2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)
Virtual Conference, May 4-6, 2021

Automated, Scalable and Generalizable Deep Learning for Tracking Cortical Spreading Depression Using EEG

Alireza Chamanzar†, Xujin Liu†, Lavender Y. Jiang, Kimon A. Vogt, José M. F. Moura, and Pulkit Grover
† co-first authors

***Abstract*— We present a non-invasive deep learning approach for tracking cortical spreading depressions (CSDs) in scalp electroencephalography (EEG) signals. Our method, which we refer to as CSD spatially aware convolutional network or CSD-SpArC, combines a convolutional neural network, which extracts temporal features from the EEG signal of each electrode, with a graph neural network, which exploits the spatial structure of EEG signals on the scalp. Using high-density EEG, this combination of networks misses no CSDs, even the narrowest ones (informed by widths observed in the real world), with less than 1.3% "post-stitching" false alarm rate. We further use the network to track CSD wave propagation by detecting when the recording at each electrode is affected substantially by the propagating wave, quantifying its "spatio-temporal tracking accuracy." Tested on simulated CSD waves on real head MRI models of 4 subjects, CSD-SpArC achieves spatio-temporal tracking accuracies of up to 86.65%±0.60%, with an average false alarm rate less than 3.5%, using high-density EEG (256 electrodes). We show the scalability of our trained network to different densities of EEG and generalizability to different head models.**

I. INTRODUCTION

This paper introduces a deep learning based framework for non-invasive tracking of cortical spreading depressions (CSDs) in the brain using simulated scalp electroencephalography (EEG) signals. CSDs are waves of depression in the spontaneous neural activity due to neurochemical changes in the brain [1]. CSDs propagate slowly (1 to 8 mm/min) across the cortical surface [2], and increasing evidence suggests that they can cause secondary brain injuries after traumatic brain injuries (TBIs), stroke, and hemorrhages [3]. Early detection and continuous monitoring of CSD propagation is crucial to reduce permanent damages to the brain tissue in patients with brain injuries. Our consultations with clinicians and experts (including Dr. Hartings [4]) suggest that what would affect the clinical decision-making the most is being able to detect CSDs reliably and infer how the wave is traveling. Informed by this, the goal of this work is to *track* CSDs, i.e., a) detect when there is a CSD episode (i.e., a single wave of CSD in the brain), and b) track its propagation in the sensor (i.e., EEG electrode) space to inform which electrodes are most affected during each time interval. The latter provides the clinician with a spatial sense of the wave. We quantify the accuracy of detecting CSDs correctly at non-overlapping time intervals using "spatio-temporal tracking accuracy". This detection accuracy is defined in Section II.

Prior work: There have been limited efforts in the past decade to visually detect CSDs using scalp EEG (e.g., [2], [5]). Observing that this is labor intensive, and few people across the world have this training, recently, we developed techniques for *automating* EEG-based detection of CSDs, an algorithm we call "WAVEFRONT," that uses traditional computer vision techniques and manual parameter adjustment for offline detection [6]. Complementary to this work, we also recently developed the first technique to localize *neural silences*, which can be viewed as depressions in neural activity that *do not* spread, using EEG recordings [7], [8].

In another work that classifies waves of activity (but not spreading silences), Agrusa *et al.* presented a deep convolutional neural network (CNN) for classifying the type of gastric slow wave activity (as normal or abnormal wave) in recorded data, using high-resolution non-invasive electrogastrogram (EGG; counterpart of EEG on the surface of the abdomen) [9]. This algorithm uses a deep learning approach for classifying the slowly propagating waves. However, it cannot directly be used for CSD tracking, as defined above, because: **(i)** as parts of CSD waves move in and out of the brain sulci, in EEG recordings (which are sensitive largely to shallow brain signals), they break into smaller parts called "wavefronts" [6]. There are no such folds on the surface of the stomach, and hence the technique in [9] is not designed to detect such wavefronts; **(ii)** as the brain signal goes through different layers of the head (namely, the cerebrospinal fluid, the low-conductivity skull, and the scalp), the changing conductivity adds to spatial low-pass filtering and blurring of the signals. This blurring makes the localizing CSD wavefronts challenging, especially the ones with a narrow width of depression. The blurring is less challenging in EGG recordings since there is no low-conductive layer similar to the skull to cause additional blurring; **(iii)** the CNN network of [9] outputs only a binary label for the whole recording to classify the type of slow waves (i.e., normal or abnormal), and it does so using projections of EGG activity on a regular 2D grid (see the definition of regular grids in [10]). Our goal, as noted above, is to detect when the CSD propagates and to track which EEG electrodes are most affected by CSDs when they are present. The algorithm in [9], being designed for regular 2D grids, does not easily extend to tracking in the irregular 3D grid of EEG electrodes. **(iv)** Using kernels with a fixed size in [9] makes the network *unscalable* (in a sense defined below) to different densities of electrodes.

Our contributions: In this work, we obtain an automated, generalizable algorithm for tracking CSDs using EEG. We define two notions of generalizability. First, generalizability to different patients (head models), which means that the trained algorithm can be applied to new patients without requiring to be re-trained. Second, generalizability to dif-

*This work was supported, in part, by grants from the Chuck Noll Foundation for Brain Injury Research, CMU Center for Machine Learning and Health (CMLH), and a CMU Swartz Center Innovation Fellowship.
*Lori Shutter, Jonathan Elmer, Jed Hartings provided helpful insights.

2021 10th International IEEE/EMBS Conference on Neural Engineering (NER) | 978-1-7281-4337-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/NER49283.2021.9441333

ferent densities of the electrode grid, which means that the algorithm trained on a specific electrode density can be directly applied to other densities to track CSDs. We refer to this generalizability as "scalability."

Our algorithm combines CNNs and graph neural networks (GNNs). CNNs are networks with local connections and shared weights that extract localized multi-scale features and construct highly expressive representations based on regular grids of data (e.g., images, texts, and time series) [11], [12]. GNNs can be considered as a generalization of CNNs to irregular grids (i.e., graph structures such as 3D meshes) [11], [13]. EEG recording is temporally on a regular grid (1D time series), where a CNN can be used to extract temporal features for each electrode, and spatially on an irregular grid (3D electrode mesh), where a GNN can be utilized.

In this work, we exploit the spatio-temporal information of CSD waves in EEG signals using a deep learning framework. It comprises: **(i)** a CNN to extract temporal features from the preprocessed signals of each electrode; **(ii)** a GNN to aggregate the temporal features and to exploit the spatial structure of electrode locations on the scalp to extract spatial information; and **(iii)** a multilayer perceptron (MLP) to classify the presence or absence of CSD wavefronts at each electrode location in small non-overlapping time intervals. The scalability of our proposed method, which we refer to as CSD spatially aware convolutional network or CSD-SpArC, is tested through rigorous simulations on a wide range of EEG electrode densities, from a low-density grid of 20-electrode standard EEG caps, to higher densities with up to 256 electrodes at 10-5 standard locations [14]. This makes it possible to exploit the large amount of low-density EEG data from the continuous recordings at intensive care units, to train the CSD-SpArC for higher densities of EEG, for which there are not many recorded CSD data. In addition, we tested the generalizability of CSD-SpArC to 4 different head models. This paves the way towards rapid detection of CSDs for new patients as they arrive at hospitals.

II. Method

In this section, we explain the dataset generation process, data labeling at each electrode, EEG preprocessing steps, and details of the CSD-SpArC algorithm.

A. Dataset generation

We closely follow the steps in [6] to simulate the EEG signals of CSD waves: **(i)** we preprocess (using *FreeSurfer*[1] software) and extract 3D head models of 4 healthy individuals (OAS1_0001, 2, 4, and 5) from an open-source magnetic resonance imaging (MRI) dataset (OASIS-1[2]); **(ii)** we simulate homogeneous annulus shaped CSD waves on a real brain model, with the origin chosen randomly on the cortical surface, while excluding the bottom of the brain where there is no EEG electrode coverage; **(iii)** we generate spontaneous brain activity based on the simplifying assumptions in [6], using a normal random process, and suppressed brain activitiy at the locations and time points where CSD wavefronts are to be located following the steps in [6]. The amplitude of the suppressed signal at the simulated CSD wave is $\sim$25% of the simulated normal brain signals. This amplitude reduction is chosen based on the average of the reported range of CSD depressions [17]; and, finally, **(iv)** we estimate leadfield matrices ($\mathbf{A}$) based on the extracted head models using *FieldTrip* [18] and applying them to the simulated brain signals to obtain scalp EEG signals.

B. Binary labeling of data at each electrode across time for training

We train our algorithm using binary labels that denote whether the signal of an electrode at a specific time point is substantially affected by CSD waves. To extract these binary labels, we look at the brain sources corresponding to the 30% largest elements in the i^{th} row of the leadfield matrix $\mathbf{A}$, and, if at least at one of these sources at time t there is a CSD depression, we assign label "1" to the signal of the i^{th} electrode at time t.

C. CSD-SpArC

1) EEG preprocessing: To improve the signal-to-noise-ratio (SNR) of CSD depressions in the simulated EEG signals, we adopt some of the preprocessing steps in [6] including Laplacian spatial filtering (for high density EEG caps with 128 and 256 electrodes), average power calculation, envelope extraction (using a sliding time window of 40s), and cross-correlation (with 2-minute negative pulse) for each signal. In addition, we normalize the distribution of training, validation, and test sets separately by mean subtraction and standard deviation (SD) division. The preprocessed signal is then downsampled (with temporal step size of 6s) before being fed into our network. This downsampling reduces the computational complexity of the algorithm, while maintaining the required temporal information of the preprocessed signal to track CSD waves. The preprocessed signal, after the envelope extraction and cross-correlation, is concentrated in a very low frequency band. In fact, the temporal width of depression is in the order of minutes [2].

2) Training algorithm: We used a deep learning framework including: **(i)** a CNN architecture called Multi Scale 1D ResNet[3] [15]; **(ii)** a GNN architecture called graph attention network (GAT) [16]; and **(iii)** a MLP classifier.

Model: Fig. 1 provides an overview of our model. The first part of our network is a Multi Scale 1D ResNet [15], which is shared between all electrodes/nodes. This CNN consists of multiple 1D convolution filters of various sizes and residual connections (known to improve time series classification [15], [19]). CSD detection at time t is performed for non-overlapping 5-min windows of the preprocessed signals of EEG electrodes, where t is the midpoint of this window. The binary labels of these midpoints are used as the CSD ground truth for the corresponding time windows. These 5-min windows from all electrodes are fed as a batch into the

[1] https://surfer.nmr.mgh.harvard.edu

[2] http://www.oasis-brains.org

[3] https://github.com/geekfeiw/Multi-Scale-1D-ResNet

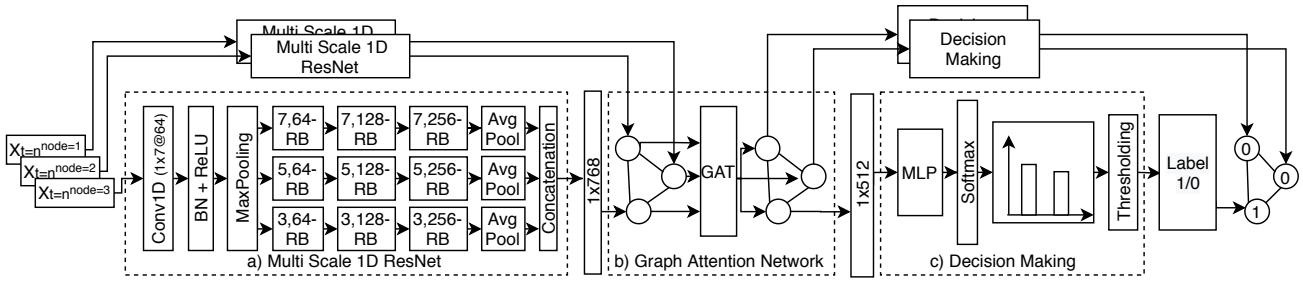

Fig. 1. Architecture overview of our CSD detection algorithm, shown for three input nodes/electrodes: a) temporal feature extraction of peprocessed scalp EEG signals using a 1D Multi Scale ResNet netwrok [15], where "m,n-RB" is a residual block of n filters, each with length of m; b) aggregation of extracted temporal features and exploiting the spatial information using a graph attention network (GAT) [16]; and c) classification of CSDs at each node using a multilayer perceptron (MLP). The final output of our network is a spatio-temporal probability map of CSD. A threshold is then applied on the output probability map during the inference time to extract the binary outputs of CSD (1) or no-CSD (0).

ResNet to extract temporal features for each electrode, which are then input to GAT. The GAT model is constructed using learnable attention layers to aggregate the incoming feature vectors to a node, where the attention score calculations are based on a *k*-nearest neighbor (*k*-NN) graph [16]. We use a GAT with 6 attention heads, residual connections, 2-4 layers, 500-600 hidden units, and LeakyReLU activation. Our *k*-NN graph is a geometric graph that thresholds the Euclidean distance between each pair of EEG electrodes on the scalp. The classification for each node/electrode is done using a MLP, followed by a softmax layer to generate the spatio-temporal probability map of CSD. A threshold of 0.75 is used to extract the binary outputs (1 for "CSD" and 0 for "no CSD") from the probability distribution of CSD produced by our network. Finally, recognizing that CSDs only substantially affect a few electrodes during each time interval, we stitch together these binary values at each electrode, using a sliding time window of 10 min (inspired by a similar stitching process in [6], and assuming each CSD spreads for at least 10 min). If there are at least 2 non-zero values in the time window, then a "temporal binary label" of 1 is assigned to the 10-min window. The consecutive 1's in the union of these temporal labels, across all electrodes, is declared as a single CSD episode.

Training process: We train CSD-SpArC on 5,850 simulated CSD episodes with varying length of 30-205 minutes, different speeds of propagation (1 to 8mm/min, with step size of 0.5mm/min), and different widths of wavefronts (0.5 to 6.5cm, with step size of 0.5cm), based on three different head models (OAS1_0001, 2, and 4), where there are 10 simulated episodes per each combination of width, speed, and head models. For each of the different densities of EEG electrodes (20, 32, 64, 128, and 256 at 10-5 standard locations [14]), a model is trained. To prevent overfitting, dropout layers are used on all layers, except CNN, with $P = 0.5$. The optimal combination of hyperparameters is found through validation using an open source hyperparameter optimization framework called *Optuna* [20]. The hyperparameters in CSD-SpArC are optimized over the following ranges: learning rate in (10^{-5},10^{-3}), k in k-NN graph for GAT in ($\lceil N/10 \rceil$,$\lceil N/2 \rceil$) where N is the number of EEG electrodes, weight decay in (5×10^{-4}, 10^{-2}), hidden layers of GAT in (2,4), and hidden units in (400,700).

Implementation details: CSD simulations and preprocessing are implemented in Matlab. The model is trained and tested in PyTorch, using a batch size of 40 windows, with total training iterations of 10 epochs with early stopping on validation loss of patience of 2 epochs. To train our model, we used a V100 NVIDIA GPU with 32GB memory.

Detection performance metrics: The "spatio-temporal tracking accuracy", defined for detecting CSDs at non-overlapping time intervals, is the degree of agreement between the binary ground truth labels of CSD and the binary outputs of our algorithm, measured using Cohen's kappa statistic (κ) [21]. We also provide the standard error of κ ($SE(\kappa)$), defined in [22]. The average false positive rate ($FPR = \frac{FP}{FP+TN}$, where FP is the number of false positives and TN is the number of true negatives) is reported. FPR is also reported based on the temporal binary labels (see the stitching process in Section II), which is referred to as "post-stitching" FPR in this paper.

III. Results and Discussion

In this section, we report the performance of CSD-SpArC for generalizability, scalability, and a range of CSD widths.

Performance on different ranges of CSD width: The width of CSD wavefronts varies in different neurological diseases and scenarios (e.g., 0.8 to 6.4cm in TBI [2]). We test the performance of CSD-SpArC on different widths of CSD wavefronts using different densities of EEG electrodes. The trained model (see Section II-C.2 for more details) is tested on a simulated dataset with 4 different ranges of CSD widths ((0.5,2), (2,3.5), (3.5,5), and (5,6.5)cm with 0.5cm step size), with the speed range of 2 to 8mm/min (with step size of 2mm/min) for each width range, across three different head models (OAS1_0001, 2, and 4), and for five different EEG densities (20 to 256 electrodes). Table II summarizes the results of CSD detection for different widths of wavefronts and different EEG densities. As expected, the tracking accuracy increases as the number of EEG electrodes increases and as the CSD waves become wider, with the best accuracy of 86.65%$\pm$0.60% for detecting the widest CSD waves ((5,6.5)cm) using the highest density EEG (256 electrodes). Although non-invasive detection of CSD is

TABLE I

SCALABILITY PERFORMANCE TO DIFFERENT ELECTRODE DENSITIES.

κ ± SE(κ)		Trained on 20	32	64	128	256
Tested on	**20**	75.73% ± 3.99	77.01% ± 3.80	76.54% ± 3.87	75.22% ± 4.00	76.31% ± 3.72
	32	75.33% ± 3.21	77.13% ± 3.03	77.06% ± 3.02	75.88% ± 3.15	77.01% ± 2.94
	64	75.69% ± 2.25	77.20% ± 2.15	77.42% ± 2.14	77.28% ± 2.16	77.88% ± 2.09
	128	75.08% ± 1.54	76.25% ± 1.54	76.54% ± 1.52	77.13% ± 1.74	76.91% ± 1.54
	256	76.69% ± 1.10	78.41% ± 1.05	78.24% ± 1.06	78.63% ± 1.07	79.69% ± 1.02

TABLE II

DETECTION PERFORMANCE FOR DIFFERENT RANGES OF CSD WIDTH.

κ ± SE(κ)		Trained on 20	32	64	128	256
CSD width (cm)	**0.5 - 2**	34.50% ± 21.87	40.52% ± 14.75	42.85% ± 10.23	47.71% ± 6.88	**57.46%** ± 4.10
	2 - 3.5	67.24% ± 6.35	72.56% ± 4.34	72.65% ± 3.07	72.73% ± 2.26	**75.62%** ± 1.46
	3.5 - 5	81.26% ± 3.00	82.40% ± 2.33	83.22% ± 1.63	83.25% ± 1.18	**83.98%** ± 0.81
	5 - 6.5	84% ± 2.26	85.94% ± 1.73	85.37% ± 1.25	85.86% ± 0.88	**86.65%** ± 0.60

challenging, the narrowest CSD waves ((0.5,2)cm) can still be detected using CSD-SpArC using the highest density EEG (256 electrodes) with a tracking accuracy of 57.46%±4.10% and average FPR of less than 1.43%. CSD-SpArC also detects CSD waves as narrow as 2cm using only 20 EEG electrodes with a tracking accuracy of 67.24%±6.35% and average FPR of less than 6.4%. Additionally, high-density EEG (256 electrodes) detects all CSD episodes, even the narrowest ones, with less than 1.3% "post-stitching" FPR. CSD-SpArC outperforms our earlier work [6] by requiring lower density of EEG to detect and localize narrow CSDs.

Generalizability: To test the generalizability of CSD-SpArC to new head models (unseen by the network), we test the model, which is trained on three head models of OAS1_0001, 2 and 4, on a simulated CSD dataset based on the head model of OAS1_0005 with different widths of wavefronts (0.5, 2, 5, and 6cm) and two different speeds of propagation (2 and 4mm/min). CSD-SpArC successfully detects and tracks CSD waves in this new head model with greater than 75% tracking accuracy and less than 3% false positive rate using 20 or more electrodes. CSD-SpArC generalizes without any retraining and/or fine-tuning the parameters in the model.

Scalability: Our model easily adapts to different densities and placement of EEG electrodes (Section II-C.2). This makes CSD-SpArC scalable to higher or lower densities of EEG electrodes. We test the scalability performance of CSD-SpArC as follows: we use the trained model on a specific density of EEG (e.g., 20-electrodes), and test it on simulated test dataset of other densities (e.g., 32, 64, 128, and 256 electrodes). Each test set includes CSD episodes with different widths (0.5, 2, 4, and 6cm) and speeds (1, 3, 5, 7mm/min), for three different head models. Table I shows the results of scalability tests, with the training EEG densities across columns, and test densities across rows. The results confirm the scalability of CSD-SpArC for any combination of train-test electrode densities with an accuracy of 76.7%±1.1% and FPR of less than 2.1% for a model which is trained on a low-density EEG (with only 20 electrodes) and tested on a high-density EEG (with 256 electrodes).

Limitations and future work: We only used an annular suppression pattern of CSD. CSD-SpArC needs to be tested on complex patterns, such as propagation on single gyrus, and semi-planar wavefronts [6]. In addition, this method algorithm needs to be tested on different levels of amplitude suppression. Perhaps most importantly, CSD-SpArC needs to be optimized and tested on real data from patients.

REFERENCES

[1] J. A. Hartings et al. The continuum of spreading depolarizations in acute cortical lesion development: examining leao's legacy. *J. Cereb. Blood Flow Metab*, 37(5):1571–1594, 2017.

[2] J. A. Hartings et al. Spreading depression in continuous electroencephalography of brain trauma. *Annals of neurology*, 76(5), 2014.

[3] M. Lauritzen et al. Clinical relevance of cortical spreading depression in neurological disorders: migraine, malignant stroke, subarachnoid and intracranial hemorrhage, and traumatic brain injury. *J. Cereb. Blood Flow Metab*, 31(1):17–35, 2011.

[4] J. A. Hartings. personal communication, November 2020.

[5] Z. JR. Bastany et al. Association of cortical spreading depression and seizures in patients with medically intractable epilepsy. *Clin. Neurophysiol.*, 131(12):2861–2874, 2020.

[6] A. Chamanzar et al. An algorithm for automated, noninvasive detection of cortical spreading depolarizations based on EEG simulations. *IEEE. Trans. Biomed. Eng.*, 66(4):1115–1126, 2018.

[7] A. Chamanzar and P. Grover. Silence localization. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1155–1158. IEEE, 2019.

[8] A. Chamanzar, M. Behrmann, and P. Grover. Neural silences can be localized rapidly using noninvasive scalp eeg. *bioRxiv*, 2020.

[9] A. S. Agrusa et al. A deep convolutional neural network approach to classify normal and abnormal gastric slow wave initiation from the high resolution electrogastrogram. *IEEE. Trans. Biomed. Eng.*, 67(3):854–867, 2019.

[10] D. Uznanski. Grid. *MathWorld,* `https://mathworld.wolfram.com/Grid.html`.

[11] J. Zhou, G. Cui, Z. Zhang, et al. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*, 2018.

[12] A. Chamanzar and Y. Nie. Weakly supervised multi-task learning for cell detection and segmentation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 513–516, 2020.

[13] F. Scarselli, M. Gori, et al. The graph neural network model. *IEEE Trans. on Neural Net.*, 20(1):61–80, 2008.

[14] R. Oostenveld and P. Praamstra. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.*, 112(4):713–719, 2001.

[15] F. Wang, J. Han, et al. CSI-Net: Unified human body characterization and pose recognition. *arXiv preprint arXiv:1810.03064*, 2018.

[16] P. Veličković et al. Graph Attention Networks. *International Conference on Learning Representations (ICLR)*, 2018.

[17] M. Fabricius et al. Cortical spreading depression and peri-infarct depolarization in acutely injured human cerebral cortex. *Brain*, 129(3):778–790, 2006.

[18] R. Oostenveld et al. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.*, 2011(156869), 2011.

[19] H. I. Fawaz et al. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Disc.*, 34(6):1936–1962, 2020.

[20] T. Akiba et al. Optuna: A next-generation hyperparameter optimization framework. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2623–2631, 2019.

[21] J. Cohen. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20(1):37–46, 1960.

[22] J. L. Fleiss, J. Cohen, and B. S. Everitt. Large sample standard errors of kappa and weighted kappa. *Psychol. Bull.*, 72(5):323, 1969.