Explainable Deep Learning for Arm Classification During Deep Brain Stimulation - Towards Digital Biomarkers for Closed-Loop Stimulation

Mathias Ramm Haugland (m.haugland22@imperial.ac.uk)
Department of Bioengineering

Dr. Anastasia Borovykh (a.borovykh@imperial.ac.uk)
Department of Mathematics

Dr. Yen Tai (yen.tai@imperial.ac.uk)
Department of Brain Sciences

Dr. Shlomi Haar (s.haar@imperial.ac.uk)
Department of Brain Sciences,
Imperial College London, London, United Kingdom

Abstract

Deep brain stimulation (DBS) is an effective technique for treating motor symptoms in neurological conditions like Parkinson's disease and dystonic and essential tremor (DT and ET). The DBS delivery could be improved if reliable biomarkers could be found. We propose a deep learning (DL) framework based on EEGNet to search for digital biomarkers in EEG recordings for discriminating neural response from changes in DBS parameters. Here we present a proof-of-concept by distinguishing left and right arm movement in raw EEG recorded during a DBS programming session of a DT patient. Based on the classification of 1s segments from six-channel EEG, we achieve an average accuracy of up to 93.8%. In addition, we propose a simple, yet effective model-agnostic filtering strategy for explaining the network's performance, showing which frequency band features it mostly uses to classify the EEG.

Keywords: Deep Brain Stimulation; Tremor; Deep Learning; CNN; Artificial Intelligence; EEG; Digital Biomarkers;

Introduction

Deep brain stimulation (DBS) is an effective treatment for dystonic tremor (DT) and essential tremor (ET), where the thalamus usually is the target of stimulation (Tan et al., 2019). Tremors in these conditions are, however, primarily present during voluntary movement and sustained postures only, and today's constant DBS is therefore not ideal.

To improve DBS delivery, measurable biomarkers for both tremors and movement are needed; these are thought to be related to specific frequency bands in brain activity recordings. Tan et al. (2019) showed that in local field potentials (LFPs) from the stimulation cite of ET patients with DBS (turned off) the theta band (4-7 Hz) was the most informative for postural tremors. Similarly, postural and kinetic tremors in DT patients are characterized as being in the delta-theta range ($<$ 7 Hz) (Deuschl, Ma, Brin, & Committee, 1998). From EEG recordings of the motor cortex of healthy persons, the theta band has also been associated with movement initiation (Popovych et al., 2016) and motor learning (van der Cruijsen et al., 2021).

EEG-based brain machine interfaces (BCIs) have increased in popularity due to to the data's high temporal resolution and its non-invasiveness (Huang, Chang, Yan, Yang, & an Huayan Pei, 2022). Deep learning (DL)-based methods for decoding these signals achieve high performance in, for instance, motor task classification, and are advantageous for their use of the raw EEG directly (Lawhern et al., 2018).

We have recorded EEG from a person with DT performing various arm movements and sustained postures while receiving DBS. We use a DL model based on EEGNet to classify left vs. right arm activity from recordings under varying DBS; this serves as a first step in the search for reliable cortical biomarkers that discriminate neural response from DBS changes. We then propose a simple, yet effective model-agnostic method to explain what the trained network uses to achieve high classification accuracy; we show that it extracts and uses spectral features from multiple bands, most notably the theta band.

Methods

The dataset was collected at Charing Cross Hospital. It consists of EEG measurements from a patient with DT attending a $\sim$1h-long programming session of the DBS parameters. During the session, the patient is performing various arm tasks while a clinician varies the stimulation. These tasks mostly consisted of sustained postures with one arm stretched or bended but also movement like nose tapping. Each session was continuously recorded using the DSI-7 dry electrode EEG headset from Wearable Sensing, containing seven electrodes Pz (ref.), P3, C3, F3, F4, C4, and P4, according to the 10-20 system, with a sample frequency of 300 Hz. The whole session was also recorded with a video camera. The video was uploaded to Captiv L-7000 Premier and manually annotated with DBS parameter information and movement tasks. By synchronizing the video and the EEG recordings, these annotations were then transferred to the EEG data.

Pre-processing First, samples with five or more non-changing ("dead") channel values were removed. Non-movement samples were discarded, the remaining data was split into trials based on time gaps in the data above 0.1s or changes in movement or stimulation. Each trial was linearly interpolated, notch filtered at 50 Hz to remove line noise, and bandpass (BP) filtered in three different ways: i) full range: 5-90 Hz, ii) theta: 4-8 Hz, and iii) beta: 13-30 Hz.

Training As a baseline model, the CNN-based EEGNet-8,2 by Lawhern et al. (2018) was chosen based on its design that enables learning of both spectral and spatial features in EEG recordings that then can be used to classify the data.

The EEGNet was trained on continuous 1s segments of EEG. The stimulation parameters were constant within each segment but varied between segments. In a cross-validation fashion, the model was trained for and averaged over 60 folds. For each fold, a validation set containing 16 segments of each class was randomly chosen. Then, for 600 epochs, a training set of 64 segments from each class, not overlapping with any of the validation data, was randomly chosen, and random Gaussian noise with amplitude of 0.01σ was added to each EEG channel in each segment to make the net more robust. The epoch where the validation loss was at its lowest was chosen from each fold when averaging.

Explanability To examine whether the net was able to learn spectral information, a filtering technique was proposed. By filtering the validation data with various cutoff frequencies and examining its effect on the classification loss and accuracy, we were able to demonstrate which specific frequency bands the model uses. To validate the results we also re-trained the model on the narrow band passed data (in the theta and beta bands) and compared the model's performance.

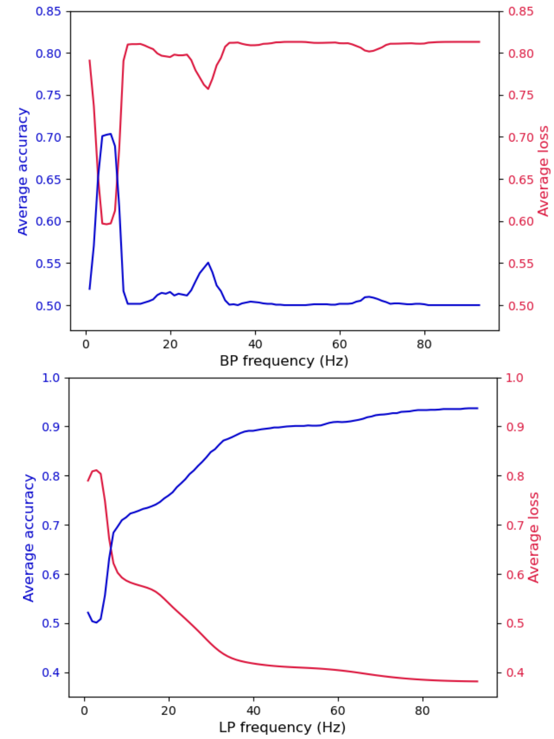

Figure 1: For the net trained on the full range (5-90 Hz), the plots show the average validation loss and accuracy when filtering the validation data with a 4 Hz-wide BP filter (top) and a LP filter with increasing cutoff (bottom).

Results

The net was trained separately on the three different (BP) filterings of the raw data and the average validation accuracies were i) 93.8%, ii) 88.7%, and iii) 92.4%. Figure 1 shows the accuracy and loss from a net trained on i) when sliding a BP filter of width 4 Hz across the full frequency range of the validation data, and lowpass filtering (LP) the same data at different cutoffs. In a similar fashion Figure 2 and Figure 3 show the results from when a BP filter of width 4 Hz was run across all frequencies, for ii) and iii), respectively.

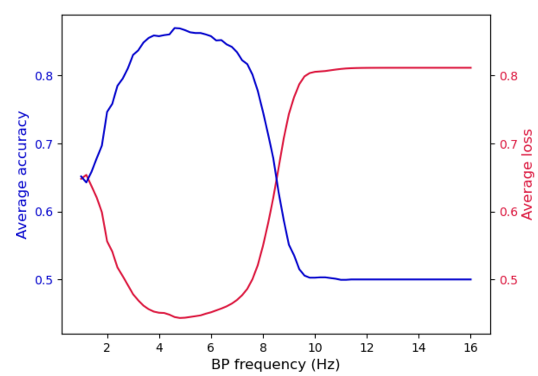

Figure 2: For the net trained on the frequencies in the theta range (4-8 Hz), the plot shows the average validation loss and accuracy when filtering the validation data with a BP filter of width 4 Hz.

Discussion

Our results suggest that when trained on the full frequency range (5-90 Hz), the net uses mostly features in the theta band (4-8 Hz) to classify arm activity (Fig 1). Since tremors in DT patients lie in this range we hypothesize that the net mainly uses features related to the arm tremors (Deuschl et al., 1998) and not the voluntary movement. A peak around 20-30 Hz (beta) is also visible, indicating that voluntary movement-related features are included to improve the classification. Some information also seems to be located around 65 Hz, which is half the stimulation frequency. Oscillations at this frequency have been related to cortical non-movement DBS response in patients with dyskinesia by (Swann et al., 2016).

As the average classification accuracies and Figures 2 and 3 show, good performance is also reached when using only the theta or the beta band data to train the net. This suggests that both bands contain relevant information for the discrimination but, when exposed to the full range the network learns more from the theta band (Fig 1). This is presumably due to the tremor, but, higher power in this band might also be the reason hence more experimenting is needed.

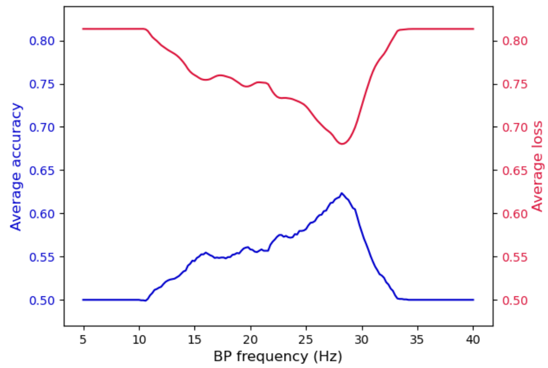

Figure 3: For the net trained on the frequencies in the beta range (13-30 Hz), the plot shows the average validation loss and accuracy when filtering the validation data with a BP filter of width 4 Hz.

Acknowledgments

S.H. is supported by the Edmond and Lily Safra Fellowship and the UK Dementia Research Institute, Care Research & Technology Centre.

References

Deuschl, G., Ma, P. B., Brin, M., & Committee, A. H. S. (1998). Consensus statement of the movement disorder society on tremor. *Movement Disorders*, *13*, 2–23.

Huang, W., Chang, W., Yan, G., Yang, Z., & an Huayan Pei, H. L. (2022). EEG-based motor imagery classification using convolutional neural networks with local reparameterization trick. *Expert Systems with Applications*, *187*, 115968.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), 056013.

Popovych, S., Rosjat, N., Toth, T., Wang, B., L.Liu, Abdollahi, R., ... Duan, S. (2016). Movement-related phase locking in the delta-theta frequency band. *NeuroImage*, *139*, 439–449.

Swann, N. C., de Hemptinne, C., Miocinovic, S., Qasim, S., Wang, S. S., Ziman, N., ... Starr, P. A. (2016). Gamma oscillations in the hyperkinetic state detected with chronic human brain recordings in parkinson's disease. *The Journal of Neuroscience*, *36*, 6445–6458.

Tan, H., Debarros, J., He, S., Pogosyan, A., Aziz, T. Z., Huang, Y., ... Brown, P. (2019). Decoding voluntary movements and postural tremor based on thalamic LFPs as a basis for closed-loop stimulation for essential tremor. *Brain Stimulation*, *12*, 858–867.

van der Cruijsen, J., Manoochehri, M., Jonker, Z. D., Andrinopoulou, E.-R., Frens, M. A., Ribbers, G. M., ... Selles, R. W. (2021). Theta but not beta power is positively associated with better explicit motor taks learning. *NeuroImage*, *240*, 118373.