



HAL
open science

Extending Convolutional Neural Networks for Localizing the Subthalamic Nucleus from Micro-Electrode Recordings in Parkinson's Disease

Thibault Martin, Maxime Peralta, Greydon Gilmore, Paul Sauleau, Claire Haegelen, Pierre Jannin, John S H Baxter

► To cite this version:

Thibault Martin, Maxime Peralta, Greydon Gilmore, Paul Sauleau, Claire Haegelen, et al.. Extending Convolutional Neural Networks for Localizing the Subthalamic Nucleus from Micro-Electrode Recordings in Parkinson's Disease. *Biomedical Signal Processing and Control*, 2021, 67, pp.102529. 10.1016/j.bspc.2021.102529 . hal-03181612

HAL Id: hal-03181612

<https://hal.science/hal-03181612>

Submitted on 25 Mar 2021

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

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

Extending Convolutional Neural Networks for Localizing the Subthalamic Nucleus from Micro-Electrode Recordings in Parkinson’s Disease

Thibault Martin¹, Maxime Peralta¹, Greydon Gilmore², Paul Sauleau³,
Claire Haegelen³, Pierre Jannin¹ and John S.H. Baxter^{1*}

Abstract—Deep brain stimulation (DBS) is an interventional treatment for Parkinson’s disease which involves the precise positioning of stimulated electrodes within deep brain structures, such as the SubThalamic Nucleus (STN). Although originally identified via imaging, additional inter-operative guidance is necessary to localize the target anatomy. Analysis of Micro-Electrode Recordings (MERs) allows for a trained neurophysiologist to infer the underlying anatomy at a particular electrode position using human audition, although it is subjective and requires a high degree of expertise. Various approaches to assist MER analysis during DBS are proposed in the literature, including deep learning methods, which rely on a *static* input description, that is, a pre-defined number of features or input size. In this paper, we propose two *dynamic* deep learning approaches adaptable to the complexity of MERs signal, by using an arbitrary long listening time (in 1s chunks), while providing feedback to the neurophysiologist as to the model’s certainty. We evaluated five different deep learning based classifiers which can use arbitrary length MERs for STN segmentation. We found that a Bayesian extension using the high-level features from SepaConvNet performed the best, increasing the balanced accuracy to 83.5%. This work represents a step forward in integrating automated analysis of MERs into the DBS surgical workflow by automatically finding and exploiting possible efficiencies in MER acquisition.

Index Terms—Bayesian Inference, Deep Brain Stimulation, Intraoperative STN detection, Micro-Electrode Recordings, Recurrent Neural Network.

I. INTRODUCTION

PARKINSON’S DISEASE (PD) is a neurodegenerative disorder which affects the basal ganglia, leading to motor symptoms including difficulty initiating and controlling movements. Main causes are still not well understood yet, which many potential environmental and genetic factors. This lack of a full physiological understanding of the disorder and therefore a lack of etiological treatment, means that symptomatic treatments, whether pharmacological or interventional, are crucial to the management of the disease. In addition, PD is currently the second most common neurodegenerative disorder after Alzheimer’s disease [1], which justifies the interest in further developing a more robust treatment of its primary symptoms.

Deep Brain Stimulation (DBS) is used to reduce motor symptoms in cases where pharmacological approaches are

limited. DBS is currently the dominant surgical intervention for PD [2]. DBS consists in high-frequency electrical stimulation of a target region, usually in the medial globus pallidus or the Subthalamic Nucleus (STN), which is delivered by a permanently implanted electrode capable of continuous stimulation. As with many neurosurgical interventions, accurate positioning during implantation is crucial and can determine the success of the procedure. Stereotactic frames are often used to help ensure precise movement of electrodes to a defined target during their implantation [2].

In the current clinical workflow, patients eligible for DBS undergo a structural Magnetic Resonance Imaging (MRI) sequence which allows for the target region and potential electrode trajectories to be identified during the pre-operative planning phase. Other pre-operative imaging modalities, especially functional neuro-imaging modalities such as Positron Emission Tomography (PET) and functional MRI, could be used to identify the STN from its functional signature, rather than solely its structural characteristics. This information is then used to determine the stereotactic frame angles and estimated depth of the electrode trajectory which are used during the intervention. However, the use of pre-operative MRI as the sole guidance modality does not guarantee the correct positioning of the electrode in the STN. According to Lozano *et al.* [3], the use of MRI alone would lead to a 20% sub-optimality rate for intra-operative STN localization. This sub-optimality results from several different sources of errors including: (1) the low resolution and image distortion which can lead to an erroneous determination of the anatomical target, (2) the possibility of inducing a brain shift with the craniotomy, which can shift the preoperative defined coordinates of the target, and (3) the presence of mechanical errors related to stereotactic frame positioning [4].

By introducing an intra-operative data modality, corrections can be made to account for some of the sources of error, such as errors in the positioning of the stereotactic frame [5] and brain shift [6]. This allows for some real-time adjustments to be made to the electrode trajectory, although they may still have drawbacks in terms of the visibility of underlying neuroanatomical regions of interest. For cortical regions, electroencephalography (EEG) is often used, allowing for the distinct electrophysiological signature of the target to be identified. However, the STN is a very deep nucleus, meaning that a prohibitively large and invasive EEG grid, as well as intensive computation, would be required to localise it. Intraoperative imaging modalities, such as

¹Univ. Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France.

²Biomedical Engineering Graduate Program, Western University, London, Canada.

³Univ. Rennes, CHU Rennes, F-35000 Rennes, France.

*John S.H. Baxter is the corresponding author. E-mail: jbxaster@univ-rennes1.fr.

intraoperative computed-tomography (iCT), could be used to assist in the intraoperative electrode positioning itself making DBS electrode placement an example of an image-guided intervention [7]. The benefit of this modality is that the electrode's position with respect to bony landmarks can be easily determined. However, the lack of soft-tissue contrast renders it difficult to distinguish between the subcortical white and gray matter. Indeed, an anatomical segmentation of the STN from this modality is limited, and in order to reduce Parkinsonian symptoms, it is important to further confirm the location of the STN, by testing the stimulation and observing changes in the patient's symptomatology [8].

In order to determine if these real-time adjustments are correct, it is common to keep the patient under only local anesthesia in what is known as "awake" brain surgery. By keeping the patient in a conscious state, the precise positioning of the electrode can be inferred by the ability of test stimulation to immediately control motor disorders or to induce side-effects. This method, although necessary, results in discomfort for the patient, and may not always be possible for patients with extreme motor symptoms that could jeopardize the safety of the surgery. Thus, it is still necessary to consider guidance modalities that are effective at assisting the precise implantation of the DBS electrode while allowing the patient to be in an unconscious state in order to improve the safety and comfort of the procedure.

Currently, two approaches have been successful: Micro Electrode Recording (MER), and interventional MRI (iMRI). Both methods have been found to have similarly strong clinical results [9][10], although the latter, as with all other neuroimaging approaches, requires extensive additional expense and a specialized MR-safe operating theatre.

The MER protocol for DBS consists of capturing an electrophysiological signal at defined intervals, similar to EEG, but along the electrode's trajectory rather than across the cortical surface. As the neuronal activity of the targeted nucleus is different from the activity of its environment, the recorded signal should reflect the neural population at the electrode's position. As the density of active neuronal bodies is higher in the STN than in the surrounding areas [11], there is higher background noise on the electrophysiological recording within the STN than outside of it. In addition, the presence of neurons with spontaneous discharge activity in the STN results in the appearance of peaks in the MER recording. This particular pattern allows the localization of the STN by human audition, but requires the intra-operative presence of a trained neurophysiologist for the identification of these signals. Despite the finding of a different electrophysiological pattern in the STN, the subjective segmentation of the STN by ear, as well as the time required to verify each coordinate during the operation, are limiting factors for the DBS protocol.

II. RELATED WORK

Therefore, in order to improve the efficiency and speed up this critical intra-operative localization step, many studies have focused on the automation of STN detection using

MER data. These studies extracted temporal or frequency relevant features [12][13][14][15][16], leaving out potential additional informative features. However, the use of hand-designed descriptive features is limited by the need for normalization and extraction from the raw data, which can be problematic across centers and MER acquisition protocols. Furthermore, these features are not representative of how an electrophysiologist analyses MER data, which involves identifying and integrating more nuanced features from the raw MER signals.

Recently, Khosravi *et al.* [17] presented an early implementation of a Deep Neural Network (DNN) to classify MER data in the context of STN localization. This work demonstrates the efficiency of using a Discrete Fourier Transform (DFT) to extract enough discriminant features, and the use of a DNN binary classifier with this data representation. DFT made it possible to obtain a spectral power density (SPD), and thus the frequency profile of a signal. Their early experiments have shown that such frequency information, especially at the 500 – 3000 Hz interval was strongly indicative of the presence of the STN. The structure of their DNN was based on a standard feed-forward Artificial Neuron Network (ANN), coupled with the use of Dropout and a weight decay term based on Ridge regression as a regularizer, to avoid overfitting. However, a possible leakage of data from identical patients between the training and validation sets may have potentially induced a bias, and thus limit the integration of this work in this field of research. In order to confront this preliminary approach with the state of the art in the field of automatic MER analysis, the results of the study need to be replicated, as the validation presented is limited by its methodology.

In parallel, the work of Peralta *et al.* [18] has led to the development of a convolutional neural network, SepaConvNET, capable of providing a prediction from one-second MER spectrograms. The motivation behind reducing the length of the MER signal acquired was to minimize the effect of using such a network during an intervention and to possibly improve its efficiency by reducing the time spent listening to the MER, which we call the *listening length*. However, it is necessary to keep a long enough listening time to detect a larger of discriminating features or patterns within which are indicative of different anatomies. According to the Peralta *et al.* [18], the choice of a one-second signal is a good balance between brevity and the capacity to obtain meaningfully accurate predictions.

SepaConvNet's motivation was to replicate the auditory analysis of the raw signal starting with the human ear, and thus to avoid the use of feature engineering. It is well known the human cochlea performs a sort of frequency decomposition over time similar to the spectrograms generated from a Short Term Fourier Transform (STFT). In contrast to representing the entire signal simultaneously through a DFT, an STFT allows for temporal information to be retained, and for particular signal frequencies to be temporally localized. This network therefore uses a spectrogram representation of the signal as a pre-processing step, providing the network

with a more descriptive input. The high-dimensionality of spectrogram data would be overwhelming for traditional machine learning approaches, but adheres perfectly with the motivation behind Deep Learning (DL) and has previously been used for auditory signal analysis [19].

The novelty of this network lies in the use of 1D separable convolutions, which makes it possible to manage a smaller number of parameters, and therefore train a lighter model better adapted to the data available. In addition, the nature of the MER data has shown that the use of 1D separable convolutions gives better results than 2D convolutions. This result was explained by the relationship between the input axes, which is not as obvious as for conventional image classification [18]. The time and the frequency axes of the spectrograms are fundamentally different, which renders shift-invariance in the frequency direction potentially restrictive in this domain.

Despite SepaConvNet’s robust results, it still processes signals with a fixed length. In order to improve the performance of this predictive method, we assume that extending SepaConvNet for a dynamic listening time, longer than one second, may allow better integration of temporal features, thus making the model more certain.

Recurrent neural networks

In an ideal case, the amount of MER acquired for use in the classifier should adapt itself as a function of the certainty of the classifier. This would allow for shorter MERs to be processed, extending them only if the classifier is uncertain and would benefit from the additional data.

Recurrent neural networks have the ability to process temporal data of variable size by using “loops” in the network architecture, referring to their previous state. In the case of a temporal analysis, this allows some information to be remembered and used at a later time-point. In particular, Long Term Short Term Memory (LSTM) cells initially proposed by Hochreiter *et al.* [20] manages, through the use of several gates, a long term, short term memory, and the combination of the latter with observations in order to obtain a predictive output. Also, Gated Recurrent Unit (GRU) layers, developed by Cho *et al.* [21], have recently demonstrated similar capabilities as LSTM layers, with a similar but simplified architecture.

Contributions

In this study, we propose a real-time extension of a Deep Learning (DL) approach for the intra-operative localization of the STN based on MER analysis. This extension aims to improve the model’s performance by integrating temporal-specific information from the signal, and then provides an analysis tool that is able to improve its certainty by analyzing longer time periods (longer than 1 second). In order to process a signal of several seconds long, a naive approach would be to return for each consecutive block of one second of listening a prediction of the predictive model. In this paper, the goal is to extend the system more in a more smart way by ensuring a best integration of discriminating features for each

analyzed second, while reflecting the network behavior into a level of network certainty as the listening progresses. The development of such a framework mainly aims at improving the decision making of the clinical team, by increasing the efficiency of electrode implantation, as well as improving surgical outcomes.

III. MATERIAL AND METHODS

A. Data acquisition

For this study, MER data was collected from 57 patients with PD during a single or bilateral DBS surgery at the London Health Sciences Center at Western University Hospital (London, Canada). Data acquisition was carried out using a configuration in which the electrode trajectory is advanced through 5 channels (anterior, posterior, medial, lateral and central). After estimating the zero point (the intersection of the channel and the side of STN) with a pre-operative segmented MRI, MER recordings were made from 10.0mm to 4.0/5.0mm after target estimation. Recording was performed at 1.0 mm increments outside the pre-operatively estimated STN, and 0.5 mm increments within. For each position, a 10s recording was sampled at 24kHz (8 bits), amplified (gain: 10,000) and digitally filtered (bandpass: 500-5000Hz, notch: 60Hz) using a *Leadpoint 5* recording station (Medtronic). The database contains a total of 11,162 10-second MER samples. The authorization to use this data follows the collaborative agreement covered by ethical clearance DSA 109045, and was approved by the Research Ethics Board at the University of Western Ontario (REB # 109045).

B. Database construction

The acquisition of the 11,162 signals was done in parallel with their annotation as either ‘in’ the STN or ‘out.’ This annotation was performed by an expert neurosurgeon at Western University Hospital (London, Canada). The database thus contains 8,588 signals annotated outside the STN, for 2,574 signals annotated as being inside the STN.

In order to provide a balanced dataset for learning approaches, two methods were used for class balancing. The training of SepaConvNet network has been reproduced according to the sampling procedure described by Peralta *et al.* [18]. The architecture allows for listening lengths of 1 second, and thus allowing for over-sampling of the under-represented class, providing class balancing as well as more data augmentation. Three randomly non-overlapping one second windows were chosen for signals annotated inside the STN, against one randomly chosen window of one second for the signals annotated outside. We obtained a training dataset composed of 7,722 one-second signals annotated inside, for 8,588 one-second signals annotated outside. For the implementation of the SepaConvNet combined to a recurrent neural network, the need to use the whole signal length has led us to perform an over-sampling approach. In this case, the signals labeled ‘in’ were triplicated in the training database, thus obtaining the same number of observations per class as the previous method.

For the purpose of performance evaluation, the data was split into a training and validation set, using a Cross-Validation (CV) method. The use of k -fold CV enhanced the estimation of the performance of a network by dividing the dataset into k sub-datasets of equal size. In this way, each network was trained on one with the combinations of $k-1$ folds, and its performance was validated on the unseen fold. For the study, 10-fold CV was used to evaluate all the networks. In addition, during the splitting process, all recordings belonging to the same patient were assigned to the same fold. In this way, networks evaluation is always performed on MERs belonging to an unseen patient.

C. Signal pre-processing

After analysis of the amplitude distribution of the signals, a threshold was determined to eliminate any artifacts related to the data acquisition method. Thus, signal amplitude was limited to the threshold values [-249:250]. After this processing, feature extraction was performed. This step is critical, as it defines the shape of the features proposed to the classifiers, and seeks to highlight any information discriminating the classes presented. Two different approaches have been reproduced here. The Fast Fourier Transform of each 10-second signal was acquired in order to replicate the feature extraction presented by Khosravi *et al.* [17]. Spectrograms of the signals were generated using an STFT as proposed by Peralta *et al.* [18], with a Hann window of 512 samples, and a hop length of 10 samples. These parameters were determined to best fit the compromise between the time and the frequency domain resolutions [18]. The signals were determined to have artifacts on the first and last time steps which were discarded by removing the first and last half-second of each recording. At this stage, each spectrogram represents nine seconds of signal, for 21,600 time points, and 257 frequency bands. Finally, each frequency band was bounded to the 5th and 95th percentiles to eliminate residual artifacts and the intensities rescaled using robust min-max normalization.

D. DL classifiers

1) *SepaConvNet*: SepaConvNet (SCN) network has been implemented according to the architecture, as well as the optimized hyperparameters proposed by Peralta *et al.* [18]. The network is composed of 4 successive blocks of convolutions, followed by a Global Average Pooling layer along the time axis to generate a series of high-level features. These features are then fed through a linear layer to a single sigmoid activation neuron to determine the network output. Each convolution block is exactly composed of a succession of 4 operations. Respectively in the order of the operations, each block begins with (1) a 10% Dropout operation. The Dropout layer allows us to limit the overfitting behaviour by randomly 'switching off' a percentage of the neurons of the previous layer during training. The set of switched-off neurons is randomly updated between each weight update. This layer is followed by (2) a 1D separable convolution operation. This type of convolution is performed in two phases. During

the first step, a one-dimensional kernel shape of length 15 is chosen, and applied along the time axis, in time-wise convolution for each frequency band (the equivalent of a depth-wise convolution). This convolution uses a zero padding technique in order to maintain an identical time dimension at the output. In a second step, 32 pointwise convolutions mix all the features along the frequency axis, for each time point, returning 32 feature maps at the output. Thus, the separable convolution output depth is driven by a matrix convolution with a few values, which allows a faster training of the network. The output layers are sent to (3) a linear rectifying unit (ReLU) activation function, before performing (4) a 1D MaxPooling operation along the time axis. This downsampling operation reduces the information along the time axis keeping only the main information for classification. For SepaConvNet, each MaxPooling step reduces the number of time-points by a factor of 2, keeping only the highest value within each pair of consecutive points. The last of the 4 convolution blocks does not contain a MaxPooling layer, being replaced by the global average pooling layer. The latter has been set up to return the time average of each feature map.

The light architecture that is achieved by the separable convolutions limits the number of training parameters to 16,752.

After the development of this architecture, SepaConvNet, as well as all other networks in this study were optimized using binary cross-entropy as a cost function and Adam as the optimizer with a learning rate of 0.005. Training was performed on a total of 5 epochs without callbacks, with a batch size of 32 samples.

2) *Bayesian SepaConvNet*: The Bayesian extension of SepaConvNet follows the equation:

$$P_{(t)}(X = p) \propto P_{(t-1)}(X = p) \times P(f_{(t)}|X = p) \quad (1)$$

with the variables:

X	the electrodes location as either 'in' or 'out' of the STN.
p	the classes to be distinguished, specifically {in, out}.
t	the length of MER used by the extension which is an integer in the range $\llbracket 1; 9 \rrbracket$.
$P_{(t)}(X = p)$	is the probability distribution of X , given past predictions up to and including time t .
$P(f_{(t)} X = p)$	is the probability of SepaConvNet generating some feature at time t (represented by $f_{(t)}$) conditioned on what anatomy the recording arose from $X = p$.

Using Bayes' Theorem, $P_{(t)}(X = p)$ can be incrementally updated over time, being proportional the product of two simpler terms: the prediction of the Bayesian network at time $t - 1$, and the probability of SepaConvNet generating some feature at time t . The Bayesian model's initial prior, $P_{(0)}(X = p)$, is defined according to the proportion of the class contained in the database. Three different versions are proposed which differ solely in terms of how $f(t)$ is defined:

- The *Binary Bayesian SCN* version using the binary output (i.e. $f_{(t)} \in \{\text{in}, \text{out}\}$) of SepaConvNet, calculat-

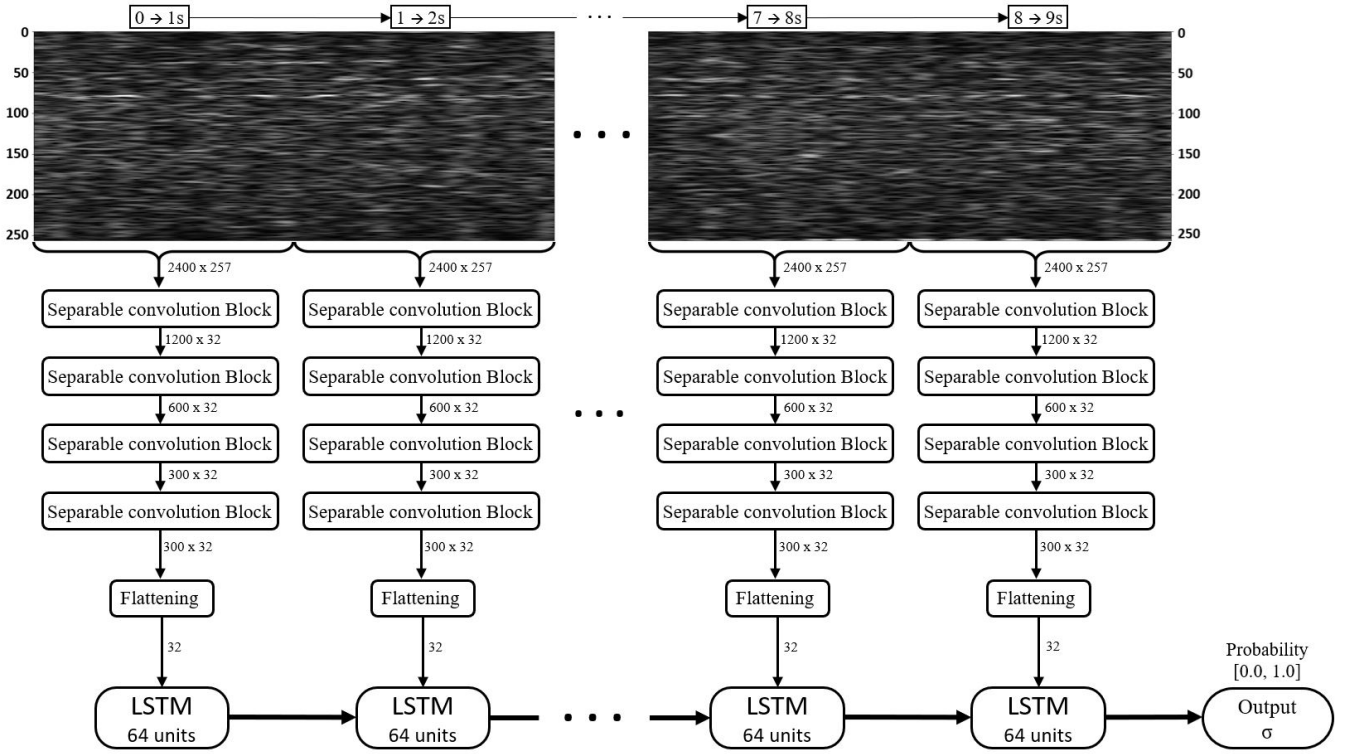


Fig. 1: Proposed structure of Recurrent SepaConvNet using LSTM as the recurrent layer.

ing $P(f_{(t)}|X = p)$ from a confusion matrix computed over the validation database of one second signals. This approach, which is the lightest, has only 4 additional parameters beyond the regular SepaConvNet.

- The *Simple GMM* of *Bayesian SCN* uses the non-thresholded output of SepaConvNet, i.e. $f_{(t)} \in [0, 1]$, and $P(f_{(t)}|X = p)$ is estimated using a Gaussian Mixture Model (GMM), by modeling the output distribution for each class, with 2 Gaussian components per class. The latter better allows the Bayesian process to reflect the underlying certainty of SepaConvNet at each time-point. This non-thresholded approach has only 12 additional parameters in addition to SepaConvNet, making it also a very lightweight extensions.
- Finally, the *Advanced GMM* of *Bayesian SCN* uses the non-negative activation vector from the second to last layer of SepaConvNet (containing 32 neurons). As with the *Simple GMM Bayesian SCN*, $P(f_{(t)}|X = p)$ is estimated using a GMM (with 4 Gaussian components per class). This approach involves an additional 8,456 parameters beyond the regular SepaConvNet which although heavier than the previous Bayesian extensions, still has far fewer weights than a neural network.

3) *Recurrent SepaConvNet*: The use of trainable end-to-end recurrent convolutional architectures has already been described in the literature, often called Long-term Recurrent Convolutional Networks (LRCNs) [22]. Figure 1 shows Recurrent SepaConvNet, which has SepaConvNet embedded within it, which consecutively integrates each one-second

signal across time, injecting the detected features into a 64-neuron recurrent layer. This layer determines which features to save from CNN at each time step, before passing these features to a single sigmoid activation neuron. Two variants of Recurrent SepaConvNet were tested, Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) as the recurrent layer. SepaConvNet network with the GRU layer has 35,408 weights and the slightly more complex LSTM layer increases that number to 41,616. All other hyper-parameters have been set to the same values used in the comparative SepaConvNet method [18].

4) *Simple DNN*: In order to provide a comparative approach to MER signal processing over ten seconds length MER, the DNN implementation presented by Khosravi *et al.* [17] was replicated. This network is based on a standard feed-forward artificial neural network, which has 10 hidden layers each composed of 50 nodes, leading to 6,025,601 weights. A weight decay term based on Ridge regression ($\lambda = 0.0285$) coupled with the use of Dropout (30%) was used on each hidden layer. The implementation of this network also uses a binary cross-entropy cost function, and an Adam optimizer with a learning rate of 0.001. Training was performed over 20 epochs, with a batch size of 256.

E. Performance evaluation and metrics

Despite the use of a sampling method to approximately balance the classes, the use of an adapted accuracy and loss metric was chosen to avoid any influence related to a residual bias. The performance of the models was calculated

using both the accuracy (ACC) and the balanced accuracy (BACC) metrics (Eq 2), the latter being less susceptible to dataset imbalance. The training of the different networks was performed using weights inversely proportional to the proportion of each class in the dataset, in order to ameliorate the effects of dataset imbalance in training.

$$BACC = \frac{1}{2} \left(\frac{True\ Positive}{Positive} + \frac{True\ Negative}{Negative} \right) \quad (2)$$

The implementations of each network were evaluated using 10-fold cross-validation on a database of 11,162 10-second samples for 57 patients with PD. The training and testing datasets for each fold were constructed to ensure that all the data from a single patient appeared in a single fold, either testing or training but not both, in order to prevent data leakage (i.e. each fold consisted of all signals from 6 or 7 patients). This results in ten identical networks being trained for each comparative method, and the performance of the method is estimated as the average of these ten networks. Since each implementation is trained and evaluated on the same data for each fold, their results can be directly compared using a pairwise Student's *t*-test.

F. Hardware & software environment

The networks were trained using on a NVIDIA GTX 1080 Ti chip, and developed with Keras, with Tensorflow as a backend (Python 3.8). STFT was conducted with the Librosa module, and DFT with Scipy module.

IV. RESULTS

The results for SepaConvNet as well as its Bayesian and Recurrent versions (in subfigures (a) and (b) respectively) are shown for two different MER signals. It should be noted that the output of the Bayesian SepaConvNets can be directly interpreted as a likelihood, which facilitates knowing precisely how confident the networks are. This is not the case for the Recurrent SepaConvNets where the connection between the output values and a likelihood is more tenuous.

The dynamics of the three Bayesian approaches varied significantly as shown in Figure 2(a). The main motivation for the development of a Bayesian framework is the hypothesis that the certainty of the network is related to the output value expressed by the latter. For the graph on the right, consecutive independent SepaConvNet outputs converge to a much higher average, which shows a higher average output, and thus a high level of certainty. This higher certainty is not reflected by the binary Bayesian network which shows an identical prediction evolution in both cases, indicating that it is likely unusable in a clinical scenario. Conversely, the Simple and Advanced GMM Bayesian extensions explicitly show the certainty level of SepaConvNet differing between the two signs and adapting their convergence speed.

Due to their more black-box nature, it is more difficult to decipher the dynamics of the Recurrent SepaConvNets with respect to time due to the lack of inherently meaningful

units with their output. It is however worth noting that they should generally trend upwards with more certainty in the underlying SepaConvNet, although not as cleanly defined as the Bayesian models. In addition, they appear to have higher certainty with respect to a higher baseline certainty in the SepaConvNet, indicating that their dynamics, with some mapping to familiar units, may be usable in a clinical context.

The quantitative results of the comparative study are presented in Table I with its mean performance across time shown in Figure 3 for SepaConvNet and the four real-time extension networks with adequate qualitative performance. (The results for SepaConvNet were based on averaging the

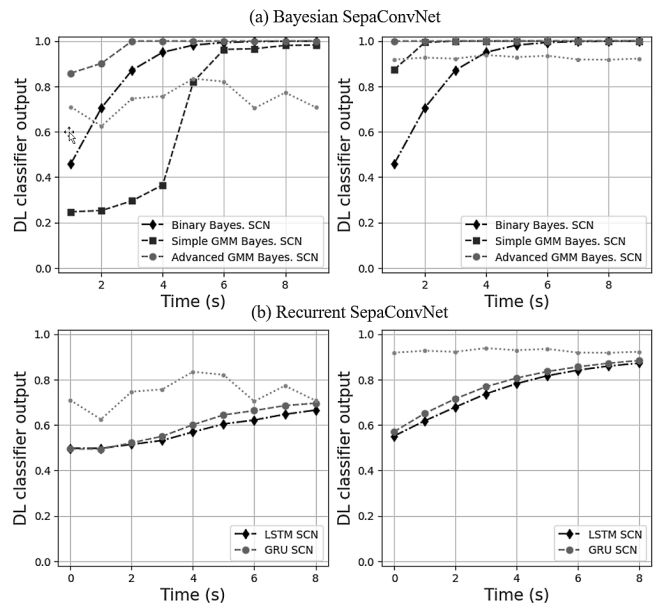


Fig. 2: SepaConvNet extensions behaviour for 2 MER signals from within the STN. Figure (a) presents the evolution of the output of Bayesian SepaConvNet and Figure (b) shows that of Recurrent SepaConvNet. The output of SepaConvNet for each one second segment is shown in light grey.

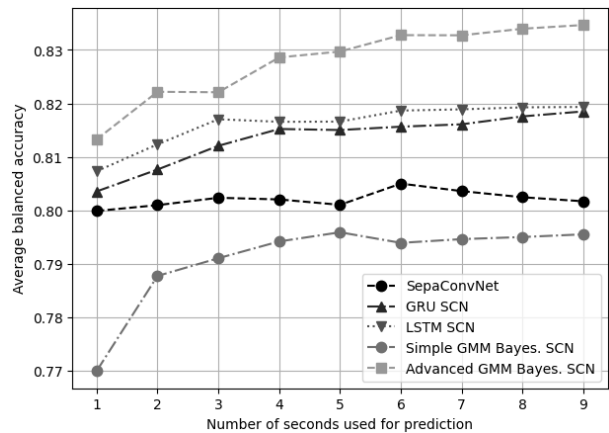


Fig. 3: Average BACC from 10-fold cross-validation according to the time of MER analyzed for SepaConvNet and its extensions.

Deep Learning Networks	Accuracy	Specificity	Sensitivity	BACC	Min BACC
(a) Simple DNN	69.7±9.0%	68.1±18.8%	73.5±28.5	72.6±7.5%	49.8%
(b) SepaConvNet	83.2±7.1%	86.0±8.4%	73.9±5.8	80.0±5.9%	64.8%
(c) Binary Bayes. SepaConvNet	84.1±6.7%	87.5±8.1%	72.6±6.7	80.1±5.8%	64.9%
(d) Simple GMM Bayes. SepaConvNet	80.8±3.9%	82.0±5.1%	77.2±10.1	79.7±5.1%	66.7%
(e) Advanced GMM Bayes. SepaConvNet	85.0±3.0%	86.4±4.8%	80.6±8.4	83.5±3.7%	72.8%
(f) LSTM SepaConvNet	84.2±2.6%	86.1±2.8%	77.8±8.6	81.9±4.4%	70.9%
(g) GRU SepaConvNet	84.2±2.3%	86.2±2.6%	77.5±9.9	81.8±4.6%	70.1%

TABLE I: Classification performances across folds given the full 9 second MER listening length. Each metric is associated with the standard deviation calculated through the validation of all 10-CV folds. This table presents the overall classification results for the networks : (a) reproducing the method presented by Khosravi *et al.* [17], (b) reproducing the SepaConvNet network, (c) the binary Bayesian extension of SepaConvNet, (d) the Simple GMM Bayesian extension using SepaConvNet’s non-thresholded output, (e) the Advanced GMM Bayesian extension using SepaConvNet’s activation vectors, (f)+(g) Recurrent SepaConvNet, using LSTM or GRU layers respectively.

output over all of the one second listening lengths included in the signal, hence the slightly improving performance over time.) Recurrent SepaConvNet with GRU and LSTM layers demonstrated improved performance over time, converging to 81.9% BACC, which outperformed the SepaConvNet average BACC (80.0%) at each one-second segment. Moreover, Bayesian approaches based on GMM modeling present an significant accuracy variation depending on the methodology used. On the one hand, the advanced approach surpasses all other methods, converging to 83.5% BACC after 9 seconds of listening, and demonstrating the interest of using a probabilistic method for task given sufficiently descriptive information, i.e. SepaConvNet’s activation vectors rather than its output.

One known limitation of Bayesian approaches is the initial reliance on a prior distribution. The lower accuracy of the Simple GMM Bayesian extension for short signals (first few seconds) is explained by the weight of the initial naive prior which necessarily biases the results towards the most common class. This initial bias is compensated for quickly, noting the boost in balanced accuracy for a listening time between 1 and 2 seconds. A similar although more modest increase can also be observed for the Advanced GMM Bayesian extension.

As shown in Table I, full 9-second MER signal listening provides the highest BACC with the advanced GMM Bayesian SepaConvNet, improving BACC up to 83.5%. Recurrent approaches also outperformed SepaConvNet alone, providing up to 81.9% for LSTM SepaConvNet. The Simple DNN approach had the lowest overall performance, although it is consistent with the performance reported by Khosravi *et al.* [17]. The two lightest Bayesian approaches in terms of parameters, i.e. Binary and Simple GMM methods, do not show accuracy improvement compared the standard network, and we measure a quite similar BACC regardless of the $f_{(t)}$ computation method.

Table I also shows an increase in sensitivity among SepaConvNet’s recurrent, and Bayesian non-threshold extension approaches, up to 80.6%, compared to SepaConvNet alone (73.9%). Conversely, the best specificity is obtained with

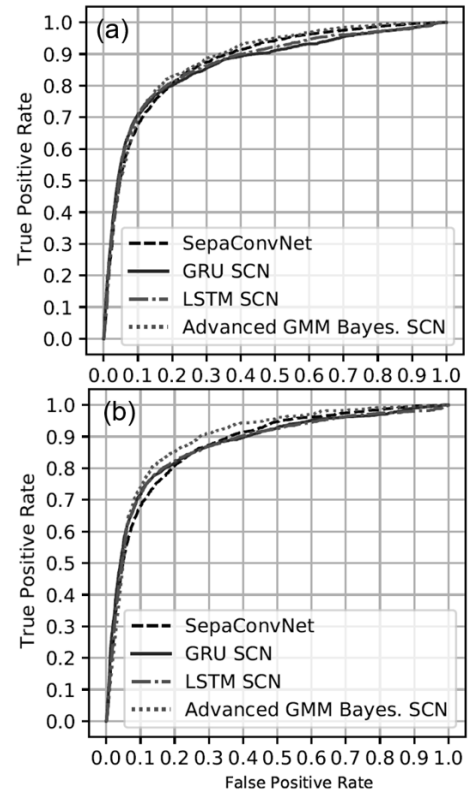


Fig. 4: ROC plot from 10-fold cross-validation for the classification results of 3 SepaConvNet extensions. Figure (a) shows the behavior of the networks after the one second analysis, and Figure (b) after the 9 seconds of MER listening.

the binary Bayesian approach, with 87.5%, while the other extensions do not outperform the baseline. Moreover, Figure 4 reveals that, abstracting from the discrimination threshold, the 3 most accurate network extensions show almost similar predictive behaviour. Figure 4 (b), in combination with Figure 5 also illustrate that all extensions tend to improve classification as listening time is increased. As for the previous results, the Advanced GMM Bayesian SCN approach obtains the best results in term of AUC.

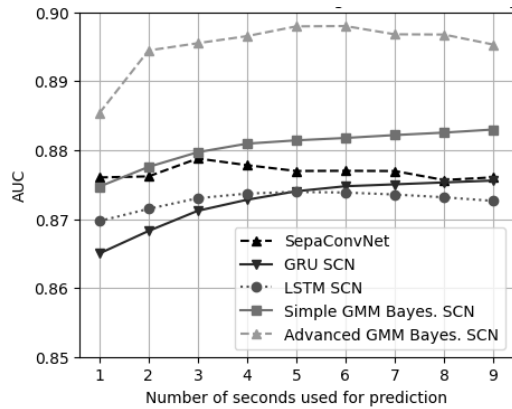


Fig. 5: AUC plot from 10-fold cross-validation according to the time of MER analyzed for SepaConvNet and its main extensions.

The standard deviation represented in Table I shows that this variability confirms the outperformance of SepaConvNet networks compared to the use of the simple DNN, the latter returning a BACC score of 72.6%, compared to recurrent extensions and the advanced GMM Bayesian SepaConvNet. Classification results per patient between SepaConvNet, simple DNN, both Recurrent SepaConvNets and Advanced GMM Bayesian SCN were statistically significant with a paired sample t-test ($p < 0.01$). In contrast, the classification of each patient by the Binary and Simple GMM methods compared to SepaConvNet did not show a significant difference.

V. DISCUSSION

The original SepaConvNet implementation represented a preliminary step in the development of an intra-operative framework for MER analysis, showing that short signal segments could be meaningfully used to streamline the DBS electrode placement workflow. For the critical task of STN localization during the DBS intervention itself, we wanted to highlight a robust predictive system that could integrate a maximum of discriminating features from the data. This paper therefore focuses on extending the predictive system to a dynamic listening range of several seconds of MERs, focusing on the combining sequential classifications produced by SepaConvNet for one-second long signals. Thus, if a single second is not enough to produce sufficient certainty in the neural network, an additional second can be considered, then another and so on until the network has converged. We have investigated two general frameworks for this combination: Bayesian inference and the use of recurrent neural layers.

The primary limitation of the Bayesian framework which works with SepaConvNet output values is the assumption that the various one second segments are independent of each other with respect to the output. This is not the case, and in fact, the output of SepaConvNet for different segments coming from the same nine second signal are highly correlated, which can push Bayesian methods to become increasingly confident in an incorrect answer rather than to approach a more uncertain middle ground for these cases.

However, this effect has been avoided by the implementation of the advanced GMM Bayesian SCN, as the latter models the probability on the second-to-last SepaConvNet layer distribution, i.e. a 32 vector space. The use of the GMM allows to capture more variations within the vector of 32 characteristics than after the linear combination made by the output which already reduces the information in terms of predictive weight. In the case where the convolution analysis of each spectrogram of one second returns different features, this method is capable of translating even more transparently the evolution of the signal processing in terms of classification probability. Among the three implemented sub-methodologies, which only differ in the likelihood estimation, it has been revealed that the simple GMM Bayesian version reflects more clearly the certainty returned by the network at each second. From a clinical implementation perspective, this method can provide the clinical team with the machine’s confidence allowing for early termination of MER acquisition and thus a more efficient workflow.

On the other hand, we have shown experimentally that the recurrent approaches significantly improved SepaConvNet’s predictions. The consistent BACC increased by the addition of recurrent layers also confirmed that relevant discriminating patterns can be extended over more than one second. This is also a potential improvement over the Binary and Simple GMM Bayesian architectures which have more limited access to the underlying features used by SepaConvNet. Recurrent networks extension have the capability of temporally extending these features, determining more descriptive versions than can be identified in a 1 second window. For all methods however the prediction was improved with each added second to the listening range, but it is the Advanced GMM method that shows the best prediction quality, increasing the BACC score to 83.5%.

Recent approaches developed in the field of MER data analysis have demonstrated the possibility of implementing a predictive model from raw MER signals with Deep Neural Networks, and thus overwhelming the performance of standard Machine Learning approaches, the latter depending on non-exhaustive lists of engineering designed features extractions [17]. The results of this study reinforce this fact, and further confirm the relevance of the SepaConvNet network architecture. Indeed, the comparison of SepaConvNet with the DNN presented by Khosravi *et al.* [17] shows that the exploitation of features after the frequency decomposition of MER signals is more efficient when using one-dimensional separable convolution layers, by increasing the BACC from 72.6% to 80.0%. In addition, the dynamic feature extraction integration achieved by hybridizing SepaConvNet with recurrent layers, or even the advanced GMM Bayesian extension, categorizes this implementation further improves its performance and the network comes to represent an advanced algorithm system in this field of research.

Overall, the accuracy of SepaConvNet has been improved by the methodologies presented here. The results show that the real-time extension of SepaConvNet (which originally processes only one second long signals) results in a respon-

sive system able to improve the network prediction with every additional second of signal. In terms of application, the main benefit is to reduce the STN search phase as much as possible by reducing to the necessary minimum the MERs listening time for each position of a trajectory. However, the signal classification improvement of the Bayesian or recurrent extensions over the original SepaConvNet, however significant, remains relatively slight compared to that of SepaConvNet over prior feature-based methods. Thus improving performance is still a valid motive for future development.

Future work

In order to integrate this predictive system into a clinical one, it will be necessary to improve its robustness by directing future work on: (1) using a larger database to consider a broader diversity of patients during training, (2) collecting MER data recorded by other acquisition systems, and therefore from multiple medical centers in an effort to standardize the predictive model with a view to making it as universal as possible, and (3) improving the interface of the predictive system with the clinical team in order to better integrate the knowledge provided by the model, as well as its limitations, e.g. real-time data visualization.

In addition, we would like to leverage DL methods such as those presented in this article to analyse the connection between MER signals at the electrode's final position with surgical outcomes. By correlating detected neural features with clinical ones, we intend to improve not only the localisation of pre-identified anatomy, but also aid in determining the precise positioning of DBS electrodes cognisant of the precise underlying neural circuitry.

VI. CONCLUSIONS

In this paper, we presented two general methods for extending Convolutional Neural Networks for MER analysis to handle arbitrary length recordings, fulfilling the future work set out by both Khosravi *et al.* [17] and Peralta *et al.* [18]. This allows for surgical workflow to be more readily streamlined, reducing the amount of time necessary for collecting electrophysiological recordings while also rendering their analysis less subjective.

This paper shows that integrating recurrent layers into SepaConvNet can significantly improve its predictions over time. The consistent BACC increase by the addition of recurrent layers confirmed that relevant discriminating features can be extended over more than one second, although without overcome the issue of interpreted the results in terms of likelihoods and confidences.

Results showed that a lightweight Bayesian approach to extending SepaConvNet network could return a level of certainty of the underlying predictive model in real-time, while maintaining the same performance as the native SepaConvNet. However, it is limited by the strong co-dependence of SepaConvNet network predictions within a single signal. The development of this framework highlighted a non-thresholded version using a simple GMM, giving representative prediction likelihoods of SepaConvNet at each

second, and making it a method that could be considered from a clinical implementation perspective. The performance limitations of this simple approach have been overcome by the implementation of an advanced approach which access more information from the network rather than a single signal. This method attains a BACC of 83.5%, the highest of all the methods evaluated here. Because of its cross-domain characteristics, the Advanced GMM Bayesian extension benefits from the advantages of Bayesian frameworks in terms of probability interface, and from the advantages of recurrent methods in terms of managing individual features. This approach is the most promising path for real-time integration into a surgical decision support system for DBS interventions.

ACKNOWLEDGMENTS

Thibault Martin is supported through a Doctoral Research Grant from Association France Parkinson. John S.H. Baxter is supported by the Institut des Neurosciences Cliniques de Rennes (INCR) and the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Post-Doctoral Fellowship (PDF) program.

REFERENCES

- [1] A. Delamarre and W. G. Meissner, "Epidemiology, environmental risk factors and genetics of Parkinson's disease," *La Presse Médicale*, vol. 46, no. 2, pp. 175–181, Mar. 2017.
- [2] C. S. Lozano, J. Tam, and A. M. Lozano, "The changing landscape of surgery for Parkinson's Disease," *Movement Disorders*, vol. 33, no. 1, pp. 36–47, 2018.
- [3] C. S. Lozano, M. Ranjan, A. Boutet, D. S. Xu, W. Kucharczyk, A. Fasano, and A. M. Lozano, "Imaging alone versus microelectrode recording-guided targeting of the STN in patients with Parkinson's disease," *Journal of Neurosurgery*, vol. 130, no. 6, pp. 1847–1852, Aug. 2018.
- [4] Z. Li, J.-G. Zhang, Y. Ye, and X. Li, "Review on Factors Affecting Targeting Accuracy of Deep Brain Stimulation Electrode Implantation between 2001 and 2015," *Stereotactic and Functional Neurosurgery*, vol. 94, no. 6, pp. 351–362, 2016.
- [5] M. Lanotte, M. Rizzone, B. Bergamasco, G. Faccani, A. Melcarne, and L. Lopiano, "Deep brain stimulation of the subthalamic nucleus: anatomical, neurophysiological, and outcome correlations with the effects of stimulation," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 72, no. 1, pp. 53–58, Jan. 2002.
- [6] M. F. Khan, K. Mewes, R. E. Gross, and O. Škrinjar, "Assessment of brain shift related to deep brain stimulation surgery," *Stereotactic and functional neurosurgery*, vol. 86, no. 1, pp. 44–53, 2008.
- [7] Z. Mirzadeh, K. Chapple, M. Lambert, R. Dhall, and F. A. Ponce, "Validation of ct-mri fusion for intraoperative assessment of stereotactic accuracy in dbs surgery," *Movement Disorders*, vol. 29, no. 14, pp. 1788–1795, 2014.
- [8] D. Sterio, M. Zonenshayn, A. Y. Mogilner, A. R. Rezai, K. Kiprovski, P. J. Kelly, and A. Beric, "Neurophysiological Refinement of Subthalamic Nucleus Targeting," *Neurosurgery*, vol. 50, no. 1, pp. 58–69, Jan. 2002.
- [9] X. Liu, J. Zhang, K. Fu, R. Gong, J. Chen, and J. 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*, vol. 107, pp. 900–905, Nov. 2017.
- [10] P. S. Lee, G. M. Weiner, D. Corson, J. Kappel, Y.-F. Chang, V. R. Suski, S. B. Berman, H. Homayoun, A. D. Van Laar, D. J. Crammond, and R. M. Richardson, "Outcomes of Interventional-MRI Versus Microelectrode Recording-Guided Subthalamic Deep Brain Stimulation," *Frontiers in Neurology*, vol. 9, 2018.

- [11] A. Benazzouz, S. Breit, A. Koudsie, P. Pollak, P. Krack, and A.-L. Benabid, "Intraoperative microrecordings of the subthalamic nucleus in Parkinson's disease," *Movement Disorders*, vol. 17, no. S3, pp. S145–S149, Mar. 2002.
- [12] K. R. Wan, T. Maszczyk, A. A. Q. See, J. Dauwels, and N. K. K. King, "A review on microelectrode recording selection of features for machine learning in deep brain stimulation surgery for Parkinson's disease," *Clinical Neurophysiology*, vol. 130, pp. 145–154, 2018.
- [13] D. Valsky, O. Marmor-Levin, M. Deffains, R. Eitan, K. Blackwell, H. Bergman, and Z. Israel, "Stop! Border Ahead: Automatic detection of subthalamic exit during deep brain stimulation surgery," *Movement disorders : official journal of the Movement Disorder Society*, vol. 32, no. 1, pp. 70–79, Jan. 2017.
- [14] L. Schiaffino, A. R. Muñoz, J. G. Martínez, J. F. Villora, A. Gutiérrez, I. M. Torres, and y. D. R. Kohan, "STN area detection using K-NN classifiers for MER recordings in Parkinson patients during neurostimulator implant surgery," *Journal of Physics: Conference Series*, vol. 705, Apr. 2016.
- [15] S. Wong, G. H. Baltuch, J. L. Jaggi, and S. F. Danish, "Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during DBS surgery with unsupervised machine learning," *Journal of Neural Engineering*, vol. 6, no. 2, Mar. 2009.
- [16] S. Coelli, V. Levi, J. D. V. D. Vecchio, E. Mailland, S. Rinaldo, R. Eleopra, and A. M. Bianchi, "Characterization of Microelectrode Recordings for the Subthalamic Nucleus identification in Parkinson's disease," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, July 2020, pp. 3485–3488.
- [17] M. Khosravi, S. F. Atashzar, G. Gilmore, M. S. Jog, and R. V. Patel, "Intraoperative Localization of STN During DBS Surgery Using a Data-Driven Model," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–9, 2020.
- [18] M. Peralta, A. Quoc, A. Ackaouy, T. Martin, G. Gilmore, C. Haegelen, P. Sauleau, J. Baxter, and P. Jannin, "SepaConvNet for Localizing the Subthalamic Nucleus using One Second Micro-Electrode Recordings," in *42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society in conjunction with the 43rd, Annual Conference of the Canadian Medical and Biological Engineering Society*, Montreal, Canada, July 2020.
- [19] Y. M. Costa, L. S. Oliveira, and C. N. Silla Jr, "An evaluation of convolutional neural networks for music classification using spectrograms," *Applied soft computing*, vol. 52, pp. 28–38, 2017.
- [20] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [21] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *Association for Computational Linguistics*, pp. 1724–1734, Sept. 2014.
- [22] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, Apr. 2017.