Deep Learning-based EEG Epilepsy Detection and Analysis

Jérôme Cartailler
Lariboisière - Saint Louis Hospitals
University of Paris
cartailler.jerome@gmail.com

Qitong Wang
University of Paris, LIPADE
qitong.wang@etu.u-paris.fr

Stephen Whitmarsh
Institut du Cerveau (ICM)
stephen.whitmarsh@icm-institute.org

Themis Palpanas
University of Paris, LIPADE
Institut Universitaire de France (IUF)
themis@mi.parisdescartes.fr

[*Motivation*] Electroencephalogram (EEG) is one of the most common and essential medical signals collected by neural scientists for the analysis of nerve diseases. With the rapid development of medical instruments and data collection techniques, EEG analysis has also been witnessed a dramatic progress. One important problem of EEG analysis is epileptic pattern detection and analysis. Epilepsy is a brain disease generally associated with seizures, deteriorating the life quality of many patients. This internship targets to design effective deployment schemes of modern deep learning techniques on EEG Epilepsy detection, with a focus on real-world applications for neural scientists [1].

[*Objectives*] The mission of the overall project is to build automatic tools to facilitate the real-world EEG analysis of neuroscientists. Typical analytical tasks for neuroscientists include to identify patterns of interest (epileptic spike, fast activities, slow waves, etc.) and their roots of cause in terms of the time-domain or frequency-domain morphologies, i.e., interpretability of the tools. The developed tools are also expected to benefit other scientific areas with similar data layouts and applications, such as in astrophysics (for gravitational wave detection [6]) and in finance (for market event detection).

The target outputs of the whole project are 1). to publish high-level research paper in venues of both neural science and computer science, and 2). to release and maintain open-source software to benefit the EEG analysis communities.

[*Challenges*] The main challenges of the project lie in the complexity of real-world EEG datasets. The epileptic spike and other patterns possess different length in time domain, exist across several different channels or frequency ranges, and vary among different patients, deteriorating the performance of most existing heuristic methods. For examples, traditional Local Field Potential (LFP) parameter-based epileptic spike detection methods are not very robust across different patients [2].

The abilities of deep neural networks in representation learning have the potential to mitigate these problems, since they provide unified latent features extracted from various information sources. Successes have been witnessed for capturing frequency-domain [3] and other useful information for epileptic spike detection. However, their successful deployment demands sophisticated designs on top of EEG domain knowledge, and still far away from fulfilling the real-world needs for neural scientists.

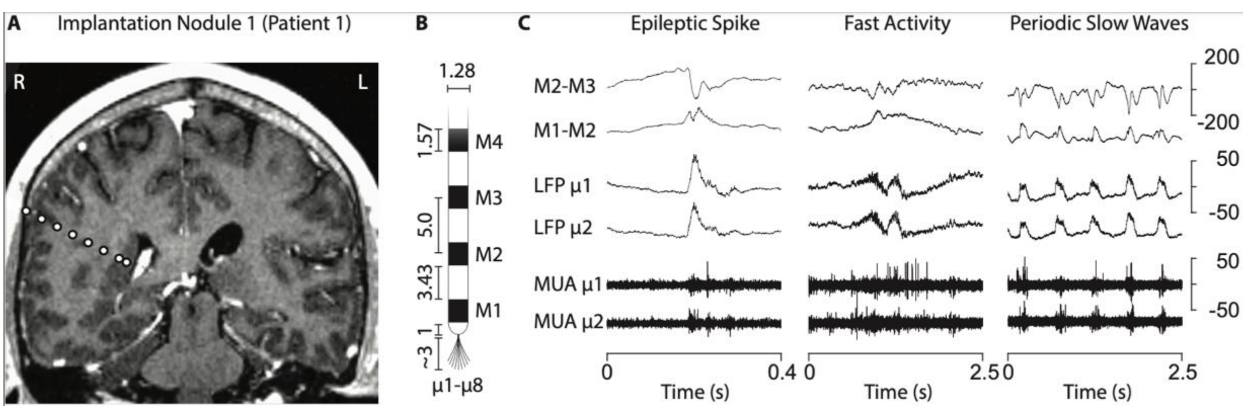

Figure 1. An illustration of our real-world dataset [1].

[*Proposed Work and Implementation*] The selected intern will first compare the differences between publicly released datasets and real-world EEG datasets, to profile the dataset characteristics and benchmark algorithm designs. Algorithm design will start with representation learning of convolutional neural networks for epilepsy detection with different lengths. Frequency-domain information will also be involved into the detection procedure, with novel architectures to be designed. Specifications are as following.

The datasets involved in this project will be three public EEG datasets (Bonn, MIT-CHB, and AES) [4]. These datasets might change based on findings in the progress. We will also use real-world datasets from our neuroscience collaborators, who are part of this project's team. For example, an ICM dataset we can use is a 3-day recording (32kHz, 20+ channels) from seven patients that carry implanted sensors, as shown in Figure 1. The total size of our datasets is 1.7TB. The selected intern will also be involved in our regular discussions with the neuroscience experts in order to exploit their domain knowledge and facilitate the results analysis.

The first step is to build a neural network which identifies the epileptic spikes in our datasets with state-of-the-art classifiers. A starting architecture could be residual convolutional neural networks [5]. The mixture of these different datasets is expected to give a sense on intra-patient epileptic pattern changes. The different properties between the datasets are also of interest to reveal the real-world requirements for the developed algorithm to be useful for the neural scientists.

With a designed model in Step 1, the second step is to involve more information into the decision making of the model, including frequency-domain information and cross-channel relationships, both where latent patterns unaware in the time domain could exist. Time-domain cross-channel relationships could naturally be taken into consideration since its straightforward for neural models to take 2-dimensional inputs (instead of 1-dimensional EEG subsequences). Frequency-domain information and frequency-domain cross-channel relationships are more complicated and require novel design of network architectures and fusion techniques.

There are two additional follow-up steps, which we will pursue (depending on the time we will have). Step 3 is to facilitate the intra-patient problem with active learning techniques or transfer learning techniques. Transfer learning techniques focus on to make the best advantage of the existing learned models, and active learning techniques focus on to make the minimal requirements of manual labels from the neurol scientists. Which technical branch to choose depends on the results of Steps 1 & 2, and would also consider the suggestions of neural scientists. Step 4 is to provide interpretability of the learned models and their predictions, based on the Class Activation Map (CAM) techniques. CAM identifies the regions (either their positions in time domain, frequency domain or channels) in an input EEG subsequence which contribute the most to its predicted label.

The technical contributions of the project lie in the novel design of the deep learning models, which involves the frequency-domain information and cross-channel relationships, and its success in real-world EEG epileptic pattern detection applications. The optional steps would provide more technical and pioneering contributions in their corresponding subareas.

[*Impact*] If successful, the proposed approach will significantly speedup the analysis of EEG data, which is now very cumbersome, because it requires lots of human effort. Moreover, the same approach could be used for the analysis of other sequence collections, such as, sensor data coming from manufacturing and engineering processes, where we are interested in detecting rare patterns of interest whose shapes may very.

References

[1]. Valerio Frazzini, Stephen Whitmarsh, et al. "In vivo Interictal Neuronal Signatures of Human Periventricular Nodular Heterotopia". bioRxiv (2020): 816173.

[2]. Petr Klimes, Jan Cimbalnik, Milan Brazdil, Jeffery Hall, François Dubeau, Jean Gotman, and Birgit Frauscher. "NREM Sleep Is the State of Vigilance That Best Identifies the Epileptogenic Zone in the Interictal Electroencephalogram." *Epilepsia* 60, no. 12 (December 2019): 2404–15.

[3]. Jialin Wang, Shen Liang, Ye Wang, Yanchun Zhang, Dake He, Jiangang Ma, Chunyang Ruan, Yingpei Wu, Xiaoyuan Hong, and Jiaqiu Shen. "A Weighted Overlook Graph Representation of EEG Data for Absence Epilepsy Detection." In 2020 IEEE International Conference on Data Mining (ICDM), pp. 581-590. IEEE, 2020.

[4]. Palak Handa, Monika Mathur, and Nidhi Goel. "Open and free EEG datasets for epilepsy diagnosis." arXiv preprint arXiv:2108.01030 (2021).

[5]. Diyuan Lu, and Jochen Triesch. "Residual deep convolutional neural network for eeg signal classification in epilepsy." arXiv preprint arXiv:1903.08100 (2019).

[6]. Sara Bahaadini, Vahid Noroozi, Neda Rohani, Scott Coughlin, Michael Zevin, Joshua R. Smith, Vicky Kalogera, and A. Katsaggelos. "Machine learning for Gravity Spy: Glitch classification and dataset." Information Sciences 444 (2018): 172-186.