

Research Paper

Graph neural networks in EEG spike detection

Ahmed Hossam Mohammed ^{a,*}, Mercedes Cabrerizo ^a, Alberto Pinzon ^b, Ilker Yaylali ^c,
Prasanna Jayakar ^d, Malek Adjouadi ^a

^a Department of Electrical and Computer Engineering, Florida International University, 10555 W Flagler St, Miami, 33174, FL, USA

^b Epilepsy Center, Baptist Hospital of Miami, 9090 SW 87th Ct Suite 201, Miami, 33176, FL, USA

^c Department of Neurology, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Portland, 97239, OR, USA

^d Brain Institute, Nicklaus Children's Hospital, 3100 SW 62nd Ave, Miami, FL 33155, USA

ARTICLE INFO

Keywords:

Scalp EEG

Interictal epileptiform discharge

Graph neural networks

Attention

Functional connectivity

Weighted phase lag index

ABSTRACT

Objective: This study develops new machine learning architectures that are more adept at detecting interictal epileptiform discharges (IEDs) in scalp EEG. A comparison of results using the average precision (AP) metric is made with the proposed models on two datasets obtained from Baptist Hospital of Miami and Temple University Hospital.

Methods: Applying graph neural networks (GNNs) on functional connectivity (FC) maps of different frequency sub-bands to yield a novel architecture we call FC-GNN. Attention mechanism is applied on a complete graph to let the neural network select its important edges, hence bypassing the extraction of features, a model we refer to as CA-GNN.

Results: On the Baptist Hospital dataset, the results were as follows: Vanilla Self-Attention $\rightarrow 0.9029 \pm 0.0431$, Hierarchical Attention $\rightarrow 0.8546 \pm 0.0587$, Vanilla Visual Geometry Group (VGG) $\rightarrow 0.92 \pm 0.0618$, Satelight $\rightarrow 0.9219 \pm 0.046$, FC-GNN $\rightarrow 0.9731 \pm 0.0187$, and CA-GNN $\rightarrow 0.9788 \pm 0.0125$. In the same order, the results on the Temple University Hospital dataset are 0.9692, 0.9113, 0.97, 0.9575, 0.963, and 0.9879.

Conclusion: Based on the good results they yield, GNNs prove to have a strong potential in detecting epileptogenic activity.

Significance: This study opens the door for the discovery of the powerful role played by GNNs in capturing IEDs, which is an essential step for identifying the epileptogenic networks of the affected brain and hence improving the prospects for more accurate 3D source localization.

1. Introduction

Epilepsy is a chronic brain disorder that affects 65 million people worldwide and 3.4 million in the United States, according to the Epilepsy Foundation. For 50% of this population, no identifiable cause is known and 30% of them do not benefit from any available treatment. Scalp electroencephalography (EEG) remains the most widely used modality for recording seizures and for understanding brain dynamics in epilepsy patients. The relevance of EEG recordings is due to the ability to explore brain electric activity at a high temporal resolution while also being relatively simple to administer, non-invasive, and cost-effective.

Although seizures can occur in non-epileptic brains, their existence is still considered to be a strong epilepsy-indicator. Hence, seizure detection has been widely studied to early-identify the occurrence of epilepsy for anti-epileptic treatment purposes. For example, an entropy-based dynamic graph embedding model was proposed in [1] where

the graph structure is inferred from the correlation among the signals of the multi-channel scalp EEG. The detection of seizures is based on the notion that the graph entropy during the seizure time interval is different from other time intervals. A progressive deep wavelet cascade classification (PDWC) model was proposed in [2] where recognition is performed in cycles. In each of those cycles, wavelet energy features are extracted by discrete wavelet transform (DWT) that are then fed into a set of cascade random forest (RF) classifiers. A technique based on reservoir computing (RC) was used to detect seizures in intracranial EEG (iEEG) of rats in [3]. The detection of such seizures could aid in the process of localizing the seizure onset zone (SOZ) [4]. It is notable that EEG is typically used along with other modalities such as magnetoencephalogram (MEG) to identify the epileptogenic zone with higher accuracy as in [5] where high-density EEG (HD-EEG) was utilized for its high spatial resolution. Magnetic resonance imaging (MRI) is a generic imaging modality that has several uses such as

* Corresponding author.

E-mail address: amoha107@fiu.edu (A.H. Mohammed).

<https://doi.org/10.1016/j.artmed.2023.102663>

Received 12 October 2022; Received in revised form 6 June 2023; Accepted 14 September 2023

Available online 19 September 2023

0933-3657/© 2023 Elsevier B.V. All rights reserved.

diagnosing breast tumors and predicting chemotherapy responses [6]. One particular application of interest is combining EEG with MRI to accurately localize the SOZ.

Despite the advancements in scalp-EEG seizure detection systems, which have been proven to be helpful in clinical settings, the continuous supervision of patients and complete visual review by specially trained personnel remain a necessity [7]. Several studies have shifted their focus on epileptogenic activity in the form of interictal epileptiform discharges (IEDs) occurring in between seizures. This shift is due to three main reasons: **First**, the high sensitivity of EEG signals to the presence of movement artifacts (eye blinks, tongue movement, etc.), which are abrupt, complicates their delineation despite the success of several artifact removal algorithms [8]. **Second**, seizures are generally random and unpredictable, while IEDs are relatively consistent morphologically and spatially, which could provide better insight for building and modeling the so-called epileptogenic network [9]. Identifying such networks is crucial for localizing the focal source from which the epileptogenic activity may originate. Accurate source localization is critical for surgical interventions that become inevitable when patients with focal epilepsy face drug-resistant and recurring seizures that could ultimately lead to abnormal neuronal death [10–12]. For patients with focal epilepsy, it is clinically essential to consider multiple IEDs to average the random effects of regular background activity to locate the SOZ more accurately. The **third** reason is that IEDs are more frequent than epileptic seizures. Waiting for seizures to happen requires extended EEG recording sessions, which require added cost and more time and effort spent monitoring these recording sessions.

The visual inspection of EEG by neurologists with the purpose of identifying IEDs is a highly subjective task, exceedingly tedious, and time-consuming. Hence, the automation of such a process has been the main focus of numerous studies [13–18]. Several procedures are followed to be able to detect IEDs from artifacts or non-IEDs (NIEDs). Early studies relied mainly on the morphological features of the potential epileptogenic discharges (such as the rising and falling edges) in context of the containing EEG segment (such as the average peak value) to determine whether a specific discharge is an IED or NIED using a rule-based approach. For instance, in [19], epileptogenic EEG sharp transients were detected and false positives were eliminated based on the spatial and temporal context information available on 16 channels of EEG. The feasibility of using the Walsh transformation to detect EEG interictal spikes was evaluated in [20]. A complete real-time system with software and hardware specifications to detect spikes was described in [21] using the wavelet transform. Differentiating the geometrical characteristics between spikes and normal EEG activity was done using mathematical morphology in [22].

More recently, the identification process was subdivided into a multi-step machine learning (ML)-based system. To cite some, a non-linear digital filter that uses adaptive autoregressive (AR) model [23] to eliminate trivial non-spikes was used as a pre-classifier in [24]. This stems from the convention that EEG is formed by the combination of stationary background waves (trivial non-spikes) and non-stationary spike waves (spikes and spike like non-spikes). Hence, the stationary waves could be mathematically expressed as an AR process. Support vector machine (SVM) was then used as a post-classifier to extract actual spikes. The generic spike template proposed by Chatrjian [25] was used in [26] to collect more templates which are then clustered with the k-means algorithm. Clusters having less than 5% of the collected templates were dropped and hence the number of clusters was automatically determined. The centroids for the remaining clusters are used as the final templates to detect IEDs. The study presented in [27] extended the previous approach by extracting duration, amplitude, slope, and area features from the average channel (on which the candidate spike was identified) as well as the two related bipolar channels. Each channel provided 10 such features and hence, an RF classifier was used to detect spikes using the 30 features from the three channels. In a study by Heers et al. semi-automated and automated detections of

IED types in long-term video-EEG monitoring (LTM) were in agreement with visually detected IED types and the SOZ [28].

One of the major challenges in such architectures is the high imbalance in the dataset. This is caused by the high sensitivity of the pre-filtering step (in order not to miss IEDs) which collects a large number of artifacts that ought to be rejected by the subsequent ML step. Synthetic minority oversampling technique (SMOTE) was adopted in [29] to synthesize additional spikes and hence mitigate the problem. Extracting features from the EEG segments can also be useful with scarce data. However, hand-crafting features requires domain-specific knowledge and opens the door for extended research in order to identify the best set of features for the ML problem at hand. For instance, the structuring element (pertaining to the mathematical morphology analysis) used in [29] is different from the one used in [30]. Hence, recent studies try to identify spikes from raw EEG. For instance, in [31], a simple VGG convolutional neural network (CNN) was used to detect IEDs in 2-second EEG epochs. CNNs were used in what is called as SpikeDeeptector that was introduced in [32] to identify the EEG channels that contain neural spikes. Spikes, ripples, and ripples-on-spikes present in iEEG were detected in [33] using a long short-term memory (LSTM) neural network.

A similar multi-step approach is considered in this study. Within a few minutes of scalp EEG, candidate IEDs are collected using a rule-based algorithm depending on the morphological features of the discharge in context of the containing EEG segment. Those candidates are investigated by an expert neurologist to distinguish IEDs from potential false positives. These labeled EEG segments are used to train different neural network (NN) architectures.

Our main contribution lies in developing two graph neural network (GNN) architectures, namely functional connectivity (FC)-based GNN (FC-GNN) and complete attention GNN (CA-GNN). The two proposed models are compared against Vanilla Self-Attention, Hierarchical Attention [34], Vanilla VGG [31], and Satelight [35] on two datasets. The first dataset is obtained from Baptist Hospital of Miami. The second dataset is the Temple University Hospital EEG Events Corpus (TUEV) [36]. The improved results based on average precision (AP) call for more investigation into the role GNNs can play in identifying IEDs. We also outline in the Discussion section a list of suggestions for future work and summarize the strengths and weaknesses of the proposed method.

The development of the FC-GNN architecture is inspired by the findings of a previous study by our research group [37] where it was emphasized that IEDs yield FC maps that are statistically distinct from those of NIEDs. This structural disproportionality could be exploited by GNNs in favor of our task. The role of graph convolutional networks (GCNs) applied on transverse and longitudinal bipolar EEG montages for IED detection is investigated in [38]. This study was the only graph-based architecture designed for IED detection reported in recent reviews [39–41]. Hence, this field remains widely unexplored. Instead of defining the graph edges through EEG montages, FC-GNN relies on FC analysis to establish the edges. A relatively similar approach was recently published in [42] and in [43] for the detection of seizures in neonatal EEG and for person identification, respectively. In [44], EEG-GNN was introduced and it was shown to outperform the standard CNN classifiers on both the ErrP, and RSVP brain-computer interface (BCI) datasets. Pearson correlation coefficient was used as the FC metric establishing the graph edges. In [45], GNNs applied on 8 different FC measures were compared against each other and against 3 baseline models for the classification of Alzheimer's Disease (AD) based on EEG. Despite no certain FC measure outperformed the other FC measures, the GNN models were shown to significantly perform better than the baseline models. Results show that GNNs applied to graphs estimated through FC analysis performed better than when applied to graphs whose edges are determined using the spatial distance between the EEG electrodes. This observation adds another motivation for using an FC-based GNN architecture to detect epileptic spikes in EEG.

FC maps visualize the coupling between different brain regions. This information could hardly be realized by mere visual inspection of the EEG. We use the weighted phase lag index (wPLI) metric [46] for its high immunity to the direct linear mixing effects, albeit no FC measure resolves the volume conduction problem entirely [47]. Moreover, wPLI provides the FC measure at several frequency sub-bands of interest, namely delta-band (δ : 0.5–4 Hz), theta-band (θ : 4–8 Hz), alpha-band (α : 8–13 Hz), and beta-band (β : 13–30 Hz). We exclude the gamma-band (γ : 30–48 Hz) for its strong susceptibility to muscle artifact in scalp EEG recordings as reported in [48,49]. It was also reported that the γ -band yields low-reliability graph metrics [50]. At the beginning of FC-GNN, a CNN is used to extract representations of each EEG channel separately. Positional encoding on the channels' embeddings is then performed before they are input to the GNN of each frequency sub-band FC map. Each GNN produces its own encoding and thus an attention mechanism [51] is used to combine these encodings. Eventually, a multi-layer perceptron (MLP) is used to output the final classification.

In order to extend this architecture further without extracting hand-crafted features (FC maps), the CA-GNN applies attention to the nodes' dimension of a complete graph (a single graph where all nodes are connected to each other). This is done to let the neural network choose the most important graph edges as directed by the downstream classification training task.

The rest of the paper is organized as follows: Section 2 presents the data collection/preprocessing procedures of the Baptist dataset, and the pre-filtering algorithm used to capture candidate discharges. Section 3 introduces the different NN architectures used. For the sake of self-sufficiency within this paper, the necessary background concepts for understanding the architectures are described in detail assuming that the reader may not be readily familiar with such concepts. Results, discussions, limitations, and suggested future work are presented in Section 4. Concluding remarks are finally presented in Section 5.

2. Collected data

2.1. Data preprocessing

Recordings from 19 scalp electrodes placed according to the 10–20 standard EEG system were collected from 39 patients at Baptist Hospital of Miami. The study process was approved by the Institutional Review Board of Florida International University (protocol number: IRB-150247). Patients were told to relax and avoid movement whenever possible during the recording session. Subjects are all adults ranging from 40 to 80 years old. The recording sampling frequencies were 512, 256, and 200 Hz. In order to minimize the processing time, all data were downsampled to 100 samples/s. This was done after the AC line noise was removed with a 60 Hz notch filter and applying a 4th order Butterworth band-pass filter with 0.75–38 Hz passing frequency range. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN) categorizes IEDs into four groups: sharp waves, spikes, spike-wave complexes, and polyspike-wave complexes [52]. Most of the collected IEDs are sharp waves and spikes with few instances of spike-wave complexes. The filtered EEG data were divided into 3-second segments as suggested by the neurologists in order to provide temporal context around the discharge of interest. A brief description of the TUEV dataset is given in Section 4.2 before presenting its results.

2.2. Pre-filtering

The whole EEG recording is divided into 3-second segments. Note that the final classification is done on the level of the segment as a whole. Therefore, a segment is labeled as a candidate IED-segment if it contains at least one candidate discharge. In reference to Fig. 1, all the following conditions have to be met in a candidate IED-segment:

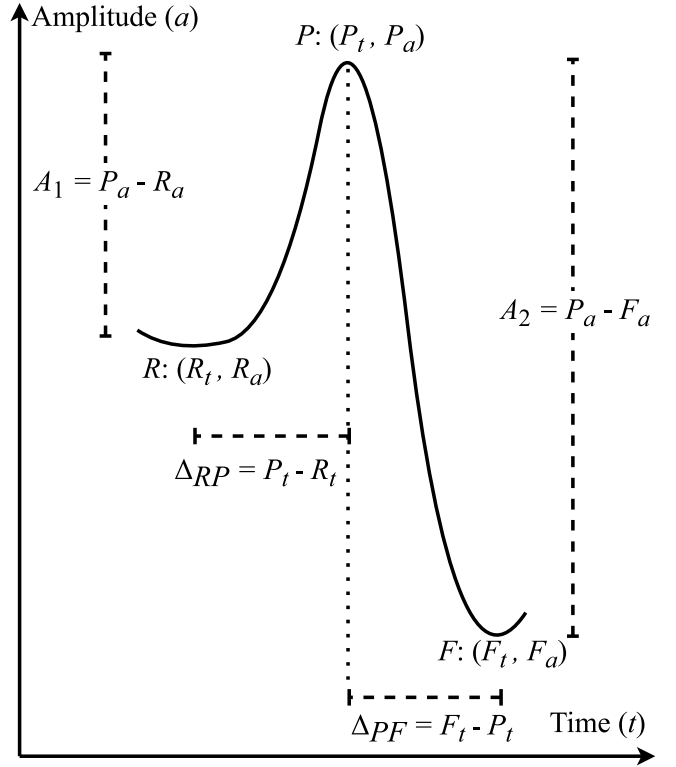


Fig. 1. Template spike with RPF waveform.

- For each local maximum, P , within every electrode channel of the segment, both the preceding local minimum, R , and the following local minimum, F , are identified. The average, μ_A , and the sample standard deviation, s_A , of the quantity $A = (A_1 + A_2)/2$ is computed. A peak is considered if the following conditions are met: (i) $A \geq \mu_A + 2s_A$, (ii) $A \geq 20 \mu V$, and (iii) $0.1 \leq A_1/A_2 \leq 6$.
- $20 \text{ ms} \leq \Delta_{RP} + \Delta_{PF} \leq 240 \text{ ms}$
- $|\Delta_{RP} - \Delta_{PF}| \leq 0.75 (\Delta_{RP} + \Delta_{PF})$
- P_a has to be at least 1.25 times larger than all the amplitudes within 140 ms prior to R .
- Within every RPF waveform, the absolute of the slopes of the steepest rising edge and the steepest falling edge should be at least 0.8 mV/s . In addition, the maximum absolute curvature between the rising and falling edges should be at least 0.55 mV/s^2 .
- It was shown in [53] that independent component analysis (ICA) has higher ability to concentrate spike events (as expressed through a spikiness index) into fewer number of components when compared to both raw MEG and singular value decomposition (SVD). The spikiness index of the m th component, $z_m(t)$, is given by: $I_m = \max |z_m(t)| / \frac{1}{T} \int |z_m(t)| dt$, where the maximum is computed over the entire observation interval T (3 s in our case). Instead of performing ICA followed by selecting $\mathfrak{F}$ components with the highest spikiness indices, we use reconstruction ICA (RICA) to extract $\mathfrak{F}$ (9 in our MATLAB implementation) sources directly. In its simplest form, RICA learns an unmixing matrix, $W \in \mathbf{R}^{\mathfrak{F} \times 19}$, that minimizes $\lambda \|Wx\|_1 + \frac{1}{2} \|W^T Wx - x\|_2^2$, where $x \in \mathbf{R}^{19 \times 300}$ represents the whole EEG segment recording from the 19 scalp electrode channels. The term λ is a hyperparameter that determines how important the minimization of the first part of the objective function is compared to the second part. The second part tries to enforce the orthonormality across the sources (not mere linear independence) by trying to enforce $W^T W$ to be

equal to the identity matrix I . We hint out that this is not feasible due to the given under-completeness (i.e., $\mathfrak{F} < 19$). However, the algorithm is encouraged to enforce this orthonormality through the soft reconstruction L_2 -norm loss. L_2 -norm is avoided in the first part of the objective function due to its undesired property of discouraging sparsity in the resulting sources. Hence, L_1 -norm is used instead as our aim is to enforce the sparsity of the sources, i.e., the reading at a specific electrode is only affected by a few sources. A new time course $d(t)$ is then computed and is given by $\max|z_m(t)|$, where the maximization is done across the $\mathfrak{F}$ components. The $d(t)$ time course is then simply thresholded. For a certain RPF waveform that satisfied all the previous conditions, if $d(P_i) < 0.01$, then this waveform will be omitted.

- (g) It is important to observe that strong artifacts can easily produce high spikiness index (I_m defined in the previous step). Thus, to enhance the selectivity of the algorithm further, we fit a single current dipole using the multiple signal classification (MUSIC) algorithm within ± 20 ms around each P satisfying all the conditions above. Around each P , the EEG recording, $\mathbf{X} \in \mathbf{R}^{19 \times L}$, is given as: $\mathbf{A}\mathbf{s} + \mathbf{n}$. L is the length of the cropped 40 ms recording in samples, $\mathbf{A} \in \mathbf{R}^{19 \times r}$ is the mixing matrix of the r sources responsible for producing the interictal activity present in $\mathbf{s} \in \mathbf{R}^{r \times L}$. r is set to 4 as recommended in [53] and $\mathbf{n} \in \mathbf{R}^{19 \times L}$ denotes the noise. Each column of the matrix $\mathbf{A}$ is denoted by $\mathbf{a}(\theta) = \mathbf{G}(\rho)\phi$ where $\mathbf{G}(\rho) \in \mathbf{R}^{19 \times 3}$ is the gain matrix. $\mathbf{a}(\theta)$ represents the effect of a certain source (located at position ρ in the brain and with orientation ϕ) on the 19 scalp electrodes. Hence, θ is characterized by the $\{\rho, \phi\}$ pair. In practice, the gain matrix is formed as follows: The MRI is segmented using any of the available techniques like the one in [54] which makes use of a CNN model. As the electromagnetic properties of the segmented regions (skin, skull, grey/white matter) are known, the effect of a source at location ρ on the electrodes can be determined by solving the forward model which results in the head model. Since we do not have MRIs for the patients, we acquired a template head model in Brainstorm software [55]. The covariance matrix of the data, $\mathbf{X}\mathbf{X}^T$, has eigenvectors Φ which can be written as $[\Phi_s, \Phi_n]$ where Φ_s and Φ_n represent orthonormal bases for the epileptogenic signal subspace and the noise subspace, respectively. The MUSIC algorithm finds the source locations as those for which the corresponding $\mathbf{a}(\theta)$ projects almost entirely into the estimated signal subspace [56], i.e., high subspace correlation (denoted by *subcorr*). Hence, it is desired to find θ that maximizes $\text{subcorr}(\mathbf{a}(\theta), \Phi_s)_1^2$ which is equivalent to finding ρ that maximizes $\text{subcorr}(\mathbf{G}(\rho), \Phi_s)_1^2$ since $\mathbf{a}(\theta)$ is in the span of $\mathbf{G}(\rho)$. Thus, we want to find the maximum dot product of two normalized vectors (one in the column space of $\mathbf{G}(\rho)$ and the other in the column space of Φ_s). This is how $\text{subcorr}(\mathbf{G}(\rho), \Phi_s)_1$ is defined in [57] which is the maximum subspace correlation coefficient. Note that we are only fitting a single dipole. Therefore, there is no need to find the corresponding ϕ that is necessary to fit the subsequent dipoles as described in [56]. To find $\text{subcorr}(\mathbf{G}(\rho), \Phi_s)_1$, we first need to orthogonalize the two matrices by carrying out singular value decomposition (SVD). Since Φ_s is already orthogonalized, we only decompose $\mathbf{G}(\rho)$ as $\mathbf{U}_{\mathbf{G}(\rho)}\Sigma_{\mathbf{G}(\rho)}\mathbf{V}_{\mathbf{G}(\rho)}^T$. $\text{subcorr}(\mathbf{G}(\rho), \Phi_s)_1$ is the highest singular value of the matrix $\mathbf{U}_{\mathbf{G}(\rho)}^T\Phi_s$ and therefore $\text{subcorr}(\mathbf{G}(\rho), \Phi_s)_1^2 = \lambda_{\max}(\mathbf{U}_{\mathbf{G}(\rho)}^T\Phi_s\Phi_s^T\mathbf{U}_{\mathbf{G}(\rho)})$ where $\lambda_{\max}()$ is the maximum eigenvalue of the enclosed square matrix. The single dipole is fitted if the maximum subspace correlation is above 0.925 for any ρ .

Note that all the previously mentioned rules are inspired by our earlier work in [20] and from Ossadtchi et al. in [53]. However, the thresholds were relaxed in a way that would make the algorithm

sensitive enough to capture all IEDs that we previously had. Indeed, this came with the price of having potential false positives. From the recordings of the 39 patients, 759 flagged (by the pre-filtering algorithm described above) EEG segments were checked by an expert and were split into 166 IED segments and 593 NIED segments. Hence, we observe a high imbalance in the data which is quite typical in IED detection tasks. It is the responsibility of the subsequent ML algorithms (described in the following section) to filter out the NIED segments. Note that the amount of segments is somehow limited considering the amount of data needed to train modern NN architectures. To alleviate this problem, random right-left reflections of the EEG segments are performed during the training of the different models to help them observe more variations of IED/NIED segments. This approach is motivated by the widely used image augmentation techniques. We start by trying several architectures in an attempt to discover the one with the best performance. For this, we split the data into 5 folds (with almost the same number of IED and NIED segments) and then trained the different models in an inter-patient fashion. To maintain fairness in the comparison, the number of trainable parameters is kept nearly the same (10–11 K) for all models.

3. Methods

3.1. Vanilla self-attention

Since EEG data is sequential in nature, recurrent neural networks (RNNs) were extensively used with EEG in different applications. For instance, epileptic seizures and spikes were detected in [29,58–60], respectively using variations of long short-term memory (LSTM) [61]. It should be noted that for EEG data, the sequence can be very long depending on the sampling rate and the duration of the analyzed segment which would raise the problem of vanishing gradients. Although LSTM and gated recurrent unit (GRU) [62] architectures were able to mitigate this problem which is very prominent in vanilla RNN, it was shown that the self-attention mechanism is more adept at solving it entirely. This is because self-attention gives the output at each time step the ability to peek at the input of each time step [51]. In [63], self-attention was applied along with LSTM to detect epileptic spikes. Although the attention mechanism could be used with the different variations of RNN, it was shown in [51] that with the help of positional encoding of the time samples and masking, the attention mechanism is sufficient to handle the ordered nature of the input. We briefly describe the architecture in the following enumerated steps:

1. For each input sample, i , across the attention dimension (time in this model) and each attention head, h , among the H attention heads (set as a hyperparameter), a linear dense layer generates a key-value pair, i.e., $(\mathbf{k}_{h,i}, \mathbf{v}_{h,i})$. The lengths of $\mathbf{k}_{h,i}$ and $\mathbf{v}_{h,i}$ vectors are d_{qk} and d_v , respectively.
2. Another dense layer is used to generate a query vector, $\mathbf{q}_{h,j}$ for each output sample, j , and each attention head. In this model, each input sample, i , has a corresponding output sample, j . Hence, the name self-attention.
3. The term $e_{h,i,j}$ represents the effect input, i , will have on a certain output, j , at a certain attention head, h , and is given by $\mathbf{q}_{h,j}^T \mathbf{k}_{h,i} / \sqrt{d_{qk}}$ (scaled dot-product). The normalized version of those effects, $\chi_{h,i,j}$, is obtained by applying a *softmax* function across the input dimension to yield $\sum_i \chi_{h,i,j} = 1$. For a certain j , the output would be the concatenation of the quantity $\sum_i \chi_{h,i,j} \mathbf{v}_{h,i}$ across the different attention heads.

Positional Encoding: Since the attention mechanism described above includes only linear layers whose inputs have no inherent ordering, positional encoding is needed in order to distinguish between the input features at different i values. Instead of concatenating the absolute time to the features at each sample, the features at i are added to the vector $\mathbf{pe}(i) = [\sin(i/100^{2 \times 1/E}), \cos(i/100^{2 \times 1/E}), \sin(i/100^{2 \times 2/E}), \cos(i/100^{2 \times 2/E})]$.

$\cos(i/100^{2 \times 2/E}), \sin(i/100^{2 \times 3/E}), \cos(i/100^{2 \times 3/E}), \dots]^T$ where E is the number of features (19 electrodes), which is also the length of the $pe(i)$ vector.

Similar to the encoder block in [51], dropout is applied to the output of the attention layer, followed by addition with the original input before layer normalization [64] is applied. The output from such a process is inputted to a feed-forward leaky rectified linear unit (ReLU) layer on top of which the same steps are applied (dropout, additive skip-connection, and finally layer normalization). Hence, the encoder block consists of two sub-blocks of similar architectures except that one has the attention layer while the other introduces non-linearity through the feed-forward layer.

This model is formed by stacking the following layers: Positional Encoding, Encoder Block ($H = 7, d_{qk} = 7, d_v = 7$), 1D-MaxPooling (6), Positional Encoding, Encoder Block ($H = 6, d_{qk} = 6, d_v = 6$), 1D-MaxPooling (5), Positional Encoding, Encoder Block ($H = 4, d_{qk} = 4, d_v = 4$), 1D-MaxPooling (2). The output tensor from that sequence has a shape of $(B, 19, 5)$. Conv1D layer with 19 input channels (Note: The number 19 here does not refer to the number of EEG scalp electrodes whereas B denotes the batch size), 1 output channel, kernel size of 1, and a leaky-ReLU activation function is then applied. Finally a dense layer with a sigmoid activation function is applied to obtain the classification of the EEG segment.

3.2. Hierarchical attention

One of the elegant approaches that was applied on EEG for the task of human decision prediction was the hierarchical LSTM model with attention presented in [34]. This approach addresses the non-stationarity problem of the long EEG segment by dividing it into smaller epochs on which LSTM networks are applied separately. By doing so, the vanishing gradient problem is also better addressed by reducing the length of the sequence the LSTM network is applied on. Another LSTM network is applied on the encodings of the epochs. Hence, the name hierarchical LSTM. We adopt a slight variation of this model by replacing the LSTM with attention (again to address the problem of vanishing gradients entirely). The description of the model is as follows:

1. The EEG segment is divided into 0.5-second epochs with 50% overlap.
2. Positional encoding is applied on the samples of each epoch in the same manner described in the previous subsection. A separate attention layer is then applied on the encoded samples of each epoch. However, it is desired to have one vector embedding representing the encoding of the whole epoch. For that, we do not generate a query vector from each sample as was the case in the vanilla self-attention model. Instead, each epoch is assigned a single trainable query vector per attention head. This approach is somehow similar to the attention mechanism proposed in the state-of-art *Fastformer* architecture [65]. A dot product between the epoch query vector and sample keys is then carried out to obtain the weight coefficients that will be multiplied with the sample values in the same way described previously. We refer to this as the sample attention. The used hyperparameters are: $H_{sample} = 3, d_{qk, sample} = 5, d_{v, sample} = 5$. The middle dense layer maps $15 (H_{sample} \times d_{v, sample})$ input features to 15 output features on which leaky-ReLU is applied.
3. After step 2, we have a new sequence of embeddings, each representing the encoding of its corresponding epoch. Hence, another positional encoding and attention layer (again with a single trainable query vector per attention head) are applied to this epoch sequence. The hyperparameters of this epoch attention are: $H_{epoch} = 3, d_{qk, epoch} = 5, d_{v, epoch} = 5$.
4. Finally, the output of the epoch attention (step 3) is fed to an MLP to obtain the classification of the segment. Leaky-ReLU activations are used for all layers of the MLP except for the final one where sigmoid function is used. The number of neurons within the layers are as follows: $15 (H_{epoch} \times d_{v, epoch})$, 10, 5, and finally 1.

3.3. Vanilla VGG

Despite their design simplicity, CNN architectures have shown many successes in many applications. These successes are not only limited to the conventional image/video input data types, but also extend to sequential data such as EEG signals for spike detection [32] and textual data for emotion analysis [66]. The VGG CNN architecture described in [31] was able to perform well on our task as will be presented in the Results section. In our implementation, we stack the following layers: Conv1D (19, 32, 3, 1), Leaky-ReLU, Conv1D (32, 32, 3, 1), Leaky-ReLU, Conv1D (32, 32, 2, 2), Leaky-ReLU, Conv1D (32, 16, 3, 1), Leaky-ReLU, Conv1D (16, 16, 3, 1), Leaky-ReLU, Conv1D (16, 16, 2, 2), Leaky-ReLU, Conv1D (16, 8, 3, 1), Leaky-ReLU, Conv1D (8, 8, 3, 1), Leaky-ReLU, Conv1D (8, 8, 2, 2), Leaky-ReLU, Conv1D (8, 4, 3, 1), Leaky-ReLU, Conv1D (4, 4, 3, 1), Leaky-ReLU, Conv1D (4, 4, 2, 2), Leaky-ReLU, Conv1D (4, 2, 3, 1), Leaky-ReLU, Conv1D (2, 2, 3, 1), Leaky-ReLU, Conv1D (2, 1, 3, 1), Leaky-ReLU, Conv1D (1, 1, 3, 1), Leaky-ReLU, Conv1D (1, 1, 3, 1), Leaky-ReLU, Conv1D (1, 1, 3, 1), Leaky-ReLU, Conv1D (1, 1, 3, 1), Sigmoid. The arguments of the Conv1D layers are the number of input channels, number of output channels, kernel size, and finally the stride. Valid padding is used for all the Conv1D layers. The Conv1D layers with kernel size of 2 and a stride of 2 are placed to reduce the dimension significantly (similar to pooling layers).

3.4. Satelight

In this subsection, we provide a brief overview of the architecture described in [35]. In the beginning, 15 temporal 1-D convolutional kernels (each of length 50 time samples) are applied to the EEG channels. For each of the resulting 15 channels (depth channels not EEG electrodes), 2 kernels encode the readings from all the 19 EEG channels at each time sample. The output is then flattened to yield a tensor of shape $(B, \text{time length}, 30 = 15 \times 2)$. Batch normalization, ReLU, Dropout ($p = 0.2$), and MaxPooling (with a kernel of size 4 to reduce the time dimension) are then applied.

Two consecutive blocks are then introduced. In reference to [35], each block has two main sub-blocks and a MaxPooling layer. The first sub-block has the following structure: Attention Layer ($H = 1, d_{qk} = 30, d_v = 30$), Batch Normalization, Dropout ($p = 0.2$). The output from these three operations is added to the input, i.e., skip connection. The output of the first sub-block is then forwarded to the second which has the following structure: Linear Layer (30 input features, 30 output features), Batch Normalization, ReLU, Dropout ($p = 0.2$). The output of these operations is also added with the input to the second sub-block. Eventually, 1-D MaxPooling is then applied to reduce the time dimension. The kernel sizes of the MaxPooling layers of the first and second blocks are 4 and 5, respectively. The final output is then flattened and forwarded to a linear layer with a single output neuron and a sigmoid activation function.

We emphasize that although the general architecture of the Satelight is adopted, the number of layers and kernels were modified to fit our data. This is done for two reasons: First, our input EEG segment has a different time length and sampling rate, and the number of electrodes used is not the same as those used in the original Satelight paper. Second, we attempted to force all the models to have roughly the same number of trainable parameters (10k–11k) for a fair comparison.

3.5. Functional connectivity GNN (FC-GNN)

In a previous study [37] by our group, a statistically significant difference was found in the graph structures produced by the FC maps when comparing IED-segments to NIED-segments. This naturally led us to conclude that GNNs could be applied on FC maps due to their ability to harness such graph composition dynamics and subtle changes in favor of our application.

Among a wide variety of FC measures, our choice of wPLI, as introduced originally in [46], was attributed to its frequency-detailed design nature where the connectivity could be inspected under different frequency sub-bands independently as well as for its ability to overcome volume conduction. This wPLI measure at a certain frequency, f , between two time series, x and y , collected from two different electrodes is given by:

$$wPLI_{x,y}(f) = \frac{\left| \mathbb{E} \left[\Im \{ X(f) Y^*(f) \} \right] \right|}{\mathbb{E} \left[\left| \Im \{ X(f) Y^*(f) \} \right| \right]}, \quad (1)$$

where $|\cdot|$, $\mathbb{E}[\cdot]$, $\Im[\cdot]$, and $*$ represent the absolute, expectation, imaginary part, and the complex conjugate functions, respectively. Moreover, $X(f)$ and $Y(f)$ are the frequency representations of x and y , respectively. The computation of the above formula is intractable as it contains an expectation operator as well as a Fourier transform. To empirically compute it, each discretized 3-second EEG segment is divided into a sufficiently large number of overlapping windows, N , where $N \geq 30$ [46] in order to avoid large estimator bias [67]. This is done to replace the expectation operator by the arithmetic average; i.e., $\frac{1}{N} \sum_{n=1}^N$, where n is the index of the window. $X(f)$ in Eq. (1) is replaced by $X_n(f)$ which is the Fast Fourier Transform (FFT) of the n th window of x . Similarly, $Y(f)$ is replaced by $Y_n(f)$.

To compute the coupling between the two series within a specific frequency sub-band, S , an integral of the above formula needs to be carried out as such [68]:

$$wPLI_{x,y}^S = \frac{\int_{f_L}^{f_U} wPLI_{x,y}(f) df}{f_U - f_L}, \quad (2)$$

where f_L and f_U are the lower frequency bound and upper frequency bound of S , respectively. Note that f ranges from 0 Hz to $f_s/2$ Hz where f_s is the sampling frequency (100 samples/s).

In order to have binary edges within the FC graphs of the four frequency sub-bands of interest (δ , θ , α , and β), thresholding is done on two steps: (i) A threshold value between 0.5 and 0.6 was recommended in [69] to indicate the beginning of physical connectivity between two regions. Through empirical evaluations, our choice was set at a slightly higher threshold of 0.75 as the minimum wPLI measure below which connectivity links were omitted. (ii) Relying on the minimum FC value in i alone could yield very dense connectivity matrices in instances where a strong background activity surpassing the threshold level happens to exist. Hence, we adopt an adaptive threshold that would consider the highest 25% of all the connectivities provided that they are already above the 0.75 specified minimum threshold.

The proposed architecture shown in Fig. 2 is structured as follows:

1. Inspired by the good results demonstrated by the simple CNN in the Vanilla VGG subsection, our aim was to take advantage of it in building a new architecture to further enhance the performance. Instead of relying on a CNN to take-in the EEG segment as a whole while considering all the channels simultaneously, we use the CNN to produce the embedding of each EEG channel separately while disregarding others. To do this efficiently, we pass the EEG segment to a two-dimensional CNN with one-dimensional kernels that handle each channel independently. Hence, the kernels only observe the time samples of a certain channel while producing its embedding. The structure of the channel embedding CNN is as follows: Conv2D (1, 2, (1, 3)), Leaky-ReLU, Conv2D (2, 3, (1, 3)), Leaky-ReLU, MaxPool2D ((1,2)), Conv2D (3, 4, (1, 3)), Leaky-ReLU, Conv2D (4, 5, (1, 3)), Leaky-ReLU, MaxPool2D ((1,2)), Conv2D (5, 6, (1, 3)), Leaky-ReLU, Conv2D (6, 7, (1, 3)), Leaky-ReLU, Conv2D (7, 8, (1, 3)), Leaky-ReLU, Conv2D (8, 9, (1, 3)), Leaky-ReLU, MaxPool2D ((1,2)), Conv2D (9, 10, (1, 3)), Leaky-ReLU, Conv2D (10, 11, (1,

3)), Leaky-ReLU, Conv2D (11, 12, (1, 3)), Leaky-ReLU, Conv2D (12, 13, (1, 3)), Leaky-ReLU, MaxPool2D ((1,2)), Conv2D (13, 13, (1, 3)), Leaky-ReLU, Conv2D (13, 13, (1, 3)), Leaky-ReLU, Conv2D (13, 13, (1, 3)), Leaky-ReLU, Conv2D (13, 13, (1, 3)), Leaky-ReLU, Conv2D (13, 13, (1, 3)), Leaky-ReLU, Conv2D (13, 13, (1, 2)), Leaky-ReLU. The arguments to the Conv2D layers in order are the number of input channels, the number of output channels, and the kernel size. A stride of 1 and valid padding are used for all the Conv2D layers.

2. Positional encoding is applied to each of the 13-dimensional embeddings before they are passed to the GNNs (next step). This is done to help the GNNs which are inherently order invariant distinguish between different EEG channels. Hence, positional encoding here is done along the EEG channels dimension, not the time dimension as was the case in the Vanilla Self-Attention Section 3.1.
3. From the EEG segment, the wPLI FC maps of the different sub-bands are obtained as described above. The thresholded connectivities can be thought of as edges between EEG channel nodes whose embeddings have been obtained as in the previous step. Hence, the embeddings of the channels are forwarded to the four graphs to obtain a separate embedding for each graph. The embedding from each GNN is extracted according to the following: The input before getting processed by the GNN is referred to as $F^{(0)} \in \mathbf{R}^{19 \times 13}$ where each row corresponds to the positionally-encoded EEG channel embedding. In the first layer of the GNN, each node receives the embeddings from its neighbors and adds them together. Mathematically, this is done by multiplying the highly sparse FC adjacency matrix with $F^{(0)}$. This can be carried out efficiently using sparse matrix multiplication in deep learning (DL) frameworks such as pytorch. Each node then adds its own embedding to the accumulated embeddings of the neighbors. Note that it was found in [70] that the addition of a trainable ϵ parameter did not result in an enhanced performance and therefore we did not experiment its usage for simplicity. Each node finally passes the vector formed by the addition of its own embedding and the embeddings of the neighbors to a dense layer (13 input & 13 output neurons) followed by a leaky-ReLU activation. This yields $F^{(1)} \in \mathbf{R}^{19 \times 13}$ with each row representing the embeddings of a node after the processing of one layer. We add two more layers to obtain $F^{(2)} \in \mathbf{R}^{19 \times 13}$ and $F^{(3)} \in \mathbf{R}^{19 \times 13}$ using the same sparse FC adjacency matrix. $F^{(0)}$, $F^{(1)}$, $F^{(2)}$, and $F^{(3)}$ are finally concatenated together along the node feature dimension to form $F \in \mathbf{R}^{19 \times (4 \times 13)}$ so that each row represents the aggregated embedding across all layers. The graph embedding is then formed by summing all the rows of F together to form a 52-dimensional (4×13) vector.
4. Since each frequency sub-band generates its own graph embedding, we need to combine them before forwarding the new overall embedding to a classification MLP. One option would be to simply concatenate the embeddings together. Instead, we chose to apply an attention mechanism with a single attention head across the four embeddings. Note that the *softmax* function applied in an attention head to compute the χ parameters is a differentiable approximation of the maximization function which would simply attend fully to one of the embeddings. Therefore, the decision to apply attention was basically to observe the frequency embedding that the NN would learn on its own to pay the most attention to. We shall elaborate on this further in the results section. Hence, the input index, i , here is an index of the frequency sub-band. Since we only have a single attention head with a single output, we set the query to be a trainable vector (of hyperparameter length 15) that will generate the normalized attention weights (χ_δ , χ_θ , χ_α , χ_β) after being dot multiplied with the keys (k_δ , k_θ , k_α , k_β) generated from the embeddings of the graphs.

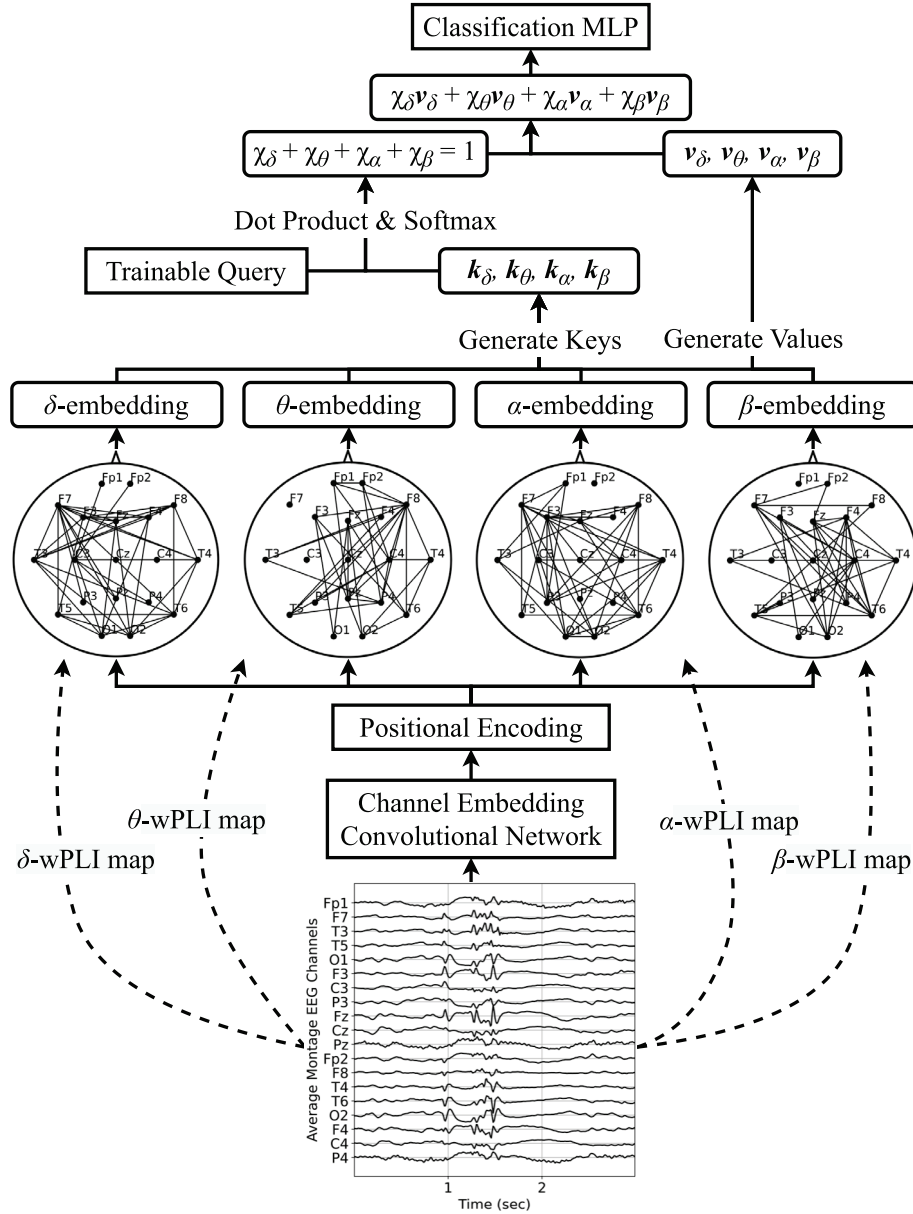


Fig. 2. FC-GNN model: Attention is used to combine graph embeddings of the four sub-bands. Positional encoding gives the GNNs the ability to distinguish between channels.

5. Finally, a weighted average of the values (v_δ , v_θ , v_α , v_β) also generated from the embeddings (each of length 15) is obtained and forwarded to a classification MLP with dense layers of sizes: 15, 10, 5, and finally 1 with a sigmoid function. All other dense layers are followed by leaky-ReLU activation.

3.6. Complete attention GNN (CA-GNN)

As can be seen from the results presented in Section 4 below, the model described in the previous subsection (FC-GNN) is able to yield better results compared to other existing architectures on the collected Baptist dataset. However, one of its shortcomings is its reliance on the computation of FC maps for the different frequency sub-bands. Hence, the need to develop a similar graph-based model that requires no feature engineering (represented in the FC analysis in FC-GNN). Therefore, we start by using a CNN to extract embeddings for each EEG channel separately followed by positional embedding in the same way described in steps 1 and 2 in the previous subsection. Again this would yield the embedding matrix $F^{(0)} \in \mathbb{R}^{19 \times 13}$.

However, without carrying out the FC analysis (Eqs. (1) and (2) are not used), it was unclear how the EEG segment can be reflected into a meaningful graph structure with defined edges. It is important to observe the similarity between the self-attention when applied across the nodes dimension and the message passing mechanism adopted in GNN, where each node passes its embeddings to its neighbors. The only difference is that in the self-attention case, the receiving node (EEG channel) gets to choose the most important channels to focus on (i.e., the attention heads). Indeed, this comes with a quadratic computational complexity cost. We analyze further this issue in Section 4. Similarly, the additive residual connection in the self-attention encoder block resembles the addition of the node's own embedding to the received embeddings from the neighbors. Overall, the encoder block with all of its constituents (self-attention, residual connection, and non-linearity) is similar to the GNN layer presented in the previous subsection.

The CA-GNN architecture follows the assumption that all the nodes of the graph (single graph not 4 as in FC-GNN) are directly connected to each other yielding a complete graph. We then use self-attention

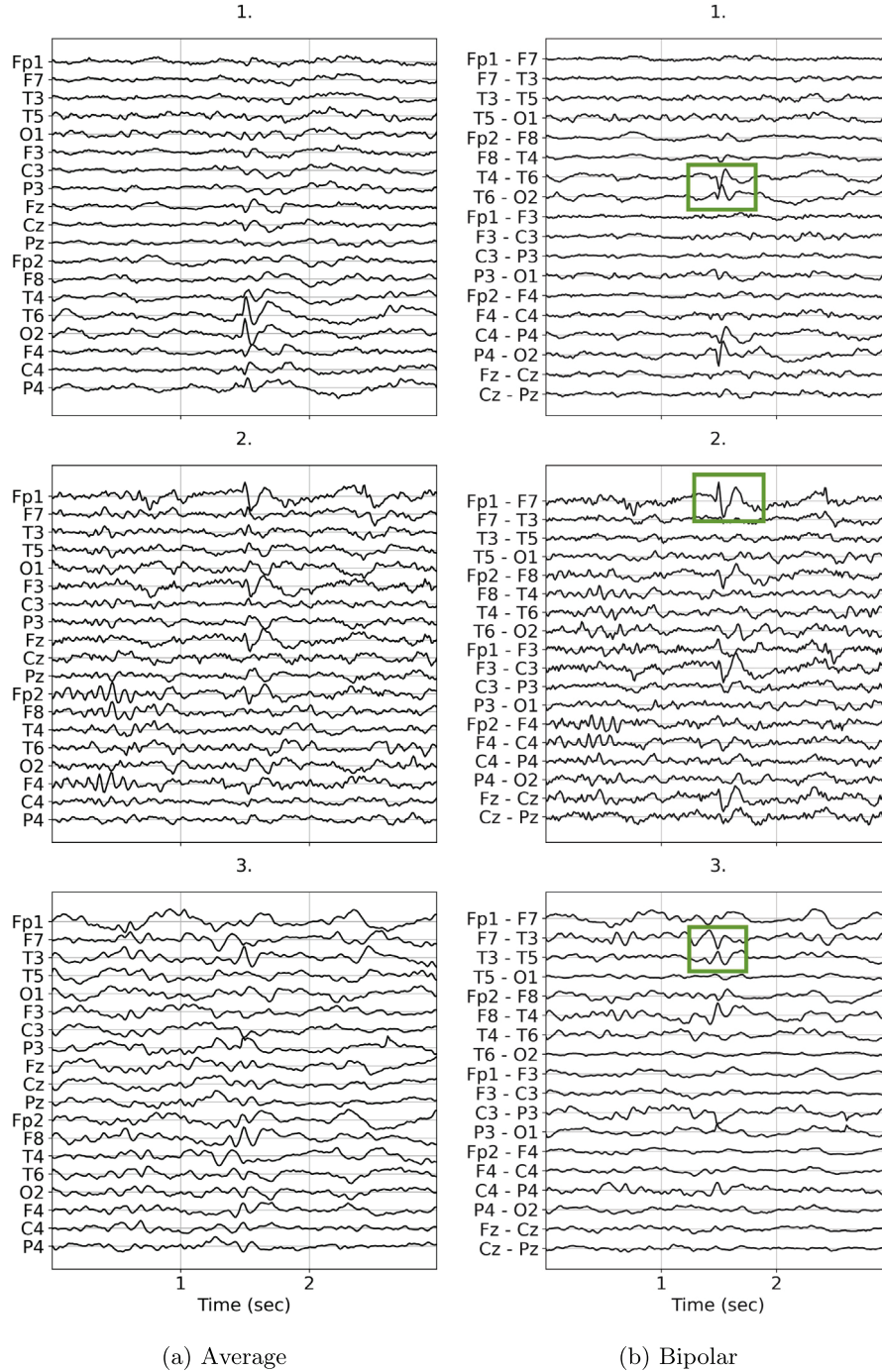


Fig. 3. TP with respect to both FC-GNN and CA-GNN.

with 3 attention heads in order to let each node learn the most important 3 neighbors it should softly attend to. Hence, the same classical self-attention mechanism is applied here but on the nodes dimension instead of the time dimension as was the case in the Vanilla Self-Attention model.

As per the explanation above, we apply an encoder block on $F^{(0)}$ followed by positional encoding to obtain $F^{(1)} \in \mathbf{R}^{19 \times 13}$. As done previously, we add two more layers to obtain $F^{(2)} \in \mathbf{R}^{19 \times 13}$ and $F^{(3)} \in \mathbf{R}^{19 \times 13}$. For all the encoder blocks, H , d_{qk} , and d_v were set to 3, 5, and 7, respectively. The embedding matrices are concatenated together to form $F \in \mathbf{R}^{19 \times (4 \times 13)}$ and eventually all the rows are added up to yield a 52-dimensional graph embedding vector. This vector is then passed to an MLP with dense layers of the following sizes: 52, 25, 10, 5, and

finally 1. All layers are followed by leaky-ReLU activation except for the output layer that has sigmoid activation function.

4. Results & discussion

4.1. Baptist dataset results

The implementation code of the six models described in the previous section (Self-Attention, Hierarchical Attention, Vanilla VGG, Satelight, FC-GNN, and CA-GNN) is made public for the research community¹.

¹ https://github.com/ahmedmohammed107/GNNs_EEG_Spike_Detection-

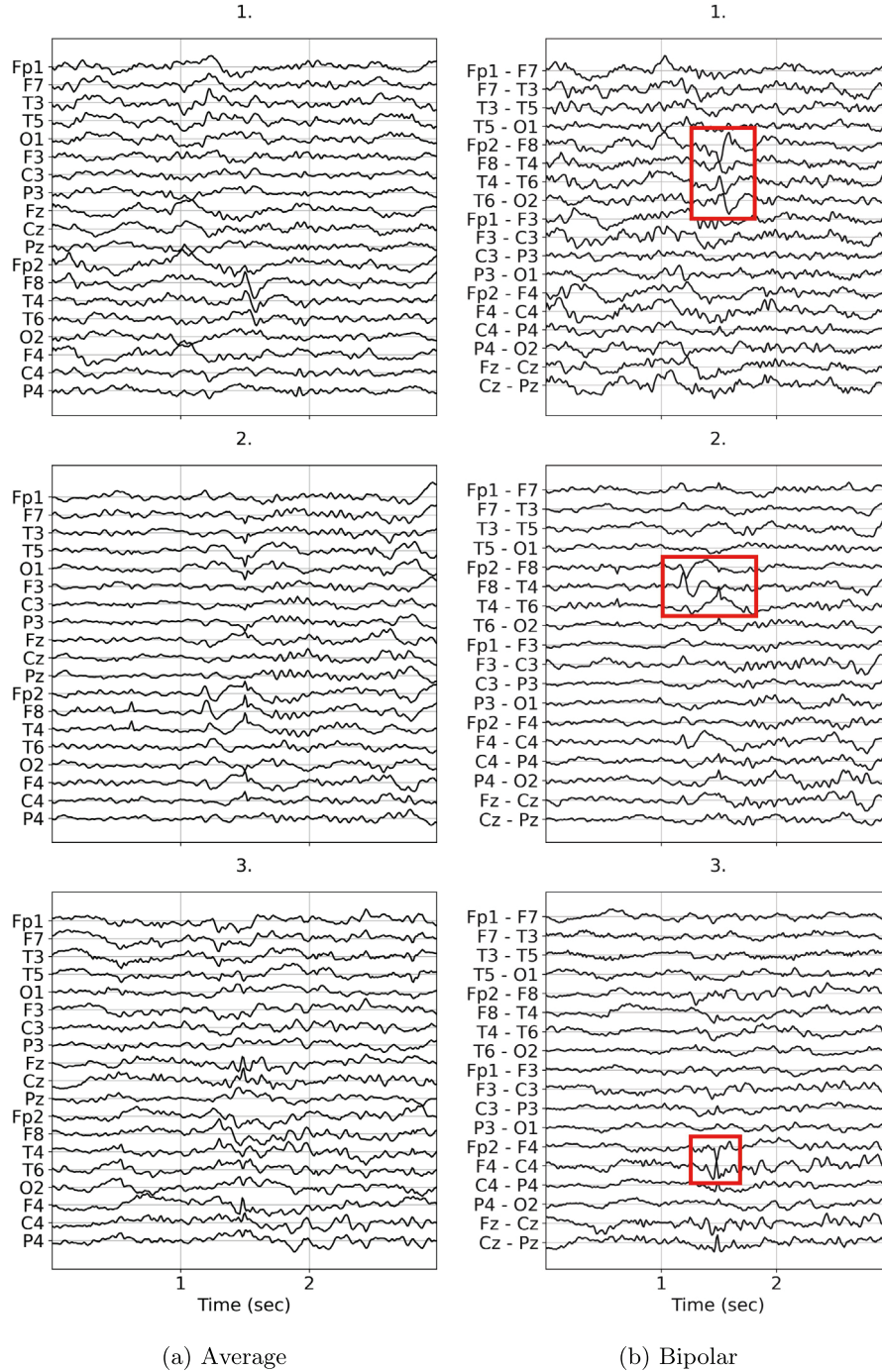


Fig. 4. FN with respect to both FC-GNN and CA-GNN.

As mentioned before, the results presented in this section are obtained by testing the performance of the different models on data from patients that were never considered during the training phase. Shown in Figs. 3, 4, 5, and 6 are 12 examples of EEG segments (from 12 different patients) in the dataset which contained discharges that were captured by the pre-filtering algorithm and thus were labeled as candidate segments. Due to its high sensitivity, the pre-filtering algorithm could capture multiple discharges in the segments. For simplicity, we only show one such detection per segment. Each figure shows three examples of EEG segments belonging to the same category plotted in average montage and bipolar montage. The four categories are true positive (TP), false negative (FN), false positive (FP), and true negative (TN). This labeling is made with respect to the two proposed models

(FC-GNN and CA-GNN). All detections for the different models are made with a threshold of 0.5. It is important to emphasize that the pre-filtering algorithm as well as the different models are applied to the average montage. However, we also present the segments in bipolar montage to facilitate the recognition of IEDs for experts. The IEDs as identified by the expert and captured by the pre-filtering algorithm are marked with green boxes for TP (in Fig. 3(b)) and red boxes for FN (in Fig. 4(b)) on the bipolar plots. Both Self-Attention and Vanilla VGG failed to identify the IED present in the first TP example. Only FC-GNN and CA-GNN classified the second TP example correctly. Satelight was able to detect the IED in the third TP example. Hierarchical Attention correctly detected the IEDs in the first two FN examples. Self-Attention and Satelight were able to detect only that in the first FN example.

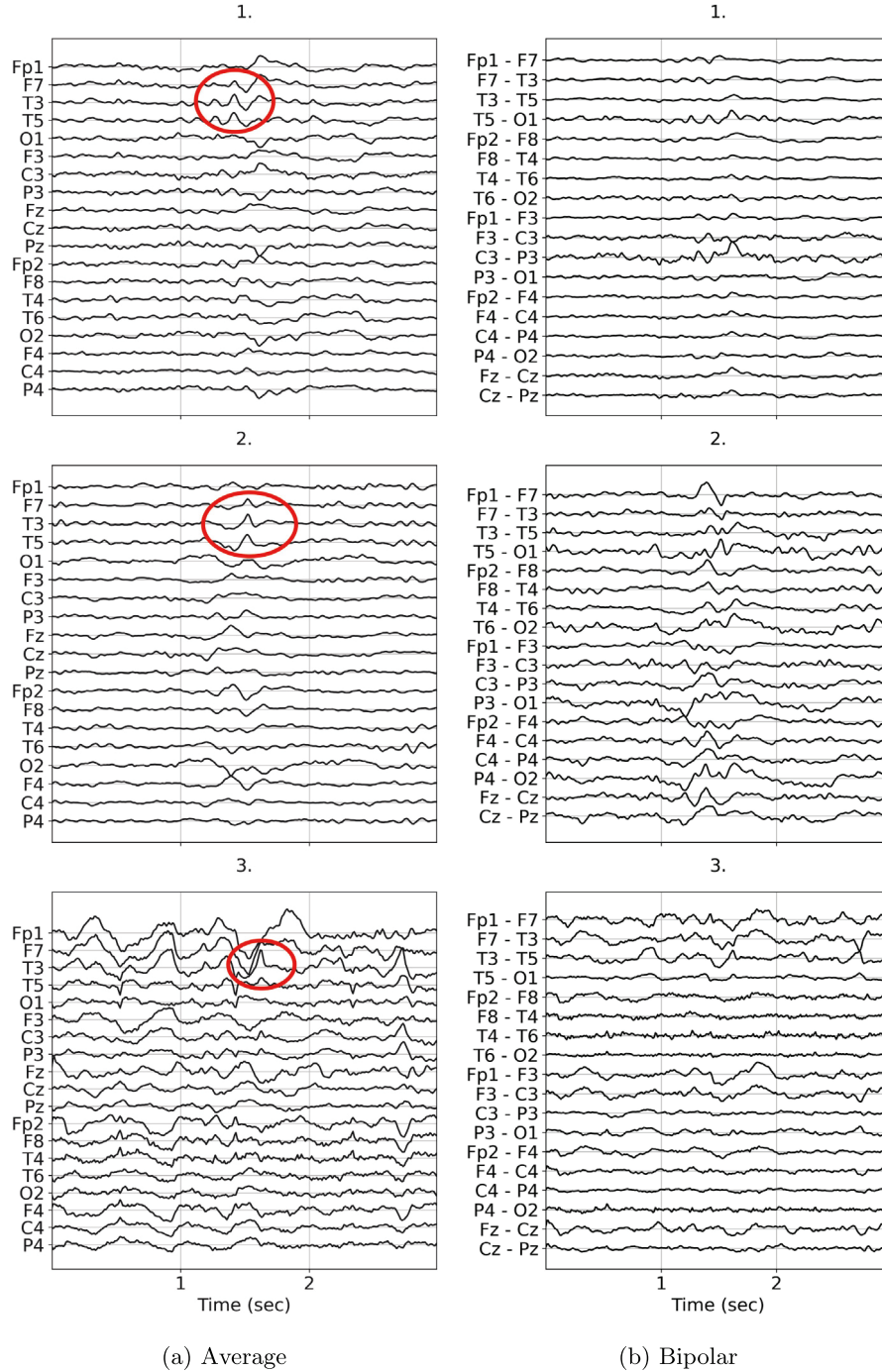


Fig. 5. FP with respect to both FC-GNN and CA-GNN.

We show NIED-segments with discharges that were captured by the pre-filtering algorithm marked in red ellipses for FP (in Fig. 5(a)) and green ellipses for TN (in Fig. 6(a)). While we use the colors green/red to indicate correct/incorrect classification, we stress on the fact that the classification is done on the level of the whole EEG segment. The Vanilla VGG model classified all the presented FP examples correctly. The first and third examples were also correctly classified by the Satelight model. Self-Attention and Hierarchical Attention were able to correctly classify the second and third FP examples, respectively. Self-Attention and Hierarchical Attention misclassified all TN examples presented in Fig. 6. The third TN example was also misclassified by the vanilla VGG model.

Due to the imbalanced nature of the datasets used to train IED detection algorithms, several studies present the overall performance in terms of sensitivity or recall (of all IED-segments, how many are captured by the algorithm?) and precision (how accurate is the algorithm when it does make a detection?) instead of the overall accuracy. These measures however, require a predefined specific threshold and therefore vary with its change. The change in the confidence threshold represents the trade-off between the desired high precision and high recall.

One of the widely used metrics in classification tasks that integrates out the confidence threshold dependence is the area under the receiver operating characteristics curve (ROC-AUC). Another metric that also takes into account several possible threshold levels and is more

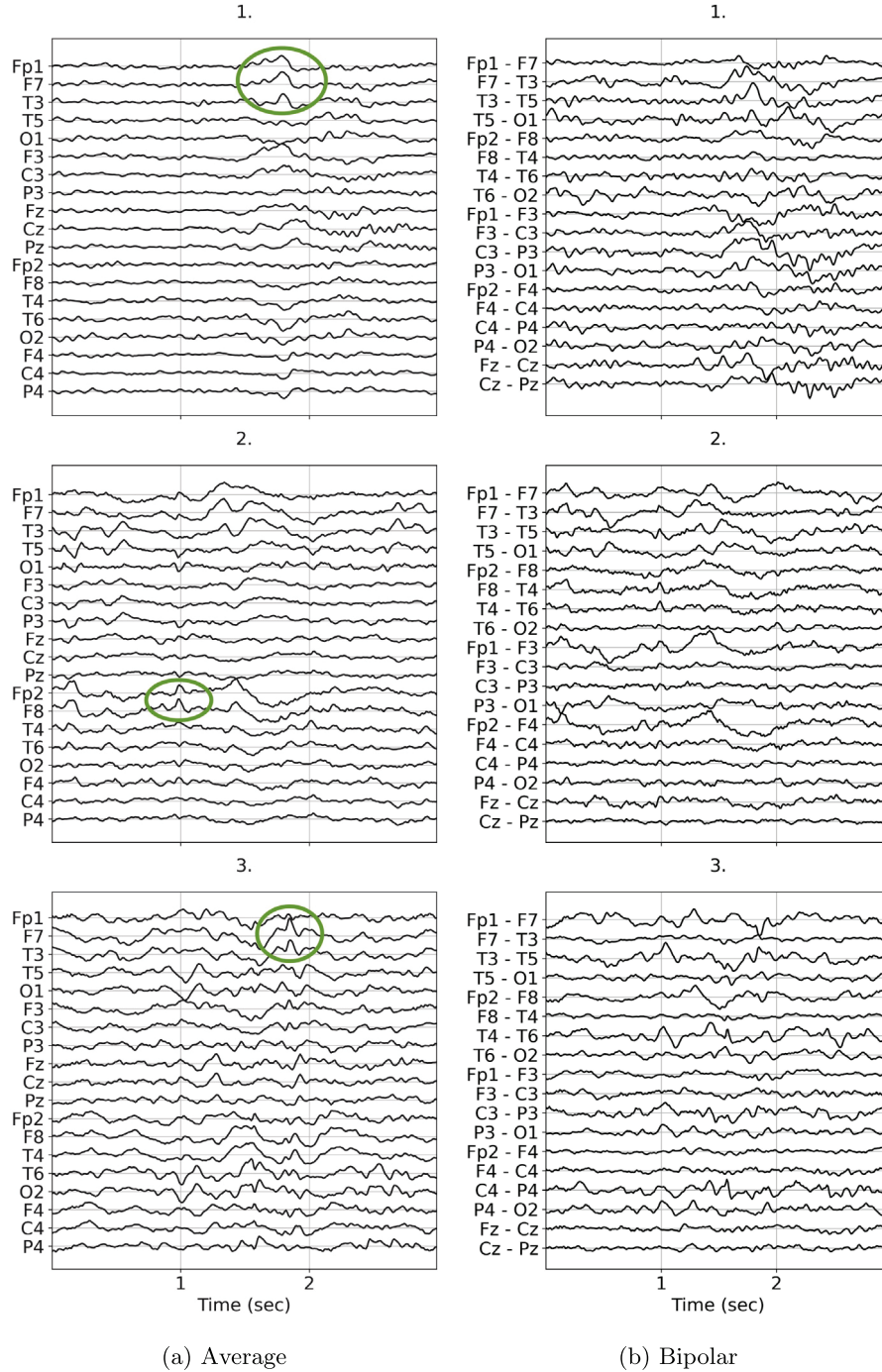


Fig. 6. TN with respect to both FC-GNN and CA-GNN.

appropriate for highly imbalanced datasets is the average precision (AP). This AP measure is computed from the precision-recall curve which is obtained by plotting the precision on the vertical axis and recall on the horizontal axis for different confidence threshold values. Note that calculating the area under the precision-recall curve using the trapezoidal rule could yield extra optimistic results. Instead, AP is computed as the weighted mean of precision values achieved at each threshold, with the increase in the recall used as the weight. Hence, we show the overall performance of the models in terms of the AP metric.

The results presented in Table 1 are the best observed among several training trials (each with a different random initialization). This was done because it was observed that for some trials, the training was stuck at sub-optimal points. This retraining approach was adopted for

all models to maintain a fair comparison. It is clear how the proposed FC-GNN yields higher average AP with less standard deviation than those of the Vanilla Self-Attention, Hierarchical Attention, Vanilla VGG, and Satalight. It was anticipated to observe a high attention that the network would pay to the embeddings of specific frequency sub-bands. This is due to the role played by the θ -band in highlighting epileptogenic activity as reported in several studies. For instance, IED spikes were shown in [71] to have an impact on the theta rhythm for patients with temporal lobe epilepsy (TLE). Patients with mesial TLE (MTLE) showed significantly less θ -band power compared to the non-MTLE (NMTLE) group in [72]. In a previous study by our group [73], the θ -band required the least amount of power penalization among other frequency sub-bands to delineate the region of interest (ROI)

Table 1

Validation Average Precision (AP) scores on the Baptist dataset for the 5 groups of patients. The results are presented in the same order of subsections 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6.

	Self attention	Hierarchical attention	Vanilla VGG	Satelight	FC-GNN	CA-GNN
Group 1	0.8587	0.8524	0.9327	0.8601	0.952	0.9706
Group 2	0.9538	0.8998	0.9329	0.8991	0.9706	0.9747
Group 3	0.8578	0.7555	0.8125	0.9174	0.9603	0.9647
Group 4	0.9138	0.8692	0.9568	0.9586	0.9991	0.9927
Group 5	0.9302	0.896	0.9649	0.9745	0.9837	0.9912
Avg.	0.9029	0.8546	0.92	0.9219	0.9731	0.9788
Std. Dev.	0.0431	0.0587	0.0618	0.046	0.0187	0.0125

significantly. Surprisingly, the χ_θ parameter was not consistently higher than the other attention weights for the different groups of patients. Even within the same group, the frequency sub-band with the highest χ parameter was different from trial to trial. It is unclear for us why the model did not pay special high attention to the θ -band despite the aforementioned results of previous studies. Perhaps, this behavior could be attributed to the relatively limited available training dataset and should be further investigated once a larger dataset becomes available. We discuss this further in the discussion subsection. On this dataset, the CA-GNN architecture offers a slight improvement over the FC-GNN.

4.2. TUEV dataset results

In this subsection, we briefly describe the preparation procedure of the open-source TUEV dataset and present the results obtained when applying our method on this dataset. The same preprocessing described earlier for the collected Baptist dataset is adopted here. We consider only the average montage of the 19 electrodes of interest, applied notch and band-pass filters, and downsampled the EEG to 100 samples/s.

The TUEV dataset is divided into training and evaluation sets containing labeled EEG recordings. Six labels are assigned to 1-second waveforms of the EEG electrodes. The six labels are: spike and slow wave (1: spsw), generalized periodic epileptiform discharge (2: gped), periodic lateralized epileptiform discharge (3: pled), eye movement (4: eyem), artifact (5: artf), and background (6: bckg). Following [74], we assign the first three labels to the IED class whereas the last three labels are assigned to the NIED class. From the training set, data of some patients were kept as a hold-out validation set. No patient existed in both sets. The evaluation set was used for testing. Since our method processes 3-second EEG segments, we extend the considered IED segments around the original 1-second labeled waveforms. Extended NIED segments are considered if no overlap occurs with any of the extended IED segments.

This dataset is balanced for IED detection task. In particular, the number of IED/NIED segments is 3509/3648 in the training set, 890/990 in the validation set, and 1832/1649 in the test set. It is explicitly stated in the TUEV dataset documentation that the background label is a catch-all label that is given to a waveform if it clearly belongs to none of the first five labels. This could explain the relatively balanced nature of the dataset. This was the case in the dataset used in [35] with 16008 IEDs and 15478 artifacts collected from 50 children. Those discharges were collected by a rule-based architecture and then used to train the Satelight model. The resulting testing AP for the different models are as follows: Vanilla Self-Attention $\rightarrow$ 0.9692, Hierarchical Attention $\rightarrow$ 0.9113, Vanilla VGG $\rightarrow$ 0.97, Satelight $\rightarrow$ 0.9575, FC-GNN $\rightarrow$ 0.963, and CA-GNN $\rightarrow$ 0.9879. We note how the CA-GNN architecture outperforms all other approaches including FC-GNN on both datasets (Baptist and TUEV). This is probably due to the freedom granted to each node to pay attention to other nodes as directed by backpropagation instead of dictating the neighboring nodes by FC analysis.

4.3. Discussion

Computational Complexity: For the FC-GNN architecture, the overall complexity is dominated by the FC analysis which consists of two steps. First, we compute the FFT for all windows of all electrode recordings. The complexity of this part is $O(E * N * l * \log(l))$ where E , N , and l represent the number of electrodes, the number of windows, and the length of each window, respectively. The second step involves computing the wPLI metric between all possible pairs of channels as described in Eq. (1) which takes $O(E^2 * N * l)$ complexity. Hence, the overall complexity is $O(E * N * l * \log(l) + E^2 * N * l)$. For the CA-GNN architecture, the computation is dominated by the attention part across the electrodes which requires quadratic complexity (i.e., E^2). Note that the number of electrodes is relatively limited even in the state-of-the-art HD-EEG (typically below 512). The significance of the E^2 factor is thus limited. For vanilla self-attention and Satelight, the computation is quadratic in the length of the EEG segment. With the high sampling rate of modern EEG modalities of up to 10,000 samples/s, the quadratic factor can significantly slow the computation. Note that both architectures rely on max pooling to reduce the time dimension. The hierarchical attention architecture reduces the time dimension by hierarchically using attention. Indeed, as the vanilla VGG architecture does not include any attention across any dimension, it could be viewed as the simplest architecture. However, it is important to consider the enhanced performance that comes with the limited complexity of attention across the electrodes.

In the proposed deep learning architectures, the channel embedding CNN is responsible for extracting the temporal features of the discharge within the isolated EEG channel. The spatial features, on the other hand, are embedded with the original encoding through the adopted positional encoding scheme. For instance, in the case of an eyeblink spike-like artifact affecting the two frontal electrodes (Fp1 and Fp2), the GNN can identify these two electrodes with the help of positional encoding. Theoretically, this added spatial information helps the GNN recognize eyeblinks and differentiate them from similar actual spikes that could occur in other electrodes' readings. Hence, positional encoding would differentiate the embedding of the graphs in both scenarios. In brief, the CNN handles the temporal features while the GNN with the positional encoding handles the spatial interaction between the EEG channels.

Some limitations and suggestions for future work that could extend the application scope of the proposed method are provided below. Moreover, strengths and weaknesses of the proposed method are summarized at the end of these suggestions.

- Instead of having a CNN shared by all electrodes to generate an embedding for each electrode channel in both FC-GNN and CA-GNN, a separate CNN could be assigned for each electrode as in Satelight. Doing so could reduce the overall number of kernels required to generate a good representative embedding for each channel and thus reduce the overall number of trainable network parameters.

- (b) When deploying the proposed architectures with HD-EEG systems, the global summation pooling strategy used to generate an embedding of a graph (i.e., summing all rows of the matrix F) might not be the best option. The hierarchical pooling technique becomes more suitable for designs that involve a large number of electrodes. For example, the study in [75] utilizes a so-called DIFFPOOL layer that maps a given graph with given nodes' embeddings and edges to another graph of super nodes (or clusters) along with their own embeddings and edges. An assignment matrix mapping each node accomplishes this task. The algorithm is trained end-to-end without recourse to on-the-shelf community detection algorithms such as Louvain or BigCLAM.
- (c) It is important to note that both FC-GNN and CA-GNN have CNN channel encoders. However, FC-GNN bypasses the electrodes attention step done in CA-GNN by running FC analysis to define the edges and perform standard message passing. An edge defined by the FC metric represents an explainable entity. On the other hand, the attention done in CA-GNN grants each node the complete freedom to choose the other nodes to attend to. An intermediate approach would combine the best of the two worlds by first defining the edges using any available FC metric (not necessarily wPLI as in our study), and then letting the neural network decide on those edges deemed most important. Different FC metrics could also be used as a hyperparameter as described in [45].
- (d) Another way to increase the explainability of such graph-based models for detecting EEG epileptic spikes would be to add a fake node connected to all other nodes. After a few GNN layers, attention could be applied only at the fake node through multiple attention heads. The classification output of the segment would depend only on the resulting embedding of that fake node. We expect that for a true positive (TP) example (EEG segment with a spike successfully detected by the algorithm) in the test set, the fake node will learn to give more attention to the nodes corresponding to the electrodes affected by the spike. However, this expected behavior would decay as the number of GNN layers increase due to the over-smoothing issue of GNNs (nodes' embeddings become more like each other as we stack more GNN layers). Indeed, this behavior needs to be validated. Another possible method to tackle this issue is to formulate the problem as a node classification problem instead of a graph classification problem. This approach will force the GNN to recognize the EEG channels affected by the spike. However, getting labels for such data becomes more challenging than simple EEG segment labeling. Labeling groups of electrodes could be a feasible workaround. For example, for a patient with focal epilepsy in the left frontal (LF) lobe, only the LF group of electrodes (including Fp1, F7, and F3) are given the IED label within a spike segment.
- (e) To expand on the previous point (d), through the adopted attention mechanism, several studies attempt to investigate which parts of the input the neural network attended to when generating its output. For example, in [35,63], the model automatically attended to EEG waveforms of specific interest to predict the existence of spikes. Similarly, in [34], the proposed model highlighted the time points contributing to the classification based on a human decision. The channel encoder part of the architecture can thus take advantage of such an approach. Hence, we would know the time of interest at each EEG channel from the channel encoder, and then we would determine the channel(s) the fake node attended to as discussed in the previous point.
- (f) The main limitation of most ML-based solutions in the medical field remains the scarcity of data available for training reliable ML systems, with only a limited amount of labeled open-source EEG records accessible to the research community. We also contend with other issues related to de-identified data to ensure patients' anonymity and compliance with HIPAA requirements. It is important to stress the enormous benefits and breakthroughs that would come with large standardized publicly available datasets. The training phase of machine learning would be significantly enhanced, with plenty more data available to challenge the testing and validation phases. This process can be monitored initially by expert(s) in the early stages to validate the detection results of the model. In [32], for example, the labeling was done semi-automatically using the principal components to label the discharges. Discharges at the boundaries in between clusters were re-assessed by experts. For our study, the classification is done on relatively long 3-second segments with multiple candidate discharges to provide useful temporal context information to the classifier. This would make it harder to rely on the first two principal components for clustering or to adopt SMOTE before training the models.
- (g) Among the methods used to tackle the data scarcity problem without overfitting, as mentioned in the last point (e), is to adopt Bayesian neural networks (BNNs), where parameters of the deep learning model are assumed to be random variables. Instead of learning the model parameters directly through backpropagation, these parameters -in whole or in part- can be assumed to follow a Gaussian distribution with learnable mean and standard deviation. BNNs have several advantages, including robustness to overfitting and estimating uncertainty in predictions. However, this comes at the cost of scalability. This drawback is not as significant when dealing with relatively small models like the ones at hand. Hence, we believe that adopting BNNs is worth investigating. Another approach would be to train models as generative adversarial networks (GAN) to synthesize more data for training. For example, an auxiliary classifier GAN (AC-GAN) was used in [76] to feed the developed LSTM network architecture, called IEDnet, with more samples to improve the detection performance. We believe that this approach would be a better option compared to tackling the problem as an anomaly detection problem. This recommendation is made since supervised learning architectures yield better performance in general.
- Strengths:** (i) Performance: In terms of average precision, the introduced models outperformed other existing architectures on the collected Baptist hospital dataset. On the TUEV dataset, the CA-GNN architecture produced the best results. (ii) For the CA-GNN architecture, no domain knowledge is required since there are no features that are explicitly extracted. This is a strength and a weakness simultaneously since the model explainability becomes an issue. We outlined some methods to tackle this issue in parts (c), (d), and (e) of the suggested future work listed above.
- Weaknesses:** (i) Explainability: It is generally harder to interpret the results of deep learning architectures compared to standard ML algorithms such as decision trees that are trained on hand-crafted features. (ii) Data limitation: We elaborated on this issue extensively in parts (f) and (g) in the above list.

5. Conclusion

In this study, we proposed a multi-stage algorithm for IED detection. First, a rule-based morphological algorithm was used to collect candidate IEDs. To guarantee high enough sensitivity of this first stage, all the used thresholds were tweaked until the algorithm was able to successfully capture all IEDs that were at hand previously. For the second stage, two graph-based architectures are proposed and examined along with four other architectures for comparative purposes.

As presented in the results section, the proposed CA-GNN outperformed the other approaches on both the Baptist Hospital dataset and the Temple University Hospital dataset. In particular, the validation

AP results for the Baptist Hospital dataset is as follows: Vanilla Self-Attention $\rightarrow 0.9029 \pm 0.0431$, Hierarchical Attention $\rightarrow 0.8546 \pm 0.0587$, Vanilla VGG $\rightarrow 0.92 \pm 0.0618$, Satelight $\rightarrow 0.9219 \pm 0.046$, FC-GNN $\rightarrow 0.9731 \pm 0.0187$, CA-GNN $\rightarrow 0.9788 \pm 0.0125$. For the Temple University Hospital dataset, the testing AP results in the same order are 0.9692, 0.9113, 0.97, 0.9575, 0.963, and 0.9879.

In summary, this study sheds light on the role played by GNNs applied on FC maps of different frequency sub-bands combined with attention for the purpose of IED detection. Furthermore, another model is proposed that relies entirely on attention to deduce the underlying graph structure which could enhance the overall average precision. As previously mentioned in the introduction section, the role of GNNs in detecting epileptogenic spikes has been investigated in only one previous study [38] where the edges were defined through the EEG montage. Hence, the approach followed in this manuscript for the same purpose opens the door widely for further investigation to combine FC analysis, attention mechanism, and GNNs to develop novel robust architectures. The guidelines outlined in points (a) through (e) in the suggested future work could be a starting point.

Declaration of competing interest

None of the authors have potential conflicts of interest to be disclosed.

Acknowledgments

Funding

This work was supported by the National Science Foundation, USA [grant numbers CNS-1920182 and CNS-2018611].

Role of funding source

The study sponsor had in no way involvement in this manuscript.

References

- [1] Li G, Jung JJ. Seizure detection from multi-channel EEG using entropy-based dynamic graph embedding. *Artif Intell Med* 2021;122:102201. <http://dx.doi.org/10.1016/j.artmed.2021.102201>.
- [2] He H, Liu X, Hao Y. A progressive deep wavelet cascade classification model for epilepsy detection. *Artif Intell Med* 2021;118:102117. <http://dx.doi.org/10.1016/j.artmed.2021.102117>.
- [3] Buteneers P, Verstraeten D, van Mierlo P, Wyckhuys T, Stroobandt D, Raedt R, et al. Automatic detection of epileptic seizures on the intra-cranial electroencephalogram of rats using reservoir computing. *Artif Intell Med* 2011;53(3):215–23. <http://dx.doi.org/10.1016/j.artmed.2011.08.006>.
- [4] Staljanens W, Strobbe G, Holen RV, Birot G, Gschwind M, Seeck M, et al. Seizure onset zone localization from ictal high-density EEG in refractory focal epilepsy. *Brain Topogr* 2017;30(2):257–71.
- [5] Plummer C, Vogrin SJ, Woods WP, Murphy MA, Cook MJ, Liley DTJ. Interictal and ictal source localization for epilepsy surgery using high-density EEG with MEG: A prospective long-term study. *Brain* 2019;142(4):932–51. <http://dx.doi.org/10.1093/brain/awz015>, [arXiv:https://academic.oup.com/brain/article-pdf/142/4/932/28272863/awz015_supplementary_figures_and_tables.pdf](https://academic.oup.com/brain/article-pdf/142/4/932/28272863/awz015_supplementary_figures_and_tables.pdf).
- [6] Yin X-X, Hadjilucas S, Zhang Y, Tian Z. MRI radiogenomics for intelligent diagnosis of breast tumors and accurate prediction of neoadjuvant chemotherapy responses-a review. *Comput Methods Programs Biomed* 2022;214:106510. <http://dx.doi.org/10.1016/j.cmpb.2021.106510>, URL <https://www.sciencedirect.com/science/article/pii/S0169260721005848>.
- [7] Baumgartner C, Koren JP. Seizure detection using scalp-EEG. *Epilepsia* 2018;59(S1):14–22. <http://dx.doi.org/10.1111/epi.14052>, [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.14052](https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.14052).
- [8] Nemtas P, Birot G, Pittau F, Michel CM, Schaller K, Vulliemoz S, et al. Source localization of ictal epileptic activity based on high-density scalp EEG data. *Epilepsia* 2017;58(6):1027–36.
- [9] Conrad EC, Tomlinson SB, Wong JN, Oechsl KF, Shinohara RT, Litt B, et al. Spatial distribution of interictal spikes fluctuates over time and localizes seizure onset. *Brain* 2020;143(2):554–69.
- [10] Bartolomei F, Trébuccon A, Bonini F, Lambert I, Gavaret M, Woodman M, et al. What is the concordance between the seizure onset zone and the irritative zone? A SEEG quantified study. *Clin Neurophysiol* 2016;127(2):1157–62.
- [11] Mégevand P, Spinelli L, Genetti M, Brodbeck V, Momjian S, Schaller K, et al. Electric source imaging of interictal activity accurately localises the seizure onset zone. *J Neurol Neurosurg Psychiatry* 2014;85(1):38–43.
- [12] Tamila E, AlHilani M, Tanaka N, Tsuboyama M, Peters JM, Grant PE, et al. Assessing the localization accuracy and clinical utility of electric and magnetic source imaging in children with epilepsy. *Clin Neurophysiol* 2019;130(4):491–504.
- [13] Abd El-Samie FE, Alotaiby TN, Khalid MI, Alshebeili SA, Aldosari SA. A review of EEG and MEG epileptic spike detection algorithms. *IEEE Access* 2018;6:60673–88.
- [14] da Silva Lourenço C, Tjepkema-Cloostermans MC, van Putten MJ. Machine learning for detection of interictal epileptiform discharges. *Clin Neurophysiol* 2021;132(7):1433–43.
- [15] Wilson SB, Emerson R. Spike detection: A review and comparison of algorithms. *Clin Neurophysiol* 2002;113(12):1873–81.
- [16] Diykh M, Abdulla S, Saleh K, Deo RC. Fractal dimension undirected correlation graph-based support vector machine model for identification of focal and non-focal electroencephalography signals. *Biomed Signal Process Control* 2019;54:101611.
- [17] Gabor AJ, Seyal M. Automated interictal EEG spike detection using artificial neural networks. *Electroencephalogr Clin Neurophysiol* 1992;83(5):271–80.
- [18] Özdağ Ö, Zhu G, Yaylali I, Jayakar P. Real-time detection of EEG spikes using neural networks. In: 1992 14th Annual international conference of the IEEE engineering in medicine and biology society. Vol. 3. IEEE; 1992, p. 1022–3.
- [19] Glover J, Raghaven N, Ktonas P, Frost J. Context-based automated detection of epileptogenic sharp transients in the EEG: Elimination of false positives. *IEEE Trans Biomed Eng* 1989;36(5):519–27. <http://dx.doi.org/10.1109/10.24253>.
- [20] Adjouadi M, Sanchez D, Cabrerizo M, Ayala M, Jayakar P, Yaylali I, et al. Interictal spike detection using the Walsh transform. *IEEE Trans Biomed Eng* 2004;51(5):868–72.
- [21] Clarençon D, Renaudin M, Gourmelon P, Kerckhove A, Catérini R, Boivin E, et al. Real-time spike detection in EEG signals using the wavelet transform and a dedicated digital signal processor card. *J Neurosci Methods* 1996;70(1):5–14.
- [22] sen Pon L, Sun M, Scabassi R. The bi-directional spike detection in EEG using mathematical morphology and wavelet transform. In: 6th International conference on signal processing, 2002. Vol. 2. 2002, p. 1512–5. <http://dx.doi.org/10.1109/ICOSP.2002.1180082>.
- [23] Akay M. Biomedical signal processing. Academic Press; 2012.
- [24] Acir N, Güzeliz C. Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Comput Biol Med* 2004;34(7):561–75.
- [25] Chatrian G. A glossary of terms most commonly used by clinical electroencephalographers. *Electroenceph Clin Neurophysiol* 1974;37:538–48.
- [26] Nonclercq A, Foulon M, Verheulpen D, De Cock C, Buzatu M, Mathys P, et al. Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology. *J Neurosci Methods* 2012;210(2):259–65.
- [27] Wang Z, Wu D, Dong F, Cao J, Jiang T, Liu J. A novel spike detection algorithm based on multi-channel of BECT EEG signals. *IEEE Trans Circuits Syst II* 2020;67(12):3592–6.
- [28] Heers M, Böttcher S, Kalina A, Katletz S, Altenmüller D-M, Baroumand AG, et al. Detection of interictal epileptiform discharges in an extended scalp EEG array and high-density EEG—A prospective multicenter study. *Epilepsia* 2022;63(7):1619–29. <http://dx.doi.org/10.1111/epi.17246>, [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.17246](https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.17246).
- [29] Xu Z, Wang T, Cao J, Bao Z, Jiang T, Gao F. BECT spike detection based on novel EEG sequence features and LSTM algorithms. *IEEE Trans Neural Syst Rehabil Eng* 2021;29:1734–43. <http://dx.doi.org/10.1109/TNSRE.2021.3107142>.
- [30] Geng X, Hu G, Tian X. Neural spike sorting using mathematical morphology, multiwavelets transform and hierarchical clustering. *Neurocomputing* 2010;73(4):707–15. <http://dx.doi.org/10.1016/j.neucom.2008.11.034>, URL <https://www.sciencedirect.com/science/article/pii/S0925231209004330>.
- [31] Lourenço C, Tjepkema-Cloostermans MC, Teixeira LF, van Putten MJ. Deep learning for interictal epileptiform discharge detection from scalp EEG recordings. In: Mediterranean conference on medical and biological engineering and computing. Springer; 2019, p. 1984–97.
- [32] Saifur Rehman M, Lienkämper R, Parpaley Y, Wellmer J, Liu C, Lee B, et al. SpikeDeeptector: A deep-learning based method for detection of neural spiking activity. *J Neural Eng* 2019;16(5):056003.
- [33] Medvedev AV, Agoureeva GI, Murro AM. A long short-term memory neural network for the detection of epileptiform spikes and high frequency oscillations. *Sci Rep* 2019;9(1):1–10.
- [34] Hasib MM, Nayak T, Huang Y. A hierarchical LSTM model with attention for modeling EEG non-stationarity for human decision prediction. In: 2018 IEEE EMBS international conference on biomedical & health informatics. 2018, p. 104–7. <http://dx.doi.org/10.1109/BHI.2018.8333380>.
- [35] Fukumori K, Yoshida N, Sugano H, Nakajima M, Tanaka T. Satelight: Self-attention-based model for epileptic spike detection from multi-electrode EEG. *J Neural Eng* 2022;19(5):055007.

- [36] Harati A, López S, Obeid I, Picone J, Jacobson MP, Tobochnik S. The TUH EEG CORPUS: A big data resource for automated EEG interpretation. In: 2014 IEEE signal processing in medicine and biology symposium. 2014, p. 1–5. <http://dx.doi.org/10.1109/SPMB.2014.7002953>.
- [37] Mohammed AH, Morar U, Cabrerizo M, Rajaei H, Pinzon A, Yaylali I, et al. Dynamics of electrical activity in epileptic brain and induced changes due to interictal epileptiform discharges. *IEEE Access* 2022;10:1276–88. <http://dx.doi.org/10.1109/ACCESS.2021.3138385>.
- [38] Nhu D, Janmohamed M, Perucca P, Gilligan A, Kwan P, O'Brien T, et al. Graph convolutional network for generalized epileptiform abnormality detection on EEG. In: 2021 IEEE signal processing in medicine and biology symposium. IEEE; 2021, p. 1–6.
- [39] Nhu DB, Janmohamed M, Antonic-Baker A, Perucca P, O'Brien TJ, Gilligan A, et al. Deep learning for automated epileptiform discharge detection from scalp EEG: A systematic review. *J Neural Eng* 2022.
- [40] Janmohamed M, Nhu D, Kuhlmann L, Gilligan A, Tan CW, Perucca P, et al. Moving the field forward: detection of epileptiform abnormalities on scalp electroencephalography using deep learning—clinical application perspectives. *Brain Commun* 2022;4(5). <http://dx.doi.org/10.1093/braincomms/fcac218>, [arXiv:https://academic.oup.com/braincomms/article-pdf/4/5/fcac218/45879970/fcac218.pdf](https://academic.oup.com/braincomms/article-pdf/4/5/fcac218/45879970/fcac218.pdf). fcac218.
- [41] Foumani NM, Miller L, Tan CW, Webb GI, Forestier G, Salehi M. Deep learning for time series classification and extrinsic regression: A current survey. 2023, [arXiv:2302.02515](https://arxiv.org/abs/2302.02515).
- [42] Raeisi K, Khazaei M, Croce P, Tamburro G, Comani S, Zappasodi F. A graph convolutional neural network for the automated detection of seizures in the neonatal EEG. *Comput Methods Programs Biomed* 2022;222:106950. <http://dx.doi.org/10.1016/j.cmpb.2022.106950>.
- [43] Wang M, El-Fiqi H, Hu J, Abbass HA. Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states. *IEEE Trans Inf Forensics Secur* 2019;14(12):3259–72. <http://dx.doi.org/10.1109/TIFS.2019.2916403>.
- [44] Demir A, Koike-Akino T, Wang Y, Haruna M, Erdogmus D. EEG-GNN: Graph neural networks for classification of electroencephalogram (EEG) signals. In: 2021 43rd Annual international conference of the IEEE engineering in medicine & biology society. 2021, p. 1061–7. <http://dx.doi.org/10.1109/EMBC46164.2021.9630194>.
- [45] Klepl D, He F, Wu M, Blackburn DJ, Sarriani P. EEG-based graph neural network classification of Alzheimer's disease: An empirical evaluation of functional connectivity methods. *IEEE Trans Neural Syst Rehabil Eng* 2022;30:2651–60. <http://dx.doi.org/10.1109/TNSRE.2022.3204913>.
- [46] Vinck N, Oostenveld R, Van Wingerden M, Battaglia F, Pennartz CM. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage* 2011;55(4):1548–65.
- [47] Palva JM, Wang SH, Palva S, Zhigalov A, Monto S, Brookes MJ, et al. Ghost interactions in MEG/EEG source space: A note of caution on inter-areal coupling measures. *Neuroimage* 2018;173:632–43.
- [48] Fraga González G, Smit DJ, Van der Molen MJ, Tijms J, Stam CJ, De Geus EJ, et al. EEG resting state functional connectivity in adult dyslexics using phase lag index and graph analysis. *Front Hum Neurosci* 2018;12:341.
- [49] Whitham EM, Pope KJ, Fitzgibbon SP, Lewis T, Clark CR, Loveless S, et al. Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20Hz are contaminated by EMG. *Clin Neurophysiol* 2007;118(8):1877–88. <http://dx.doi.org/10.1016/j.clinph.2007.04.027>, URL <https://www.sciencedirect.com/science/article/pii/S1388245707001988>.
- [50] Kuntzelman K, Miskovic V. Reliability of graph metrics derived from resting-state human EEG. *Psychophysiology* 2017;54(1):51–61.
- [51] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in neural information processing systems*. vol. 30. 2017.
- [52] Noachter S. A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. *Electroenceph clin Neurophysiol* 1999;52:21–41.
- [53] Ossadtschi A, Baillet S, Mosher J, Thyerlei D, Sutherling W, Leahy R. Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering. *Clin Neurophysiol* 2004;115(3):508–22. <http://dx.doi.org/10.1016/j.clinph.2003.10.036>.
- [54] Jiang F, Grigorev A, Rho S, Tian Z, Fu Y, Jifara W, et al. Medical image semantic segmentation based on deep learning. *Neural Comput Appl* 2018;29:1257–65.
- [55] Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci* 2011;2011.
- [56] Mosher J, Leahy R. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Trans Signal Process* 1999;47(2):332–40. <http://dx.doi.org/10.1109/78.740118>.
- [57] Mosher J, Leahy R. Recursive MUSIC: A framework for EEG and MEG source localization. *IEEE Trans Biomed Eng* 1998;45(11):1342–54. <http://dx.doi.org/10.1109/10.725331>.
- [58] Singh K, Malhotra J. Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features. *Complex Intell Syst* 2022;1–14.
- [59] Liu Y, Huang Y-X, Zhang X, Qi W, Guo J, Hu Y, et al. Deep C-LSTM neural network for epileptic seizure and tumor detection using high-dimension EEG signals. *IEEE Access* 2020;8:37495–504. <http://dx.doi.org/10.1109/ACCESS.2020.2976156>.
- [60] Xu G, Ren T, Chen Y, Che W. A one-dimensional cnn-lstm model for epileptic seizure recognition using eeg signal analysis. *Front Neurosci* 2020;14:578126.
- [61] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80.
- [62] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. 2014, [arXiv preprint arXiv:1412.3555](https://arxiv.org/abs/1412.3555).
- [63] Fukumori K, Yoshida N, Sugano H, Nakajima M, Tanaka T. Epileptic spike detection by recurrent neural networks with self-attention mechanism. In: *ICASSP 2022 - 2022 IEEE international conference on acoustics, speech and signal processing*. 2022, p. 1406–10. <http://dx.doi.org/10.1109/ICASSP43922.2022.9747560>.
- [64] Ba JL, Kiros JR, Hinton GE. Layer normalization. 2016, [arXiv:1607.06450](https://arxiv.org/abs/1607.06450). URL <https://arxiv.org/abs/1607.06450>.
- [65] Wu C, Wu F, Qi T, Huang Y, Xie X. Fastformer: Additive attention can be all you need. 2021, [arXiv:2108.09084](https://arxiv.org/abs/2108.09084). URL <https://arxiv.org/abs/2108.09084>.
- [66] Xu D, Tian Z, Lai R, Kong X, Tan Z, Shi W. Deep learning based emotion analysis of microblog texts. *Inf Fusion* 2020;64:1–11. <http://dx.doi.org/10.1016/j.inffus.2020.06.002>, URL <https://www.sciencedirect.com/science/article/pii/S156625352030302X>.
- [67] Kay SM. Fundamentals of statistical signal processing. Prentice Hall PTR; 1993.
- [68] Janwattanapong P, Cabrerizo M, Pinzon A, Gonzalez-Arias S, Barreto A, Andrian J, et al. Connectivity patterns of interictal epileptiform discharges using coherence analysis. In: 2016 IEEE signal processing in medicine and biology symposium. IEEE; 2016, p. 1–6.
- [69] Lakshmanan M, Senthilkumar DV. Transition from phase to generalized synchronization. In: *Dynamics of nonlinear time-delay systems*. Springer; 2011, p. 201–26.
- [70] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? 2018, [arXiv:1810.00826](https://arxiv.org/abs/1810.00826). URL <https://arxiv.org/abs/1810.00826>.
- [71] Fu X, Wang Y, Ge M, Wang D, Gao R, Wang L, et al. Negative effects of interictal spikes on theta rhythm in human temporal lobe epilepsy. *Epilepsy Behav* 2018;87:207–12. <http://dx.doi.org/10.1016/j.yebeh.2018.07.014>.
- [72] Bettus G, Wendling F, Guye M, Valtin L, Régis J, Chauvel P, et al. Enhanced EEG functional connectivity in mesial temporal lobe epilepsy. *Epilepsy Res* 2008;81(1):58–68. <http://dx.doi.org/10.1016/j.eplepsyres.2008.04.020>.
- [73] Mohammed AH, Cabrerizo M, Morar U, Rajaei H, Pinzon A, Yaylali I, et al. Penalized functional connectivity maps for patients with focal epilepsy. *IEEE Access* 2021;9:204–17. <http://dx.doi.org/10.1109/ACCESS.2020.3046851>.
- [74] Nhu D, Janmohamed M, Shakhathreh L, Gonen O, Perucca P, Gilligan A, et al. Automated interictal epileptiform discharge detection from scalp EEG using scalable time-series classification approaches. Cold Spring Harbor Laboratory Press; 2022, <http://dx.doi.org/10.1101/2022.07.06.22277287>, [arXiv:https://www.medrxiv.org/content/early/2022/07/10/2022.07.06.22277287.full.pdf](https://www.medrxiv.org/content/early/2022/07/10/2022.07.06.22277287.full.pdf). URL <https://www.medrxiv.org/content/early/2022/07/10/2022.07.06.22277287>.
- [75] Ying Z, You J, Morris C, Ren X, Hamilton W, Leskovec J. Hierarchical graph representation learning with differentiable pooling. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Advances in neural information processing systems*. Vol. 31. Curran Associates, Inc; 2018, URL https://proceedings.neurips.cc/paper_files/paper/2018/file/e77dbaf6759253c7c6d0efc5690369c7-Paper.pdf.
- [76] Geng D, Alkhachroum A, Bicch MAM, Jagid JR, Cajas I, Chen ZS. Deep learning for robust detection of interictal epileptiform discharges. *J Neural Eng* 2021;18(5):056015. <http://dx.doi.org/10.1088/1741-2552/abf28e>.