# SepaConvNet for Localizing the Subthalamic Nucleus using One Second Micro-Electrode Recordings

Maxime Peralta, Anh Quoc, Antoine Ackaouy, Thibault Martin, Greydon Gilmore, Claire Haegelen, Paul Sauleau, John Baxter, Pierre Jannin

# SepaConvNet for Localizing the Subthalamic Nucleus Using One Second Micro-electrode Recordings

**Maxime Peralta**
Univ Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France
maxime.peralta@univ-rennes1.fr

**Quoc Anh Bui**
Univ Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France

**Antoine Ackaouy**
Univ Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France

**Thibault Martin**
Univ Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France

**Greydon Gilmore**
School of Biomedical Engineering
Western University
London, CANADA

**Claire Haegelen**
Univ Rennes, CHU Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France

**Paul Sauleau**
Univ Rennes, CHU Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France

**John S.H. Baxter**
Univ Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France

**Pierre Jannin**
Univ Rennes
Inserm, LTSI - UMR 1099
F-35000 Rennes, France
pierre.jannin@univ-rennes1.fr

May 18, 2020

## ABSTRACT

Micro-electrode recording (MER) is a powerful way of localizing target structures during neurosurgical procedures such as the implantation of deep brain stimulation electrodes, which is a common treatment for Parkinson's disease and other neurological disorders. While Micro-electrode Recording (MER) provides adjunctive information to guidance assisted by pre-operative imaging, it is not unanimously used in the operating room. The lack of standard use of MER may be in part due to its long duration, which can lead to complications during the operation, or due to high degree of expertise required for their interpretation. Over the past decade, various approaches addressing automating MER analysis for target localization have been proposed, which have mainly focused on feature engineering. While the accuracies obtained are acceptable in certain configurations, one issue with handcrafted MER features is that they do not necessarily capture more subtle differences in MER that could be detected auditorily by an expert neurophysiologist. In this paper, we propose and validate a deep learning-based pipeline for subthalamic nucleus (STN) localization with micro-electrode recordings motivated by the human auditory system. Our proposed Convolutional Neural Network (CNN), referred as SepaConvNet, shows improved accuracy over two comparative networks for locating the STN from one second MER samples.

## 1 Introduction

Deep brain stimulation (DBS) is a common neurosurgical procedure in which electrodes are introduced into particular regions of the brain in order to correct for abnormal neural behaviour. The most common disorder treated using DBS is Parkinson's Disease (PD), a quickly evolving neuro-degenerative disease that results in the loss of motor control. Although DBS is not technically a cure for PD, it can greatly enhance the quality of life of the patient by limiting the

motor symptoms. There are three common targets for DBS in PD patients, the bilateral Ventral Intermediate nucleus of the thalamus (VIM), medial Globus Pallidus (mGP) and Subthalamic Nucleus (STN) [1].

Besides the adequate choice of structure, the most determinant factor of the surgery's success is the accurate positioning of the stimulating electrodes relative to the structures of interest. This positioning is currently performed in multiple phases. As pre-operative MR imaging is a routine part of clinical care for this procedure, a surgical planning phase is performed involving the identification of the target and the appropriate craniotomy site on the patient's skull, and the corresponding electrode trajectory. The second phase involves mounting a stereotactic frame onto the patient's skull, re-imaging them and using image registration to find the target co-ordinates in the space of the frame, allowing for the electrode to be implanted guided by the stereotactic frame in the operating room. However, preoperative imaging does not guarantee correct positioning of the electrode during the operation for various reasons including the limited resolution of the images and the potential registration error between the patient space and the stereotactic frame space. Additional deformations during surgery due to brain shift resulting from the craniotomy imply that the selected stimulation target coordinates determined in the surgical planning phase may no longer correspond exactly with the selected stimulation target location [2]. Lozano *et al.* [3] estimated the frequency of sub-optimal electrode implementation to be 20% for DBS performed solely with the preoperative imaging. Thus, an additional "fine-tuning" phase is required to ensure a correct implantation.

This "fine-tuning" stage is performed via "trial-and-error" when a small amount of stimulation is applied to the electrodes and the effects of said stimulation are used to infer whether or not the electrodes are in the correct position. This process is necessary as the ideal target for stimulation is defined functionally rather than purely anatomically, that is, the ideal target is one in which the stimulation has the desired effect. This process however can be made complicated by the use of different anesthesia regimens during the procedure. Local anesthesia maintains the wakefulness of the patient and thus allows the clinical team to test the clinical effects of various electrode positions more easily, incorporating additional behavioral and guided motor activities. However, performing the procedure with the patient awake can be a source of discomfort for the patient and additional patient movement can be a source of risk. Global or general anesthesia implies that the patient is not awake during the surgery, making the surgery more comfortable and controlled. However, this limits the amount of clinically useful information accessible to the surgical team during the procedure. Regardless of the type of anesthesia used, a second source of information that informs the position of the electrode relative to the targeted structure is considered mandatory either to limit the number of "trial-and-error" in the case of local anesthesia or to ensure a higher success rate in the case of global anesthesia.

One possibility is to use interventional MRI (iMRI) to refine the anatomical localization of the electrode during the surgery, but this can be prohibitively expensive and requires a specially designed operating theatre. An alternative approch is MER, which uses the electrical components of the stimulation electrode to record electrophysiological signals along the trajectory selected prior to the surgery. Using a segment of these signals, the clinical team listens to and/or visualizes the raw MER signal to determine if the micro-electrode is within the surgical target nuclei at the time of recording. Performing this assessment at several depths along the implant trajectory provides the surgical team enough evidence to delineate the borders of the surgical target nuclei. Whether to use intraoperative imaging or MER remains under debate. Lee *et al.* [4] and Liu *et al.* [5] compared the usage of iMRI versus MER, showing equivalent clinical outcomes.

We identified two limits of MER that can be leveraged by our approach. Firstly, the MER procedure is time consuming as many depths along the electrode trajectory are to be tested (each depth is an average of 10 seconds in duration). Secondly, this procedure is subjective as whether or not the signal originated from within the desired anatomy is determined by the neuro-physiologist based on their qualitative visual and auditory assessment of the signal. Quantitative criteria used to identify the surgical target are specific neuronal firing patterns and an increase of background noise. However, these criteria are not always obvious and dependant one the interpretation by the clinical team, which requires significant training and expertise in this domain.

## Contributions

In this paper, we propose a deep learning framework named SepaConvNet (Separable Convolution-based neural Network) for the detection of MER signals arising from the STN, differentiating them from nearby structures, using 1 second MER recordings. This state-of-the-art method shows high accuracy using class balanced accuracy metrics and outperforms ResNeXt, a state-of-the-art Convolutional Neural Network (CNN) architecture. This framework can work in real-time augmenting the clinical teams capability in terms of DBS electrode implantation for Parkinson's Disease (PD).

## 2   Theory and Previous Work

Automating STN localization with MER is an active field of research. Originally, these strategies were based on developing descriptive features characterizing the MER data and applying a classifier to these features. (A detailed review of such methods has been performed by Wan *et al.* [6].) These characteristics fall into two large classes: spike independent features [7, 8, 9] (e.g. power spectral density) many of which can be computed or approximated with an appropriate filter-bank, and spike dependent features [10, 11, 12, 13, 14, 15, 16] (e.g. standard deviation of interspike intervals). While adding spike dependant features can provide additional information about the signal, it also adds several constraints about signal acquisition [6].

The large amount of work that has been performed on MER feature engineering for the past decade is indicative that optimal features are likely tedious to develop. However, the fact that trained neurophysiologists can audibly distinguish different MER signals (as it can be in the current clinical workflow) hints that learned features arising from the raw signal itself may provide better outcomes. However, the extent and variety of features investigated demonstrates the hidden complexity of MER signal analysis. To this extent, using a CNN to automatically learn non-linear combinations of convolution filters to address this problem may be considerable easier than using hand-designed ones given the large span of linear and non-linear features that have been prevously been considered.

### Spectrogram representation of MER

In current clinical practice, MER signals can be interpreted by a trained neurophysiologist by listening to them, indicating that the patterns of interest within these signals are detectable using the human ear. It is well-known in many fields such as perceptual psychology that the ear performs a frequency decomposition of incoming auditory signals and that this decomposition is the data provided to the central nervous system. Spectrograms have previously been used for this application; Karthick *et al.* [17] successfully classified STN MER signals by designing features in a time-frequency domain. Thus, we chose a spectrogram representation of these signals as an initial stream of input features. This representation allows us to provide a more complex and predictive input data representation without the complex feature engineering and can be applied in real-time facilitating its integration into clinical workflow.

### Convolutional Neural Networks

Originally developed by LeCun *et al.* [18], CNNs are a type of deep neural network based on convolution layers. These layers consist of several tunable kernels operating on the input space by a linear convolution operation usually followed by a point-wise non-linear activation. Stacking these convolution layers allows the neural network to learn to construct higher level features from the original input space which are more readily used for classification tasks. Following the convolutions layers, typically one or two fully connected layers are added, which use the high level features to perform the classification. During the training process, these kernels and the weights in the fully connected layers are iteratively updated through a process of optimizing a loss function which acts as a surrogate for optimizing accuracy.

### ResNeXt

ResNeXt [19] is an extension of ResNet [20], which are both well known architectures in the computer vision community. Aside from its good performance, the structure of ResNeXt is highly scalable. Indeed, one of the motivations of the original study [19] was to be able to control the shape of the structure of the network with only three parameters, where former well performing CNN models needed almost each layer to be tuned manually. Therefore, it is common to reuse pretrained models and finetune them, but there is no major publicly available CNN model pretrained on MER spectrograms. The ability of ResNeXt to be scalable motivated us to use it as a CNN baseline, to compare the performances of our proposed model.

## 3   Material and Methods

### 3.1   Data acquisition

The MER signals in this study were recorded from 57 PD patients having undergone DBS surgery at London Health Sciences Centre at Western University hospital. The Bens gun was used to advance 3-5 microelectrodes as a unit. Signals were recorded from 10.0 mm above the pre-operatively determined target zero point to 4.0/5.0 mm below the zero-point, the end-point being detectable activity from the Substantia Nigra. The drive was advanced in 1.0 mm increments and 0.5 mm increments within the target nucleus. At each depth, a 10-second recording was obtained once a clean recording was observed (i.e. free from drive noise). The signals were sampled (24kHz, 8 bit), amplified (gain:

10,000) and digitally filtered (bandpass: 500-5000 Hz, notch: 60Hz) using the Leadpoint recording station (Leadpoint 5, Medtronic). The ground truths were determined by the electrophysiologist and neurosurgeon at the time of the recording in the operating room. The database was approved by the local ethics board at Western University (REB 1090485).

## 3.2 Database construction

By observing the signals individually, we noticed that the signals were highly temporally redundant. Usually, the neurophysiologist can be certain of the location of the electrode by listening to several seconds of signal. We hypothesised that a deep learning algorithm can exhibit good performance with far less signal (i.e only one second) which could improve clinical workflow by reducing procedure time. Reducing the duration of the signal has three positive effects for the training. Firstly, a reduced input dimensionality makes the deep neural network less prone to overfitting, thus allowing better generalisation performance with a limited number of training samples. Secondly, a smaller input limits the memory amount required to store the input, the gradient, the inference tensors, etc. This smaller memory consumption allows to train the neural network with a larger batch size, and thus train the network more quickly. Lastly, not using the whole 10 second signal allows us to take several shorter windows of each signal, thus increasing the database size.

We took advantage of the third property to balance our databases. Indeed, in our database, there are approximately three times more signals labeled as outside of the STN ('0') than signals labeled as inside of the STN ('1'). To overcome this class imbalance, for each signal labeled as '0' one 1 second window was extracted, were three different 1 second windows were used for each signal labeled as '1', thus balancing the dataset.

To split the data into training, validation/test sets, we used a stratified 5-fold cross-validation (CV). This data splitting technique consists in splitting the data in 5 different folds, with approximately the same number of signals in each fold, with the constraint that all signals coming from a single patient are assigned to the same fold. This constraint is mandatory to ensure that the estimation of network performance is representative of new patients as no patient appeared in both the testing and training sets for either class.

Once the stratified fold splitting is done, 4 folds were used as training data, and the last one as testing/validation data. This process was repeated 5 times in total in order to have every fold used as testing/validation once.

We obtained a database with 9055 one second signals labelled as '0' (outside the STN) and 8255 labelled as '1' (inside the STN), each signal having a length of 24000 samples.

## 3.3 Signal preprocessing

The first step of signal preprocessing was to threshold the signal, in order to limit the effects of non-informative acquisition artifacts. That being done, the temporal input signal is converted into a spectrogram representation. To compute the spectrogram, a Short Term Fourier Transform (STFT) was used, which applies a sliding window to the signal and computes the Fourier transformation for each window. From this Fourier transform, only the magnitudes of the vector were retained. Using STFT leads to a fundamental tradeoff: if the sliding window is wide, the spectrogram will have a higher accuracy in the frequency dimension, but will lose resolution in the time dimension. The STFT was performed with a Hann window of 512 samples width and a hop length of 10 samples, giving us input spectograms with 257 frequency bands and 2400 time points. An example MER spectrogram is shown in Fig. 1.

## 3.4 Proposed Convolutional Neural Network

In the literature, convolution kernels are often 2D as most CNNs are used to address computer vision problems, in which the input space is composed of two spatial dimensions. Spectrogram classification with 2D CNNs has already been successfully performed on music [21, 22] and doppler radar [23] which present at least some informative frequency shift-invariant properties. However, for MER signals, the two axes of our spectrograms are fundamentally different and there is no practical reason that a pattern detected in the low frequency bands would be equivalent to one in the high frequency bands. Thus we claim that in our use-case, 1D kernels are theoretically more adequate and our input signals are not treated as 2D images, but as 1D vector-valued signal with each element being the intensity in a frequency band.

Furthermore, in MER analysis, the amount of data available is limited. Thus, we replaced traditional convolutions with separable ones in our proposed CNN structure. Separable convolutions have fewer trainable weights and are therefore faster to train for a similar theoretical computational power, allowing to achieve similar results with a lighter network [24]. For small datasets with high dimensionality, a lighter model can limit the impact of overfitting, thus leading to better classification results and training stability. Separable convolution works in two steps. First, a timewise convolution is performed, using $K$ kernels per feature. Second, a pointwise convolution is used in order to mix channels
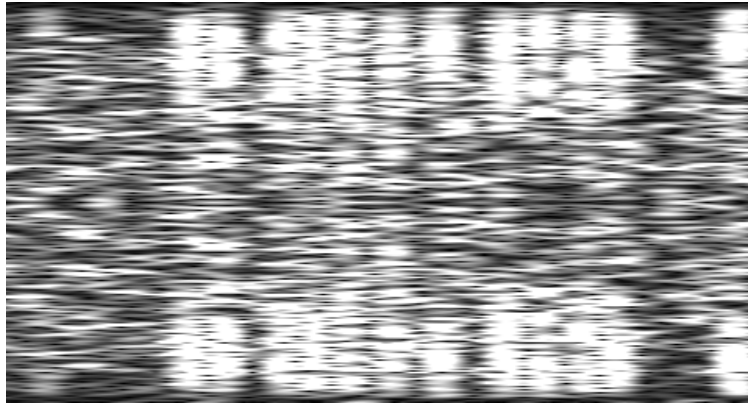
Figure 1: Raw spectrogram of 0.2 seconds of MER signal inside the STN after normalization.
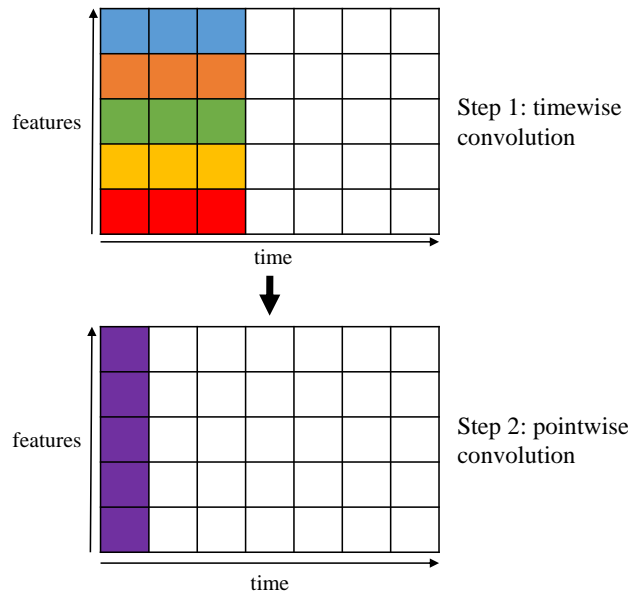


Figure 2: Example of a separable convolution. On the first step, 5 kernels (one for each feature band) of size 3 are convolving along the time axis. The resulting matrix is passed as input of the second step, which performs a convolution along the time axis with $N$ kernel of size 1, mixing all the features together at each timestep. The dimensions of the output matrix are the number of timesteps times $N$.

to produce $N$ new features. Given an appropriate choice of $K$ and sufficiently large kernels, this can largely reduce the number of parameters in the total convolution operator. Figure 2 shows these two stages of a separable convolution operation with $K = 1$ and $N = 1$ and would output a 1D-vector. In our paper, the hyperparameters $K$ and $N$ have been optimized through a Bayesian process, described in Section 3.6.

Our proposed CNN structure, called SepaConvNet (Separable Convolution-based neural Network) is presented in Figure 3. It consists of $D$ successive computational blocks followed by a global average pooling layer (along the time-axis) and a fully-connected neuron with sigmoid activation, in order to perform the binary classification. Each computational block is composed of a spatial dropout with a drop rate of 0.1, followed by a 1D separable convolution with kernel size of 15 with ReLU activation, a concatenation layer and a 1D max pooling layer of size 2 and stride 2 on the time axis. All computational blocks are identical in parameterization, except the last one, which doesn't have a max pooling layer.

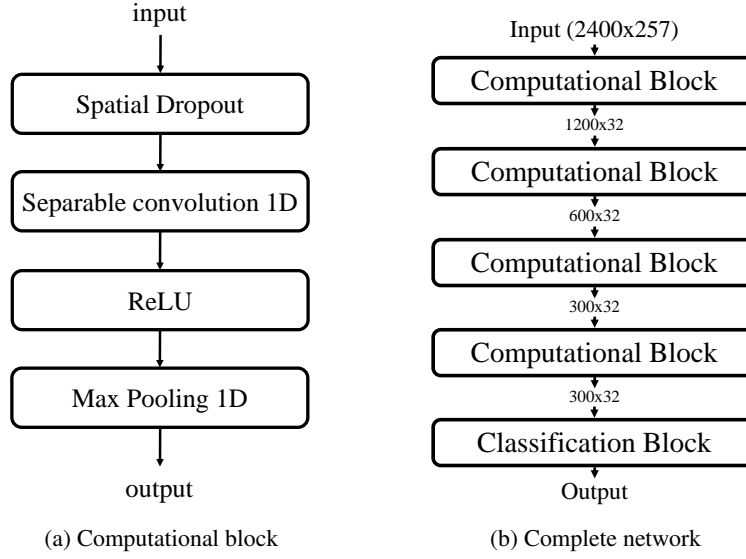(a) Computational block      (b) Complete network

Figure 3: SepaConvNet structure. Figure (a) presents the composition of a computational block. Figure (b) presents the global network, the classification block being an average time-pooling layer followed by a fully-connected neuron with sigmoid activation.

### 3.5 Accuracy and loss metrics

As classes remain slightly unbalanced, we used balanced accuracy (Eq. 1) to measure the classification performance and to compare methods. Each classifier was trained with class weighting, giving a weight to training sample inversely proportional to the representation of the belonging class in the training set, thus assigning the same total weight to both classes regardless of the number of samples in each.

$$BACC = \frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right) \tag{1}$$

### 3.6 Hyperparameters Optimization (HPO)

An HPO has been done using Bayesian optimization, with Gaussian processes as a surrogate model, and expected improvement as a criteria, with a 5-fold cross validation. The number of points tested was the square of the number of hyper-parameters to optimize plus one.

### 3.7 Software environment

All neural networks have been developed in Python with Keras (version 2.2.4), with Tensorflow (version 1.12) as a backend. STFT has been performed with Librosa (version 0.7.0).

## 4 Experiment

Our network was compared against two baselines: ResNeXt 2D, which is the standard version of ResNeXt, and ResNeXt 1D, with every 2D operation replaced by its 1D equivalent. Each CNN has been trained and tested with the same 5-fold CV partitions. We used a learning rate of 0.001 for both ResNeXt versions, 0.005 for SepaConvNet, with Adam optimizer and a batch size of 32.

For SepaConvNet, we performed HPO on the parameters $D$, $K$ and $N$ as presented in Section 3.4. For both versions of ResNeXt, we optimized the number of filters in each convolution layer and the depth of the network, and fixed the cardinality to 16. It is to note that the time resolution of input spectrograms had to be decreased for ResNeXt 2D due to an excessive memory consumption.
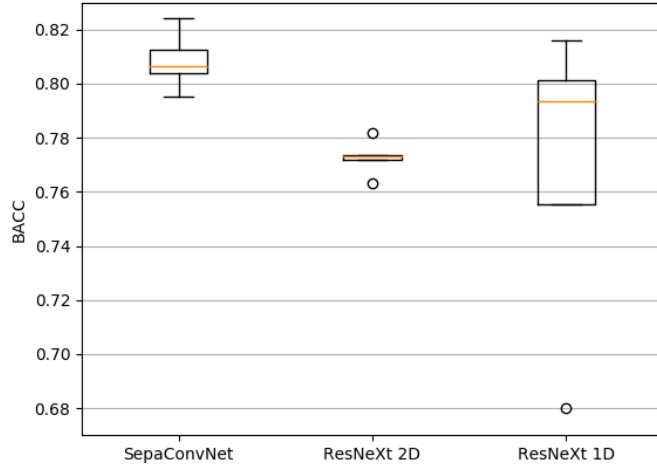
Figure 4: Quantitative results showing the distribution of 5-fold CV balanced accuracies for the three methods investigated.

| Network | BACC | Sens. | Spec. | f1 |
|---|---|---|---|---|
| SepaConvNet | 80.9% | 75.6% | 86.1% | 79.2% |
| ResNeXt 2D | 77.3% | 70.1% | 84.5% | 74.9% |
| ResNeXt 1D | 76.8% | 66.0% | 87.6% | 73.4% |

Table 1: Mean classification performances for the three methods investigated. 'Inside the STN' is considered the positive class.

Results of the 5-fold CV are reported in Figure 4 and Table 1. Classification results between SepaConvNet and both ResNeXt versions were statistically significant with a paired sample t-test ($p < 0.05$).

## 5  Discussion

The results presented in Section 4 confirm our hypothesis that for this particular problem, the spectrograms should be treated as 1D vector signals. Despite the increased capability for parameter re-use, the ResNeXt 2D version did not significantly outperform its 1D counterpart, suggesting that MER signals do not have any significant non-trivial frequency invariant features (i.e. those in which the shift invariance in the frequency domain would provide a benefit in detecting).

Secondarily, SepaConvNet significantly outperformed both baselines. After HPO, our proposed CNN used fewer trainable parameters than both 2D and 1D ResNeXt (resp. 16.752, 373.953 and 240.449). This decrease in parameters however is not a sign that the ResNeXt networks were overfit as HPO did cover configurations with lower parameterization. However, it does indicate that generic but lighter weight models (largely using separable convolutions) can detect important features in neurological MER signals.

**Future work**

For future work, we would like to test SepaConvNet on other MER databases to determine if these networks can transfer well to different institutions with different MER hardware.

We would like to investigate on the impact of input signal length on the performance of the system. We would also like our system to allow arbitrary signal lengths to be used, in order to enhance the system ability to work in real time.

Finally, we would further define the classification block of SepaConvNet. Currently, we are using time average pooling followed by a fully-connected neuron. We hypothesize that replacing it with a recurrent layer could enhance the classification performances, as the network could learn the optimal time-pooling strategy.

## 6 Conclusions

We proposed a custom CNN architecture, referred as SepaConvNet, based on 1D separable convolution to analyze STFT spectrograms of MER signals. We experimentally showed that, for MER spectrogram analysis, the SepaConvNet structure outperformed one state-of-the-art structure for image analysis. Our results suggest that deep learnt spectrogram-based features allow the achievement of comparable, or more optimal classification results, than commonly used time domain features. Our system can accurately detect the STN with one second of MER signal. An automatic analysis of MER signal with our system could thus greatly enhance the DBS procedure by significantly reducing its duration.

## Acknowledgments

## References

[1] Wolfgang Hamel, Johannes A Köppen, François Alesch, Angelo Antonini, Juan A Barcia, Hagai Bergman, Stephan Chabardes, Maria Fiorella Contarino, Philippe Cornu, Walter Demmel, et al. Targeting of the subthalamic nucleus for deep brain stimulation: a survey among parkinson disease specialists. *World neurosurgery*, 99:41–46, 2017.

[2] Seong-Cheol Park, Jung Kyo Lee, Seok Min Kim, Eu Jene Choi, and Chong Sik Lee. Systematic stereotactic error reduction using a calibration technique in single-brain-pass and multitrack deep brain stimulations. *Operative Neurosurgery*, 15(1):72–80, 2018.

[3] Christopher S Lozano, Manish Ranjan, Alexandre Boutet, David S Xu, Walter Kucharczyk, Alfonso Fasano, and Andres M Lozano. Imaging alone versus microelectrode recording–guided targeting of the stn in patients with parkinson's disease. *Journal of neurosurgery*, 130(6):1847–1852, 2018.

[4] Philip S Lee, Gregory M Weiner, Danielle Corson, Jessica Kappel, Yue-Fang Chang, Valerie R Suski, Sarah B Berman, Houman Homayoun, Amber D Van Laar, Donald J Crammond, et al. Outcomes of interventional-mri versus microelectrode recording-guided subthalamic deep brain stimulation. *Frontiers in neurology*, 9:241, 2018.

[5] Xuemeng Liu, Jibo Zhang, Kai Fu, Rui Gong, Jincao Chen, and Jie Zhang. Microelectrode recording–guided versus intraoperative magnetic resonance imaging–guided subthalamic nucleus deep brain stimulation surgery for parkinson disease: a 1-year follow-up study. *World neurosurgery*, 107:900–905, 2017.

[6] Kai Rui Wan, Tomasz Maszczyk, Angela An Qi See, Justin Dauwels, and Nicolas Kon Kam King. A review on microelectrode recording selection of features for machine learning in deep brain stimulation surgery for parkinson's disease. *Clinical Neurophysiology*, 130(1):145–154, 2019.

[7] Anan Moran, Izhar Bar-Gad, Hagai Bergman, and Zvi Israel. Real-time refinement of subthalamic nucleus targeting using bayesian decision-making on the root mean square measure. *Movement disorders: official journal of the Movement Disorder Society*, 21(9):1425–1431, 2006.

[8] Adam Zaidel, Alexander Spivak, Lavi Shpigelman, Hagai Bergman, and Zvi Israel. Delimiting subterritories of the human subthalamic nucleus by means of microelectrode recordings and a hidden markov model. *Movement disorders*, 24(12):1785–1793, 2009.

[9] Dan Valsky, Odeya Marmor-Levin, Marc Deffains, Renana Eitan, Kim T Blackwell, Hagai Bergman, and Zvi Israel. S top! border ahead: A utomatic detection of subthalamic exit during deep brain stimulation surgery. *Movement Disorders*, 32(1):70–79, 2017.

[10] Stephen Wong, GH Baltuch, JL Jaggi, and SF Danish. Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during dbs surgery with unsupervised machine learning. *Journal of neural engineering*, 6(2):026006, 2009.

[11] Hayriye Cagnan, Kevin Dolan, Xuan He, Maria Fiorella Contarino, Richard Schuurman, Pepijn van den Munckhof, Wytse J Wadman, Lo Bour, and Hubert CF Martens. Automatic subthalamic nucleus detection from microelectrode recordings based on noise level and neuronal activity. *Journal of neural engineering*, 8(4):046006, 2011.

[12] Wanpracha Chaovalitwongse, Youngseon Jeong, Myong K Jeong, Shabbar Danish, and Stephen Wong. Pattern recognition approaches for identifying subcortical targets during deep brain stimulation surgery. *IEEE intelligent systems*, 26(5):54–63, 2011.

[13] Hernán Darío Vargas Cardona, Jose Bestier Padilla, Ramiro Arango, Hans Carmona, Mauricio A Alvarez, Enrique Guijarro Estellés, and Alvaro Angel Orozco. Neurozone: On-line recognition of brain structures in stereotactic surgery-application to parkinson's disease. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2219–2222. IEEE, 2012.

[14] Konrad Ciecierski, Tomasz Mandat, Rafał Rola, Zbigniew W Raś, and Andrzej W Przybyszewski. Computer aided subthalamic nucleus (stn) localization during deep brain stimulation (dbs) surgery in parkinson's patients. In *Annales Academiae Medicae Silesiensis*, volume 5, pages 275–283, 2014.

[15] Vikram Rajpurohit, Shabbar F Danish, Eric L Hargreaves, and Stephen Wong. Optimizing computational feature sets for subthalamic nucleus localization in dbs surgery with feature selection. *Clinical Neurophysiology*, 126(5):975–982, 2015.

[16] L Schiaffino, A Rosado Muñoz, J Guerrero Martínez, J Francés Villora, A Gutiérrez, I Martínez Torres, et al. Stn area detection using k-nn classifiers for mer recordings in parkinson patients during neurostimulator implant surgery. In *Journal of Physics: Conference Series*, volume 705, page 012050. IOP Publishing, 2016.

[17] PA Karthick, Kai Rui Wan, R Yuvaraj, Angela AQ See, Nicolas Kon Kam King, and Justin Dauwels. Detection of subthalamic nucleus using time-frequency features of microelectrode recordings and random forest classifier. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4164–4167. IEEE, 2019.

[18] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[19] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[20] Sam Gross and Michael Wilber. Training and investigating residual nets. *Facebook AI Research*, 2016.

[21] Youngwook Kim and Taesup Moon. Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks. *IEEE geoscience and remote sensing letters*, 13(1):8–12, 2015.

[22] Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al. Cnn architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)*, pages 131–135. IEEE, 2017.

[23] Yandre MG Costa, Luiz S Oliveira, and Carlos N Silla Jr. An evaluation of convolutional neural networks for music classification using spectrograms. *Applied soft computing*, 52:28–38, 2017.

[24] Franck Mamalet and Christophe Garcia. Simplifying convnets for fast learning. In *International Conference on Artificial Neural Networks*, pages 58–65. Springer, 2012.