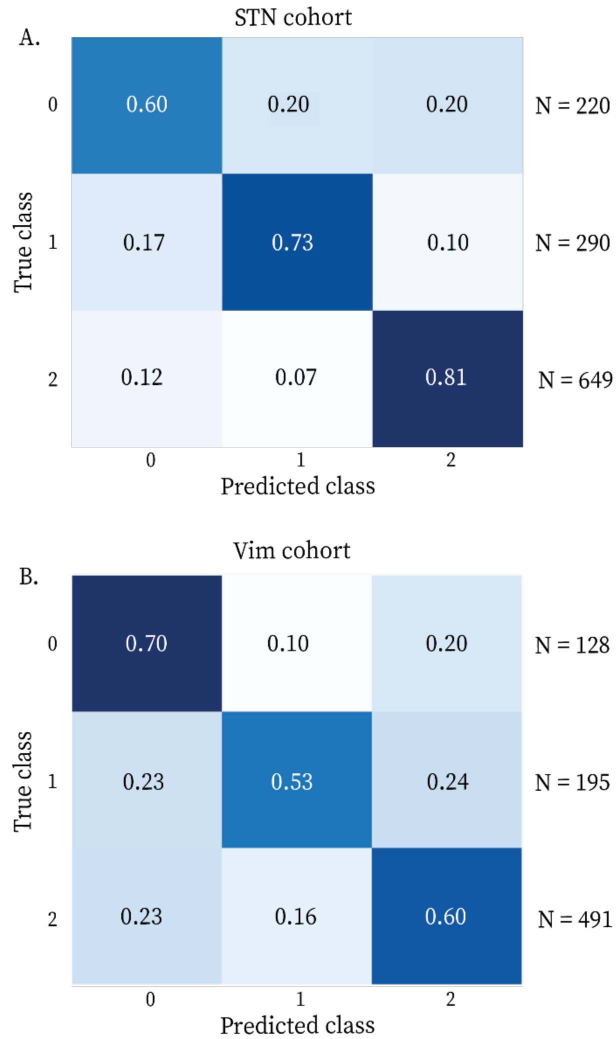

Figure 2: Confusion matrices showing the ratios of classified and misclassified samples for each class with the Adaboost model for the STN (A) and the Vim (B) cohort. N indicates the number of records available for each class.

Table 2: Precision, recall and f1-scores for the 3 implemented machine learning models, for the STN cohort (left) and Vim cohort (right).

	STN cohort			Vim cohort		
Model	Precision	Recall	F1	Precision	Recall	F1
Ada	0.76	0.75	0.76	0.67	0.60	0.62
SVC	0.75	0.73	0.74	0.66	0.62	0.63
NB	0.73	0.68	0.70	0.64	0.51	0.53

4 Discussion

Developing a reliable model for DBS effect prediction is crucial for optimizing parameter programming. This study demonstrated that machine learning-based prediction approaches can be effectively applied to intra-operative stimulation data. Specifically, we successfully classified three effect categories using a combination of clinical variables, stimulation-related features, and probabilistic map-derived data. We benchmarked three machine learning models based on distinct principles: Adaboost, Support Vector Classifier, and Naïve Bayes. While performance differences among the models were minor, Adaboost demonstrated the highest accuracy, likely due to its ability to better capture complex relationships within the data. Our best model achieved a maximum balanced accuracy of 0.71 on the STN cohort. The accuracy score is comparable to or slightly higher than those reported in most previous studies [3], [4]. This suggests that, while probabilistic maps and clinical data provide valuable information, they are not the sole predictors of DBS effects in intra-operative data, consistently with the findings of post-operative analyses. Notably, one prior work achieved over 92% accuracy in a four-class classification [6]; however, their higher performance may be attributed to a substantially larger patient cohort (275 individuals). A broader patient cohort, in fact, mitigates the impact of individual electrode positions on probabilistic maps, improving their generalizability and enhancing the robustness of derived features. Furthermore, dividing patients into efficacy quartiles helped mitigate class imbalance. This issue is particularly evident in the results for our STN cohort, where the minority classes exhibit lower classification accuracy compared to the majority class.

All three models exhibited higher performance in the STN cohort compared to the Vim cohort (~0.70 accuracy vs. ~0.60). This could be due to the larger number of patients and stimulations available for the STN cohort, which enabled a better refinement of PSS and probabilistic adverse effects areas. Additionally, the nature of the predicted effect (rigidity for the STN cohort, tremor for the Vim cohort) may have also influenced the performance. The performance drop between STN and Vim cohort was particularly noticeable in the recall score. This suggests that the models were less effective in capturing the true positives in the Vim cohort. Despite it being still an exploratory study, this work underlines how intra-operative stimulation test data allow to predict DBS effects in the same fashion, and with comparable results, as post-operative monopolar review data. Finding an optimal prediction model and an established set of predictors is, nonetheless, still an open challenge. Future works should, therefore, focus on these aspects with the support of comprehensive datasets including fiber tracts information and electrophysiological recordings.

Author Statement

Research funding: research was funded by the Swiss National Science Foundation (Grant No. 205320-207491). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study.

Data availability: data will be made available on request.

References

[1] A. M. Lozano *et al.*, “Deep brain stimulation: current challenges and future directions,” *Nat. Rev. Neurol.*, vol. 15, no. 3, Art. no. 3, Mar. 2019, doi: 10.1038/s41582-018-0128-2.

[2] K. Wårdell *et al.*, “Deep Brain Stimulation: Emerging Tools for Simulation, Data Analysis, and Visualization,” *Front. Neurosci.*, vol. 16, 2022, Accessed: Oct. 19, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2022.834026

[3] A. Nowacki *et al.*, “Probabilistic Mapping Reveals Optimal Stimulation Site in Essential Tremor,” *Ann. Neurol.*, vol. 91, no. 5, pp. 602–612, 2022, doi: 10.1002/ana.26324.

[4] R. R. Shamir, T. Dolber, A. M. Noecker, B. L. Walter, and C. C. McIntyre, “Machine learning approach to optimizing combined stimulation and medication therapies for Parkinson’s disease,” *Brain Stimulat.*, vol. 8, no. 6, pp. 1025–1032, 2015, doi: 10.1016/j.brs.2015.06.003.

[5] T. A. Dembek *et al.*, “Probabilistic sweet spots predict motor outcome for deep brain stimulation in Parkinson disease,” *Ann. Neurol.*, vol. 86, no. 4, pp. 527–538, Oct. 2019, doi: 10.1002/ana.25567.

[6] A. Boutet *et al.*, “Sign-specific stimulation ‘hot’ and ‘cold’ spots in Parkinson’s disease validated with machine learning,” *Brain Commun.*, vol. 3, no. 2, p. fcab027, Apr. 2021, doi: 10.1093/braincomms/fcab027.

[7] D. Vogel *et al.*, “Probabilistic stimulation mapping from intra-operative thalamic deep brain stimulation data in essential tremor,” *J. Neural Eng.*, vol. 21, no. 3, p. 036017, Jun. 2024, doi: 10.1088/1741-2552/ad4742.

[8] J. D. Johansson, F. Alonso, and K. Wardell, “Patient-Specific Simulations of Deep Brain Stimulation Electric Field with Aid of In-house Software ELMA,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2019, pp. 5212–5216, Jul. 2019, doi: 10.1109/EMBC.2019.8856307.

[9] M. Åström, E. Diczfalusy, H. Martens, and K. Wårdell, “Relationship between Neural Activation and Electric Field Distribution during Deep Brain Stimulation,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 664–672, Feb. 2015, doi: 10.1109/TBME.2014.2363494.

[10] D. Vogel, A. Shah, J. Coste, J.-J. Lemaire, K. Wårdell, and S. Hemm, “Anatomical brain structures normalization for deep brain stimulation in movement disorders,” *NeuroImage Clin.*, vol. 27, p. 102271, Apr. 2020, doi: 10.1016/j.nicl.2020.102271.