Graph Convolutional Network For Generalized Epileptiform Abnormality Detection On EEG

D. Nhu[1], M. Janmohamed[2], P. Perucca[2], A. Gilligan[3], P. Kwan[2], T. O'Brien[2], C. W. Tan[1] and L. Kuhlmann[1]

1. Department of Data Science and AI, Faculty of IT, Monash University, Victoria, Australia
2. Department of Neuroscience, Central Clinical School, Monash University, Victoria, Australia
3. Neurosciences Clinical Institute, Epworth Healthcare Hospital, Victoria, Australia
{duong.nhu1, mubeen.janmohamed, piero.perucca, patrick.kwan}@monash.edu
{terence.obrien, chang.tan, levin.kuhlmann}@monash.edu, agilligan@optusnet.com.au

Abstract— Epilepsy diagnostic investigation involving manual visual analysis of electroencephalogram (EEG) is a time-consuming process. Deep neural networks, especially the convolutional network (CNN), have been applied to interictal epileptiform discharge (IED) detection and have achieved promising results. However, these networks do not incorporate clinical features of EEG montages. In recent years, graph convolution has succeeded in learning features from structural graph-like data. In this paper, we explore the novel application of different architectures of graph convolutions with Chebyshev polynomial filters which learn spatio-temporal features from EEG montages. We conducted a number of experiments with transverse and longitudinal montages on a set of routine EEG recordings from patients with idiopathic generalized epilepsy. We split these EEG recordings into 2s windows with or without IED and evaluated different architectures in terms of how well they classified these windows. We achieved the best AUC of 0.92. Furthermore, we explored different thresholds of the output probability and observed that at 0.8, based on the selection of collected data, we achieved a mean false-positive rate per minute of 0.44 and still preserved a reasonable mean sensitivity of 0.64 across all EEG recordings. The results indicate that our approaches could produce clinically useful performance levels. Our work could be extended to improve the interpretability of the automated software in a clinical environment.

I. Introduction

Epilepsy is the most common serious, chronic neurological disorder across the lifespan, affecting approximately 50 million people worldwide [1]. It is usually diagnosed by a clinical neurologist using Electroencephalography (EEG) to record the voltage fluctuations resulting from neuronal post-synaptic potentials within the brain via surface scalp electrodes. An EEG recording of an epileptic patient may show interictal epileptiform discharges (IEDs) which can be in the form of spikes or sharp waves, often associated with slow frequency waveforms which disrupt the normal background. In genetic/idiopathic generalized epilepsy (GGE), accounting for 15-20% of all epilepsies [2], IEDs often manifest as broad-field, frontally predominant symmetric, synchronous, 2.5-6Hz spike/polyspike-wave discharges, and are called generalized epileptiform discharges. These IEDs can range from 0.5s to 16s long.

Detecting IEDs is a challenging task that requires manual analysis of EEG recordings. Researchers have been applying various machine learning methods on automating IED detection [3–5]. These methods pass hand-picked features from the EEG to a classifier such as SVM, KNN, or decision tree to discriminate features of IEDs from those of artifacts and background activities. The features can be grouped into 3 domains, time, frequency, and wavelet [5]. As the features selection is manual and time-consuming, these methods could only be performed on small datasets ($\leq$ 50 patients), resulting in poor generalisability. Of all methods, the only software with approval from the United States Food and Drug Administration (FDA) for automated IED detection is Persyst [6], developed by Persyst Corporation. Persyst has been shown to have comparable performance to skilled senior EEG technologists [7].

With the success of deep learning in recent years [8, 9], researchers also demonstrated that the convolutional neural network (CNN) could automatically extract latent features from large datasets ($\geq$ 100 patients) and has had promising performance in IED detection [10–13]. In these studies, EEG recordings are split into small windows of 1s or 2s and considered as multivariate time series with the signal from all electrodes as features at each time step. Typically, a stack of multiple 1D convolutional layers, followed by pooling layers, are trained to classify these windows into either abnormal (containing IED) or normal (without IED). In addition, it is common to have more normal windows than windows with IEDs which might cause the model to be biased towards the majority class. Authors in [10] trained CNN models with triplet loss to address this problem. Focal loss and over-sampling IED windows were also explored in [13]. One drawback of these deep learning methods is that the structural linkages of electrode pairs are not taken into account. An electrode linkage design connects signals from different regions of the brain to derive a multi-channel montage. Neurologists often use different montages to visualize logical patterns to help them interpret the EEG accurately.

To better learn the features of structural graph-like data, graph convolution networks (GCN) have been studied and shown promising results [14, 15]. In GCN, the network consists of multiple convolutional layers on the Laplacian matrix of a graph. The application of GCN on EEG data has been studied in epileptic seizure

prediction [16, 17]. In these studies, the Chebyshev polynomials spectral filtering was used across all convolutional layers. Time and frequency domain features per electrode, such as standard deviation and power spectral density, are manually extracted as inputs to the GCN.

In this paper, we consider an EEG montage (Figure 1) as a graph and explore different graph convolutional architectures to learn features of brain activities from the entire sensor array as well as local information of a region. The longitudinal and transverse bipolar montages were studied as they covered the parasagittal and temporal regions bilaterally. We also use an embedding layer to automatically learn temporal features from an electrode signal. This will make the interpretation of detected IEDs more comprehensive. The experiments were conducted on a large set of routine EEG recordings from patients with idiopathic generalized epilepsy (IGE). As routine EEG is a clinical standard step in epilepsy diagnosis, we implement general architectures which were invariant to the diversity of patients.

II. Methods

A graph is a pair $G = (V,E)$ where V is a set of vertices and E is a set of edges connecting paired vertices. An EEG montage measures the potential difference among a set of electrode pairs. A graph can be derived from this by considered each electrode as a vertex, and the linkage between 2 electrodes as an edge. As such, a GCN can be applied to this derivation of EEG. In the following, we briefly explain graph convolution and describe our proposed architecture in detail.

II-A. Graph Convolution

Suppose we have graph G and degree of the vertex v, d_v, the Laplacian matrix of G [18] is

$$L(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0, \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

This can be written as $L = T^{-\frac{1}{2}} L T^{-\frac{1}{2}} = I - T^{-\frac{1}{2}} A T^{-\frac{1}{2}}$, where T is the diagonal matrix with the (v,v)-th entry of d_v, I is an identity matrix, and A is the adjacent matrix.

The Laplacian can be written in the Fourier domain as $L = U\Lambda U^T$, with the Fourier basis $U = [u_0, ..., u_{n-1}] \in \mathbb{R}^{\text{n x n}}$ and $\Lambda = diag([\lambda_0, ..., \lambda_1] \in \mathbb{R}^{\text{n x n}}$ [14]. As such, the Fourier transform of a graph signal x is $\hat{x} = U^T x$. The spectral filtering operation, g_θ, is then defined as

$$g_\theta(L) * x = g_\theta(U\Lambda U^T)x = U g_\theta(\Lambda) U^T x \quad (2)$$

The computation of $g_\theta(\Lambda) = diag(\theta)$ is expensive, where θ is a vector of Fourier coefficients [14]. Instead, We can approximate $g_\theta(\Lambda)$ with Chebyshev polynomials as follows

$$g_\theta(\Lambda) \approx \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\Lambda}) \quad (3)$$

where $T_k(\widetilde{\Lambda}) = \cos(k\widetilde{\Lambda})$ is the Chebyshev polynomials of order k, θ are polynomial coefficients, and $\widetilde{\Lambda} = 2\Lambda/\lambda_{max} - I_n$ is a diagonal matrix of scaled eigenvalues in [-1, 1]. The Chebyshev recurrence relation can be used to compute $T_k(\widetilde{\Lambda})$, $T_k(\widetilde{\Lambda}) = 2\widetilde{\Lambda} T_{k-1}(\widetilde{\Lambda}) - 2\widetilde{\Lambda} T_{k-2}(\widetilde{\Lambda})$, $T_0(\widetilde{\Lambda}) = 1$, $T_1(\widetilde{\Lambda}) = \widetilde{\Lambda}$.

In general, we can design a Chebyshev convolutional layer for a graph signal x as follows

$$y = g_\theta * x = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(L)x \quad (4)$$

where θ_k is the vector of trainable coefficients at the Chebyshev polynomials of order k.

II-B. Graph Convolution Network On EEG

An EEG montage is constructed from an arrangement of multiple EEG channels with each montage channel derived from the linkage of electrode pairs to measure the potential difference between them. Different montages show different left-right hemispheric or anterior-posterior topographic patterns of the EEG signals and help clinicians interpret them more accurately. A commonly used montage is bipolar. There are 2 types of bipolar montage, longitudinal and transverse, splitting the brain vertically and horizontally, respectively. Visualisations of these in the 10-20 system are demonstrated in Figure 1. The advantages of a bipolar montage are subtle on a referential montage [19], such as determining lateralization of IEDs and allowing phase reversals as a means to localize maximally involved electrodes when amplitude changes across adjacent electrodes. The transverse montage provides useful recognition of anterior-posterior gradient network discharges commonly seen in generalized epileptiform discharges as well as easy recognition of sleep. Both montages also serve as good identifiers of ocular artifacts.

A Laplacian matrix could be constructed from an EEG montage by viewing it as a graph, in which, each electrode is a vertex, and the linkages of electrode pairs are edges. A graph convolution network on a montage can be formed by stacking multiple layers of Eq. 4. This will enable the network to learn local features from an electrode and its neighbors. Global features of the graph could be captured by summing or averaging these features.

II-C. Architecture Design

Our architectures include an embedding block, a stack of multiple Chebyshev convolutional layers, followed by 2 fully connected layers.

II-C1. Embedding

The signal from an electrode can be viewed as a univariate time series. Two 1D convolutional blocks are applied to the individual electrode to extract temporal features. Each convolutional block consists of 1 convolution layer

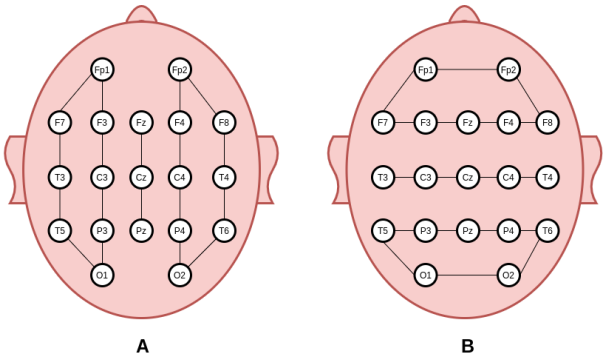

Figure 1. A. Longitudinal bipolar montage. B. Transverse bipolar montage. The connection from an electrode to another means the potential difference between these two is measured.

with Leaky ReLU activation [20], followed by layer normalisation [21] and spatial dropout [22] operations. The convolution layers have the same kernel size of 3 and stride of 1 and use 32 and 64 filters respectively. The dropout rate was set to 0.2. We then apply a global average pooling layer to the outputs of the last convolution block. This is followed by a fully-connected layer with 256 hidden units.

II-C2. Graph Convolution Network

In our work, we studied the GCN with 2 montages – longitudinal and transverse montages. The core of our architectures is a Chebyshev block, illustrated in Figure 2, that can be stacked to form different architectures described in Figure 3. Each Chebyshev block consists of 2 Chebyshev polynomial convolutional layers with the residual connection. We also normalize the output features of each channel with layer normalization [21].

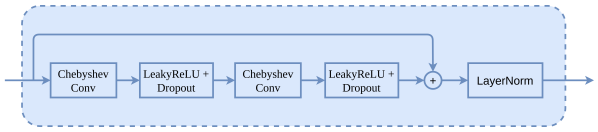

Figure 2. Chebyshev block with residual connection and layer normalisation

We designed 3 different architectures and studied the performance with the 2 mentioned montages (Figure 3). The Architecture A consists of 4 blocks of Chebyshev convolutional layers, followed by a global sum pooling and 2 fully connected layers. This architecture could be trained separately with each of the 2 montages.

As each montage displays different logical patterns of signals, a combination of both montages will help the network extract features from 2 hemispheres of the brain. Following the work in [17] that has shown successes in GCN for seizure detection, we designed 2 architectures combining the 2 mentioned montages. We modified the designs by using our convolution blocks, adding the outputs from 2 branches, and normalizing the summation with layer normalization. Architecture B has 4 blocks of 2 convolutional branches learning latent features from the 2 montages simultaneously. The outputs of these branches are then added, followed by a dropout and a layer normalization layers. Architecture C extends architecture B by alternating the Chebyshev convolution block on each montage in each branch. In the 2 architectures, the convolutional blocks are followed by a global sum pooling layer to reduce the size of feature maps, and 2 fully connected layers. A combination of both montages will help the network extract bilateral features from the longitudinal montage and allow for more accurate field recognition of frontal epileptiform discharges via the transverse montage. Our experiment later shows that this is a better approach, producing the highest F1 score and AUC score in terms of windows classification and whole EEG recording classification, respectively.

The number of filters in all Chebyshev convolutonal layers are 256. The order of polynomials is 2. All fully connected layers have 512 hidden units with Leaky ReLu as the activation function. The dropout rates in all convolution blocks are 0.2 and those of the fully-connected layers are 0.5.

III. Experimental Setup

Training deep learning network on the whole routine EEG recording (e.g: 30 minutes with 256 data points per second) requires huge computational resources. A common solution is to split the EEG recording into smaller windows and classify them into abnormal (containing IED) and normal (without IED). An EEG recording is considered to be epileptic if it has at least 1 window with IED. In this section, we first describe the dataset and pre-processing step in our experiments. Then, we explain the configurations of the training process. All GCN models were implemented with Spektral [23].

III-A. Dataset

We collected 110 routine EEG recordings from patients with idiopathic generalized epilepsy and 116 EEG recordings without IED, seen at the Alfred Health Hospital, Melbourne, Australia. All EEG recordings were recorded with the 10-20 system and annotated by 3 certified neurologists with accredited training in EEG reporting. The start and end of annotation were estimated at onset or just preceding the earliest spike/slow-wave or rhythmic change and after the end of the estimated spike/wave rhythmic activity. There are 1,413 IEDs in total with an average duration of 2s. We randomly split these into train and test sets. The train set consists of 80 EEG recordings with IED and 92 normal recordings. The numbers of recordings in the test set are 30 and 24, respectively. All EEG recordings were recorded with 10-20 system and sampling rates of 250 Hz and 256 Hz. This study was conducted with approval from the Alfred Health Ethics Committee.

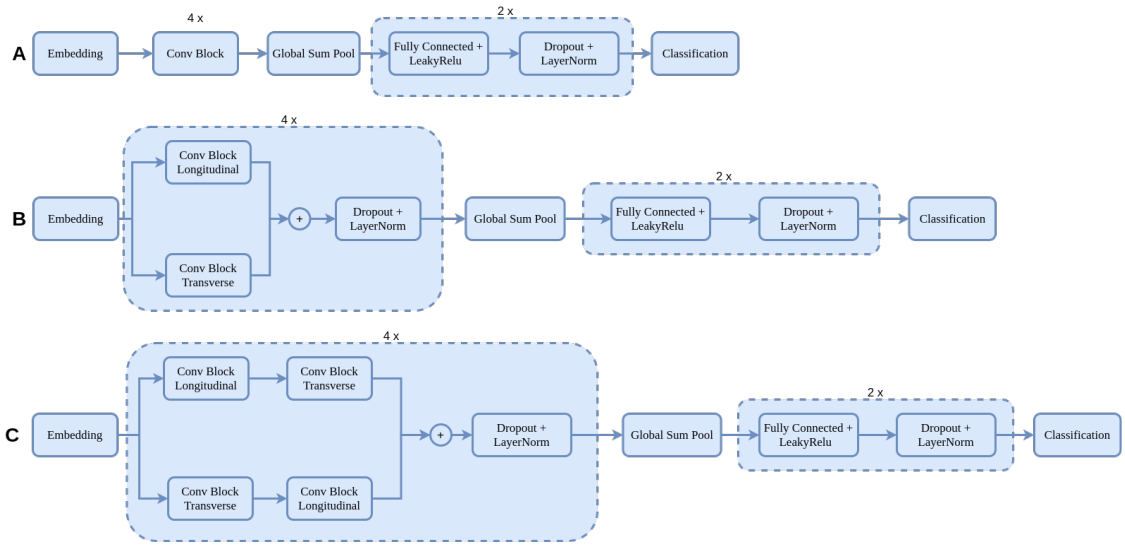

Figure 3. Three different architectures in our experiments

III-B. Pre-processing

These EEG recordings needed to be pre-processed before they could be used by a machine learning model. We standardized the number of electrodes in our dataset by excluding the auricular electrodes M1 and M2, as they were not present in some of the EEG recordings. This left us with 19 electrodes (channels) for each EEG recording. Then, all EEG recordings were resampled to 256 Hz using polyphase filtering. We also applied band-pass filters of 0.5 - 50 Hz to remove muscle artifacts and slow-frequency artifacts which do not contribute any information to the diagnosis of epilepsy [10]. The average duration of the IED annotations in our dataset is 2s. As such, we split the EEG into 2s windows of epileptic discharges and normal with a 50% overlap. This has been found to be effective in [10, 13]. Finally, this gives us input of 19×512 for our models.

In our experiments, we only used normal windows from normal EEG recordings (without IED). This prevents the use of any IEDs that might have been inadvertently missed by the neurologists during the labeling process and might confuse the models. This yields 1,934 and 615 windows with IEDs for training and testing, respectively. As windows without IED significantly outnumber windows with IED, we randomly over-sampled the minor class to balance the dataset. All windows were z-score normalized electrode-wise with global mean and standard deviation of each electrode.

III-C. Training Configuration

Our architectures were trained using Adam optimizer with weight decay [8] and a mini-batch size of 128. The learning rate was set as 10^{-3} and the weight decay was chosen as 10^{-4}. We reduced the learning rate by a factor of 0.5 if there was no improvement after 5 epochs.

IV. Results

Our experiments are divided into 2 parts, windows classification, and whole EEG classification. In terms of windows classification, a randomly given window is classified as abnormal if it contains any IED. We trained architecture A separately with transverse and longitudinal montages. Results of the average of the outputs from these models were also examined. We compared our models with the ResNet models, which we had previously shown to work well for IED windows classification [13]. The ResNet models consist of 3 residual blocks with 3 convolutional layers in each block, followed by a global average pooling layer. We refer interested readers to the papers [13, 24] for more detailed descriptions. This also explored different approaches to address the imbalanced dataset problem, focal loss, triplet loss, and over-sampling. Regarding the whole EEG recording, if there are any abnormal windows, the recording will be considered epileptic.

IV-A. Windows classification

Recall that the goal is to classify a randomly given EEG window into normal or abnormal. We evaluated our models with a fixed clean test set of windows, described in the pre-processing step. This would help us calculate all metrics precisely without any contamination by bad samples. Having too many false positives makes the clinician's work tedious and more time-consuming given the need to navigate and review all suggestions. As our data is imbalanced, the AUC score (commonly used in the literature) becomes a poor measure. Table 1 shows that all models have similar AUC. Hence, we evaluate the models using the F1 score which provides better clinical value as it combines sensitivity and precision. The failure of clinical utility is thus reliably reflected by the F1 score. Overall, our models had lower

sensitivity scores than the ResNet models, however, they achieved higher F1 scores. Architecture C had the highest F1 score of 0.42, followed by the average of the outputs of architecture A with transverse and longitudinal montages, 0.37. All metrics were calculated at the probability threshold of 0.5. This indicates that our models might have better clinical utility.

Table 1. Results of GCN and ResNet in window classification. All metrics were calculated at the probability threshold of 0.5.

Model	Sens	Spec	Prec	F1	AUC
A - Trans (1)	0.51	0.98	0.24	0.32	0.91
A - Long (2)	0.64	0.95	0.16	0.26	0.91
Architecture B	0.60	0.98	0.32	**0.42**	0.92
Architecture C	0.62	0.97	0.14	0.25	0.91
Average of 1 & 2	0.39	0.99	0.36	0.37	0.92
ResNet - Oversampling	0.73	0.83	0.14	0.23	0.92
ResNet - Focal	0.84	0.82	0.06	0.12	0.92
ResNet - Focal & Oversampling	0.64	0.97	0.23	0.34	0.91
ResNet - Triplet	0.75	0.93	0.14	0.24	0.94

IV-B. Whole EEG recording classification

We showed that our models had higher F1 scores than ResNet models and might have better clinical utility. To verify this, we further measured how well our models classify the whole EEG recordings as normal or abnormal. We considered an EEG recording as abnormal if it had at least 1 IED. As we have a similar number of normal and abnormal EEG recordings, AUC is acceptable here. By computing the true-positive rate and false-positive rate of whole EEG recording classification at different thresholds of output probability from window classification, we calculated the AUC scores using the composite trapezoidal rule and summarized the findings in Table 2. Architecture B had the best AUC score of 0.84. The AUC scores suggest that an optimal threshold could be used to produce clinically useful performance levels. We explored different thresholds and observed that at the threshold of 0.80, these models had reasonable mean false positives (FP) per minute overall EEG recordings and still preserved a good sensitivity. The results of mean FP per minute and mean sensitivity overall EEG recordings are shown in Table 3. We included both normal and abnormal windows from epileptic EEG recordings in this task. We observed that the average of architecture A with the 2 montages had the most reasonable results, whose mean FP per minute was 0.44, less than 1 while still maintaining a reasonable mean sensitivity of 0.64. Architecture B had the highest sensitivity of 0.73 but suffered from the highest mean FP per minute of 5. This is opposite to architecture A with transverse montage whose mean FP per minute was the lowest but mean sensitivity was also the worst.

V. Discussion

In terms of IED and normal window classification, all models achieved high AUC scores. Despite having lower sensitivity scores than ResNet models, our models

Table 2. Results of whole EEG recording classification

Model	AUC
Architecture A - Trans (1)	0.45
Architecture A - Long (2)	0.80
Architecture B	0.84
Architecture C	0.77
Average of 1 & 2	0.72

Table 3. Mean FP/minute and mean sensitivity across all EEG recordings in test set at 0.8 probability threshold

Model	Mean FP/minute	Mean Sensitivity
Architecture A - Trans (1)	0.35	0.43
Architecture A - Long (2)	2.59	0.71
Architecture B	5.0	0.73
Architecture C	2.44	0.68
Average of 1 & 2	0.44	0.64

achieved better F1 scores. We investigated individual windows in each EEG and found that the majority of the misclassified windows as IED windows were eye movements or slow-frequency artifacts. This means that the band-pass filtering method in our pre-processing step is inefficient. On the other hand, all models achieved high AUC scores in the whole EEG recording classification task. Our results showed that optimizing the threshold for the output probability would make the models more clinically useful by reducing false-positive rates while maintaining an adequate yield of positive events.

Interpretability of results from automated software is crucial in a clinical environment. Our work could be extended with a similar method as in [16] to visualize the contribution of an electrode signal to IEDs. This would be useful in localizing the source of focal IEDs. Focal IEDs are maximal in one region and GCN method would allow automated topographic heat maps to be generated, making interpretation of localization easier for clinicians. In addition, we could also train the proposed architectures to automatically tag electrodes with more clinically meaningful labels if adequate annotated data is available (e.g: subcategories of IED). This will make the outputs more comprehensive to neurologists and will be part of our future work.

VI. Conclusions

This paper explored different GCN architectures for IED detection and compared the performances between transverse and longitudinal bipolar montages. In window classification, despite achieving high AUC scores, all models had low F1 scores due to artifacts which we plan to investigate further with artifact removal methods. However, the AUC scores in the whole EEG recording classification were high, indicating the possibility of thresholding the output probability of window classification to achieve clinical usefulness. With the integration of montages, our work might be extended to improve the interpretability and clinical relevance of the automated software. Although the current and previous work on

automated IED detection shows some promise, the low F1 scores do not allow for integration as yet into a ready algorithm for clinical usage. Further exploration of additional models is required for improving precision whilst maintaining a high sensitivity. Cross-validation across datasets acquired in different hospitals will be conducted to verify the generalisability of these models.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the Graduate Research Industry Scholarship (GRIP) at Monash University, Australia.

REFERENCES

[1] WHO, "Epilepsy," 2021. (available at: *https://www.who.int/news-room/fact-sheets/detail/epilepsy*).

[2] P. Jallon and P. Latour, "Epidemiology of idiopathic generalized epilepsies," *Epilepsia*, vol. 46 Suppl 9, pp. 10–14, 2005. (available at: *https://onlinelibrary.wiley.com/doi/10.1111/j.1528-1167.2005.00309.x*).

[3] A. Isaksson, A. Wennberg, and L. H. Zetterberg, "Computer analysis of EEG signals with parametric models," *Proceedings of the IEEE*, vol. 69, no. 4, pp. 451–461, Apr. 1981. (available at: *https://ieeexplore.ieee.org/document/1456256*).

[4] A. T., M. G., D. G., E. C., L. Astrakas, S. Konitsiotis, and M. Tzaphlidou, "Automated Epileptic Seizure Detection Methods: A Review Study," *Epilepsy - Histological, Electroencephalographic and Psychological Aspects*, D. Stevanovic, Ed. InTech, Feb. 2012. (available at: *http://www.intechopen.com/books/epilepsy-histological-electroencephalographic-and-psychological-aspects/automated-epileptic-seizure-detection-methods-a-review-study*).

[5] F. E. Abd El-Samie, T. N. Alotaiby, M. I. Khalid, S. A. Alshebeili, and S. A. Aldosari, "A Review of EEG and MEG Epileptic Spike Detection Algorithms," *IEEE Access*, vol. 6, pp. 60 673–60 688, 2018. (available at: *https://ieeexplore.ieee.org/document/8489863*).

[6] P. Corporate, "Persyst," 2021. (available at: *https://www.persyst.com/*).

[7] M. Scheuer, "Spike detection: Inter-reader agreement and a statistical Turing test on a large data set," *Clinical Neurophysiology*, vol. 128, no. 1, pp. 243–250, 2017. (available at: *https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=85007206710&origin=inward*).

[8] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *arXiv:1711.05101 [cs, math]*, Jan. 2019. (available at: *http://arxiv.org/abs/1711.05101*).

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]*, May 2015. (available at: *http://arxiv.org/abs/1505.04597*).

[10] Y. Hao, H. M. Khoo, N. von Ellenrieder, N. Zazubovits, and J. Gotman, "DeepIED: An epileptic discharge detector for EEG-fMRI based on deep learning," *NeuroImage: Clinical*, vol. 17, no. June 2017, pp. 962–975, 2018. (available at: *https://doi.org/10.1016/j.nicl.2017.12.005*).

[11] J. Jing, H. Sun, J. A. Kim, A. Herlopian, I. Karakis, M. Ng, J. J. Halford, D. Maus, F. Chan, M. Dolatshahi, C. Muniz, C. Chu, V. Sacca, J. Pathmanathan, W. Ge, J. Dauwels, A. Lam, A. J. Cole, S. S. Cash, and M. B. Westover, "Development of Expert-Level Automated Detection of Epileptiform Discharges During Electroencephalogram Interpretation," *JAMA Neurology*, Oct. 2019. (available at: *https://jamanetwork.com/journals/jamaneurology/fullarticle/2752666*).

[12] S. Clarke, "Computer-assisted EEG diagnostic review for idiopathic generalized epilepsy," *Epilepsy and Behavior*, 2019. (available at: *https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=85074486818&origin=inward*).

[13] D. Nhu, M. Janmohamed, L. Shakhatreh, O. Gonen, P. Kwan, A. Gilligan, C. Wei Tan, and L. Kuhlmann, "Automated Inter-Ictal Epileptiform Discharge Detection from Routine EEG," *Healthier Lives, Digitally Enabled*, 2021, pp. 65–71. (available at: *https://ebooks.iospress.nl/doi/10.3233/SHTI210012*).

[14] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," *arXiv:1606.09375 [cs, stat]*, Feb. 2017. (available at: *http://arxiv.org/abs/1606.09375*).

[15] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *arXiv:1609.02907 [cs, stat]*, Feb. 2017. (available at: *http://arxiv.org/abs/1609.02907*).

[16] I. Covert, B. Krishnan, I. Najm, J. Zhan, M. Shore, J. Hixson, and M. J. Po, "Temporal Graph Convolutional Networks for Automatic Seizure Detection," *arXiv:1905.01375 [cs, eess, stat]*, May 2019. (available at: *http://arxiv.org/abs/1905.01375*).

[17] D. Zeng, K. Huang, C. Xu, H. Shen, and Z. Chen, "Hierarchy Graph Convolution Network and Tree Classification for Epileptic Detection on Electroencephalography Signals," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2020. (available at: *https://ieeexplore.ieee.org/document/9149886*).

[18] C. Fan, "Spectral Graph Theory," 1997. (available at: *http://www.math.ucsd.edu/~fan/research/revised.html*).

[19] J. N. Acharya and V. J. Acharya, "Overview of EEG Montages and Principles of Localization," *Journal of Clinical Neurophysiology*, vol. 36, no. 5, pp. 325–329, Sep. 2019. (available at: *http://journals.lww.com/00004691-201909000-00002*).

[20] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *Proceedings of the 30 th International Conference on Machine Learning*, vol. 28, p. 6, 2013. (available at: *https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf*).

[21] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization," *arXiv:1607.06450 [cs, stat]*, Jul. 2016. (available at: *http://arxiv.org/abs/1607.06450*).

[22] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using Convolutional Networks," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, Jun. 2015, pp. 648–656. (available at: *http://ieeexplore.ieee.org/document/7298664/*).

[23] D. Grattarola and C. Alippi, "Graph Neural Networks in TensorFlow and Keras with Spektral," *arXiv:2006.12138 [cs, stat]*, Jun. 2020. (available at: *http://arxiv.org/abs/2006.12138*).

[24] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, Jul. 2019. (available at: *http://arxiv.org/abs/1809.04356*).