Automated Seizure Detection and Seizure Type Classification From Electroencephalography With a Graph Neural Network and Self-Supervised Pre-Training

Siyi Tang[1,*], Jared A. Dunnmon[2], Khaled Saab[1], Xuan Zhang[3,†], Qianying Huang[3,†]
Florian Dubost[4], Daniel L. Rubin[4,5,‡], Christopher Lee-Messer[6,‡,*]

[1]Department of Electrical Engineering, Stanford University, Stanford, CA, USA
[2]Department of Computer Science, Stanford University, Stanford, CA, USA
[3]Department of Statistics, Stanford University, Stanford, CA, USA
[4]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
[5]Department of Radiology, Stanford University, Stanford, CA, USA
[6]Department of Child Neurology, Stanford University, Stanford, CA, USA

[†]Equal contributions. Work done during study at Stanford University.
[‡]Equal contributions.

[*]Address correspondence to:

Siyi Tang
siyitang@stanford.edu
Department of Electrical Engineering
Stanford University, CA, USA

Dr. Christopher Lee-Messer
cleemess@stanford.edu
Department of Child Neurology
Stanford University, CA, USA

Abstract

Automated seizure detection and classification from electroencephalography (EEG) can greatly improve the diagnosis and treatment of seizures. While prior studies mainly used convolutional neural networks (CNNs) that assume image-like structure in EEG signals or spectrograms, this modeling choice does not reflect the natural geometry of or connectivity between EEG electrodes. In this study, we propose modeling EEGs as graphs and present a graph neural network for automated seizure detection and classification. In addition, we leverage unlabeled EEG data using a self-supervised pre-training strategy. Our graph model with self-supervised pre-training significantly outperforms previous state-of-the-art CNN and Long Short-Term Memory (LSTM) models by 6.3 points (7.8%) in Area Under the Receiver Operating Characteristic curve (AUROC) for seizure detection and 6.3 points (9.2%) in weighted F1-score for seizure type classification. Ablation studies show that our graph-based modeling approach significantly outperforms existing CNN or LSTM models, and that self-supervision helps further improve the model performance. Moreover, we find that self-supervised pre-training substantially improves model performance on combined tonic seizures, a low-prevalence seizure type. Furthermore, our model interpretability analysis suggests that our model is better at identifying seizure regions compared to an existing CNN. In summary, our graph-based modeling approach integrates domain knowledge about EEG, sets a new state-of-the-art for seizure detection and classification on a large public dataset (5,499 EEG files), and provides better ability to identify seizure regions.

Introduction

Up to 10% of people will experience at least one seizure over their lifetime[1]. Seizures commonly result from a brain insult such as traumatic brain injury, cerebrovascular disease, or drug intoxication[1–3], and are among the most common neurological emergencies[4,5]. Seizures can also be chronic as in the case of epilepsy[6], a neurological disease that affects 50 million people worldwide[7]. In addition to physical problems such as injuries, loss of consciousness, and disturbances of movement resulting from seizures, people with epilepsy also experience higher rates of psychological conditions, social stigma, and unemployment[8,9].

For patients, definitive detection of a seizure is only the first step in their diagnostic journey. An important subsequent step is to classify seizures into finer-grained types, such as focal seizure versus generalized seizure or absence seizure versus tonic seizure, in order to identify targeted therapies, comorbidities, and prognosis[10,11].

Electroencephalography (EEG) plays a critical role in this process of seizure detection and classification. Common scalp EEG (referred to as simply "EEG" hereafter) measures the electrical activity of the brain using multiple electrodes placed on an individual's scalp. Clinically, EEG-based seizure detection and seizure type classification are performed by a board-certified EEG reader who visually examines a patient's EEG signals over time periods ranging from hours to days. This manual analysis of EEG signals is extremely resource- and time-consuming, and is often impractical for real-time continuous seizure monitoring in hospital units where an expert EEG reader is often not available around the clock[12]. Automated algorithms for seizure detection and seizure type classification can greatly mitigate these limitations, accelerate epilepsy diagnosis, and improve outcomes for acutely ill patients.

Substantial research has been conducted on developing EEG-based automated seizure detection algorithms. Earlier studies extracted hand-crafted features and applied classical machine learning models, such as support vector machines and Bayesian networks, for automated seizure detection[13–20]. More recent studies leveraged deep learning methods—such as convolutional neural networks (CNNs)[21–25], hybrid architectures combining CNNs and recurrent neural networks (RNNs)[26,27], or graph neural networks (GNNs)[28]—to automatically learn useful features for seizure detection. Importantly, deep learning-based methods have shown promise across populations ranging from adults to neonates[24,29].

In contrast to automated seizure detection, automated seizure type classification remains largely unaddressed, and many open questions remain outstanding. First of all, although there is a standard International League Against Epilepsy (ILAE) seizure classification system, this classification system makes use of (the first) clinically observed symptoms as the basis for classification. Indeed, to our knowledge, there is no consensus seizure nomenclature based purely on EEG morphology. One major limitation in producing such a consensus classification system and using it for automated seizure type classification has been the limited availability of large EEG databases. In 2018, Shah et al.[30] released the Temple University Hospital Seizure Detection Corpus (TUSZ), an EEG database consisting of 5,612 EEG files, with 3,050 annotated seizures from clinical recordings. Though this corpus does not capture every known seizure type, it provides an important step towards automating seizure classification by providing a large open set of baseline examples of seizures. The Temple University Hospital group additionally proposed their own seizure classification scheme using terms from 1981 ILAE schema[31] by dividing the seizures in their database into eight seizure types (e.g. focal non-specific seizure, generalized non-specific seizure, etc).

The release of TUSZ database has motivated several studies on developing automated seizure type classification algorithms. Based upon the results so far[32–36], the seizure type classification problem appears to be substantially more difficult than seizure detection alone. Investigators have used classical machine learning methods[32,33], deep-learning-based CNNs[34–36], or hybrid CNN-RNN models[37]. In most

of these prior studies on seizure detection or seizure type classification, the same patients' EEGs were used for both training and evaluating the algorithms, which does not demonstrate the model's ability to generalize to unseen patients. Like some other prior studies[16,26–28,35], we aim to evaluate our model's generalizability to unseen patients' data, which is motivated by real-world scenarios where new patients' data are continuously being collected and analyzed.

We note that prior state-of-the-art deep learning-based methods for automated seizure detection[24,29] or seizure type classification[35,36] used CNNs that assume image-like structures in EEG signals or their spectrograms. However, CNN-based approaches do not reflect the natural geometry of EEG electrodes. Standard EEG electrode placements (see Figure 1a for the layout of electrodes in the 10-20 international EEG configuration) represent a graph structure rather than a two-dimensional image structure in EEG signals. Thus, a more natural way of modeling EEGs is representing them as graphs. A few studies have explored modeling EEGs as graphs in other domains, such as EEG-based identification of visual stimulus[38], emotion recognition[39,40], and distinguishing neurological patients' normal EEGs from healthy individuals' EEGs[41]. To the best of our knowledge, only two studies have examined modeling EEGs with graphs for seizure detection or localization. Varatharajah et al.[42] designed a factor graph model based on significant domain knowledge to encode the spatial and temporal relationships in EEGs for seizure onset localization. Covert et al.[28] proposed a Temporal Graph Convolutional Network that performs feature extraction that is localized and shared over time and space for automated seizure detection. Nevertheless, these two studies were only limited to seizure localization or detection on internal datasets, and did not use self-supervised pre-training approaches. Our study is the first to date that investigates seizure detection and seizure type classification with graph-based modeling and self-supervised pre-training on the public TUSZ dataset.

Since expert-labeled medical data are extremely expensive to obtain, label-free methods that improve the model's performance, such as self-supervised pre-training, have gained particular interest in medical domains. Self-supervised pre-training involves pre-training a model on large amounts of unlabeled data using an alternative objective that is related to the target tasks of interest. In medical domains, self-supervised pre-training has been widely applied to medical image analysis[43–48]. To the best of our knowledge, only three studies have investigated self-supervised learning for EEG signals. Banville et al.[49] proposed three self-supervised pre-training tasks—contrastive predictive coding[50], relative positioning, and temporal shuffling—to learn useful representations of EEG signals for sleep staging from EEGs. Inspired by the SimCLR[51] framework from computer vision, Mohsenvand et al.[52] handcrafted a few data augmentation methods for EEG signals, and presented a framework for emotion recognition, normal/abnormal classification, and sleep-stage scoring from EEGs. Kostas et al.[53] adopted a self-supervised contrastive pre-training method that was initially used for automatic speech recognition, and fine-tuned the model on downstream EEG-based sleep stage classification and brain-computer interface tasks. However, these self-supervised learning studies did not model EEGs as graphs, neither did they address seizure detection or classification.

In this study, we present a graph-based modeling method with self-supervised pre-training for automated seizure detection and seizure type classification. We model EEG signals as graphs, where the nodes in the graphs represent EEG electrodes, and the edges represent either the physical distance or connectivity between the electrodes. More precisely, we propose two different graph structures: (a) modeling the edges by the pairwise physical distance between EEG electrodes in the standard 10-20 system (see Figure 1b), or (b) modeling the edges by the pairwise cross-correlation between the signals in the EEG channels (see Figure 1c). We adapt Diffusion Convolutional Recurrent Neural Network (DCRNN)[54], a recurrent neural network with graph diffusion convolutions, to model the spatiotemporal dependencies in EEGs. Our model achieves an Area Under the Receiver Operating Characteristic curve (AUROC) value of 0.824 for seizure detection, and a weighted F1-score of 0.710 for seizure type classi-

fication. Compared to state-of-the-art CNN and Long Short-Term Memory (LSTM) models, our graph model without self-supervisied pre-training provides 1.2 points increase (1.5%) in AUROC for seizure detection, and 2.4 points increase (3.5%) in weighted F1-score for seizure type classification. Furthermore, we propose a novel self-supervised pre-training strategy that pre-trains the model for predicting the next 12-s EEGs given the previous 60-s/12-s EEGs. This pre-training strategy leverages the large amounts of unlabeled EEGs already present in our dataset, and thus does not require additional data or labels for pre-training. Compared to our non-pre-trained model, self-supervised pre-training significantly improves the model performance by 5.1 points (6.2%) in AUROC for seizure detection, and 3.9 points (5.5%) in weighted F1-score for seizure type classification. Consequently, by leveraging both the graph structure of the data and self-supervised pre-training strategy, our graph model significantly outperforms previous state-of-the-art CNN and LSTM models by 6.3 points (7.8%) in AUROC for seizure detection and 6.3 points (9.2%) in weighted F1-score for seizure type classification. Importantly, our model with self-supervised pre-training achieves an AUROC value of 0.875 on seizure detection, matching the performance of the previous state-of-the-art CNN[24] (AUROC = 0.88) that was pre-trained using supervised learning on a dataset that is an order of magnitude larger than TUSZ. In addition, detailed analyses of the individual seizure type predictions suggest that our model significantly improves the performance on low-prevalence seizure classes. Specifically, our model without self-supervised pre-training performs 10 points (12.0%) better than the baselines in prediction accuracy on the low-prevalence absence seizure class, and our self-supervised pre-training provides 48 points increase (184.6%) in prediction accuracy on the low-prevalence combined tonic seizure class. Lastly, to understand how our model reaches the final predictions, we examine methods for model interpretability by occluding small segments of EEGs and observing the change in the model output. We quantitatively show that our graph-based modeling approach, particularly when combined with self-supervised pre-training, is better at identifying seizure regions than an existing CNN.

Our main contributions are summarized as follows. First, our graph-based modeling approach significantly improves model performance on automated seizure detection and seizure type classification on the public TUSZ dataset compared to existing methods. Second, by leveraging large amounts of unlabeled EEGs, our self-supervised pre-training strategy further improves the model performance without the need of additional data or labels. Third, our graph neural network is flexible to the EEG electrode placements and sequence lengths, and thus can be easily applied to EEGs with different electrode geometry or length. Finally, our graph-based modeling approach, particularly when combined with self-supervised pre-training, substantially improves the ability to identify seizure regions over an existing CNN, suggesting that our model is better at capturing spatiotemporal dependencies in EEG signals.

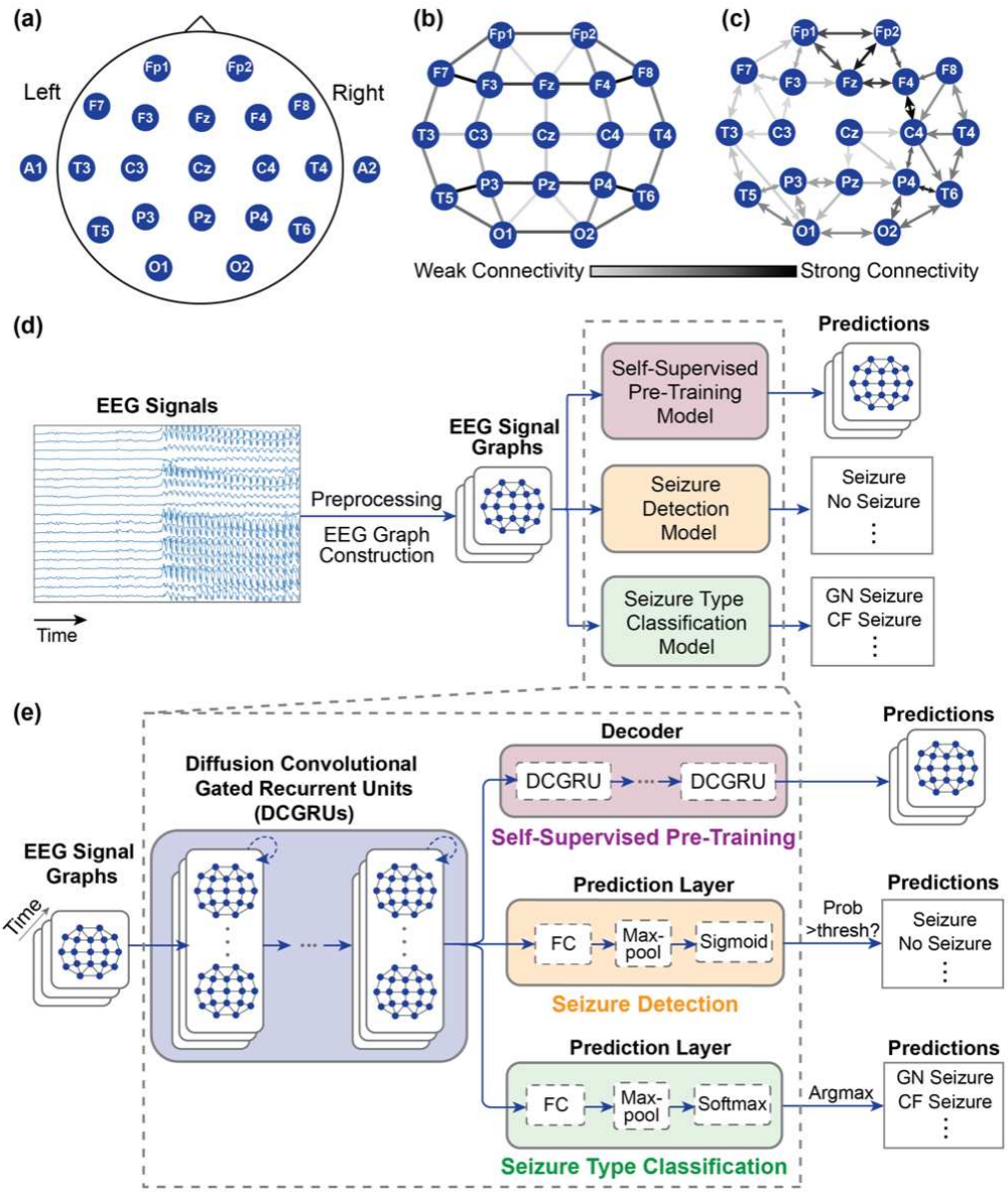


Figure 1: **(a) Electrode placement in the international 10-20 system. (b) Distance EEG graph.** Edge weights are computed based on the pairwise Euclidean distance between EEG electrodes in the international 10-20 system. This results in an undirected, weighted graph. Self-edges are not shown here. Darker (lighter) edges indicate stronger (weaker) connectivity. **(c) An example correlation EEG graph.** Edge weights are computed based on the pairwise cross-correlation between preprocessed signals in EEG channels. Only the edges of the top-3 neighbors of each node are kept, resulting in a directed, weighted graph. **(d) Overview of our methods.** We preprocess the EEG clips and construct the EEG graphs (see Methods) before inputting them to the corresponding model for seizure detection, seizure type classification, or self-supervised pre-training. **(e) Details of our seizure detection, seizure type classification, and self-supervised pre-training models.** We employ Diffusion Convolutional Gated Recurrent Units (DCGRUs)[54], a variant of Gated Recurrent Units where the matrix multiplications are replaced by graph diffusion convolutions. For seizure detection and seizure type classification, the models consist of several layers of DCGRUs and a prediction layer. For self-supervised pre-training, the model consists of an encoder and a decoder, each having the same number of DCGRUs. Abbreviation: FC, fully connected layer.

Results

Overview

In this study, we model EEG signals as graphs, where the nodes in the graphs represent EEG electrodes, and the edges represent the physical distance or connectivity between electrodes. We investigate two novel methods of computing the edge weights. The first approach is based on the pairwise physical distance between the electrodes according to the standard 10-20 system (referred to as "distance graph" hereafter), which results in a universal undirected, weighted graph for all the EEG clips (Figure 1b). The second approach is based on the pairwise cross-correlation between the preprocessed signals (referred to as "correlation graph" hereafter), which results in a unique directed, weighted graph for each EEG clip (Figure 1c). We adapt a graph neural network, Diffusion Convolutional Recurrent Neural Network (DCRNN), to model the spatiotemporal dependencies in EEG signals. Following the experimental setting of a prior study[24], we examine our model's capability of fast detection/classification over 12-s EEG clips, as well as slow detection/classification over 60-s EEG clips. The seizure detection task is to predict whether or not a seizure event exists within each clip, the seizure type classification task is to predict the seizure type given a seizure EEG clip, and the self-supervised pre-training task is to predict the next 12-s EEG clip given the previous 12-s (or 60-s) EEG clip. For additional details on EEG graph construction, model architecture, and model training procedures, see Methods.

Data

We use the public TUSZ v1.5.2 dataset in this study. To ensure that the held-out test set only consists of unseen patients, we exclude five patients from the test set who exist in both of the TUSZ v1.5.2 official train and test sets. This results in 4,599 EEG files (18.9% seizure) in the TUSZ train set and 900 EEG files (25.6% seizure) in the TUSZ test set (Table 1a). Supplementary Figure 1 shows the distributions of EEG file duration, patients' age and sex, and four seizure classes used in our study.

For seizure detection and self-supervised pre-training, we use all the EEGs with and without seizure events in TUSZ train/test set, and adopt a non-overlapping sliding window approach to obtain the EEG clips (see Methods). Whereas for seizure type classification, we use only EEGs with seizure events in TUSZ train/test set. We exclude myoclonic seizures because there are only two myoclonic seizures in the dataset. Furthermore, because simple partial seizures and complex partial seizures are focal seizures characterized by consciousness and impaired consciousness respectively[55,56], they are not distinguishable from other types of focal seizures by EEG signals alone. Hence, we combine focal non-specific seizures, simple partial seizures, and complex partial seizures to form a class named combined focal (CF) seizures. In addition, since tonic and tonic-clonic seizures are rare in the dataset (there are only 18 tonic seizures and 30 tonic-clonic seizures in TUSZ train set), and tonic-clonic seizures always start with a tonic phase[55], in order to maximize the training data for tonic/tonic-clonic seizures, we combine tonic-clonic seizures with tonic seizures to form a class named combined tonic (CT) seizures. This results in four seizure classes in total (Table 1b): combined focal (CF) seizures, generalized non-specific (GN) seizures, absence (AB) seizures, and combined tonic (CT) seizures. We obtain one 12-s (or 60-s) EEG clip for each seizure event.

We further randomly split the official TUSZ train set by patients into train and validation sets with a ratio of 90%/10% for model training and hyperparameter tuning respectively, and we hold-out the TUSZ test set for model evaluation. The train, validation, and test sets consist of EEGs from distinct patients. The resulting sample sizes are shown in Table 1c. Details on data preprocessing can be found in the Methods section.

	EEG Files (% Seizure)	Seizure Events	Patients (% Seizure)	Total Duration (% Seizure)
TUSZ Train Set	4,599 (18.9%)	2,377	592 (34.1%)	45174.72 min (6.3%)
TUSZ Test Set	900 (25.6%)	522	45 (77.8%)	9031.58 min (9.8%)

(a) Train and test sets of TUSZ v1.5.2.

	CF Seizure		GN Seizure		AB Seizure		CT Seizure	
	Seizure Events	Patients	Seizure Events	Patients	Seizure Events	Patients	Seizure Events	Patients
TUSZ Train Set	1,868	148	409	68	50	7	48	11
TUSZ Test Set	297	24	114	11	49	5	61	4

(b) Number of seizure events and patients in each seizure class in our study.

	EEG Clip Length (Secs)	Train Set		Validation Set		Test Set	
		EEG Clips (% Seizure)	Patients (% Seizure)	EEG Clips (% Seizure)	Patients (% Seizure)	EEG Clips (% Seizure)	Patients (% Seizure)
Pre-training & Seizure Detection	60-s	38,613 (9.3%)	530 (34.0%)	5,503 (11.4%)	61 (36.1%)	8,848 (14.7%)	45 (77.8%)
	12-s	196,646 (6.9%)	531 (33.9%)	28,057 (8.7%)	61 (36.1%)	44,959 (10.9%)	45 (77.8%)
Seizure Type Classification	60-s & 12-s	1,925 (100.0%)	179 (100.0%)	450 (100.0%)	22 (100.0%)	521 (100.0%)	34 (100.0%)

(c) Number of EEG clips and patients in the train, validation, and test sets in our study.

Table 1: **Overview of the data used in this study. (a)** Summary of data in the official train and test sets of Temple University Seizure Corpus (TUSZ) v1.5.2. To ensure the test set only consists of unseen patients, we exclude five patients from the test set who also appear in the train set. **(b)** Number of seizure events and patients in TUSZ v1.5.2 in the following four seizure classes used in our study: combined focal (CF), generalized non-specific (GN), absence (AB), and combined tonic (CT) seizures. The CF class is formed by combining focal non-specific, simple partial, and complex partial seizures. The CT class is formed by combining tonic and tonic-clonic seizures. We exclude myoclonic seizures because there are only two myoclonic seizure events in the dataset. **(c)** Sample sizes after data preprocessing. We further randomly split the TUSZ v1.5.2 official train set into train and validation sets by 90%/10% across patients. Hence, our train, validation, and test sets consist of EEGs from distinct patients. For self-supervised pre-training and seizure detection tasks, we use all the available EEG files (see (a)) in each train/validation/test set. Because we adopt a non-overlapping sliding window approach (see Methods) for self-supervised pre-training and seizure detection tasks, the number of samples for 60-s and 12-s EEG clips are different. For seizure type classification, we use only EEG files containing seizure events (see (b)) in each train/validation/test set. The number of samples for 60-s and 12-s clips for seizure type classification are the same because only one EEG clip is obtained from each seizure event.

Graph neural network performance

To validate that our graph-based modeling approach improves seizure detection and seizure type classification compared to traditional CNN- or RNN-based methods, we compare our DCRNN (without self-supervised pre-training) to two baseline models: (a) Dense-CNN[24], a previous state-of-the-art CNN with densely connected inception modules, and (b) Long Short-Term Memory (LSTM)[57], a recurrent neural network with feedback connections. Furthermore, we compare our model's performance to results of two other CNNs reported in the literature[35,36]. Following standards established by recent studies on seizure detection[24,28,29] and seizure type classification[33,35,37], we compare the AUROC scores on the test set between the baselines and our DCRNN models for seizure detection, and weighted F1-scores for seizure type classification. As a fair comparison to the CNN-based ensemble model SeizureNet[35], we trained and evaluated our DCRNN using a new train and test split: the same 3-fold patient-wise split

reported by Asif et al.[35]. Asif et al. did not aggregate the seizure types in the TUSZ dtaset and used the original labels (excluding myoclonic seizures), resulting in seven seizure types instead of four in our other seizure classification experiments. The seven seizure classes are focal non-specific seizure, simple partial seizure, complex partial seizure, generalized non-specific seizure, absence seizure, tonic seizure, and tonic-clonic seizure. Table 2a (third to sixth rows) shows the performance of our DCRNN model (without self-supervised pre-training) and the baselines for seizure detection and seizure type classification. Table 2b shows the performance of our DCRNN model and SeizureNet on seven-class seizure type classification. The mean and standard deviation were calculated from five runs with different random seeds for each model.

For seizure detection, both of our distance/correlation graph-based DCRNN models (without self-supervised pre-training) perform on par with Dense-CNN, and significantly better ($p < 0.05$) than LSTM on 12-s/60-s EEG clips. Note that the seizure/non-seizure ratios are different between 60-s and 12-s EEG clips in the test set, and that there are many more training samples for 12-s seizure detection (see Table 1c), which likely results in higher AUROC scores for 12-s EEG clips than 60-s EEG clips across all the models (without self-supervised pre-training). See Supplementary Table 1 for additional model evaluation scores (F1-score, Area Under the Prevision-Recall curve, sensitivity, and specificity) for seizure detection.

For seizure type classification, both of our distance/correlation graph-based DCRNN models significantly outperform ($p < 0.05$) the two baseline models on 12-s EEG clips, and perform on par with the baselines on 60-s EEG clips. Notably, our correlation graph-based DCRNN achieves a weighted F1-score of 0.710, providing 13.4 points (23.3%) increase over Dense-CNN and 5.8 points (8.9%) increase over LSTM. Compared to the weighted F1-score reported for SeizureNet[35], our 60-s correlation graph-based model provides 3 points (4.8%) improvement on seven-class seizure type classification (Table 2b). Note that seizure type classification in Asif et al.[35] is performed independently on 1-s EEG clips obtained by sliding a 1-s window (0 or 0.5-s overlap) on EEGs with seizures, and one seizure event may be associated with multiple EEG clips and predictions. In contrast, our seizure type classification is performed on 12-s/60-s EEG clips from different seizure events.

Moreover, to compare our model's performance on seizure type classification to Iesmantas et al.[36], we report AUROC values of individual seizure classes in Supplementary Table 2. For GN and AB seizure classes that are common between our studies, our DCRNN (without self-supervised pre-training) achieves substantially higher AUROC scores than Iesmantas et al., providing 3.5 points increase (4.5%) for GN seizures, and 26.3 points increase (36.5%) for AB seizures.

In addition, we conduct an ablation study to examine the effectiveness of Fourier transform in our preprocessing step for our graph-based methods. We find that frequency-domain inputs with Fourier transform result in a large and significant improvement of model performance compared to time-domain inputs on both seizure detection and seizure type classification. See Supplementary Results and Supplementary Table 3 for details.

Self-supervised pre-training improves graph neural network performance

To assess the effectiveness of self-supervised pre-training in improving our model's performance on seizure detection and seizure type classification, we compare the results of our DCRNN model without and with self-supervised pre-training. Specifically, we pre-train our DCRNN model for predicting next 12-s preprocessed EEG clips given the previous 60-s/12-s preprocessed EEG clips. We fine-tune our models for seizure detection and seizure type classification by initializing the model parameters with the self-supervised model weights.

Table 2a (seventh and eighth rows) shows the model performance with self-supervised pre-

Model	Seizure Detection AUROC (mean ± std)		Seizure Type Classification Weighted F1-score (mean ± std)	
	12-s	60-s	12-s	60-s
Dense-CNN	0.812 ± 0.014	0.796 ± 0.014	0.576 ± 0.101	0.626 ± 0.073
LSTM	0.786 ± 0.014	0.715 ± 0.016	0.652 ± 0.019	0.686 ± 0.020
DCRNN Correlation Graph Without Pre-training	**0.812 ± 0.012***	**0.804 ± 0.015***	**0.710 ± 0.023***	0.701 ± 0.030
DCRNN Distance Graph Without Pre-training	**0.824 ± 0.020***	**0.793 ± 0.022***	**0.703 ± 0.025***	0.690 ± 0.035
DCRNN Correlation Graph With Pre-training	**0.861 ± 0.005**$^{+}$	**0.850 ± 0.014**$^{+}$	0.723 ± 0.017	**0.749 ± 0.017**$^{+}$
DCRNN Distance Graph With Pre-training	**0.866 ± 0.016**$^{+}$	**0.875 ± 0.016**$^{+}$	**0.746 ± 0.024**$^{+}$	**0.749 ± 0.028**$^{+}$

(a) Seizure detection and seizure type classification results.

Model	Seven-Class Seizure Type Classification Weighted F1-score (mean ± std)
SeizureNet[35]	0.620^{++}
DCRNN Correlation Graph, 12-s	0.619 ± 0.006
DCRNN Distance Graph, 12-s	0.585 ± 0.006
DCRNN Correlation Graph, 60-s	**0.650 ± 0.008**
DCRNN Distance Graph, 60-s	0.606 ± 0.009

(b) Seven-class seizure type classification results for models without self-supervised pre-training.

Table 2: **(a) Results of seizure detection and seizure type classification for the baselines, DCRNN models without and with self-supervised pre-training.** Our graph-based DCRNN models perform on par with or significantly better than the baseline Dense-CNN and LSTM on both seizure detection and seizure type classification tasks. Moreover, self-supervised pre-training significantly improves the performance of our DCRNN models on both tasks. Mean and standard deviation (std) were obtained by five runs with different random seeds for each model. Results with statistically significant improvement are highlighted in **bold**. *DCRNN model (without self-supervised pre-training) performs significantly better than at least one of the baseline models ($p < 0.05$). $^{+}$DCRNN model with self-supervised pre-training performs significantly better than DCRNN model without self-supervised pre-training ($p < 0.05$). **(b) Comparison between our DCRNN models and SeizureNet[35] on seven-class seizure type classification.** Our correlation graph-based DCRNN model trained and evaluated on 60-s EEG clips substantially outperforms the CNN-based ensemble model, SeizureNet. $^{++}$Score reported in Asif et al.[35]. Note that SeizureNet performs classification independently on 1-s EEG clips, where different EEG clips may come from the same seizure events.

training. Compared to DCRNN without self-supervised pre-training, models with self-supervised pre-training perform significantly better on both seizure detection and seizure type classification tasks. Notably, distance graph-based DCRNN with self-supervised pre-training achieves an AUROC of 0.875 for seizure detection on 60-s EEG clips, which is 8.2 points (10.3%) higher than model without self-supervised pre-training. Importantly, the 0.875 AUROC score matches the performance of Dense-CNN reported in Saab et al.[24] that was pre-trained using supervised learning on a dataset that is an order of magnitude larger than TUSZ. Similarly, distance graph-based DCRNN with self-supervised pre-training achieves a weighted F1-score of 0.749 for seizure type classification on 60-s clips, which is 5.9 points (8.6%) higher compared to its non-pre-trained counterpart. These results suggest the effectiveness of our self-supervised pre-training strategy. Importantly, the pre-trained model weights are shared by both seizure detection and seizure type classification models, which indicates that our self-supervised pre-training provides good model initialization that generalizes across tasks.

Figure 2a shows the Receiver Operating Characteristic (ROC) curves for our median DCRNN models and the median baseline models for seizure detection. At a low false positive rate (FPR) operating

point (e.g. 20% FPR), our DCRNN models with self-supervised pre-training achieve higher true positive rates than the baselines.

Figure 2b shows the normalized confusion matrices that are averaged across five random runs for our distance graph-based DCRNN model and the baselines for seizure type classification on 12-s EEG clips. We summarize two main observations here. First, our DCRNN model without self-supervised pre-training performs significantly better than the baselines on the low-prevalence AB seizure class ($p < 0.05$). Second, with self-supervised pre-training, the prediction accuracy of the low-prevalence CT seizure is substantially improved, and is significantly better than the baselines and the non-pretrained DCRNN model ($p < 0.05$).

Surprisingly, despite being a majority class, many of the GN seizures are misclassified (Figure 2b). To examine the reason of the misclassifications, a board-certified neurologist (CLM) manually analyzed the EEG signals for the GN seizures in the test set that are misclassified by all of our four best models (i.e. 12-s/60-s distance/correlation graph with self-supervised pre-training). We find that many of the misclassified GN seizures are in fact focal seizures but are mislabeled as GN seizures in the dataset. See Supplementary Results and Supplementary Figures 2–3 for details.

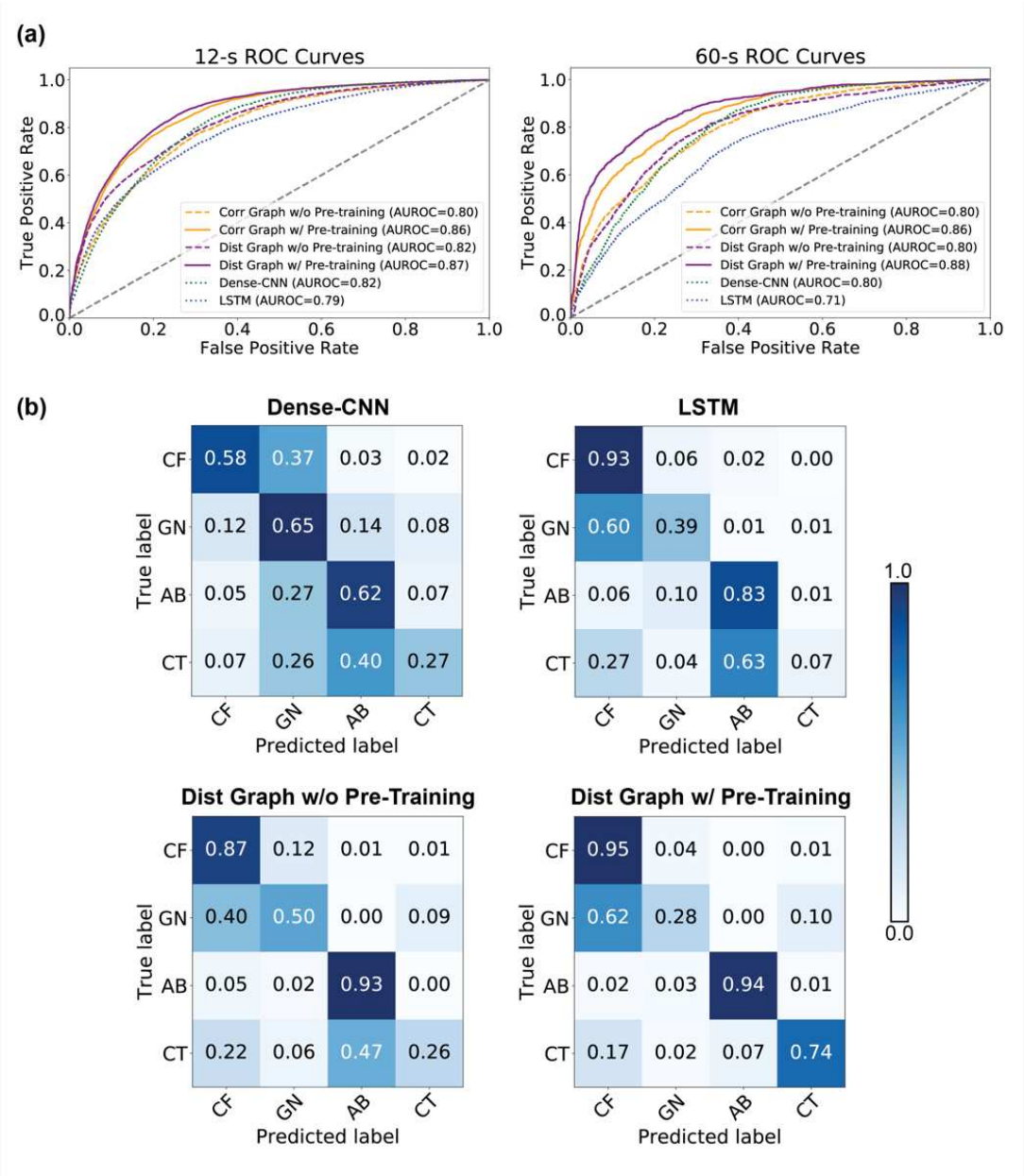


Figure 2: **(a) Receiver Operating Characteristic (ROC) curves of our median DCRNN models and the median baselines for seizure detection.** At a low false positive rate operating point (e.g. 20% FPR), our models with self-supervised pre-training achieve higher true positive rates compared to the baseline models. **(b) Confusion matrices (averaged across five runs with different random seeds) for the baselines and our DCRNN model (distance graph) without and with self-supervised pre-training for seizure type classification on 12-s EEG clips.** Each row of the confusion matrices is normalized by dividing by the number of examples in the corresponding class, such that each row sums up to one. Without self-supervised pre-training, our graph-based model (bottom left panel) significantly outperforms the baseline Dense-CNN and LSTM on the low-prevalence absence (AB) seizure class ($p < 0.05$). Furthermore, with self-supervised pre-training (bottom right panel), the prediction accuracy on the low-prevalence combined tonic (CT) seizure class is significantly improved ($p < 0.05$). Abbreviations: CF, combined focal seizure; GN, generalized non-specific seizure; AB, absence seizure; CT, combined tonic seizure.

Graph neural network is flexible to EEG sequence lengths

Unlike CNNs, our model is independent of the EEG clip length because of its recurrent architecture. To validate that this property leads to better generalization performance across different clip lengths compared to CNNs, we test the trained DCRNN models on the clip length that is different from which the models are trained on. Specifically, we test on 12-s clips with the model trained on 60-s clips, and vice versa. Because Dense-CNN model parameters are dependent on the clip length, we can only test the 60-s trained Dense-CNN on 12-s clips by imputing the remaining 48-s clips with values such as 0's. Supplementary Table 4 shows the performance when the models are tested on different clip lengths. For seizure detection tested on 12-s clips, Dense-CNN's performance decreases by 28.9 points in AUROC, whereas DCRNN's performance decreases by 8.3 points. For seizure type classification tested on 12-s clips, Dense-CNN's performance decreases by 11.6 points in weighted F1-score, while DCRNN's performance decreases by 2.8 points. This suggests that the flexibility to EEG clip lengths is an advantage of our method compared to CNN-based methods.

Graph neural network provides better ability to identify seizure regions

Visualizing spatial/temporal regions important for model prediction

To understand the spatial and/or temporal regions that are important for our model's predictions, we perform the following model interpretability analyses. See Methods for details on the experimental setup.

For seizure detection, an EEG clip may partially overlap with a seizure event or may be completely seizure free. We hypothesize that both the spatial and temporal dimensions can be important for our model's prediction of seizures. Therefore, we occlude 1-s EEG signals in one channel at a time and compute the percentage change in the model output compared to the original output. A larger (smaller) drop in the output indicates that the occluded region is more (less) important for predicting seizure. Figure 3 shows examples of percentage changes in four correctly predicted 60-s EEG clips in the test set, which are obtained from our correlation graph-based DCRNN with self-supervised pre-training. The the percentage changes are overlaid on raw EEG signals and are shown in the bipolar double banana montage. We refer to this visualization as saliency maps in the remainder of this paper.

From Figure 3, we observe that temporal regions with seizures have higher saliency than temporal regions without seizures. Moreover, for focal seizures (Figures 3a–3b), high saliency is localized in the more abnormal EEG channels. Specifically, in Figure 3a, the most abnormal regions are Fp2-F8, Fp2-F4, Cz-Pz, T4-T6, and T6-O2, which correspond to the bright regions in the saliency map. In Figure 3b, while our manual examination of the entire EEG signal (i.e. before and after this EEG clip) suggests that the seizure may start in regions involving Fp1, F3, F7, and P3, the saliency map highlights P3-O1 while ignoring Fp1-F7, Fp1-F3, and F3-C3. This is likely because of the ongoing artifact (spikes in Figure 3b) in Fp1-F7, Fp1-F3, and F3-C3. In contrast, high saliency for generalized seizures (Figures 3c–3d) is more diffused across the channels, and the regions with high-frequency artifacts have lower saliency than regions without artifacts. Such patterns are expected given that generalized seizures occur in all areas of the brain. Supplementary Figure 4 shows the same examples in the referential average montage.

Furthermore, we performed the same occlusion experiment using the baseline Dense-CNN. Supplementary Figures 5–6 show the saliency maps from Dense-CNN for the same examples visualized in the double banana and average montages respectively. Compared to our DCRNN model, the high saliency regions from Dense-CNN do not localize in any spatial or temporal seizure regions. This suggests that our graph-based model is better at identifying seizure regions than Dense-CNN, which is intuitive given that our model more accurately encodes the natural geometric structure or connectivity of the

EEG electrodes. In the next section, we provide more quantitative analyses of the model's ability to identify seizure regions.

For seizure type classification, since a seizure event spans the entire EEG clip length, we hypothesize that the spatial differences in the EEG signals are more important for the model's prediction of the seizure types than temporal differences. To leverage our model's advantage that it is independent of the number of EEG electrodes, we completely drop one EEG channel at a time and compute the percentage change in the model output. Supplementary Figure 7 shows examples of percentage changes in correctly classified 60-s EEG clips in the test set. We observe that for CF seizures, spatial regions with high saliency correspond to where the seizures are localized. See Supplementary Results and Supplementary Figures 7–8 for details.

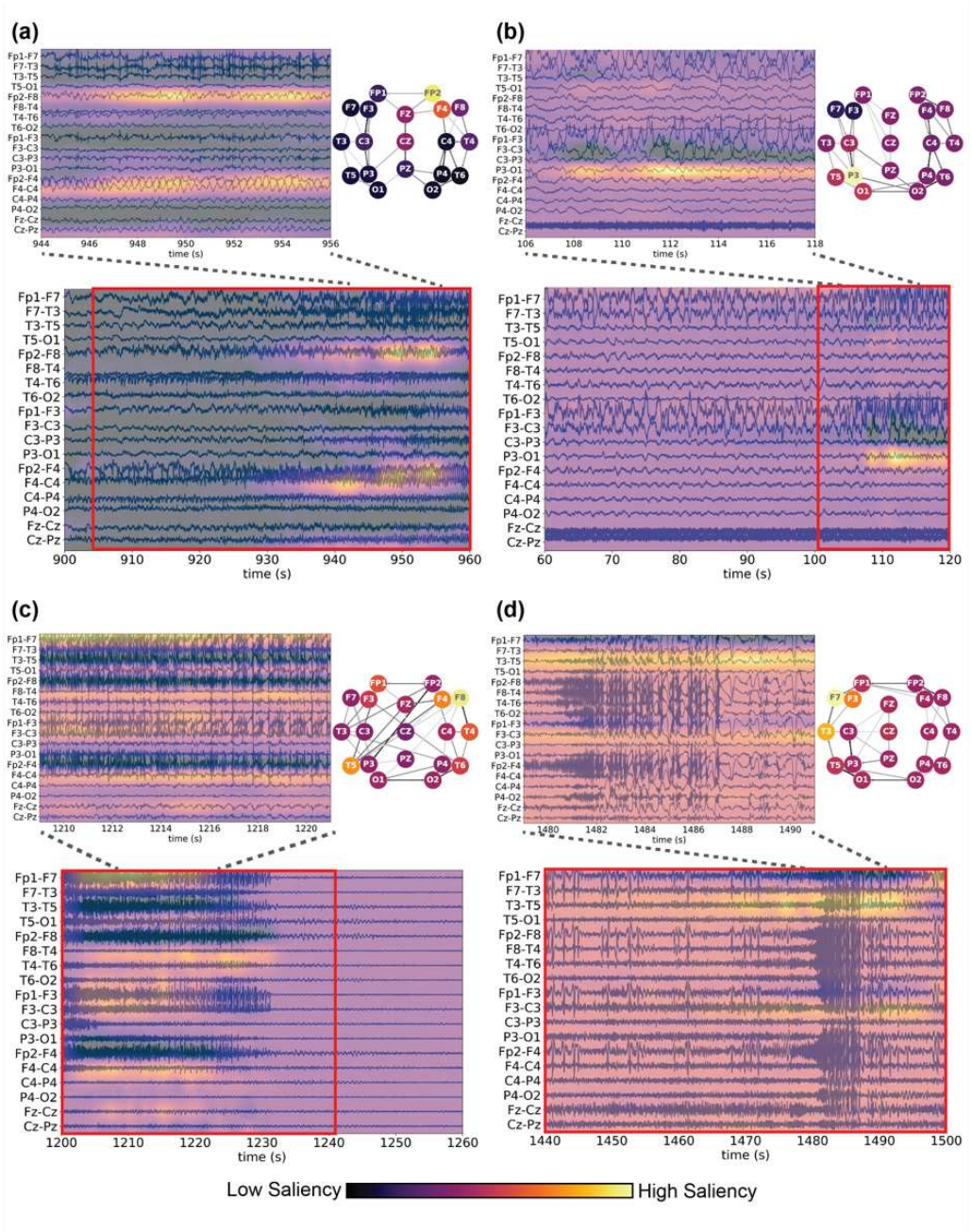


Figure 3: **Example saliency maps for seizure detection obtained from our correlation graph-based DCRNN model shown in the *double banana* montage for (a)–(b) focal seizures, and (c)–(d) generalized seizures.** The saliency maps are obtained by occluding 1-s EEG signals in one EEG channel at a time and calculating the percentage change in the model output (see Methods). To visualize the percentage changes in the double banana montage, we subtract the values between the corresponding channels in the montage. Brighter (Darker) color indicates higher (lower) saliency. In each subfigure, the bottom panel shows the 60-s saliency map, the top left panel shows the saliency map zoomed in to the 12-s most salient time periods, the top right panel shows the saliency values that are averaged along the time dimension and overlaid on the corresponding correlation graph, and the red boxes indicate the duration of the seizures. We observe that temporal regions with seizures have higher saliency than temporal regions without seizures. Moreover, for focal seizures (a–b), high saliency is localized in the more abnormal brain regions. In contrast, for generalized seizures (c–d), high saliency is more diffused across the brain regions. The examples shown here are correctly predicted 60-s EEG clips in the test set. Note that the values within a saliency map are normalized, and thus should not be compared across different saliency maps.

Quantifying model ability to identify seizure regions

In addition to visually analyzing the saliency maps, we provide a quantitative analysis of the model's capability to identify seizure regions. Specifically, we compare the high saliency regions (i.e. normalized saliency values > 0.5) to the event-based seizure annotations in the TUSZ dataset (i.e. annotations that indicate where and when seizures occur). For each EEG clip and its corresponding saliency map, we define a coverage metric that measures how many true seizure regions in both the spatial and temporal dimensions are detected, and a localization metric that measures how many detected spatial and temporal seizure regions are true seizure regions. Note that the coverage and localization metrics are analogous to recall and precision, respectively, for binary classification problems. We only performed this analysis for seizure detection because many of the seizure EEG clips also consist of non-seizure temporal regions, which provides more insights about the coverage and localization measures. See Methods and Equations 12–13 for more details on the definitions of the coverage and localization metrics.

Figures 4a–4b show the distributions of coverage scores for baseline Dense-CNN, correlation graph-based DCRNN without and with self-supervised pre-training, where the coverage scores are computed on each model's correctly predicted 60-s test EEG clips. Similarly, Figures 4c–4d show the distributions of localization scores for the three models on their correctly predicted 60-s test EEG clips. We observe that compared to Dense-CNN, our DCRNN models have many more saliency maps with high coverage scores (e.g. coverage > 0.9 in Figures 4a–4b), suggesting that our DCRNN models are better at identifying relevant seizure regions than Dense-CNN. Similarly, there are many more saliency maps with high localization scores for DCRNN, which suggests that compared to Dense-CNN, a larger proportion of seizure regions that are detected by our model are true seizure regions. Notably, self-supervised pre-training improves coverage and localization scores for both focal and generalized seizures, where the most notable improvement is on the localization scores for focal seizures (e.g. localization > 0.9 in Figure 4c).

Here, we are particularly interested in the localization metric for focal seizures. The ability of not only detecting focal seizure regions, but also localizing the regions as precisely as possible would be highly desirable for a seizure detection model from a clinical standpoint. In Figures 4e–4f, we show example saliency maps of focal seizures in the test set whose localization scores are above 0.9, which are obtained from our correlation graph-based DCRNN model with self-supervised pre-training. Because the event-based annotations in the TUSZ dataset are based on the Temporal Central Parasagittal (TCP) montage, we visualize the saliency maps and raw EEG signals in the TCP montage here. Only high saliency regions (i.e. normalized saliency value > 0.5) are shown here, and the red boxes indicate the annotated spatial and temporal regions of the focal seizures. We can see that the high saliency regions overlap well with the annotated seizure regions.

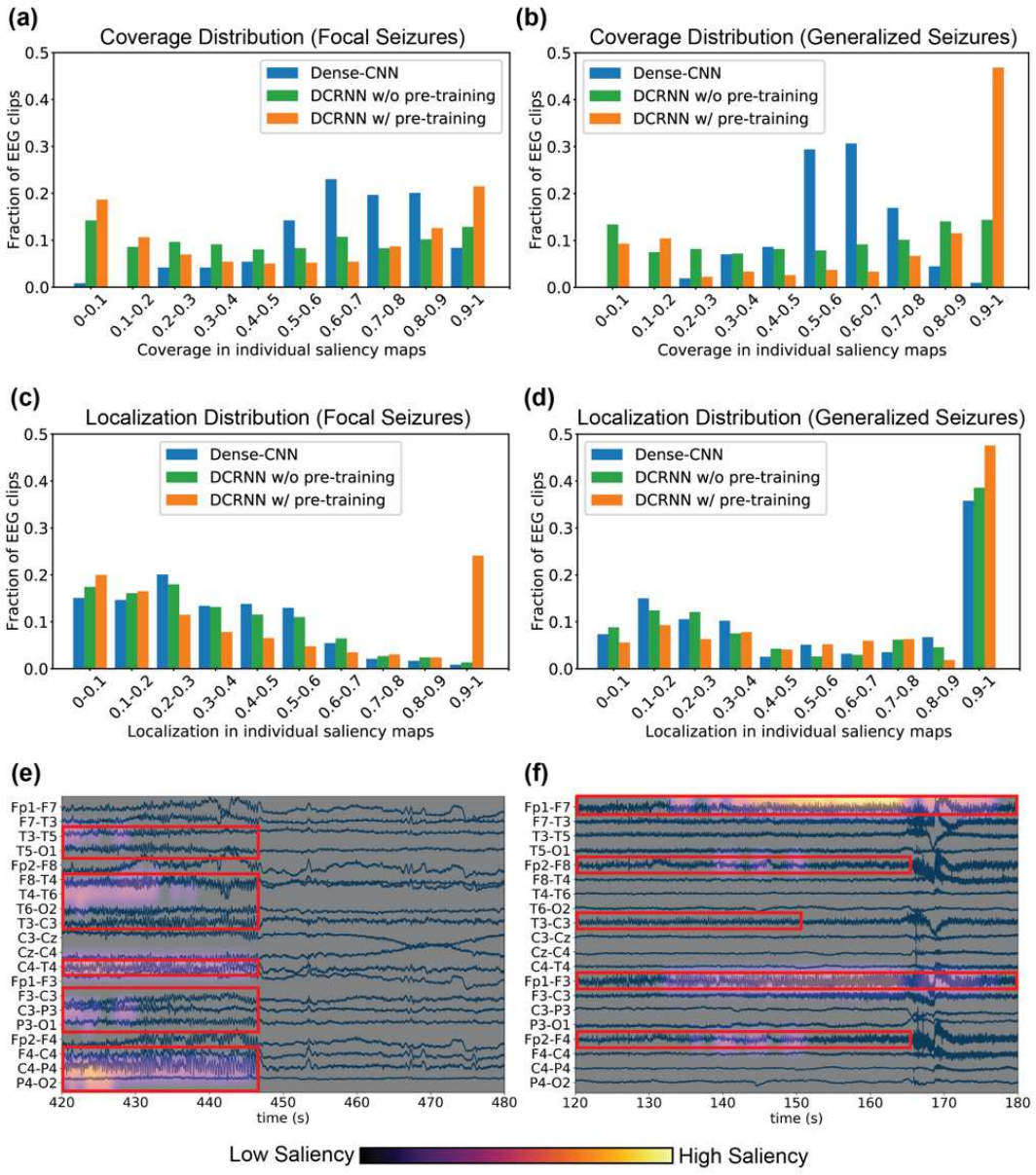


Figure 4: **(a)–(b) Distributions of coverage scores for baseline Dense-CNN, correlation graph-based DCRNN without and with self-supervised pre-training on each model's correctly predicted 60-s EEG clips for (a) focal seizures, and (b) generalized seizures.** There are many more saliency maps with high coverage scores (e.g. coverage > 0.9) from our DCRNN model without self-supervised pre-training than baseline Dense-CNN. Moreover, self-supervised pre-training substantially improves the coverage scores. **(c)–(d) Distributions of localization scores for the three models on their correctly predicted 60-s EEG clips for (c) focal seizures, and (d) generalized seizures.** There are many more saliency maps with high localization scores (e.g. localization > 0.9) from our DCRNN models compared to baseline Dense-CNN. Notably, self-supervised pre-training substantially improves the localization scores, especially for focal seizures. **(e)–(f) Examples of saliency maps with focal seizure whose localization scores are above 0.9 shown in the TCP montage.** Only high saliency regions (i.e. normalized saliency value > 0.5) are shown here, and the red boxes indicate the annotated spatial and temporal regions of the focal seizures.

Discussion

In this study, we present a novel graph-based deep learning method for seizure detection and seizure type classification. Our method has two primary technical innovations: graph-based EEG representation and self-supervised pre-training. In addition, as part of our work, we examine the classification scheme put forth by the Temple University Hospital group[30,58] and alter it in order to make it better fit the underlying classes of seizures based upon their morphology. Combining these approaches, we achieve a new state-of-the-art on both seizure detection and seizure type classification on the public TUSZ dataset.

Our DCRNN model architecture holds the promise of better capturing the underlying geometry of EEG sensors in the same way that CNNs are thought to provide a better inductive bias for images[59–61]. In addition to substantially improved performance on the public TUSZ datset, our method provides better ability to identify seizure regions than a previous state-of-the-art CNN. The saliency maps in Figures 3–4 demonstrate a focus on EEG abnormalities that much better match human approaches to EEG interpretation. Even more importantly, because these saliency maps can help explain to physicians why the algorithm makes its predictions, we believe doctors are much more likely to be able to interpret and trust the algorithm's results, which is a critical step towards clinical use of deep learning models[62,63].

The DCRNN architecture also offers increased flexibility. With its recurrent sequential architecture, the model can easily be adapted to different EEG lengths without retraining. Moreover, graph-based modeling approach allows for easy adaptation to different EEG sensor geometries. Different scalp montages and even the diverse electrode geometries of intracranial and stereo EEGs can be easily modeled. Hence, our method offers the possibility for an automated algorithm to approach the same level of flexibility as human expert EEG interpreters. Furthermore, it is possible to incorporate different methods to determine graph connectivity. While we show that a graph structure based upon Euclidean distance provides an intuitive basis for the graph topology, we additionally show that a graph topology based upon cross-correlation also gives improved performance over the baselines. In addition, this correlation-based graph topology has the following advantages. First, it can be used even when the physical locations of electrodes are unknown. Second, it captures connectivity beyond spatial information of EEG electrodes, which could also be particularly useful for intracranial EEGs where the connectivity between electrodes is not as dependent on their spatial alignment. In the future, one can investigate additional connectivity measures that have been shown to be particularly useful for predicting seizure onset or epileptogenic zones[64,65].

Self-supervised learning is a crucial element in the success of our method. By leveraging large amounts of unlabeled EEGs already present in the dataset, our self-supervised pre-training strategy significantly improves our model performance on both seizure detection and seizure type classification without the need of additional data or labels (Table 2a and Figure 2). Of note, our DCRNN model with self-supervised pre-training achieves an AUROC value of 0.875, matching the performance of Dense-CNN reported in Saab et al.[24] (AUROC = 0.88) that was pre-trained using supervised learning on a different dataset which is an order of magnitude larger than TUSZ. Moreover, our self-supervised pre-training method is general and can be applied to other sequence-to-sequence model architectures, and would be particularly useful for real-world clinical settings, where abundant, unlabeled data are continuously being generated in daily clinical routines.

There exist several limitations in this study that can be improved in the future. First, we represent the EEGs in each channel by the amplitudes of one-second Fast Fourier Transform in this study. While this preprocessing step leads to siginificantly improved model performance compared to time-domain EEG signals (see Supplementary Table 3), more sophisticated single-channel representations could be investigated in the future to improve the performance and interpretability of the method.

Second, our detailed error analysis on the misclassified GN seizures suggests that the TUSZ

dataset, annotated by a team of trained undergraduates, may contain errors in seizure type labels. A prior study[24] showed that a seizure detection model trained on large data with noisy labels greatly outperformed the model trained on small, clean data. Therefore, the noisy labels in TUSZ could still be valuable. It is also important to note that given the difficulties of EEG interpretation, there is likely intrinsic variability in seizure classification even among experts. Hence, to improve the quality of model evaluation, a test set with accurate labels that are approved by a consensus of experts would be helpful.

Third, this study is both made possible and limited by the TUSZ dataset and its seizure classification scheme[30,58]. After extensive visual analysis of the TUSZ dataset aided by early versions of our model, we modified the original classification by eliminating myoclonic seizure class that has limited representation in the dataset (i.e. only two myoclonic seizures) and combining tonic and tonic-clonic seizures to maximize the training data of tonic seizures. Furthermore, we merged focal non-specific, simple partial, and complex partial seizures into a combined focal seizure class. We justify the decision of forming combined focal seizure class based upon avoiding using clinical observations to determine electroencephalographic seizure classifications. In general, clinical seizure classification is a difficult task even when the patient's video recording and clinical history are available, and direct examination of the clinical relationship to scalp EEG appearance during a seizure is complicated. Specifically, the classes "simple partial seizures" and "complex partial seizures" are distinguished by a clinical behavior, consciousness, whose underlying neurobiology is neither easy to measure nor fully understood[66]. Therefore, while it is of profound interest to determine which EEG seizure patterns lead to impaired consciousness, we advocate strongly in this work that a system for EEG classification of seizures needs to be based as much as possible on EEG waveform features alone. To support future advances in the field, we believe that developing such an electrographic seizure classification system based on EEG morphology will be of broad interest to the research community. We imagine that a system analogous to the American Clinical Neurophysiology Society terminology for ICU EEG waveforms[67,68] would arise out of this effort. This might be done by consensus opinion or by testing competing classifications on their clinical usefulness. Furthermore, future work on EEG datasets from different populations and age groups is needed to evaluate the model's generalizability across populations and age groups.

In conclusion, we present a novel graph-based modeling method with self-supervised pre-training for automated seizure detection and seizure type classification. Our study is the first to date that examines both seizure detection and seizure type classification on the public TUSZ dataset. Our novel self-supervised pre-training approach significantly improves the model performance, and sets new state-of-the-arts on both tasks. In addition, our graph-based modeling approach is flexible to the EEG electrode placements and sequence lengths, and provides better ability to identify seizure regions. We also propose a more practical seizure type classification scheme based on EEG morphology. Lastly, our study suggests that graph-based methods are a more effective alternative to CNN-based methods for modeling EEG signals, and opens exciting future opportunities to build graph-based EEG representations for neurology applications.

Methods

Data preprocessing

In this study, we use the public TUSZ v1.5.2 dataset, which is a subset of the larger Temple University Hospital EEG Corpus[58] with seizure type labels. We use only signals from the following 19 electrodes in the 10-20 international EEG configuration: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz. Because the EEG signals are sampled at different frequencies in the dataset, we resample them to the same frequency of 200 Hz using the "resample" function in SciPy python package[69].

We perform the following preprocessing steps to obtain EEG clips in the frequency domain and their corresponding labels. First, for seizure detection, we use all the EEGs with and without seizure events in each of the TUSZ train/test set. We obtain the EEG clips by sliding a 12-s (or 60-s) window over the EEG signals without overlaps, and ignore the last window if it is shorter than the clip length. The label for each clip is $y = 1$ if at least one seizure event occurs within this clip, otherwise $y = 0$.

Second, for seizure type classification, we use only EEGs with seizure events in each of the TUSZ train/test set. We obtain one 12-s (or 60-s) EEG clip for each seizure event starting at 2-s before the annotated seizure onset time, where a 2-s offset accounts for tolerance in the annotations. If a seizure event is shorter than 12-s (or 60-s), the EEG clip is truncated at the end of the seizure to prevent a clip from having multiple seizure types. The label for each clip is the index of the corresponding seizure class, i.e. $y \in \{0, 1, 2, 3\}$, which corresponds to {CF, GN, AB, CT}.

Third, for self-supervised pre-training, we use all the EEGs in each of the TUSZ train/test set, and adopt the same sliding window approach as seizure detection to obtain the EEG clips.

Fourth, for each EEG clip in each of the seizure detection/seizure type classification/self-supervised pre-training tasks, we perform the following preprocessing steps to transform the signals in time domain to frequency domain: (a) slide a t second window over the EEG clip without overlap, where t is the time step size for networks involving recurrent layers such as LSTM or DCRNN; (b) apply Fast Fourier Transform (FFT) to each t second window, and retain the log amplitudes of the non-negative frequency components similar to prior studies[28,35,37]; (c) z-normalize the EEG clip with respect to the mean and standard deviation of the training data. Because EEG clips for seizure type classification may have variable lengths due to short seizures, we pad the clips with 0's to facilitate model training in batches. We use $t = 1$ second as a natural choice of the time step size. After preprocessing, each EEG clip can be denoted as $\mathbf{X} \in \mathbb{R}^{T \times N \times M}$, where $T = 12$ (or $T = 60$) represents the clip length, $N = 19$ represents the number of EEG channels/electrodes, and $M = 100$ represents the feature dimension after the aforementioned Fourier transform.

Representing EEGs as graphs

In this study, we propose representing an EEG clip as a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, where $\mathcal{V}$ denotes the set of nodes (i.e. 19 EEG electrodes/channels), $\mathcal{E}$ denotes the set of edges, and $\mathbf{W}$ is the adjacency matrix. Following our preliminary study[70], we examine the following two methods of constructing the EEG graph.

- **Distance graph**: The edge weight $\mathbf{W}_{ij}$ between electrodes v_i and v_j is obtained by applying a thresholded Gaussian kernel[71] to the pairwise Euclidean distance between v_i and v_j according to the standard 10-20 system:

$$\mathbf{W}_{ij} = exp\left(- \frac{dist(v_i, v_j)^2}{\sigma^2} \right) \text{ if } dist(v_i, v_j) \leq \kappa, \quad \text{otherwise } 0 \tag{1}$$

where $dist(v_i, v_j)$ is the pairwise Euclidean distance between electrodes v_i and v_j according to the 10-20 system, σ is the standard deviation of the distances, and κ is the threshold for sparsity. This method results in a universal undirected, weighted graph for all the input EEG clips. We experimented with various values of κ and found that $\kappa = 0.9$ resulted in a reasonable graph (e.g. no long-range connection), and thus used $\kappa = 0.9$ in all the experiments.

- **Correlation graph**: The edge weight $\mathbf{W}_{ij}$ between electrodes v_i and v_j is the absolute value of the normalized cross-correlation between the preprocessed signals in v_i and v_j. To introduce sparsity to the graph, we limit the number of neighbors each node can have. For each node, only the edges whose weights are among the top-τ of the node are kept:

$$\mathbf{W}_{ij} = |\mathbf{X}_{:,i,:} \star \mathbf{X}_{:,j,:}| \text{ if } v_j \in \mathcal{N}(v_i), \quad \text{otherwise } 0 \tag{2}$$

where $\star$ represents the normalized cross-correlation, and $\mathcal{N}(v_i)$ represents the top-τ neighbors of v_i. This method results in a unique directed, weighted graph for each input EEG clip. We treat τ, the number of neighbors whose edges are kept for each node, as a hyperparameter in our experiments.

Figure 1b shows the distance graph, and Figure 1c shows an example correlation graph with $\tau = 3$. In both figures, darker (lighter) edges indicate stronger (weaker) connections, and self-edges are not shown for better visualization.

Model architecture

We use Diffusion Convolutional Recurrent Neural Network[54] (DCRNN), a recurrent neural network with graph diffusion convolutions, to model the spatiotemporal dependencies in EEG signals. DCRNN was initially developed for traffic forecasting, where the dynamics of traffic flow were modeled as a diffusion process. Similarly, we can also model the spatial dependency in EEG signals as a diffusion process, because one could expect an electrode to be influenced more by electrodes in its proximity (measured by distance or correlation). Specifically, the diffusion process is characterized by a bidirectional random walk on a directed graph $\mathcal{G}$, which results in the following diffusion convolution[54]:

$$\mathbf{X}_{:,m} *_{\mathcal{G}} f_\theta = \sum_{k=0}^{K-1} \big(\theta_{k,1}(\mathbf{D}_O^{-1}\mathbf{W})^k + \theta_{k,2}(\mathbf{D}_I^{-1}\mathbf{W}^T)^k\big)\mathbf{X}_{:,m} \text{ for } m \in \{1, ..., M\} \tag{3}$$

where $\mathbf{X} \in \mathbb{R}^{N \times M}$ is the preprocessed EEG clip at time step $t \in \{1, ..., T\}$ with N nodes and M features, f_θ is the convolution filter with parameters $\theta \in \mathbb{R}^{K \times 2}$, $\mathbf{D}_O$ and $\mathbf{D}_I$ are the out-degree and in-degree diagonal matrices of the graph respectively, $\mathbf{D}_O^{-1}\mathbf{W}$ and $\mathbf{D}_I^{-1}\mathbf{W}^T$ are the state transition matrices of the outward and inward diffusion processes respectively, and K is the number of maximum diffusion steps.

For undirected graphs, the diffusion convolution is similar to ChebNet spectral graph convolution[72] up to a constant scaling factor, and thus can be computed using stable Chebyshev polynomial bases as follows[54]:

$$\mathbf{X}_{:,m} *_{\mathcal{G}} f_\theta = \mathbf{\Phi}\Big(\sum_{k=0}^{K-1} \theta_k \mathbf{\Lambda}^k\Big)\mathbf{\Phi}^T\mathbf{X}_{:,m} = \sum_{k=0}^{K-1} \theta_k \mathbf{L}^k \mathbf{X}_{:,m} = \sum_{k=0}^{K-1} \tilde{\theta}_k T_k(\tilde{\mathbf{L}})\mathbf{X}_{:,m} \quad \text{for } m \in \{1, ..., M\} \tag{4}$$

where $T_0(x) = 1, T_1(x) = x$, and $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ for $k \geq 2$ are the bases of the Chebyshev polynomial, $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-\frac{1}{2}} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T$ is the normalized graph Laplacian, and $\tilde{\mathbf{L}} = \frac{2}{\lambda_{max}}\mathbf{L} - \mathbf{I}$ is the scaled graph Laplacian mapping the eigenvalues from $[0, \lambda_{max}]$ to $[-1, 1]$.

In our study, we use Equation 3 for the directed correlation graphs, and Equation 4 for the undirected distance graph.

Next, we employ Gated Recurrent Units[73] (GRUs), a variant of RNN with gating mechanism, to model the temporal dynamics in EEGs. Specifically, the matrix multiplications in GRUs are replaced with diffusion convolutions (or ChebNet spectral graph convolutions for undirected graphs), allowing spatiotemporal modeling of EEG signals. Equations 5–8 show the resulting GRU with diffusion convolutions (or DCGRU):

$$\mathbf{r}^{(t)} = \sigma\big(\mathbf{\Theta}_{r*\mathcal{G}}[\mathbf{X}^{(t)}, \mathbf{H}^{(t-1)}] + \mathbf{b}_r\big) \tag{5}$$

$$\mathbf{u}^{(t)} = \sigma\big(\mathbf{\Theta}_{u*\mathcal{G}}[\mathbf{X}^{(t)}, \mathbf{H}^{(t-1)}] + \mathbf{b}_u\big) \tag{6}$$

$$\mathbf{C}^{(t)} = tanh\big(\mathbf{\Theta}_{C*\mathcal{G}}[\mathbf{X}^{(t)}, (\mathbf{r}^{(t)} \odot \mathbf{H}^{(t-1)}] + \mathbf{b}_C\big) \tag{7}$$

$$\mathbf{H}^{(t)} = \mathbf{u}^{(t)} \odot \mathbf{H}^{(t-1)} + (1 - \mathbf{u}^{(t)}) \odot \mathbf{C}^{(t)} \tag{8}$$

where $\mathbf{X}^{(t)}$, $\mathbf{H}^{(t)}$ denote the input and output of DCGRU at time step t respectively, σ denotes Sigmoid function, $\odot$ represents the Hadamard product, $\mathbf{r}^{(t)}$, $\mathbf{u}^{(t)}$, $\mathbf{C}^{(t)}$ denote reset gate, update gate and candidate at time step t respectively, $*\mathcal{G}$ denotes the diffusion convolution (or ChebNet spectral graph convolution), $\mathbf{\Theta}_r$, $\mathbf{b}_r$, $\mathbf{\Theta}_u$, $\mathbf{b}_u$, $\mathbf{\Theta}_C$ and $\mathbf{b}_C$ are the weights and biases for the corresponding convolutional filters.

Finally, the seizure detection and seizure type classification models consist of several DCGRU layers, followed by a fully connected layer with ReLU non-linearity, and max-pooling over nodes. We also apply dropout[74] to the fully connected layer during training to mitigate overfitting. Lastly, a Sigmoid or a Softmax activation is applied to the output of the max-pooling layer for seizure detection and seizure type classification respectively. Figure 1e illustrates the model architecture.

Self-supervised pre-training task

The self-supervised pre-training task is to predict the next w second preprocessed EEG clips given the previous preprocessed 12-s/60-s EEG clip. We experimented with different lengths of w and found that $w = 12$ resulted in low regression loss on the validation set given 12-s/60-s clips, and thus used $w = 12$ for all the self-supervised pre-training experiments.

We employ a sequence-to-sequence model architecture[75] for self-supervised pre-training, where both the encoder and decoder consist of several DCGRU layers (Figure 1e).

Baselines

We compare our DCRNN model to two baselines: (a) Dense-CNN[24], an existing CNN with densely connected inception modules, and (b) Long Short-Term Memory (LSTM)[57], a RNN with feedback connections. For Dense-CNN, we employ the same model architecture as described in Saab et al.[24]. For LSTM, we have the number of LSTM layers and hidden units the same as the number of DCGRU layers and hidden units in our DCRNN model. The baselines were trained and fine-tuned on the same preprocessed EEG clips in the train and validation sets, and the results are reported on the same held-out test set.

Model training procedure

Training for all models was accomplished using the Adam optimizer[76] in PyTorch on one NVIDIA Titan RTX GPU. Model parameters were randomly initialized for models without self-supervised pre-training,

and were initialized with the pre-trained weights for models with self-supervised pre-training. We performed the following hyperparameter search on the validation set: (a) initial learning rate within range [5e-5, 1e-3]; (b) $\tau \in \{2, 3, 4\}$, the number of neighbors to keep for each node in the correlation graphs; (c) the number of DCGRU layers within range $\{2, 3, 4, 5\}$ and hidden units within range $\{32, 64, 128\}$; (d) the maximum diffusion step $K \in \{2, 3, 4\}$; (e) dropout probability in the last fully connected layer. We used a batch size of 40 EEG clips, the maximum possible across all models and baselines on a single Titan RTX GPU. The hyperparameters were selected based on the best performance on the validation sets, and are detailed in the next sections. We used the cosine annealing learning rate scheduler[77] in PyTorch for all model training. We ran five runs with different random seeds for all models to obtain the mean and standard deviation values reported in this manuscript. During training, we performed data augmentation to reduce overfitting by (a) randomly scaling the amplitude of the raw EEG signals by a scale between 0.8 and 1.2, and (b) randomly reflecting the signals along the midline. In all experiments, model training was early stopped when the validation loss did not decrease for five consecutive epochs.

Model training for seizure detection

For seizure detection, the plentiful negative examples in the train set were undersampled such that the train set had 50% positive examples, which resulted in 27,292 training examples for 12-s clips and 7,188 training examples for 60-s clips.

We used binary cross-entropy as the loss function to train the seizure detection models. Equation 9 shows the binary cross-entropy loss for one example, where p denotes the probability output from the Sigmoid activation, and y denotes the ground truth label for that example.

$$L_{\text{binary}} = -\Big(y \log(p) + (1 - y) \log(1 - p)\Big) \tag{9}$$

The models were trained for 100 epochs with an initial learning rate of 1e-4. The top-3 neighbors' edges were kept for each node for correlation graphs, the maximum number of diffusion step was 2, and the dropout probability was 0 (i.e. no dropout). The model consists of two DCGRU layers with 64 hidden units, resulting in 168,641 trainable parameters for the distance graph and 280,769 trainable parameters for the correlation graphs. Model training for seizure detection took about 20-min for 12-s EEG clips, and about 30-min for 60-s EEG clips.

To obtain the final seizure/non-seizure prediction, we performed decision threshold search on the validation set. More specifically, to balance between precision and recall scores, we selected the decision threshold that results in the highest F1-score on the validation set. When evaluating the models on the test set, EEG clips with probabilities above this decision threshold are predicted as seizures, while clips with probabilities below this decision threshold are predicted as non-seizures.

Model training for seizure type classification

For seizure type classification, we used multi-class cross-entropy as the loss function during training. Equation 10 shows the multi-class cross-entropy loss for a single example, where $\mathcal{C}$ denotes the set of seizure classes, p_c denotes the probability output from the Softmax activation for class c, and y_c denotes the ground truth label for class c for that example.

$$L_{\text{multi-class}} = -\sum_{c \in \mathcal{C}} y_c \log(p_c) \tag{10}$$

The models were trained for 60 epochs with an initial learning rate of 3e-4. The top-3 neighbors' edges were kept for each node for correlation graphs, the maximum number of diffusion step was 2, and the

dropout probability was 0.5. The model consists of two DCGRU layers with 64 hidden units, resulting in 168,836 trainable parameters for the distance graph and 280,964 trainable parameters for the correlation graphs. Model training for seizure type classification took about 3-min for 12-s EEG clips, and about 7-min for 60-s EEG clips.

Model training for self-supervised pre-training

For self-supervised pre-training, we used mean absolute error (MAE) as the loss function. Mathematically, let $\mathbf{x} = x_1, ..., x_n$ be the ground truth values in a preprocessed EEG clip, and $\hat{\mathbf{x}} = \hat{x}_1, ..., \hat{x}_n$ be the predicted values, the MAE between $\mathbf{x}$ and $\hat{\mathbf{x}}$ is defined as follows:

$$L_{\text{MAE}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i| \tag{11}$$

The models were trained for 350 epochs with an initial learning rate of 5e-4. The top-3 neighbors' edges were kept for each node for correlation graphs, and the maximum number of diffusion step was 2. The model consists of three DCGRU layers with 64 hidden units in both the encoder and decoder, resulting in 417,572 trainable parameters for the distance graph and 690,980 trainable parameters for the correlation graphs. Model training for self-supervised prediction took about 10-h for 12-s EEG clips, and about 24-h for 60-s EEG clips.

Model interpretability

To understand how our DCRNN models reach the final predictions, we examine the spatial and/or temporal regions important for the model predictions. First, for seizure detection, an EEG clip may partially overlap with a seizure event or may be completely seizure free. We hypothesize that both the spatial and temporal dimensions can be important for our model's prediction of seizures. Therefore, we zero-fill 1-s EEG signals in one channel at a time and compute the percentage change in the model's output logit (i.e. model output before Sigmoid activation) with respect to the original non-occluded output. Furthermore, we reverse the signs of the percentage changes so that larger values indicate higher importance, and we scale the values in each clip to [0, 1] using the minimum and maximum values in that clip. This results in a matrix $\mathbf{M} \in \mathbb{R}^{N \times T}$, where $N = 19$ is the number of EEG channels, $T = 12$ (or $T = 60$) is the clip length, and $\mathbf{M}_{ij}$ indicates the relative change in the model output when the j-th second EEG clip in the i-th channel is occluded. A larger $\mathbf{M}_{ij}$ indicates that the occluded region is more important for predicting seizure, and vice versa. We visualize $\mathbf{M}$ as saliency maps overlaid on EEG signals. See Figure 3 and Supplementary Figure 4 for examples.

Next, for seizure type classification, since a seizure event spans the entire EEG clip, we hypothesize that the spatial differences in the EEG signals are more important for the model's prediction of the seizure types than temporal differences. To leverage our model's advantage that it is independent of the number of EEG channels, we completely drop one EEG channel at a time in the input EEG clips as well as the graph structure, and compute the percentage change in the model output logit with respect to the original output. This results in an array $\mathbf{M'} \in \mathbb{R}^{N}$, where $N = 19$ is the number of EEG channels, and $\mathbf{M}'_i$ indicates the relative change in the model output when the i-th channel is dropped. In order to visualize $\mathbf{M'}$ as saliency maps overlaid on EEG signals, the value for each EEG channel is replicated along the time dimension. See Supplementary Figures 7–8 for examples.

Finally, in addition to visually analyzing the aforementioned saliency maps, we quantify the model's capability of identifying seizure regions based on the saliency maps. Note that we only perform this analysis on saliency maps obtained from seizure detection models. This is because in seizure detec-

tion, many of the seizure EEG clips also consist of non-seizure temporal regions, which provides more insights about the model's ability to identify seizure regions. We leverage the event-based seizure annotations (i.e. annotations that indicate where and when seizures occur) available in the TUSZ dataset. We define the coverage metric as the fraction of the number of high saliency values that overlaps with annotated seizure regions to the number of annotated seizure regions, where high saliency values are normalized saliency values that are above 0.5. Intuitively, the coverage metric quantifies how many true seizure regions are detected. Let $\mathbf{M}^{\text{annot}} \in \mathbb{R}^{N \times T}$ be the matrix representing the event-based seizure ground truth annotations, where $\mathbf{M}_{ij}^{\text{annot}} = 1$ if there is a seizure at the j-th second in the i-th channel, otherwise $\mathbf{M}_{ij}^{\text{annot}} = 0$. Let $\mathbf{M} \in \mathbb{R}^{N \times T}$ be the saliency map obtained by occluding 1-s EEG clip from a seizure detection model described above. Mathematically, the coverage metric is defined as follows:

$$\text{Coverage} = \frac{\sum_{i,j} \mathbb{1}\{\mathbf{M}_{ij} > 0.5\}\mathbf{M}_{ij}^{\text{annot}}}{\sum_{i,j} \mathbf{M}_{ij}^{\text{annot}}} \tag{12}$$

Similarly, we define the localization metric as the fraction of the number of high saliency values that overlaps with annotated seizure regions to the number of high saliency values. Hence, the localization metric quantifies how many detected seizure regions are true seizure regions. Mathematically, the localization metric is defined as follows:

$$\text{Localization} = \frac{\sum_{i,j} \mathbb{1}\{\mathbf{M}_{ij} > 0.5\}\mathbf{M}_{ij}^{\text{annot}}}{\sum_{i,j} \mathbb{1}\{\mathbf{M}_{ij} > 0.5\}} \tag{13}$$

Performance metrics

In seizure detection, true positives (TP) are correct seizure predictions, true negatives (TN) are correct non-seizure predictions, false positives (FP) are incorrect seizure predictions, and false negatives (FN) are incorrect non-seizure predictions. True positive rate (TPR, also called recall or sensitivity), false positive rate (FPR), true negative rate (TNR, also called specificity), and precision are derived from these quantities as follows:

$$TPR \text{ (recall or sensitivity)} = \frac{TP}{TP + FN} \tag{14}$$

$$FPR = \frac{FP}{FP + TN} \tag{15}$$

$$TNR \text{ (specificity)} = \frac{TN}{TN + FP} \tag{16}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

Following recent studies on seizure detection[24,28,29], we assess seizure detection performance using the Area Under the Receiver Operating Characteristic curve (AUROC) as the main metric. The ROC curve is a plot showing TPR (i.e. sensitivity or recall) versus FPR at different decision thresholds, and thus the AUROC score provides a holistic measure for seizure detection. In addition to reporting AUROC scores (Table 2a), we also visualize the ROC curves for different models (Figure 2a).

Moreover, we report F1-score, the Area Under the Precision-Recall curve (AUPR), sensitivity, and specificity for seizure detection (Supplementary Table 1). F1-score is the harmonic mean of precision and recall defined as follows:

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{18}$$

The Precision-Recall curve is a plot showing precision versus recall at different decision thresh-

olds, and the area under the curve is summarized as the AUPR score.

Next, following recent studies on seizure type classification[13,33,35], we assess the seizure type classification performance of the different models using weighted F1-score as the main metric. In multi-class classification problems, weighted F1-score is the weighted average of F1-scores for all the classes, where the weight is the fraction of the number of examples in each class:

$$\text{Weighted F1-score} = \sum_{c \in \mathcal{C}} w_c \times \text{F1-score}(c) \tag{19}$$

where $\mathcal{C}$ denotes the set of seizure classes, w_c denotes the fraction of examples in class c, and F1-score(c) denotes the F1-score for class c. Besides weighted F1-score which shows the aggregated performance on all the seizure classes, we visualize the confusion matrices that shows the prediction accuracy for individual classes in Figure 2b, and report AUROC scores for individual seizure classes in Supplementary Table 2.

Statistical analyses

To compare the performance between models in this study, two-sample t-tests are performed on the results obtained from five runs with different random seeds. A statistical significance threshold (α) of 0.05 is used for all the reported tests in this manuscript.

Data Availability

The Temple University Hospital Seizure Detection Corpus used in this study is publicly available at `https://www.isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml`.

In addition, we provide the five patients' IDs who are excluded from the test set as well as the list of focal seizures in the test set that are mislabeled as generalized non-specific seizures at `https://github.com/tsy935/eeg-gnn-ssl`.

Code Availability

The code used in this study is publicly available at `https://github.com/tsy935/eeg-gnn-ssl`.

Acknowledgements

This work is supported by Wu Tsai Neurosciences Institute Neurotranslate Grant.

Author Contributions

ST, JAD, KS, and CLM designed the study. ST, XZ, and QH conducted the experiments. ST wrote the manuscript, and JAD, KS, FD, DLR, and CLM critically revised the manuscript for important intellectual content.

Competing Interests

The authors report no competing interests.

References

1. W Allen Hauser and Ettore Beghi. First seizure definitions and worldwide incidence and mortality. *Epilepsia*, 49:8–12, 2008.

2. Carla Bentes, Hugo Martins, Ana Rita Peralta, Carlos Casimiro, Carlos Morgado, Ana Catarina Franco, Ana Catarina Fonseca, Ruth Geraldes, Patrícia Canhão, Teresa Pinho e Melo, et al. Post-stroke seizures are clinically underestimated. *Journal of Neurology*, 264(9):1978–1985, 2017.

3. Hsien-Yi Chen, Timothy E Albertson, and Kent R Olson. Treatment of drug-induced seizures. *British journal of clinical pharmacology*, 81(3):412–419, 2016.

4. Micheal Strein, John P Holton-Burke, LaTangela R Smith, and Gretchen M Brophy. Prevention, treatment, and monitoring of seizures in the intensive care unit. *Journal of Clinical Medicine*, 8(8): 1177, 2019.

5. Nicholas S Abend, Daniel H Arndt, Jessica L Carpenter, Kevin E Chapman, Karen M Cornett, William B Gallentine, Christopher C Giza, Joshua L Goldstein, Cecil D Hahn, Jason T Lerner, et al. Electrographic seizures in pediatric icu patients: cohort study of risk factors and mortality. *Neurology*, 81(4):383–391, 2013.

6. Robert S Fisher, Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J Helen Cross, Christian E Elger, Jerome Engel Jr, Lars Forsgren, Jacqueline A French, Mike Glynn, et al. Ilae official report: a practical clinical definition of epilepsy. *Epilepsia*, 55(4):475–482, 2014.

7. WHO. Epilepsy. `https://www.who.int/news-room/fact-sheets/detail/epilepsy`, june 2019.

8. Robert S Fisher, Barbara G Vickrey, Patricia Gibson, Bruce Hermann, Patricia Penovich, Ann Scherer, and Steven Walker. The impact of epilepsy from the patient's perspective i. descriptions and subjective perceptions. *Epilepsy research*, 41(1):39–51, 2000.

9. Michael Patrick Kerr. The impact of epilepsy on patients' lives. *Acta Neurologica Scandinavica*, 126: 1–9, 2012.

10. Ingrid E Scheffer, Samuel Berkovic, Giuseppe Capovilla, Mary B Connolly, Jacqueline French, Laura Guilhoto, Edouard Hirsch, Satish Jain, Gary W Mathern, Solomon L Moshé, et al. Ilae classification of the epilepsies: position paper of the ilae commission for classification and terminology. *Epilepsia*, 58(4):512–521, 2017.

11. Robert S. Fisher, J. Helen Cross, Jacqueline A. French, Norimichi Higurashi, Edouard Hirsch, Floor E. Jansen, Lieven Lagae, Solomon L. Moshé, Jukka Peltola, Eliane Roulet Perez, Ingrid E. Scheffer, and Sameer M. Zuberi. Operational classification of seizure types by the international league against epilepsy: Position paper of the ilae commission for classification and terminology. *Epilepsia*, 58(4):522–530, 2017. doi: https://doi.org/10.1111/epi.13670.

12. Nicholas S Abend, Alexis A Topjian, and Sankey Williams. How much does it cost to identify a critically ill child experiencing electrographic seizures? *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 32(3):257, 2015.

13. Nabeel Ahammad, Thasneem Fathima, and Paul Joseph. Detection of epileptic seizure event and onset using eeg. *BioMed research international*, 2014, 2014.

14. James R Williamson, Daniel W Bliss, David W Browne, and Jaishree T Narayanan. Seizure prediction using eeg spatiotemporal correlation structure. *Epilepsy & behavior*, 25(2):230–238, 2012.

15. Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10):1761–1770, 2011.

16. Scott B Wilson, Mark L Scheuer, Ronald G Emerson, and Andrew J Gabor. Seizure detection: evaluation of the reveal algorithm. *Clinical neurophysiology*, 115(10):2280–2291, 2004.

17. F Fürbass, P Ossenblok, M Hartmann, H Perko, AM Skupch, G Lindinger, L Elezi, E Pataraia, AJ Colon, C Baumgartner, et al. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. *Clinical Neurophysiology*, 126(6):1124–1131, 2015.

18. Ali H Shoeb and John V Guttag. Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 975–982, 2010.

19. Nilufer Ozdemir and Esen Yildirim. Patient specific seizure prediction system using hilbert spectrum and bayesian networks classifiers. *Computational and mathematical methods in medicine*, 2014, 2014.

20. Ning Wang and Michael R Lyu. Extracting and selecting distinctive eeg features for efficient epileptic seizure prediction. *IEEE journal of biomedical and health informatics*, 19(5):1648–1659, 2014.

21. Mengni Zhou, Cheng Tian, Rui Cao, Bin Wang, Yan Niu, Ting Hu, Hao Guo, and Jie Xiang. Epileptic seizure detection based on eeg signals and cnn. *Frontiers in neuroinformatics*, 12:95, 2018.

22. Ihsan Ullah, Muhammad Hussain, Hatim Aboalsamh, et al. An automated system for epilepsy detection using eeg brain signals based on deep learning approach. *Expert Systems with Applications*, 107:61–71, 2018.

23. U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli. Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in biology and medicine*, 100:270–278, 2018.

24. Khaled Saab, Jared Dunnmon, Christopher Ré, Daniel Rubin, and Christopher Lee-Messer. Weak supervision as an efficient approach for automated seizure detection in electroencephalography. *npj Digital Medicine*, 3(1):1–12, 2020.

25. Md Rashed-Al-Mahfuz, Mohammad Ali Moni, Shahadat Uddin, Salem A Alyami, Matthew A Summers, and Valsamma Eapen. A deep convolutional neural network method to detect seizures and characteristic frequencies using epileptic electroencephalogram (eeg) data. *IEEE Journal of Translational Engineering in Health and Medicine*, 9:1–12, 2021.

26. Meysam Golmohammadi, Saeedeh Ziyabari, Vinit Shah, Eva Von Weltin, Christopher Campbell, Iyad Obeid, and Joseph Picone. Gated recurrent networks for seizure detection. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–5. IEEE, 2017.

27. Pierre Thodoroff, Joelle Pineau, and Andrew Lim. Learning robust features using deep learning for automatic seizure detection. In *Machine learning for healthcare conference*, pages 178–190, 2016.

28. Ian C. Covert, Balu Krishnan, Imad Najm, Jiening Zhan, Matthew Shore, John Hixson, and Ming Jack Po. Temporal graph convolutional networks for automatic seizure detection. In *Proceedings of the 4th Machine Learning for Healthcare Conference*, volume 106 of *Proceedings of Machine Learning Research*, pages 160–180, Ann Arbor, Michigan, 09–10 Aug 2019. PMLR.

29. Alison O'Shea, Gordon Lightbody, Geraldine Boylan, and Andriy Temko. Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture. *Neural Networks*, 123:12–25, 2020. ISSN 08936080. doi: 10.1016/j.neunet.2019.11.023. URL `https://linkinghub.elsevier.com/retrieve/pii/S0893608019303910`.

30. Vinit Shah, Eva von Weltin, Silvia Lopez, James Riley McHugh, Lillian Veloso, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. The temple university hospital seizure detection corpus. *Frontiers in Neuroinformatics*, 12:83, 2018. ISSN 1662-5196. doi: 10.3389/fninf.2018.00083.

31. DK Angeles et al. Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia*, 22(4):489–501, 1981.

32. Inggi Ramadhani Dwi Saputro, Nita Dwi Maryati, Siti Rizqia Solihati, Inung Wijayanto, Sugondo Hadiyoso, and Raditiana Patmasari. Seizure type classification on eeg signal using support vector machine. In *Journal of Physics: Conference Series*, volume 1201, page 012065. IOP Publishing, 2019.

33. Subhrajit Roy, Umar Asif, Jianbin Tang, and Stefan Harrer. Machine learning for seizure type classification: setting the benchmark. *arXiv preprint arXiv:1902.01012*, 2019.

34. Shivarudhrappa Raghu, Natarajan Sriraam, Yasin Temel, Shyam Vasudeva Rao, and Pieter L Kubben. Eeg based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Networks*, 124:202–212, 2020.

35. Umar Asif, Subhrajit Roy, Jianbin Tang, and Stefan Harrer. Seizurenet: Multi-spectral deep feature learning for seizure type classification. In Seyed Mostafa Kia, Hassan Mohy-ud Din, Ahmed Abdulkadir, Cher Bass, Mohamad Habes, Jane Maryam Rondina, Chantal Tax, Hongzhi Wang, Thomas Wolfers, Saima Rathore, and Madhura Ingalhalikar, editors, *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*, pages 77–87, Cham, 2020. Springer International Publishing.

36. Tomas Iesmantas and Robertas Alzbutas. Convolutional neural network for detection and classification of seizures in clinical data. *Medical & Biological Engineering & Computing*, 58, 06 2020. doi: 10.1007/s11517-020-02208-7.

37. David Ahmedt-Aristizabal, Tharindu Fernando, Simon Denman, Lars Petersson, Matthew J Aburn, and Clinton Fookes. Neural memory networks for seizure type classification. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 569–575. IEEE, 2020.

38. Soobeom Jang, Seong-Eun Moon, and Jong-Seok Lee. Eeg-based video identification using graph signal modeling and graph convolutional neural network. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3066–3070. IEEE, 2018.

39. Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 2018.

40. Xue-han Wang, Tong Zhang, Xiang-min Xu, Long Chen, Xiao-fen Xing, and CL Philip Chen. Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1240–1244. IEEE, 2018.

41. Neeraj Wagh and Yogatheesan Varatharajah. Eeg-gcnn: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network. In *Machine Learning for Health*, pages 367–378. PMLR, 2020.

42. Yogatheesan Varatharajah, Min Jin Chong, Krishnakant Saboo, Brent Berry, Benjamin Brinkmann, Gregory Worrell, and Ravishankar Iyer. Eeg-graph: A factor-graph-based model for capturing spatial, temporal, and observational relationships in electroencephalograms. In *Advances in neural information processing systems*, pages 5371–5380, 2017.

43. Hari Sowrirajan, Jingbo Yang, Andrew Y Ng, and Pranav Rajpurkar. Moco pretraining improves representation and transferability of chest x-ray models. *arXiv preprint arXiv:2010.05352*, 2020.

44. Szu-Yen Hu, Shuhang Wang, Wei-Hung Weng, JingChao Wang, XiaoHong Wang, Arinc Ozturk, Quan Li, Viksit Kumar, and Anthony E Samir. Self-supervised pretraining with dicom metadata in ultrasound imaging. In *Machine Learning for Healthcare Conference*, pages 732–749. PMLR, 2020.

45. Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020.

46. Álvaro S Hervella, Lucía Ramos, José Rouco, Jorge Novo, and Marcos Ortega. Multi-modal self-supervised pre-training for joint optic disc and cup segmentation in eye fundus images. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 961–965. IEEE, 2020.

47. Amir Jamaludin, Timor Kadir, and Andrew Zisserman. Self-supervised learning for spinal mris. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 294–302. Springer, 2017.

48. Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020.

49. Hubert Banville, Omar Chehab, Aapo Hyvarinen, Denis Engemann, and Alexandre Gramfort. Uncovering the structure of clinical eeg signals with self-supervised learning. *Journal of Neural Engineering*, 2020.

50. Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019.

51. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607, 2020.

52. Mostafa Neo Mohsenvand, Mohammad Rasool Izadi, and Pattie Maes. Contrastive representation learning for electroencephalogram classification. In *Machine Learning for Health*, pages 238–253. PMLR, 2020.

53. Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data, 2021.

54. Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

55. Robert S Fisher, J Helen Cross, Carol D'souza, Jacqueline A French, Sheryl R Haut, Norimichi Higurashi, Edouard Hirsch, Floor E Jansen, Lieven Lagae, Solomon L Moshé, et al. Instruction manual for the ilae 2017 operational classification of seizure types. *Epilepsia*, 58(4):531–542, 2017.

56. Jessica J Falco-Walter, Ingrid E Scheffer, and Robert S Fisher. The new definition and classification of seizures and epilepsy. *Epilepsy research*, 139:73–79, 2018.

57. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

58. Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in neuroscience*, 10:196, 2016.

59. Yann LeCunn and Yoshua Bengio. Convulutional Networks for Images, Speech, and Time-Series. *The Handbook of brain theory and neural networks*, 3361(10):1995, 1995. URL `http://yann.lecun.com/exdb/publis/pdf/lecun-bengio-95a.pdf`.

60. Benjamin R. Mitchell. *The Spatial Inductive Bias of Deep Learning*. PhD thesis, The Johns Hopkins University, 2017. URL `https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/40864/MITCHELL-DISSERTATION-2017.pdf`.

61. Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie q groups on arbitrary continuous data. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3165–3176. PMLR, 13–18 Jul 2020. URL `http://proceedings.mlr.press/v119/finzi20a.html`.

62. Muhammad Aurangzeb Ahmad, Carly Eckert, and Ankur Teredesai. Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, pages 559–560, 2018.

63. Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4(1):1–9, 2021.

64. Saramati Narasimhan, Keshav B. Kundassery, Kanupriya Gupta, Graham W. Johnson, Kristin E. Wills, Sarah E. Goodale, Kevin Haas, John D. Rolston, Robert P. Naftel, Victoria L. Morgan, Benoit M. Dawant, Hernán F. J. González, and Dario J. Englot. Seizure-onset regions demonstrate high inward directed connectivity during resting-state: An seeg study in focal epilepsy. *Epilepsia*, 61(11):2534–2544, 2020. doi: https://doi.org/10.1111/epi.16686. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.16686`.

65. Sarah E Goodale, Hernán F J González, Graham W Johnson, Kanupriya Gupta, William J Rodriguez, Robert Shults, Baxter P Rogers, John D Rolston, Benoit M Dawant, Victoria L Morgan, and Dario J Englot. Resting-State SEEG May Help Localize Epileptogenic Brain Regions. *Neurosurgery*, 86(6):792–801, 09 2019. ISSN 0148-396X. doi: 10.1093/neuros/nyz351.

66. Dario J. Englot and Hal Blumenfeld. Consciousness and epilepsy: Why are complex-partial seizures complex? In *Progress in Brain Research*, volume 177, pages 147–170. Elsevier, 2009. ISBN 978-0-444-53432-3. doi: 10.1016/S0079-6123(09)17711-7. URL `https://linkinghub.elsevier.com/retrieve/pii/S0079612309177117`.

67. LJ Hirsch, SM LaRoche, N Gaspard, E Gerard, A Svoronos, ST Herman, Ram Mani, H Arif, N Jette, Y Minazad, et al. American clinical neurophysiology society's standardized critical care eeg terminology: 2012 version. *Journal of clinical neurophysiology*, 30(1):1–27, 2013.

68. Lawrence J Hirsch, Michael WK Fong, Markus Leitinger, Suzette M LaRoche, Sandor Beniczky, Nicholas S Abend, Jong Woo Lee, Courtney J Wusthoff, Cecil D Hahn, M Brandon Westover, et al. American clinical neurophysiology society's standardized critical care eeg terminology: 2021 version. *Journal of Clinical Neurophysiology*, 38(1):1–29, 2021.

69. Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

70. Qianying Huang Siyi Tang, Xuan Zhang. Multiclass seizure classification from eeg with graph convolutional recurrent neural networks. `http://web.stanford.edu/class/cs224w/project/26419700.pdf`, December 2019.

71. David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

72. Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852, 2016.

73. Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-4012. URL `https://www.aclweb.org/anthology/W14-4012`.

74. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

75. Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

76. Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

77. Ilya Loshchilov and Frank Hutter. SGDR: stochastic gradient descent with warm restarts. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

Supplementary Results

Selection of the Fourier transform and its effectiveness

In our preprocessing step, in order to have a straightforward encoding of the single-channel EEG data, we choose the amplitudes of the discrete Fourier transform because of its interpretability and simplicity (see Methods).

To evaluate the effectiveness of Fourier transform, we compare the performance of our DCRNN (without self-supervised pre-training) on inputs without and with Fourier transform on both seizure detection and seizure type classification. As shown in Supplementary Table 3, frequency-domain inputs (with Fourier transform) result in significantly better ($p < 0.05$) performance than time-domain inputs (without Fourier transform). This is likely because seizures are associated with electrical activity in certain frequency bands, and thus short-time-interval frequency-domain inputs could be more informative than time-domain inputs.

Error analysis for misclassified generalized non-specific seizures

Although generalized non-specific (GN) seizures are a majority class (see Table 1b), they are largely misclassified as combined focal (CF) seizures by our models (see Figure 2b). To examine the reason of the misclassifications, a board-certified neurologist (CLM) manually analyzed the EEG signals for the GN seizures in the test set that are misclassified by all of our four best models (i.e. 12-s/60-s distance/correlation graph with self-supervised pre-training). In total, 32 GN seizures from nine patients' 21 EEG files were manually analyzed. We find that 27 seizures (from seven patients' 16 EEG files) are in fact focal seizures but are mislabeled as GN seizures in the dataset. In contrast, only five seizures (from two patients' five EEG files) are indeed generalized seizures but are misclassified by our models. This suggests that the seizure type labels in the TUSZ dataset are not fully characterized and could likely be improved. Furthermore, we notice that all of the five misclassified cases show spatial asymmetries in abnormal patterns or muscle artifacts. Therefore, our models might have exploited the asymmetric patterns, and thus misclassified them as focal seizures. See Supplementary Figures 2a–2d for examples of focal seizures that are classified as CF seizures by our models but are mislabeled as generalized seizures in the dataset, and see Supplementary Figures 2e–2f for examples of true generalized seizures that are misclassified by our models. Supplementary Figure 3 shows the same examples in the average montage.

Model interpretability for seizure type classification

Here, we examine the spatial regions that are important for our model's prediction of seizure types. To leverage our model's advantage that it is independent of the number of EEG electrodes, we completely drop one EEG channel at a time and compute the percentage change in the model output (see Methods).

Supplementary Figure 7 shows example saliency maps of correctly classified 60-s EEG clips in the test set, where the saliency maps are visualized in the double banana montage. For the combined focal seizure (CF) examples (Supplementary Figures 7a–7b), the high saliency regions correspond to brain areas where the focal seizures are localized. For the other generalized seizure types (i.e. generalized non-specific (GN), absence (AB), and combined tonic (CT) seizures), less salient regions correspond to less abnormal areas. However, it is hard to tell from the saliency maps what spatial patterns distinguish these generalized seizure types. We speculate that the reason why this analysis works better for CF seizures than generalized seizures is that focal seizures are localized in one brain region, and thus removing the corresponding EEG channels could result in a large change in the model prediction. Supplementary Figure 8 shows the same examples in the average montage.

Supplementary Tables and Figures

Model	F1-score (mean ± std)	AUPR (mean ± std)	Sensitivity (mean ± std)	Specificity (mean ± std)
Dense-CNN	0.326 ± 0.019	0.328 ± 0.043	0.293 ± 0.021	0.938 ± 0.014
LSTM	0.376 ± 0.021	0.354 ± 0.023	0.357 ± 0.045	0.934 ± 0.015
DCRNN Correlation Graph Without Pre-training	**0.392 ± 0.027***	0.370 ± 0.027	**0.373 ± 0.035***	0.935 ± 0.012
DCRNN Distance Graph Without Pre-training	**0.437 ± 0.029***	**0.411 ± 0.041***	**0.411 ± 0.038***	0.943 ± 0.006
DCRNN Correlation Graph With Pre-training	**0.484 ± 0.011**$^+$	**0.454 ± 0.020**$^+$	**0.524 ± 0.012**$^+$	0.922 ± 0.004
DCRNN Distance Graph With Pre-training	0.487 ± 0.042	0.463 ± 0.048	**0.592 ± 0.052**$^+$	0.897 ± 0.012

(a) Additional evaluation scores for seizure detection on 12-s EEG clips.

Model	F1-score (mean ± std)	AUPR (mean ± std)	Sensitivity (mean ± std)	Specificity (mean ± std)
Dense-CNN	0.404 ± 0.022	0.399 ± 0.017	0.451 ± 0.134	0.869 ± 0.071
LSTM	0.365 ± 0.009	0.287 ± 0.026	0.463 ± 0.060	0.814 ± 0.053
DCRNN Correlation Graph Without Pre-training	**0.448 ± 0.029***	**0.440 ± 0.021***	0.457 ± 0.058	**0.900 ± 0.028***
DCRNN Distance Graph Without Pre-training	0.341 ± 0.170	**0.418 ± 0.046***	0.326 ± 0.183	**0.932 ± 0.058***
DCRNN Correlation Graph With Pre-training	**0.514 ± 0.028**$^+$	**0.539 ± 0.024**$^+$	0.502 ± 0.047	0.923 ± 0.008
DCRNN Distance Graph With Pre-training	**0.571 ± 0.029**$^+$	**0.593 ± 0.031**$^+$	**0.570 ± 0.047**$^+$	0.927 ± 0.012

(b) Additional evaluation scores for seizure detection on 60-s EEG clips.

Supplementary Table 1: **Additional evaluation scores for seizure detection on (a) 12-s EEG clips, and (b) 60-s EEG clips.** Mean and standard deviation (std) are computed from five runs with different random seeds for each model. *DCRNN model (without self-supervised pre-training) performs significantly better than at least one of the baseline models ($p < 0.05$). $^+$DCRNN model with self-supervised pre-training performs significantly better than DCRNN model without self-supervised pre-training ($p < 0.05$). Abbreviation: AUPR, Area Under the Precision-Recall curve.

Model	CF (mean ± std)	GN (mean ± std)	AB (mean ± std)	CT (mean ± std)
Dense-CNN	0.893 ± 0.025	0.713 ± 0.047	0.902 ± 0.055	0.859 ± 0.056
LSTM	0.877 ± 0.013	0.747 ± 0.023	0.948 ± 0.005	0.891 ± 0.011
DCRNN Correlation Graph Without Pre-training	**0.907 ± 0.008***	**0.815 ± 0.027***	**0.972 ± 0.013***	**0.908 ± 0.005***
DCRNN Distance Graph Without Pre-training	**0.896 ± 0.011***	**0.814 ± 0.027***	**0.983 ± 0.008***	0.890 ± 0.014
DCRNN Correlation Graph With Pre-training	0.901 ± 0.008	**0.759 ± 0.025**$^{+}$	**0.991 ± 0.007**$^{+}$	0.906 ± 0.008
DCRNN Distance Graph With Pre-training	0.906 ± 0.007	0.791 ± 0.036	0.990 ± 0.003	0.905 ± 0.007

(a) AUROC scores for individual seizure classes on 12-s EEG clips.

Model	CF (mean ± std)	GN (mean ± std)	AB (mean ± std)	CT (mean ± std)
Dense-CNN	0.895 ± 0.033	0.639 ± 0.043	0.912 ± 0.056	0.677 ± 0.021
LSTM	0.888 ± 0.009	0.732 ± 0.030	0.974 ± 0.014	0.892 ± 0.025
DCRNN Correlation Graph Without Pre-training	**0.914 ± 0.007***	**0.795 ± 0.031***	0.971 ± 0.020	**0.939 ± 0.008***
DCRNN Distance Graph Without Pre-training	**0.920 ± 0.004***	**0.811 ± 0.032***	0.973 ± 0.009	**0.926 ± 0.020***
DCRNN Correlation Graph With Pre-training	0.908 ± 0.011	0.796 ± 0.055	0.992 ± 0.004	0.942 ± 0.004
DCRNN Distance Graph With Pre-training	0.922 ± 0.008	0.823 ± 0.028	**0.987 ± 0.007**$^{+}$	0.938 ± 0.010

(b) AUROC scores for individual seizure classes on 60-s EEG clips.

Supplementary Table 2: **AUROC scores for individual seizure classes obtained from seizure type classification models on (a) 12-s EEG clips, and (b) 60-s EEG clips.** Mean and standard deviation (std) are computed from five runs with different random seeds for each model. *DCRNN model (without self-supervised pre-training) performs significantly better than at least one of the baseline models ($p < 0.05$). $^{+}$DCRNN model with self-supervised pre-training performs significantly better than DCRNN model without self-supervised pre-training ($p < 0.05$). Abbreviations: CF, combined focal seizure; GN, generalized non-specific seizure; AB, absence seizure; CT, combined tonic seizure.

Model	Input Domain	AUROC (mean ± std)	
		12-s	60-s
DCRNN Correlation Graph Without Pre-training	Time	0.717 ± 0.003	0.704 ± 0.023
	Frequency	**0.812 ± 0.012***	**0.804 ± 0.015***
DCRNN Distance Graph Without Pre-training	Time	0.733 ± 0.012	0.698 ± 0.003
	Frequency	**0.824 ± 0.020***	**0.793 ± 0.022***

(a)

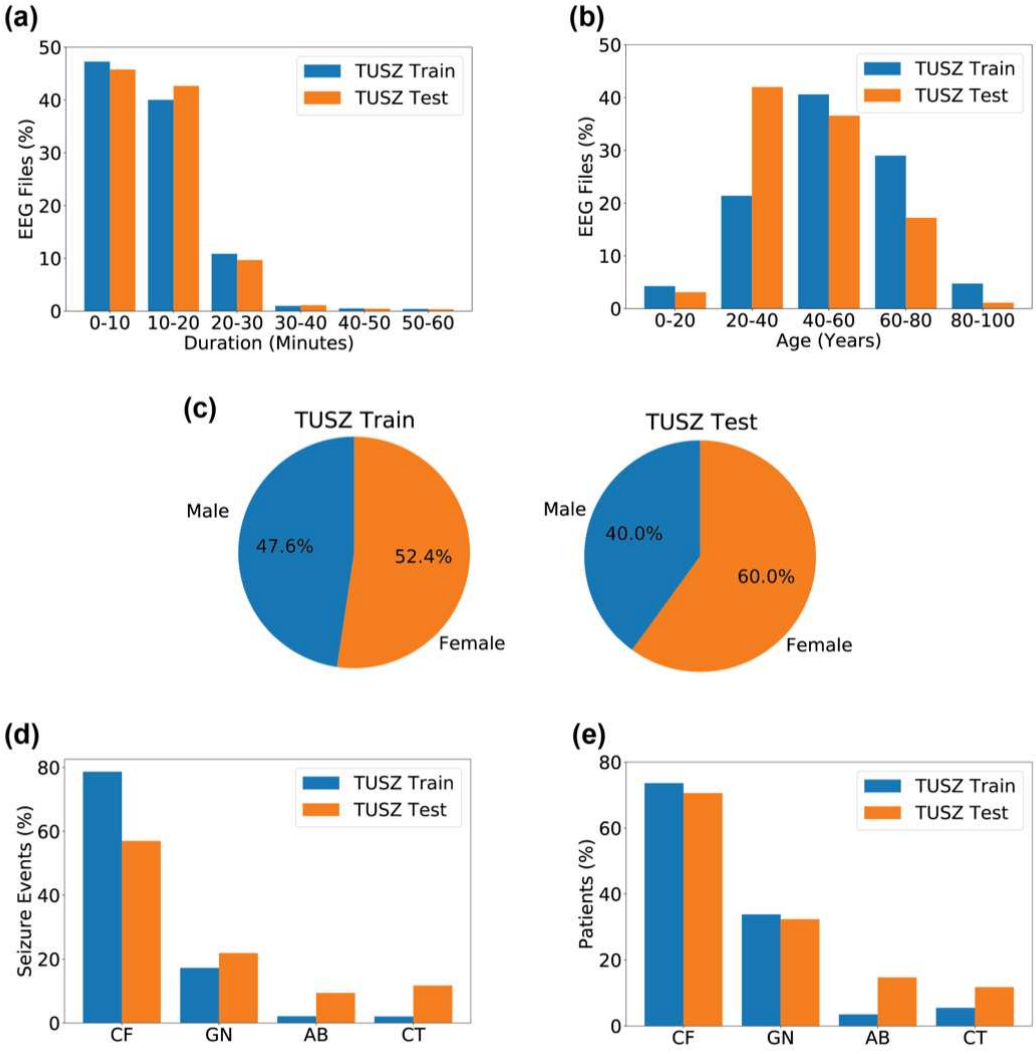
Model	Input Domain	Weighted F1-score (mean ± std)	
		12-s	60-s
DCRNN Correlation Graph Without Pre-training	Time	0.597 ± 0.023	0.618 ± 0.018
	Frequency	**0.710 ± 0.023***	**0.701 ± 0.030***
DCRNN Distance Graph Without Pre-training	Time	0.592 ± 0.022	0.608 ± 0.015
	Frequency	**0.703 ± 0.025***	**0.690 ± 0.035***

(b)

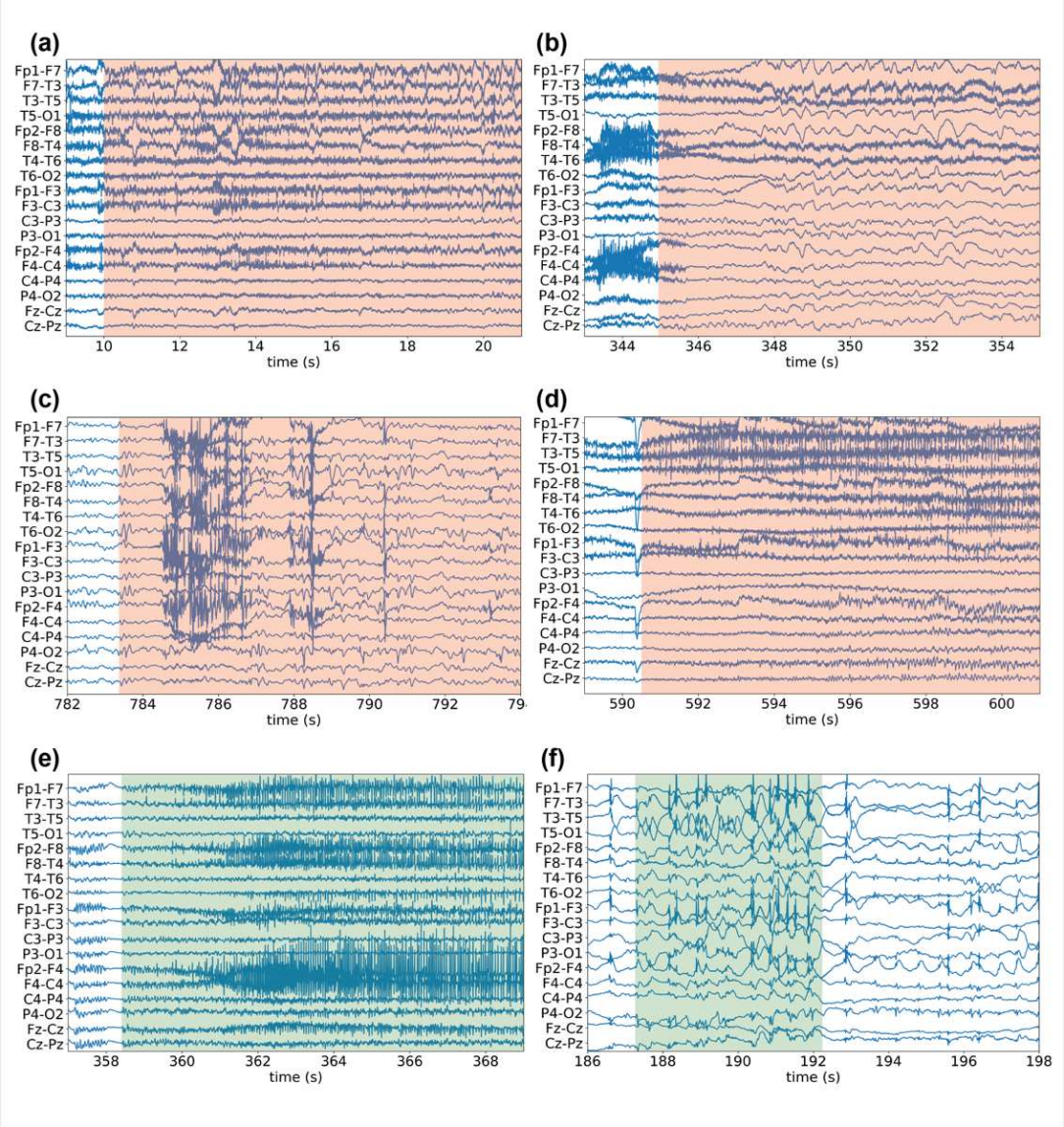
Supplementary Table 3: **Comparison of (a) seizure detection and (b) seizure type classification results from our DCRNN model (without self-supervised pre-training) on time-domain inputs and frequency-domain inputs.** Frequency-domain inputs result in significantly better performance on both seizure detection and seizure type classification tasks, indicating the effectiveness of Fourier transform in our preprocessing step. *Improvment is statistically significant ($p < 0.05$).

Model	Seizure Detection AUROC (mean ± std)		Seizure Type Classification Weighted F1-score (mean ± std)	
	60-s → 12-s (vs 12-s → 12-s)	12-s → 60-s (vs 60-s → 60-s)	60-s → 12-s (vs 12-s → 12-s)	12-s → 60-s (vs 60-s → 60-s)
Dense-CNN	0.523 ± 0.130 (↓ 0.289)	-	0.460 ± 0.079 (↓ 0.116)	-
DCRNN Correlation Graph Without Pre-training	0.770 ± 0.004 (↓ 0.042)	0.719 ± 0.038 (↓ 0.085)	0.682 ± 0.027 (↓ 0.028)	0.602 ± 0.018 (↓ 0.099)
DCRNN Distance Graph Without Pre-training	0.777 ± 0.009 (↓ 0.054)	0.760 ± 0.025 (↓ 0.033)	0.696 ± 0.040 (↓ 0.007)	0.630 ± 0.042 (↓ 0.06)
DCRNN Correlation Graph With Pre-training	0.778 ± 0.014 (↓ 0.083)	0.797 ± 0.011 (↓ 0.053)	0.704 ± 0.023 (↓ 0.019)	0.660 ± 0.013 (↓ 0.089)
DCRNN Distance Graph With Pre-training	0.791 ± 0.022 (↓ 0.075)	0.799 ± 0.027 (↓ 0.076)	0.743 ± 0.039 (↓ 0.003)	0.663 ± 0.016 (↓ 0.086)

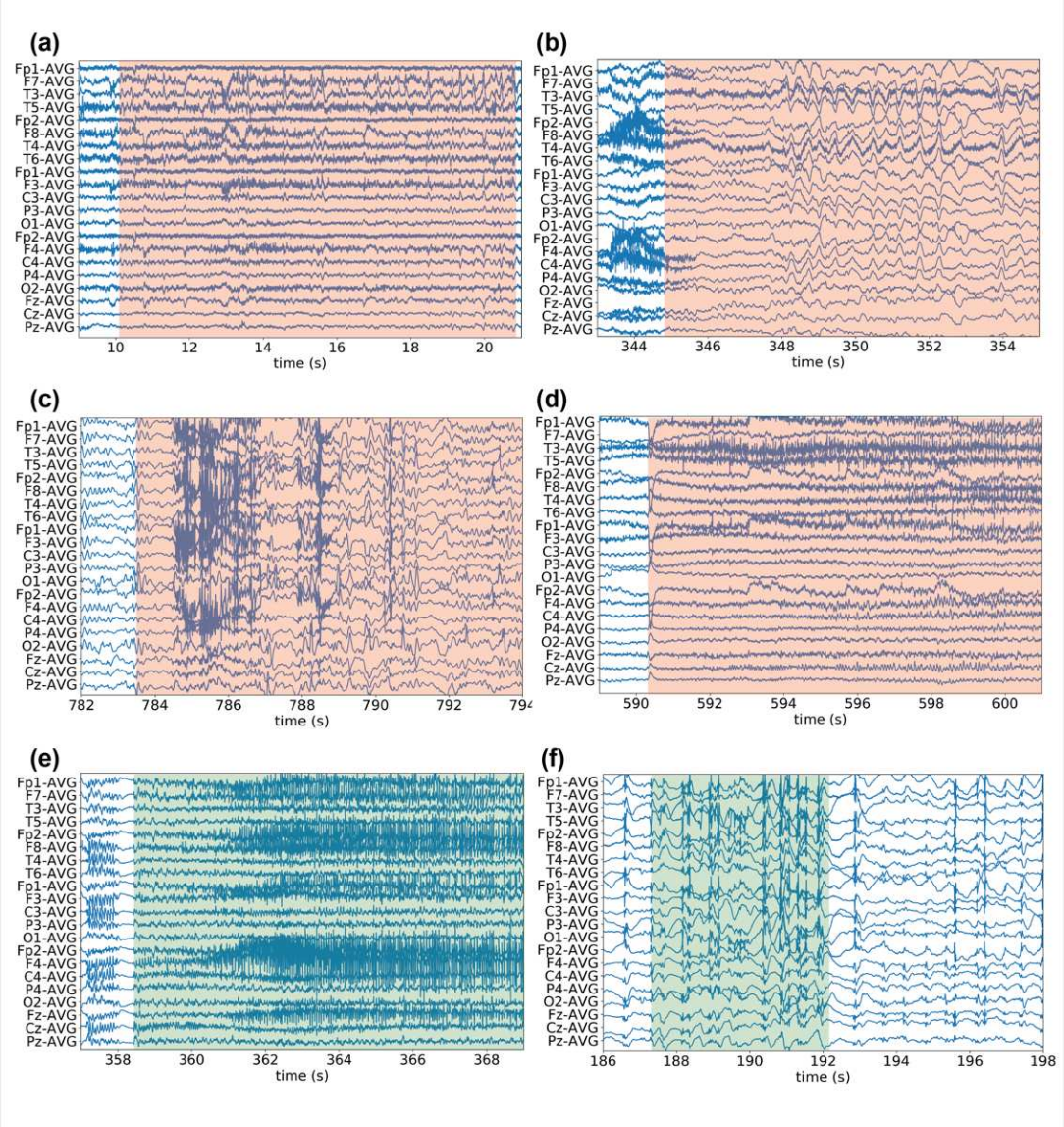
Supplementary Table 4: **Performance of Dense-CNN and our DCRNN when tested on the clip length that is different from which the models are trained on.** 60-s → 12-s (12-s → 60-s) means that the model is trained on 60-s (12-s) EEG clips and tested on 12-s (60-s) clips. ↓ shows the decrease in model performance compared to original model trained and tested on 12-s (60-s) clips. Because Dense-CNN's model parameters are dependent on the clip length, we can only test on 12-s clips with model trained on 60-s clips, where we pad the remaining 48-s with 0's to obtain 60-s clips. We observe that Dense-CNN's performance decreases by a much larger margin than our DCRNN model.



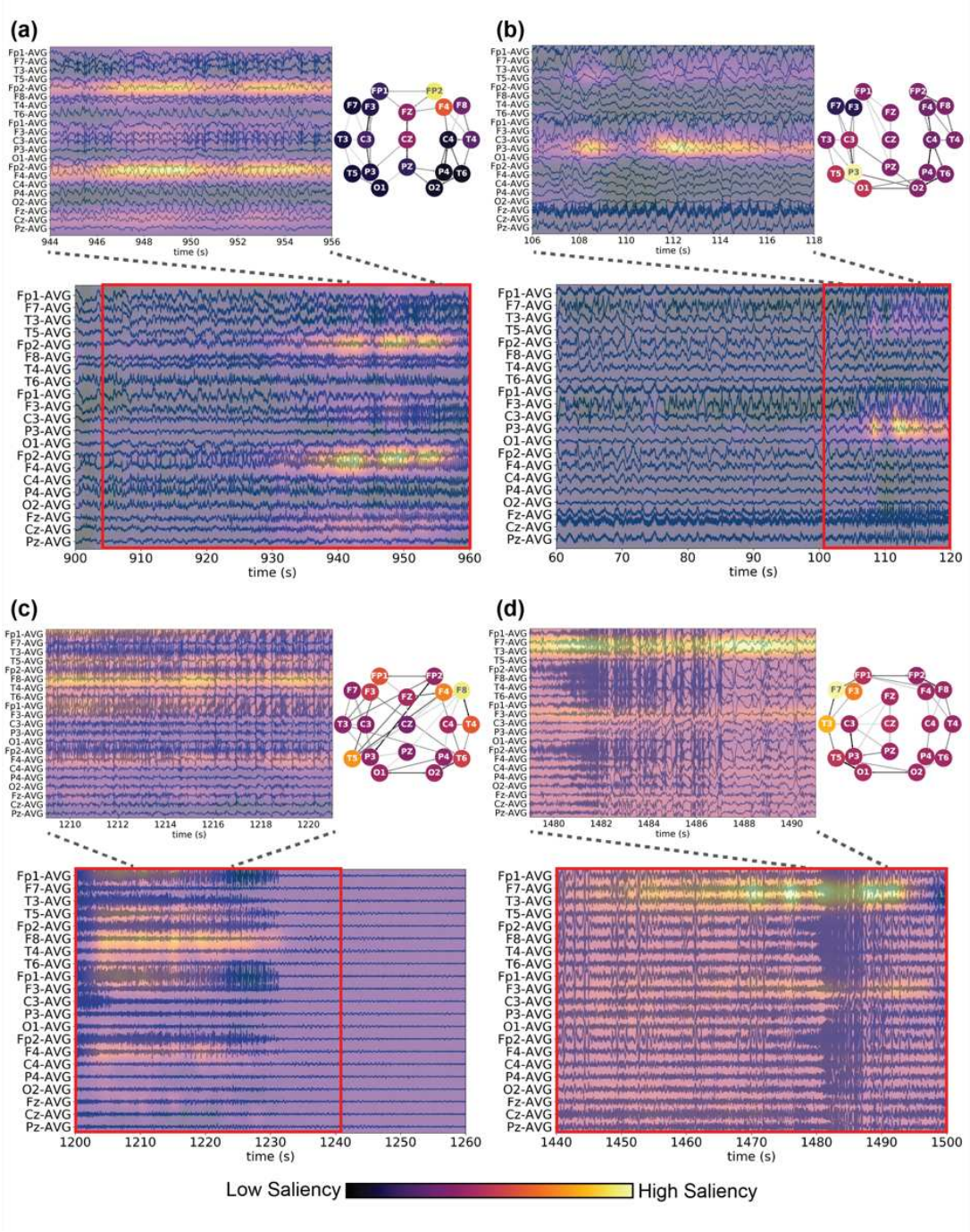
Supplementary Figure 1: **Distributions of (a) EEG file duration, (b) age, (c) sex, (d) seizure types by seizure events, and (e) seizure types by patients in the TUSZ official train and test sets used in our study.** Note that we exclude five patients from the TUSZ test set who appear in both train and test sets of TUSZ v1.5.2. Abbreviations: CF, combined focal seizure; GN, generalized non-specific seizure; AB, absence seizure; CT, combined tonic seizure.



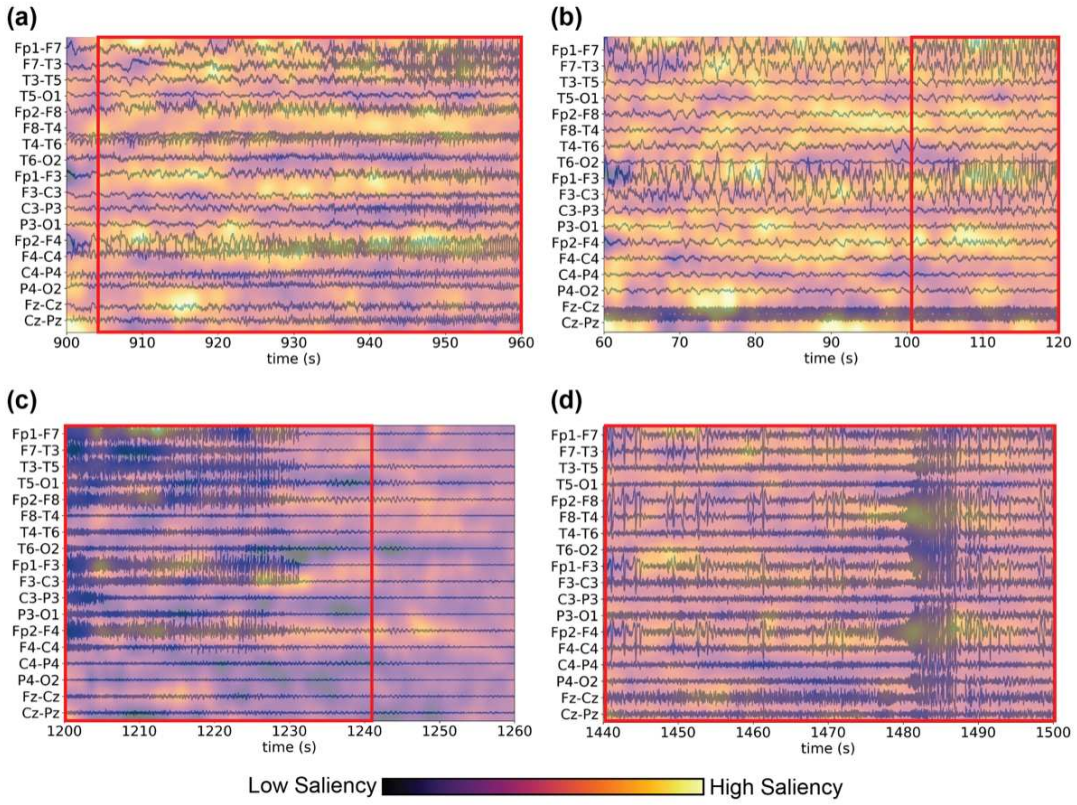
Supplementary Figure 2: **(a)–(d) Examples of focal seizures that are mislabeled as generalized non-specific seizures in TUSZ dataset shown in the *double banana* montage.** Shaded areas indicate the duration of the seizure events. These four examples are classified as combined focal (CF) seizures by all of our four best DCRNN models (i.e. 12-s/60-s distance/correlation graphs with self-supervised pre-training). Closer investigation of the EEG signals by a board-certified neurologist (CLM) suggested that these seizures are in fact focal seizures but are mislabeled as generalized seizures. Specifically, (a) shows a left frontal onset seizure, (b) shows a focal seizure that starts in the anterior temporal region, (c) shows a focal seizure with waxes and wanes in the posterior/occipital regions, and (d) shows a focal fast spiking seizure. **(e)–(f) Examples of generalized non-specific seizures that are misclassified as CF seizures by our models.** Specifically, (e) shows a generalized seizure that is bilateral but more prominent in the left frontal and right frontal areas, and (f) shows a generalized seizure with suppressed patterns, sharp waves, and spikes. Although these examples are generalized seizures, there are spatial asymmetries in the abnormalities, which could be the reason why our models misclassify them as focal seizures.



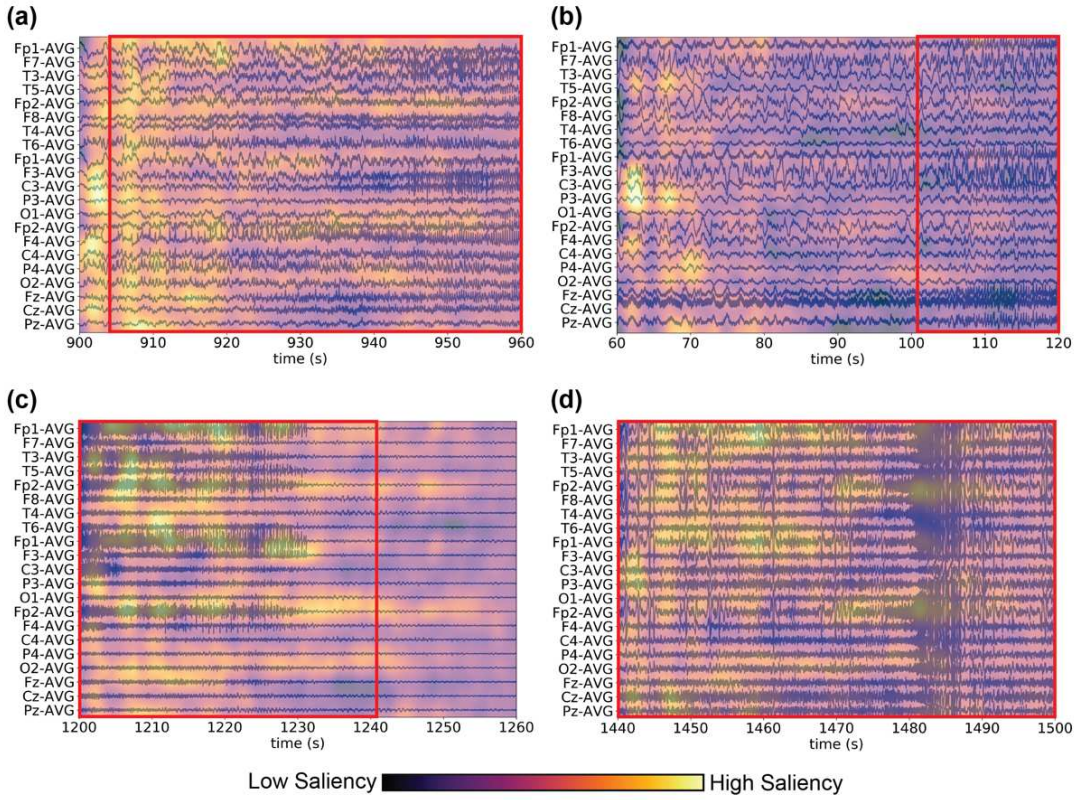
Supplementary Figure 3: **(a)–(d) Examples of focal seizures that are mislabeled as generalized non-specific seizures in TUSZ dataset shown in the *average* montage.** Shaded areas indicate the duration of the seizure events. These four examples are classified as combined focal (CF) seizures by all of our four best DCRNN models (i.e. 12-s/60-s distance/correlation graphs with self-supervised pre-training). Closer investigation of the EEG signals by a board-certified neurologist (CLM) suggested that these seizures are in fact focal seizures but are mislabeled as generalized seizures. Specifically, (a) shows a left frontal onset seizure, (b) shows a focal seizure that starts in the anterior temporal region, (c) shows a focal seizure with waxes and wanes in the posterior/occipital regions, and (d) shows a focal fast spiking seizure. **(e)–(f) Examples of generalized non-specific seizures that are misclassified as CF seizures by our models.** Specifically, (e) shows a generalized seizure that is bilateral but more prominent in the left frontal and right frontal areas, and (f) shows a generalized seizure with suppressed patterns, sharp waves, and spikes. Although these examples are generalized seizures, there are spatial asymmetries in the abnormalities, which could be the reason why our models misclassify them as focal seizures.



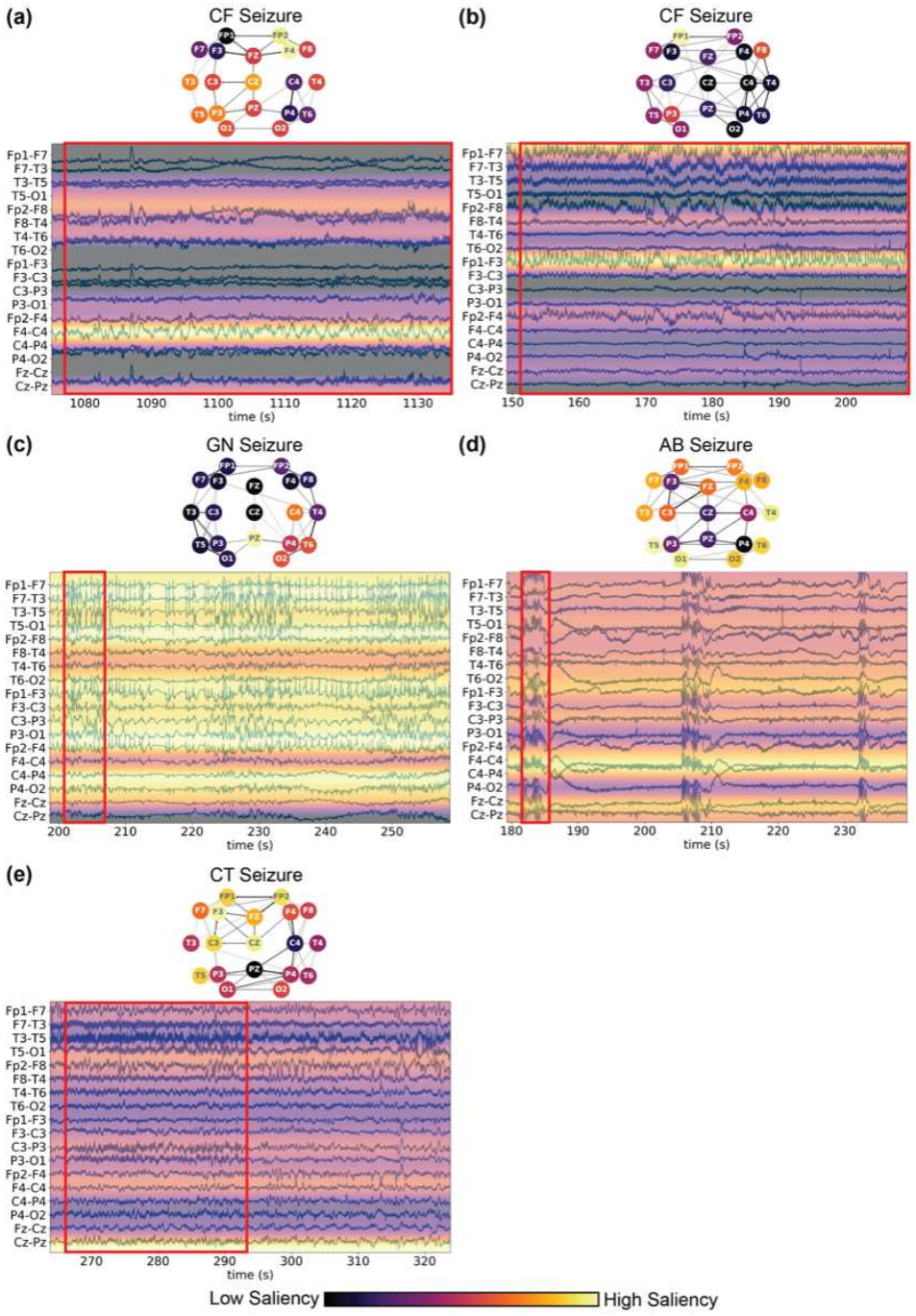
Supplementary Figure 4: **Example saliency maps for seizure detection obtained from our correlation graph-based DCRNN model shown in the *average* montage for (a)–(b) focal seizures, and (c)–(d) generalized seizures.** The saliency maps are obtained by occluding 1-s EEG signals in one EEG channel at a time and calculating the percentage change in the model output (see Methods). Brighter (Darker) color indicates higher (lower) saliency. In each subfigure, the bottom panel shows the 60-s saliency map, the top left panel shows the saliency map zoomed in to the 12-s most salient time periods, the top right panel shows the saliency values that are averaged along the time dimension and overlaid on the corresponding correlation graph, and the red boxes indicate the duration of the seizures. We observe that temporal regions with seizures have higher saliency than temporal regions without seizures. Moreover, for focal seizures (a–b), high saliency is localized in the more abnormal brain regions. In contrast, for generalized seizures (c–d), high saliency is more diffused across the brain regions. The examples shown here are correctly predicted 60-s EEG clips in the test set. Note that the values within a saliency map are normalized, and thus should not be compared across different saliency maps.



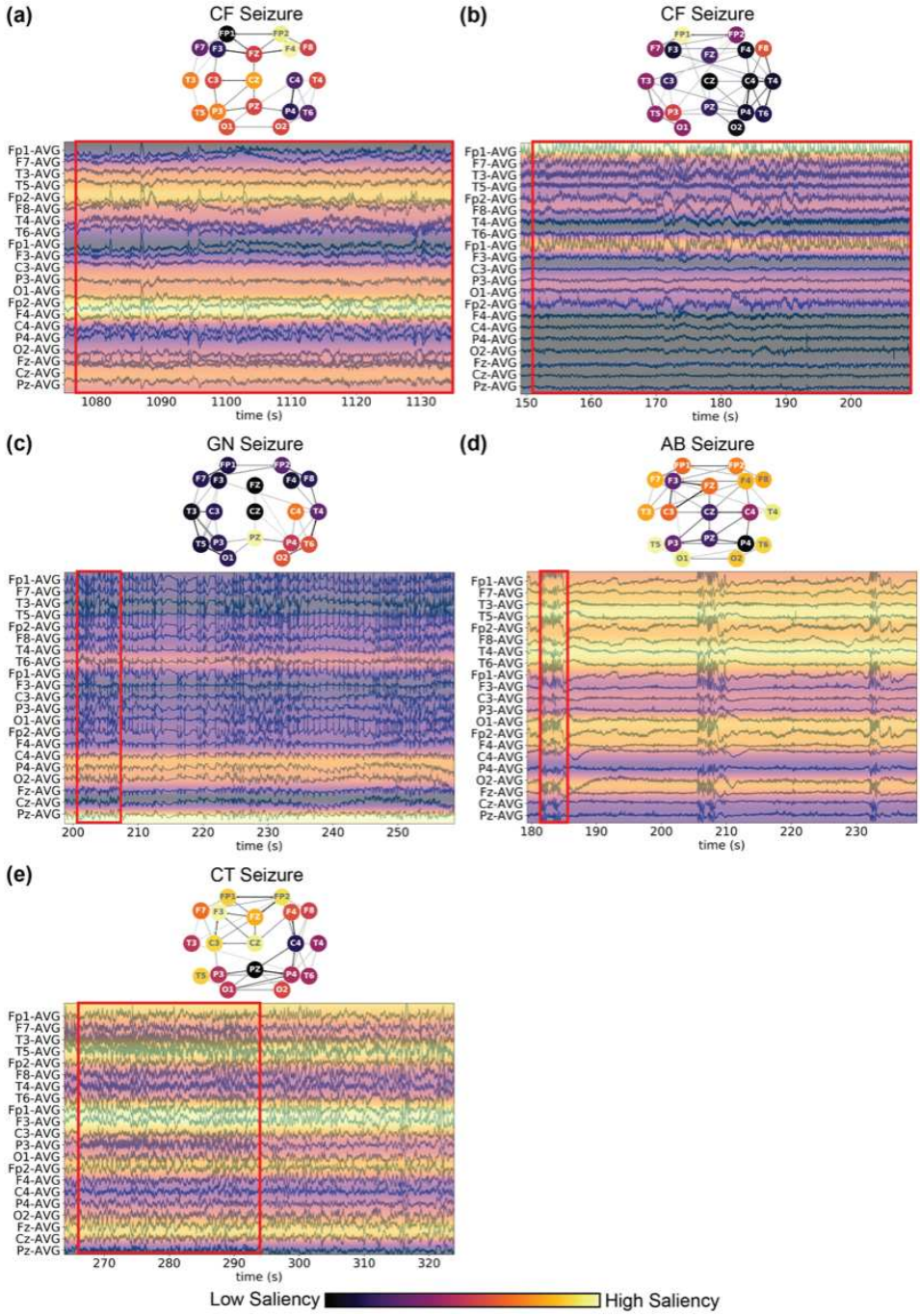
Supplementary Figure 5: **Example saliency maps for seizure detection obtained from baseline Dense-CNN shown in the *double banana* montage for (a)–(b) focal seizures, and (c)–(d) generalized seizures.** The saliency maps are obtained by occluding 1-s EEG signals in one EEG channel at a time and calculating the percentage change in the model output. To visualize the percentage changes in the double banana montage, we subtract the values between the corresponding channels in the montage. Brighter (Darker) color indicates higher (lower) saliency. Red boxes indicate the duration of the seizures. We observe that the high saliency regions do not localize in any spatial or temporal regions associated with seizures. The examples shown here are correctly predicted 60-s EEG clips in the test set. Note that the values within a saliency map are normalized, and thus should not be compared across different saliency maps.



Supplementary Figure 6: **Example saliency maps for seizure detection obtained from baseline Dense-CNN shown in the *average* montage for (a)–(b) focal seizures, and (c)–(d) generalized seizures.** The saliency maps are obtained by occluding 1-s EEG signals in one EEG channel at a time and calculating the percentage change in the model output. Brighter (Darker) color indicates higher (lower) saliency. Red boxes indicate the duration of the seizures. We observe that the high saliency regions do not localize in any spatial or temporal regions associated with seizures. The examples shown here are correctly predicted 60-s EEG clips in the test set. Note that the values within a saliency map are normalized, and thus should not be compared across different saliency maps.



Supplementary Figure 7: **Example saliency maps for seizure type classification obtained from our correlation graph-based DCRNN model for (a)–(b) CF seizures, (c) GN seizure, (d) AB seizure, and (e) CT seizure shown in the *double banana* montage.** The saliency maps are obtained by completely dropping one EEG channel at a time and calculating the percentage change in the model output. In each subfigure, the bottom panel shows the saliency values for each channel replicated along the time dimension and overlaid on 60-s EEG signals, the red boxes indicate the duration of seizures in the EEG clips, and the top panel shows the saliency values overlaid on the correlation graph structure. To visualize percentage changes in the double banana montage, we subtract the values between the corresponding channels in the montage, which results in different values between saliency maps shown on the EEG signals (bottom panels) and that shown on the graph structures (top panels). Brighter (darker) color indicates higher (lower) saliency. For CF seizures (a–b), the high saliency regions correspond to brain areas where the focal seizures are localized. For the other generalized seizure types (c–e), less salient regions correspond to less abnormal areas. The examples shown here are DCRNN's correctly classified 60-s EEG clips in the test set. Note that the values within a saliency map are normalized, and thus should not be compared across different saliency maps. Abbreviations: CF, combined focal; GN, generalized non-specific; AB, absence; CT, combined tonic.



Supplementary Figure 8: **Example saliency maps for seizure type classification obtained from our correlation graph-based DCRNN model for (a)–(b) CF seizures, (c) GN seizure, (d) AB seizure, and (e) CT seizure shown in the *average* montage.** The saliency maps are obtained by completely dropping one EEG channel at a time and calculating the percentage change in the model output. In each subfigure, the bottom panel shows the saliency values for each channel replicated along the time dimension and overlaid on 60-s EEG signals, the red boxes indicate the duration of seizures in the EEG clips, and the top panel shows the saliency values overlaid on the correlation graph structure. Brighter (Darker) color indicates higher saliency. For CF seizures (a–b), the high saliency regions correspond to brain areas where the focal seizures are localized. For the other generalized seizure types (c–e), less salient regions correspond to less abnormal areas. The examples shown here are DCRNN's correctly classified 60-s EEG clips in the test set. Note that the values within a saliency map are normalized, and thus should not be compared across different saliency maps. Abbreviations: CF, combined focal; GN, generalized non-specific; AB, absence; CT, combined tonic.