Geometric Deep Learning for Subject Independent Epileptic Seizure Prediction using Scalp EEG Signals

Theekshana Dissanayake, Tharindu Fernando Member, IEEE, Simon Denman Member, IEEE, Sridhar Sridharan Life Senior Member, IEEE and Clinton Fookes Senior Member, IEEE

Abstract—Recently, researchers in the biomedical community have introduced deep learning-based epileptic seizure prediction models using electroencephalograms (EEGs) that can anticipate an epileptic seizure by differentiating between the pre-ictal and interictal stages of the subject’s brain. Despite having the appearance of a typical anomaly detection task, this problem is complicated by subject-specific characteristics in EEG data. Therefore, studies that investigate seizure prediction widely employ subject-specific models. However, this approach is not suitable in situations where a target subject has limited (or no) data for training. Subject-independent models can address this issue by learning to predict seizures from multiple subjects, and therefore are of greater value in practice. In this study, we propose a subject-independent seizure predictor using Geometric Deep Learning (GDL). In the first stage of our GDL-based method we use graphs derived from physical connections in the EEG grid. We subsequently seek to synthesize subject-specific graphs using deep learning. The models proposed in both stages achieve state-of-the-art performance using a one-hour early seizure prediction window on two benchmark datasets (CHB-MIT-EEG: 95.38% with 23 subjects and Siena-EEG: 96.05% with 15 subjects).

Index Terms—machine learning, neural networks, signal processing, seizure prediction, electroencephalography

I. INTRODUCTION

According to recent statistics given by the World Health Organization (WHO), over 50 million people in the world suffer from epilepsy [1]. Forecasting seizures has important implications for patients, and hence there has been a growing interest in data-driven deep learning-based studies exploring epileptic seizure prediction [2], [3].

Epileptic seizure-related studies can be grouped into three classes: seizure prediction, seizure detection and seizure classification. Seizure prediction (the aim of this study) predicts the onset of a seizure in advance by differentiating between the pre-ictal and the interictal brain stages of a subject [4].

Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan and Clinton Fookes are with the Signal Processing, Artificial Intelligence and Vision Technologies (SAIVT) group, Queensland University of Technology (QUT), 4000, Australia

In contrast, seizure detection detects an epileptic seizure event by differentiating between the ictal and interictal brain stages of the subject [5]. The third category, seizure classification, seeks to recognize the seizure type [6].

Regarding seizure prediction, if the subject of interests brain stage is pre-ictal, then there is a high chance an epileptic seizure event will occur in the near future. The forecasting window of the model depends on the selected pre-ictal duration; i.e. if the pre-ictal duration is one hour, and if the model indicates the brain stage of the subject is pre-ictal, then an epileptic seizure event is forecast to occur within one hour. The pre-ictal duration is subject-specific, and is also a design choice for a given approach, i.e. the same pre-ictal duration is assumed for all subjects in the dataset [3], [4].

In general, most deep learning-based medical studies ignore inter-subject variations in data, and design global classifiers, incorporating samples from multiple subjects (ex: heart anomaly detection [7] and blood pressure estimation [8]). Such models are inherently subject-independent. In seizure prediction literature, most researchers consider a subject-specific setting due to the high inter-subject variability in EEG seizure data [2], [4]. In contrast, most seizure type classification studies and EEG anomaly detection studies in the literature are subject-independent [9].

In our analysis, we define the subject-independency of a model as its ability to perform seizure prediction regardless of the patient annotations in the dataset, i.e. a global classifier. However, being subject-independent does not imply the model can adequately generalize to unseen subjects [4]. Critically, subject-independent models offer a more viable real-world solution for seizure prediction when a selected subject within the dataset has a limited number of examples, as they obviate the need for further data collection as is required by subject-specific models [4], [10], [11].

Although EEG signals possess subject-specific characteristics [12], the active brain region for a specific event/reaction is often common across humans. For instance, the anger emotion activates the prefrontal cortex of the brain and the amygdala region of the brain in all humans [13]. Prior research has shown how a deep learning model for emotion recognition is able to learn this common pattern, in spite of other variations present between subjects. However, in the case of epilepsy epilepsy (i.e. a medical condition), along with the subject vari-
ant properties of EEG signals, variations are also encountered due to the type of seizure (i.e. focal or non-focal), adding further complexity. Depending whether the seizure is focal or non-focal, an epileptic event may originate in one region of the brain and can propagate to a different region. In the sections below, we explain how the use of Geometric Deep Learning (GDL) enables us to address this challenge and design subject-independent deep networks for epileptic seizure prediction.

Learning from graphs, termed Geometric Deep Learning (GDL), is an emerging field that deals with graph or irregularly structured data (i.e. non euclidean). GDL methods which have been designed for learning from graphs with thousands of nodes (or vertices) have been successfully employed for learning from EEG data by exploiting the natural structure of the EEG grid [14], [15]. In such studies, researchers have argued that GDL methods can improve performance for EEG-based problems given that the physical/natural representation of EEG is itself a graph-like structure [16].

A common challenge with GDL is defining the prior graph for training the GDL model. As a straightforward solution to this, researchers have relied on distance-based methods, preferably $L2$ [17]–[19], where the physical distance between electrodes determines connectivity in the graph. Other representation techniques proposed include correlation methods and mutual information gain-based methods [20]. In this study we use an $L2$ distance-based adjacency matrix formulation as it computes an interpretable representation based on physical distances.

An alternate solution to the graph definition task is synthesizing graphs conditioned on subjects. Related works in the emotion recognition literature have sought to learn a global graph for the task [21]. However, as epileptic seizure events can originate in different brain regions, learning a subject-specific graph has additional value. Furthermore, such a technique may also have other applications such as seizure localization. As such, we aim to synthesize subject-specific graphs using deep learning. Our main contributions are the following:

1) We propose a novel Graph Neural Network-based deep learning framework which uses an $L2$ distance-based adjacency matrix and graphs synthesized by a deep neural network. In both settings (distance based and synthesized graphs), we achieve state-of-the-art performance on two benchmark datasets.

2) To the best of our knowledge, we are the first study to propose a deep learning method for EEG graph synthesis. The proposed graph synthesis network can be conditioned to generate specific EEG grids based on different levels of connectivity.

3) We conduct the first subject-independent seizure prediction study using the Siena EEG dataset.

4) We demonstrate the potential value of the proposed approach to the task for seizure localization from Scalp EEGs.

The rest of the paper is organized as follows. In Section II, we discuss related work on seizure prediction and graph-based learning on EEG data. In Section III, we present our proposed methodology and provide a theoretical overview of GDL. In Section IV, we present our findings quantitatively, and use visualizations to understand the properties that reside in the synthesized graphs. In Sections V and VI, we summarize our findings and propose some future directions for deep learning-based EEG graph synthesis.

II. RELATED WORK

Epileptic seizure prediction literature can be divided into subject-independent and subject-specific seizure prediction, with subject-specific approaches being the more common method. Given the lack of subject-independent methods, we present an overview of both types of approach.

1) Subject-Independent Seizure Prediction: Our previous study developed a state-of-the-art multi-task siamese network for patient-independent epileptic seizure prediction, and can identify a seizure with a one hour early prediction window with 91.51% accuracy (for 24 subjects in the CHB-MIT-EEG database) [4]. Considering related studies, the Convolutional Neural Network (CNN) model of Hu et al. [22] achieved 86.25% accuracy on the CHB-MIT-EEG dataset. Their approach demonstrated an almost 12% accuracy improvement when identifying pre-ictal samples closer to a seizure compared to those more distant in time. The model proposed by Khan et al. [3] achieved 0.8660 ROC-AUC with a 10 minute pre-ictal window, though they used only 16 subjects from the CHB-MIT-EEG dataset.

2) Subject-Specific Seizure Prediction: A challenge when considering a subject-specific seizure model is data scarcity. A common method to overcome data limitations for subject-specific models is selecting a subset of subjects for the analysis, omitting subjects with too little data. For this, researchers have placed constraints on the number of seizure samples for a subject and the continuity of the EEG signals captured [2], [23].

Data augmentation or re-sampling is also a commonly used approach. Even though augmenting EEG data is a relatively new biomedical research problem, researchers have considered augmentation as a solution for data scarcity. The recent study by Zhang et al. [24] proposed a generative adversarial network for synthesizing EEG seizure data and demonstrated 92.2% accuracy with a 30 minute prediction window. A similar approach was proposed in [25].

Tsouiris et al. [26] overcame data scarcity challenges by extracting descriptive features from EEG signals, and achieves 99.84% accuracy on the CHB-MIT-EEG dataset using a Long-Short-Term-Memory (LSTM) network. As features, they have employed time-domain statistical, spectral power-based, autocorrelation-based and graph theory-based features. However, they fail to detail the number of datapoints used for each subject and the number of parameters in their LSTM model. While we acknowledge this method as a viable solution as extracted features contain seizure-related predictive biomarkers, we note that for deep learning model training the number of model parameters play an essential role alongside having descriptive features related to the medical event.

Finally, transfer learning can also be used to address data limitation issues. As shown in our previous analysis, prior to
being transferred into the space of a specific patient, a global classifier is required, necessitating the learning of a patient-independent model. For this setting, in our previous study, we achieved 96.67% average accuracy for all 24 subjects in the CHB-MIT-EEG dataset [4].

Considering deep learning techniques used, our previous work, which is state-of-the-art for seizure prediction, we employed a siamese-CNN architecture. The related methods of [3], [22] both used classical CNN models to address seizure prediction, albeit with different prediction windows. Despite promising results, none of these methods have been able to fully capture subject-independent patterns in the data, and thus further research is required. With respect to capturing subject-independent patterns, as discussed in [16], [27], we argue that typical CNN models cannot effectively learn from non-euclidean data available from the EEG grid. As a result, CNN models fail to capture information residing in the neighboring regions of the EEG grid. Therefore, we propose a Graph Neural Network-based (i.e. GDL) [28] approach for seizure prediction, and conduct evaluations using both physical connection-based graphs and graph synthesized using deep learning.

Few epilepsy related GDL methods exist and most have relied on subject-specific solutions [16], [29], [30], with Lian et al. [16] the only study that attempts to address subject-independent seizure prediction using GDL and iEEG data. We note that our investigation is the first study that aims to address the seizure prediction using Scalp EEG signals and GDL. We argue using scalp EEG signals adds complexity compared to using iEEG signals, given the non-invasive nature of scalp EEG signals which leads to lower fidelity signals due to factors such as scalp electrodes being affected when propagating through the human scalp. We also note that an iEEG grid is localized based on the seizure location which is determined by medical intervention. However, Scalp EEGs use a uniform grid that is placed on the head, enabling the design of general solutions from a common set of electrodes.

III. METHODOLOGY
A. Data Acquisition and Pre-processing

To evaluate the robustness of our proposed models, we use the CHB-MIT-EEG [31] and Siena EEG [32] datasets. For both databases, we extract Mel-frequency cepstrum coefficients (MFCCs) features as they contain predictive biomarkers related to epileptic seizures [4].

Following [3], [4], [26], we set the pre-ictal duration to one hour. We extract interictal samples by selecting EEG segments before and four hours after seizure onset. As in a typical anomaly detection task, we obtained more interictal recordings compared to pre-ictal recordings. To create a balanced dataset, we extracted 10s pre-ictal segments with 50% overlap, while interictal segments are non overlapping 10s segments. The majority of recordings in the CHB-MIT-EEG dataset have 23 channels, though some have only 22 channels. For consistency, we make all recordings 23-channel by using channel 22 (the physically closest channel) as the missing channel 23. The feature extraction process yields a \( \approx 300k \) sample balanced dataset extracted from 23 subjects.

For the Siena dataset, we also set the pre-ictal duration as one hour. Given the small dataset size we extract interictal samples immediately before the start of the pre-ictal duration with the exception of for subjects PN01, PN07 and PN14 for which we were able to satisfy the four hour condition followed for CHB-MIT-EEG. We used a 2s duration with 40% overlap for pre-ictal EEG samples, and non overlapping interictal samples. This change is based on the data requirements for model training. This process yields an \( \approx 180k \) balanced EEG dataset. Here, EEG recordings contain 29 channels and the dataset contains 15 subjects.

For both datasets, we were unable to find specific channel-level annotations for epileptic seizures. Therefore, even though our aim is to learn subject-specific graphs, we cannot explain the validity of our learned graphs. To further understand our results, in Section IV-D we visualize high dimensional learned embeddings using t-SNE analysis (i.e. in a 2D space). While we use two datasets to demonstrate the robustness of our method, we cannot perform cross domain evaluations as the channel information for the datasets are not consistent. The CHB-MIT-EEG dataset is composed of differential channel voltages, whereas the Siena EEG dataset contains channel level voltages. The datasets also contain different numbers of electrodes and electrode placements.

B. Chebyshev Graph Convolution.

From Spectral Graph Theory, the Graph Fourier Transform (GFT) is defined as \( \hat{X} = UX \) and the inverse GFT is \( X = U^T \hat{X} \). Here, \( X \) is the feature embedding from the time domain and \( \hat{X} \) holds values in the spectral (frequency) domain. For the GFT, \( U \) is the orthogonal eigenvector matrix of Laplacian \( L \), and \( G \) is the diagonal filter coefficient matrix \( L = D - A \), though researchers prefer the normalized Laplacian in the form of \( \hat{L} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \) (\( D \) denotes the degree matrix). Convolution (*) in the time domain is equivalent to element-wise multiplication (\( \odot \)) in the frequency domain. Therefore, intuitively, if our convolution kernel is defined as \( g \) in the spatial domain, then the convolution operation on the defined graph \( G \) becomes \( X \ast G = U^T (U \times X \odot U \times g) \). To reduce the complexity of the convolution operation, Defferrard et al. [33] proposed an approximation using Chebyshev polynomials.

\[
X \ast G \approx \sum_{i=0}^{K} \theta_i T_i(\hat{L})X, \quad \hat{L}' = \frac{2}{\lambda_{\max}} \hat{L} - 1.
\]

Here, \( \hat{L}' \) is the scaled normalized Laplacian, \( \theta \) is the learnable weight matrix \( ([K \times d_{in}, \quad d_{out}]) \), \( \lambda_{\max} \) is the maximum eigenvalue of \( \hat{L} \) and \( T_i(x) \) denotes the Chebyshev polynomials. Defining \( \hat{X}_k \) as \( \hat{X}_k = T(\hat{L})X \), and by using the recurrence form of Chebyshev polynomials, we obtain \( \hat{X}_k = 2\hat{L}\hat{X}_{k-1} - \hat{X}_{k-2} \). To \( \hat{X}_0 = X \) and \( \hat{X}_1 = \hat{L}X \). In Equation 1, \( K \) denotes the degree of the approximated polynomial and \( \lambda_{\max} \) is typically set to 2.

Graph convolution operations use a global graph structure \( G \) and a set of node specific embeddings \( (X) \). However, within the EEG domain, two main factors should be considered. Firstly, even though using the \( L2 \) distance to compute the graph structure is preferable, this places no upper bound
on the degree of a node which varies depending on the threshold selected. Secondly, given that EEG channels have highly irregular non stationary patterns, learning from raw signals may impact model performance. Even if this setting is used, depending on the sampling frequency of the EEG, parameters of the GDL layer will expand exponentially.

C. Problem Formulation

The proposed deep learning framework, shown in Figure 1, combines two components: a Classification Graph Neural Network (C-GNN), and a Graph Synthesizing Network (GSN). The C-GNN acts as a seizure prediction model and the GSN seeks to generate a graph conditioned on a subject (subject-specific). To evaluate the model’s performance variation with respect to the degree of the adjacency matrix we only use the C-GNN component. To evaluate the effectiveness of generating a graph, we use the complete system (C-GNN + GSN). The problem formulation for synthesizing a graph is discussed in Sections III-C.1 and III-C.2. In Section III-C.3 we discuss the C-GNN model architecture. Finally, we discuss the hyper-parameter tuning procedure in Section IV-A.

1) Computing Distance-based Adjacency matrices: As previously noted, the L2 distance-based adjacency matrix is a straightforward method to compute the adjacency matrix for EEG graph convolution \[17\]–\[19\]. Following \[17\]–\[19\], we obtain different adjacency matrices by thresholding the L2 distance between the physical electrodes (see Figures 2a and 2b). Then, we incrementally demonstrate how to generate adjacency matrices partially or fully, and show how this can aid in localising the seizure onset area for patients with epilepsy.

Suppose we have an \( N \) channel \((c_{x,y})\) EEG grid, which we transform to a 2D plane to compute the distance matrix,

\[
distance_{i,j} = \sum_{i=1}^{N} \sum_{j=0}^{N} L2(c_{x,y}^i, c_{x,y}^j). \tag{2}
\]

This matrix can be used to determine different adjacency matrices via a threshold. When determining these adjacency matrices, we consider the original connections that we can visually/physically observed in the EEG grid.

We argue that this setting aligns with the concept of graph learning, as if the degree of a given node is high, then the problem becomes similar to a CNN approach where we learn spatial features considering all or a stacked set of EEG channels (a 3D matrix). Since we are simply averaging all feature embeddings in the fusion operation, aggregating a higher number of embeddings may obscure important patterns associated with each node. We evaluate this through varying the degree of the distance-based graph in Section IV-B.1.

2) Graph Synthesizing Network: For graph generation, we introduce Partially-Learned and Fully-Learned settings. In the Partially-Learned setting, we aim to retain some original EEG grid connections while learning some additional connections. In the Fully-Learned setting, we extend the partially-learned connections to synthesize the entire graph using a deep learning model. The Fully-Learned method has no prior knowledge of the physical connections. Hence, the representation obtained by the method captures hidden patterns between the EEG channels, and may or may not have a direct relationship with the physical EEG grid. One property of the proposed approach is it can be constrained to a number of nodes \( (n) \) and number of edges \( (e) \) for those nodes. For both graph generation settings, we use the same CNN-Encoder network, and change other components according to the generation setting.

The architecture is developed by considering two key requirements. Firstly, we seek a model to generate subject-specific embeddings for graph generation. We use a CNN-encoder inspired by [1], with small modifications to cater for the different input feature maps used. Secondly, we need to generate a sequence of values (edge indexes and node indexes) with meaningful relationships between them. For this, we selected two LSTM networks configured with \( N \) hidden units each, which capture patterns from the resulting embedding from the CNN-Encoder. Considering the number of model parameters and limited data, we limit the number of stacked layers to one.

Figure 1 illustrates the complete architecture of the proposed framework. The Fully-Learned Graph setting is a combination of a CNN-Encoder and two LSTM branches. The CNN-Encoder acts as a subject-specific embedding generation component, which takes a feature map (MFCCs) and transforms it into an embedding of size 120 \((EM_{120})\). Then, two LSTM branches \((LSTM_N \text{ and } LSTM_E)\) are used to compute the graph. For each of those components, the hidden dimension is set to \( N \). Given \( n \) and \( e \), we use the following operations to compute the graph. From the output from \( LSTM_N \), i.e. the node attention module, we take the \( n \) indices with the highest attention values. Then, we set them as the nodes in the generated graph,

\[
\hat{N}_{[n]} = \text{top}_k(LSTM_N(Em), k = n), \hat{N} \in \mathbb{N}. \tag{3}
\]

Then, for each node, we find the \( e \) edges using the \( n \) last hidden stages from the \( LSTM_E \) network. Similar to the previous setting, we take the highest attentive indices as edges (here, \( \hat{E}_{[e \times n]} = \Lambda^{n \times e} \)),

\[
\hat{E}_{[e \times n]} = \text{top}_k(LSTM_E(Em); \neg k = e), \hat{E} \in \mathbb{E}. \tag{4}
\]

By Using \( \hat{N}_{[n]} \) and \( \hat{E}_{[e \times n]} \), the adjacency matrix can be formed using \( F_{adj}(\hat{N}_{[n]}, \hat{E}_{[e \times n]}) \), if \( A \) is initiated to \([N, N] \) zero matrix and \( \text{enu()} \) denotes an enumeration function,

\[
F_{adj} = [A_{k,j} = 1 \text{ for } j \text{ in } \hat{E}_{[i]} \text{ for } i, k \text{ in } \text{enu} (\hat{N})]. \tag{5}
\]

For a graph convolution operation, the resulting matrix should be symmetric. We make the matrix symmetric using \( A = \frac{1}{2}(A + A^T) \). To avoid self loops we multiply the generated adjacent matrix by \((1 - I_{[N, N]}), A = A \times (1 - I_{[N, N]}) \). As a result, in some cases nodes may have \( e \) – 1 connections.

For the Partially-Learned setting, we add this generated graph to the distance-based matrix \((A_d)\), and threshold the matrix back into the range of \([1, 0]\). In this setting, all overlapping connections will be set to 1.0. To train the model, we introduce the \( L_{L1-\text{matrix}} \) loss function described in Equation 6. This function implements an online mining technique that computes the \( L1 \) distance of generated graphs averaged \((\bar{A}_I)\).
Fig. 1: Proposed architecture: Classification Graph Neural Network (C-GNN) and the Graph Synthesizing Network (GSN). Conv2D : \( N(h, w) \): Convolution Layers with \( N \) # of filters sized \([h \times w]\); Chebyshev\((x_{in}, x_{out}, K)\): Chebyshev layer with features \((in, out)\) order \(K\); MFCC: Mel-frequency Cepstral coefficients; Dense: Fully-Connected. For all layers we use dropout with 0.2 probability. For all layers except the final output, we use relu activation functions. For the final output we use a sigmoid activation.

Fig. 2: Adjacency matrices computed using Equation 2 for the two EEG datasets used. Here, channels are numbered according to the annotations given in the metadata. In this setting \( A_i \subset A_{i+1} \subset \ldots \subset A_P \). Here \( P \) denotes the number of patients in the dataset,

\[
L_{L1-Matrix} = \frac{-1}{P^2} \sum_{i=1}^{P} \sum_{j=1}^{P} L1(A_i, A_j), \quad (i \neq j).
\]  

(6)

We also use Contrastive Loss \( (L_{Cons}) \) to ensure that the combined embedding produced by the two LSTM components is subject-specific. Selection of this loss function is inspired by our previous investigation and related work in [34], where \( L_{Cons} \) demonstrated an ability to generate embeddings that are conditioned on some criteria. Equation 7 describes how we can compute \( L_{Cons} \) for a given input pair \((x, x_q)\). Here, \( x_q \) is a randomly queried sample from the dataset, either from the same subject-id \((Y = 1)\) or from a different subject-id \((Y = 0)\). We set the margin to 0.001 and we use L2 distance as the distance measure \((i.e. d = L2(x, x_q))\) [4]. The complete loss function and optimization procedure is in Section IV-A,

\[
L_{Cons} = Y \times d^2 + (1 - Y) \times \max(margin - d, 0).
\]  

(7)

3) Classification Graph Neural Network (C-GNN): By definition, a GDL model deals with feature embeddings. To obtain these, flattening a 2D feature representation is not feasible as it dramatically increases the number of learnable weights in graph layers. Therefore, we used an additional dimensionality reduction network (DR-Net) to learn a subject-independent representation, which will be ultimately beneficial for the
subject-specific graph (learned or otherwise).

Preliminary experiments led us to select an LSTM layer with 64 hidden units, which was found to outperform a CNN-Encoder that returned a similar sized embedding. We believe the robustness of the LSTM network is a result of its ability to capture temporal relationships. In operation, for the CHB-MIT-EEG dataset, the input to DR-Net is shaped \([23 \times 26 \times 254]\) and it produces an output shaped \([23 \times 72]\). Here, 23 denotes the number of channels and 72 is the experimentally determined embedding size. The remainder of the model comprises a set of graph convolution layers and a simple two layer multilayer perceptron (MLP). To select optimal parameters for both DR-Net and the GDL component, we conducted a hyper-parameter tuning process which is described in Section IV-A.

IV. RESULTS

In the following subsections we present the results of our evaluations. Section IV-A presents results of hyper-parameter evaluations performed to determine the final architecture of our approach. Section IV-B.1 presents results for ten different distance-based adjacency matrices defined using the physical structure of the EEG grid (see Equation 2). In Section IV-B.2, we discuss our findings for the two proposed graph generation settings: Fully-Learned and Partially-Learned. In Section IV-C, we compare our approach with state-of-the-art models from the literature. Finally, in Section IV-D we comparatively analyse results from six different graph generation settings.

A. Hyper-parameter Tuning

Unlike classical machine learning algorithms which contain comparatively few hyper-parameters, parameter optimization for a deep neural network is a complex, time consuming and resource intensive task and exhaustive searches for optimal network designs are seldom possible. In this section, we present results from preliminary experiments conducted to determine the architecture and parameters of our framework.

The C-GNN network incorporates a stack of LSTMs and a graph neural network that uses Chebyshev graph convolution layers. Table I and II present parameters we considered during network design. Accuracy refers to the test accuracy on the CHB-MIT-EEG dataset for the network design. Accuracy refers to the test accuracy on the training configurations in Table III. For all model development hyper-parameters, we evaluated our models on different generation settings and adjacency matrices. We present our model training configurations in Table III. For all model development purposes, we used the PyTorch library.

B. Graph Network Evaluations

1) Evaluation of Distance-based Adjacency Matrices.: In this experiment, we evaluate the proposed C-GNN architecture with distance-based adjacency matrices defined using Equation 2. We evaluate if having nodes with higher degrees improves predictive performance. We conducted experiments on two different datasets, and used the ten different distance-based adjacency matrices shown in Figures 2a and 2b. For the CHB-MIT-EEG dataset, we adopt a \([23 \times 23]\) adjacency matrix, and for the Siena EEG dataset we use a \([29 \times 29]\) matrix. It should be noted that the matrices presented in Figures 2a and 2b demonstrate the property \(A_{i} \subset A_{i+1}\), as they are all derived from distance\(_{i,j}\) with different thresholds applied.

From Table IV, for the CHB-MIT-EEG dataset, having nodes with higher degrees does not improve performance, with \(A_{0}\) showing the best results with the fewest connections (average degree 1.0, see Figure 2a), however the accuracy change is negligible (\(\approx 0.20\%\)). For the Siena EEG dataset, \(A_{2}\) yields the best results having nodes with an average degree of 5.0. For the Siena dataset, we observe a slight increase in accuracy with the average degree of the matrix (average difference \(\approx 0.40\%\)). Nonetheless, these differences are small. Therefore, for the Chebyshev operation, having a dense adjacency matrix does not necessarily improve performance.
| $N_{LSTM}$ | Hidden-Size ($h_t$) | $N_{Cheb}$, ($t$) | Degree ($K_t$) | Input and Output Size ($[d_{in} \times d_{out}]_i$) | Test Accuracy |
|---|---|---|---|---|---|
| 1 | 64 | 2 | 4, 4 | $[64 \times 32], [32 \times 16]$ | 86.12% |
| 2 | 64 | 2 | 8, 4 | $[64 \times 32], [32 \times 16]$ | 87.65% |
| 2 | 64 | 2 | 4, 4 | $[64 \times 32], [32 \times 16]$ | 88.51% |
| 2 | 64 | 3 | 8, 4, 4 | $[64 \times 48], [48 \times 32], [32 \times 32]$ | 89.11% |
| 3 | 64 | 2 | 8, 4 | $[64 \times 64], [32 \times 32]$ | 90.03% |
| 3 | 72 | 2 | 8, 8 | $[72 \times 32], [32 \times 24]$ | 93.45% |
| 3 | 84 | 2 | 8, 8 | $[84 \times 32], [32 \times 16]$ | 93.29% |
| 3 | 72 | 2 | 8, 8 | $[72 \times 32], [32 \times 16]$ | 93.37% |
| 3 | 72 | 3 | 8, 4, 4 | $[72 \times 64], [32 \times 32], [16 \times 16]$ | 93.35% |
| 3 | 84 | 3 | 8, 4, 2 | $[84 \times 48], [48 \times 32], [32 \times 16]$ | 93.43% |
| 3 | 72 | 4 | 8, 3, 2 | $[72 \times 48], [48 \times 32], [32 \times 32]$ | 93.39% |

TABLE I: Hyper-parameter tuning results for the C-GNN Component (see Figure 1)

| Hyper-Parameters | Test-Acc | $L_{Cons}$ | $L_{Matrix}$ |
|---|---|---|---|
| 0.5 | 0.1 | 0.3 | 91.33% | 0.036 | -40.24 |
| 0.6 | 0.1 | 0.3 | 90.94% | 0.040 | -118.44 |
| 0.7 | 0.2 | 0.1 | 91.22% | 0.039 | -90.27 |
| 0.8 | 0.1 | 0.1 | 91.58% | 0.032 | -100.00 |

TABLE II: Loss function (Equation 8) weight tuning results.

2) Evaluation of Fully-Learned and Partially-Learned Graphs: This section presents the results from Fully-Learned (FL) and Partially-Learned (PL) graphs. Both techniques use the architecture shown in Figure 1 (highlighted in the maroon coloured box), and both models are trained using the loss function in Equation 8. In the FL graph setting, we aim to learn the entire adjacency matrix. For the PL graphs, we augment the base adjacency matrix $A_0$ given that it represents the basic underlying structure of the EEG grid.

One observation from Section IV-B.1 is having higher degree nodes does not guarantee improved classifier performance. Given this finding, for the graph synthesis procedure we limit the number of edges by varying the value $e$ closely around 4. In the generation setting, we expect the GSN to synthesize graphs that are localized to a particular region, which may help for seizure localization applications. With this objective, we also constrain $n$ to the range of 3 to 10. Given this, for the FL evaluation setting we configured the model with $[n, e] \in \{[4, 6], [6, 4], [10, 4]\}$. Since the PL graph already has a set of default connections, we evaluated the model with $[n, e] \in \{[3, 5], [4, 4], [6, 4]\}$ settings. Table V presents the results for different configurations of the GSN network on the CHB-MIT-EEG and Siena Scalp EEG datasets.

From Table V, both generation settings offer similar results to the distance-based setting in Section IV-B.1. Under the FL setting, for both databases synthesising an adjacency matrix with higher degree achieves slightly improved results, but the accuracy differences are negligible. Similarly, we observe a slight accuracy difference when comparing results for the PL learning setting for both databases. Comparing the performance of the FL and PL methods, we see similar results in accuracy and subject-specific embedding separation capability ($L_{Cons}$). Given that the generation settings are different, and different settings yield different matrices, we cannot compare losses from the $L_{L1-Matrix}$. In Section IV-D, we further analyse and visualize these learned connections.

We note that for different generation setting ($n, e$, FL and PL) the classifier produces similar results. One important observation from this analysis is that even though these graphs are synthesized, they offer similar to using the physical connectivities available in the EEG grid. As mentioned, in this architecture, we used the same prediction neural network (C-GNN), and only change the graph representation of the signal (distance-based or learned). Visualising the results, we argue that these synthesised graphs also capture some hidden associations related to subjects and seizures in the dataset.

C. Performance Comparison with Existing Models.

Table VI compares our approach with state-of-the-art models. The top-most table indicates the performance on the CHB-MIT-EEG dataset, and the bottom table shows results on the Siena EEG dataset. For completeness, we are also including state-of-the-art subject-specific (followed by *) models from the literature, though note that these are trained and evaluated in a very different manner to the models proposed in this work. The results show the two best performing models for the CHB-MIT-EEG dataset are a result of the distance-based graphs and graph synthesis approaches proposed in this paper. Both models outperform the state-of-the-art model of [4] by almost 5% accuracy. For the Siena EEG dataset, we are the first subject-independent study in the literature.

D. Analysis and Interpretation of Learned Graphs.

From Sections IV-B.1 and IV-B.2, synthesized adjacency matrices produce similar results to distance-based graphs. As these graphs are the result of a deep learning model, we cannot fully understand the internal structures in the data that lead to these results. As such, we use qualitative evaluations and high-dimensional data visualizations on the CHB-MIT-EEG dataset to unveil hidden patterns that lead to these generated graphs.

1) Fully-Learned Graphs: Figure 3 shows graphs obtained from the FL setting. Graphs are rendered using 400 randomly selected samples from each subject containing pre-ictal and ictal segments, and we show the averaged adjacency matrices. In Figure 3a we present all 24 generated graphs to demonstrate the subject-specific graph generation ability of our method. For generation, we used $n = 4$ and $e = 6$. Figure 3b illustrates synthesis results for the configuration $n = 6$, $e = 4$.

We observe variations in the connection distributions in Figure 3. For example, in Figure 3a for patients $P_0$, $P_3$ and $P_{22}$, leaned connections form a clustered set. Similarly, in
Figure 3b, graphs generated for $P_2$ and $P_3$ show the same localized pattern: although other graphs have widely-distributed connections. Patterns are not consistent across different graph creation operations, and this non-deterministic behaviour may limit the proposed system in a seizure localization setting.

We also observe similarity between generated graphs across different subjects under different generation settings. For example, in Figure 3a, generated graphs for $P_{18} \approx P_{10} \approx P_7$, $P_2 \approx P_3$ and $P_{21} \approx P_{22}$, and the graphs in Figure 3b reflects the patterns observed.

Across Figures 3a and 3b we see learned connections are widely-distributed or clustered in different settings, and from adjacency matrices we can’t conclusively determine the seizure localization setting. From Figure 4a, feature representations for $P_{10}$ and $P_{18}$ have similar values in the high dimensional space. Similar patterns can be observed for $P_{20}$, $P_{21}$ and $P_{15}$ (see top corner of Figure 4a). However features from $P_{12}$ and $P_1$ appear to be separated in the learned space. We see similar patterns in the t-SNE visualisations for the other configuration schemes in Figures 4b and 4c. These observations reflect the patterns observed in the graph synthesis results, and suggest that the deep learning model needs further improvements to fully differentiate between subjects.

2) Partially-Learned Graphs: In the Partially-Learned setting, we investigate the value of incorporating physical connections from the EEG grid. In this setting, the base adjacency matrix, $A_0$, from the distance-based approach is used (see Section IV-B.1). In doing this we provide the generation process knowledge of the physical EEG grid to help the model to learn a consistent set of connections (i.e. nodes and edges).

Figure 5 illustrates adjacency matrices for two graph synthesis settings, and Figure 5b shows a second set of results for 10-F-CV: 10 Fold Cross-Validation evaluation protocol.
Fig. 3: Fully-Learned graphs from different configurations $[23 \times 23]$. In Figure 3c we show the visualizations after mapping the adjacency matrices to the actual EEG grid. Channels: 0: FP1-F7, 1: F7-T7, 2: T7-P7, 3: P7-O1, 4: FP1-F3, 5: F3-C3, 6: C3-P3, 7: P3-O1, 8: FP2-F4, 9: F4-C4, 10: C4-P4, 11: P4-O2, 12: FP2-F8, 13: F8-T8, 14: T8-P8, 15: P8-O2, 16: FZ-CZ, 17: CZ-PZ, 18: P7-T7, 19: T7-FT9, 20: FT9-FT10, 21: FT10-T8, 22: T8-P8. All matrices are rendered for 400 randomly picked samples from each subject in the CHB-MIT-EEG dataset.

One major limitation of the FL setting (Section IV-D.1) is the lack of steady convergence; i.e. for the same subject in a different evaluation setting the GSN model creates different adjacency matrices. However, as Figure 5 shows the the
Partially-Learned setting is comparatively deterministic.

From Figure 5a and 5b, we see that the two separate generation processes with the same setting result in a similar connection arrangement being learnt for most subjects. This is in contrast to the FL setting, and we believe this illustrates the importance of incorporating the underlying structure of the EEG grid to help the model to learn a set of uniform connections associated with the physical connections.

Figure 6 presents twelve mapped visualizations for the $n = 4$ and $e = 6$ graphs for a given configuration. Here, we show all learned connections from the two generation settings on the same EEG grid (green $n = 3$, $e = 5$). As in Figure 3c, we use blue dots to mark approximate locations (taken as the centre between two EEG electrodes). From Figure 4, for a majority of subjects we see that GSN model converges to a common set of overlapping nodes, despite different generation settings. We observed this for $18$ ($\frac{18}{23}$) subjects, and omit some subjects with identical generated graphs. We had unexpected outcomes for a small set of subjects in the CHB-MIT EEG dataset. As shown in Figure 6, for $P_3$ and $P_{22}$, the learned connections for $n = 3$, $e = 5$ setting are widely-distributed compared to other subjects.

V. DISCUSSION

As presented in Section IV-C, results on both datasets exceed the current state-of-the-art; however the graph generation methods need further refinement to best capture subject-independent patterns in the data. We believe the GSN component showed robust capability in synthesising subject-specific graphs for a given configuration ($n$ and $e$). However, the feature transformation performed by the LSTM has limitations. Using t-SNE, we identify that while it can produce features that are subject invariant in the embedding space, we encounter difficulties in explaining results for some patients.

In Section IV-B.1, the distance-based evaluations showed all methods had similar results, indicating that for Chebyshev GCNs the adjacency matrix determined by the physical connections need not be dense. We believe this is due to the operation's formulation and the $K$ (or $L^K$) term in the approximation, as higher $K$ values allow the filtering operation to aggregate information across more nodes. For both FL and PL generation settings, we obtained similar results. We witnessed $\approx 0.30\%$ accuracy difference when comparing the best performing distance-based adjacency matrix result and the best generation setting result for CHB-MIT-EEG dataset, and observed similar results for the Siena EEG dataset (see Tables IV and V for more details). Although the synthesised subject-specific graphs did not outperform our distance-based method, they outperformed the existing state-of-the-art [4].

While feature embedding visualizations demonstrate that the learned patterns capture hidden relationships in the data (see Figure 4), we are unable to further investigate those patterns given that deep learning models are inherently black-box in nature. We argue this black-box nature of deep learning models hinders the understanding of hidden patterns in data, and we believe the interpretation step followed in this study helps to explain this behaviour to some extent.

VI. CONCLUSIONS

In this study we proposed two types of geometric deep learning (GDL) techniques for epileptic seizure prediction. The first GDL method used a naive graph computation approach employing the physical structure of the EEG grid, while the second method used deep neural networks to synthesise graphs. Both techniques achieved state-of-the-art results on the CHB-MIT-EEG database and Siena EEG databases. Comparing both strategies, we observed very similar results, with both achieving over 95% accuracy.

Through t-SNE visualisations, we identified that our dimensionality reduction LSTM requires further improvements in capturing subject-independent patterns in data; although our GSN component demonstrated excellent performance in generating graphs. Given that this is the first work to investigate the possibility of synthesizing graphs for GDL for seizure analysis, we encourage further research on how to effectively define feature embeddings for GDL operations that can be used along with the synthesized subject-specific graphs.

Finally, we discussed the applicability of our method to seizure localization. While we are unable to validate the proposed method due to the lack of a Scalp EEG dataset with proper electrode-seizure focused annotations, we have provided visualizations and quantitative results to demonstrate the potential of our method for model-based seizure localization. Given that the method showed promising localization ca-
Fig. 5: Partially-Learned graphs from the different configurations $[23 \times 23]$ with $A_0$ as the base adjacency matrix.

Fig. 6: Results after mapping the adjacency matrices synthesized for $n, e = 4$ (5a) and $n = 3, e = 5$ settings to the actual EEG grid. All these matrices are rendered for 400 randomly picked samples from each subject in the CHB-MIT-EEG dataset.
pability, we encourage researchers to conduct further research to validate this capability with an appropriate dataset.

The connection setting arrangement we used was inspired by the results from Section IV-B.1 regarding the degree of the adjacency matrix. However, a medical perspective is also required to optimally determine the configuration of the adjacency matrix to ensure resultant models are meaningful to medical professionals as well as being accurate for prediction.

As discussed in the Section III-A, evaluating the method’s generalizability in a cross-domain setting is currently not possible. We note that for real world situations, the generalizability of the model on completely unseen datasets is a major requirement. However, inconsistencies in the datasets available in the literature currently prevent such evaluations.

REFERENCES

[1] “World health organization-epilepsy,” https://www.who.int/news-room/fact-sheets/detail/epilepsy, accessed: 2021-05-19.

[2] H. Daoud and M. A. Bayoumi, “Efficient Epileptic Seizure Prediction Based on Deep Learning,” IEEE Transactions on Biomedical Circuits and Systems, vol. 13, no. 5, pp. 804–813, 2019.

[3] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, “Deep learning for patient-independent epileptic seizure prediction using scalp eeg signals,” IEEE Sensors Journal, pp. 1–1, 2021.

[4] C. Gómez, P. Arbeláez, M. Navarrete, C. Alvarado-Rojas, M. Le Van Quyen, and M. Valderrama, “Automatic seizure detection based on imaged-EEG signals through fully convolutional networks,” Scientific Reports, vol. 10, no. 1, Dec. 2020. [Online]. Available: https://hal.sorbonne-universite.fr/hal-03122585

[5] T. Liu, D. Truong, A. Nikpour, L. Zhou, and O. Kavehei, “Epileptic seizure classification with symmetric and hybrid bilinear models,” IEEE Journal of Biomedical and HealthInformatics, vol. 24, no. 10, pp. 2844–2851, 2020.

[6] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, “A robust interpretable deep learning classifier for heart anomaly detection without segmentation,” IEEE Journal of Biomedical and Health Informatics, pp. 1–1, 2020.

[7] J. Esmaeipoor, M. H. Moradi, and A. Kadhodahomamadi, “A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals,” Computers in Biology and Medicine, vol. 120, p. 103719, 2020.

[8] T. Fernando, S. Denman, D. Ahmedt-Aristizabal, S. Sridharan, K. R. Laurens, P. Johnstone, and C. Fookes, “Neural memory plasticity for medical anomaly detection,” Neural Networks, vol. 127, pp. 67 – 81, 2020.

[9] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O’Brien, and A. Razi, “Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review,” 2 2020.

[10] S. Yang, B. Li, Y. Zhang, M. Duan, S. Liu, Y. Zhang, X. Feng, R. Tan, L. Huang, and F. Zhou, “Selection of features for patient-independent detection of seizure events using scalp eeg signals,” Computers in Biology and Medicine, vol. 119, p. 103671, 2020.

[11] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faulbert, “Deep learning-based electroencephalography analysis: A systematic review,” p. 37, 8 2019.

[12] H. Fischer, J. Sandblom, J. Gavazzeni, P. Fransson, C. I. Wright, and L. Bäckman, “Age-differential patterns of brain activation during perception of angry faces,” NeuroImage Letters, vol. 386, no. 2, pp. 99 – 104, 2005.

[13] Z. Wang, Y. Tong, and X. Heng, “Phase-locking value based graph convolutional neural networks for emotion recognition,” IEEE Access, vol. 7, pp. 93711–93722, 2019.

[14] D. Zhang, L. Yao, K. Chen, S. Wang, P. D. Haighi, and C. Sullivan, “A graph-based hierarchical attention model for movement intention detection from eeg signals,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 11, pp. 2247–2253, 2019.

[15] Y. Liu, Y. Qu, X. Pan, and Y. Wang, “Learning graph in graph convolutional neural networks for robust seizure prediction,” Journal of Neural Engineering, vol. 17, 05 2020.