DeepEZ: A Graph Convolutional Network for Automated Epileptogenic Zone Localization from Resting-State fMRI Connectivity

Naresh Nandakumar, David Hsu, Raheel Ahmed, and Archana Venkataraman *Member, IEEE*

Abstract— *Objective:* Epileptogenic zone (EZ) localization is a crucial step during diagnostic work up and therapeutic planning in medication refractory epilepsy. In this paper, we present the first deep learning approach to localize the EZ based on resting-state fMRI (rs-fMRI) data. *Methods:* Our network, called DeepEZ , uses a cascade of graph convolutions that emphasize signal propagation along expected anatomical pathways. We also integrate domain-specific information, such as an asymmetry term on the predicted EZ and a learned subject-specific bias to mitigate environmental confounds. *Results:* We validate DeepEZ on rs-fMRI collected from 14 patients with focal epilepsy at the University of Wisconsin Madison. Using cross validation, we demonstrate that DeepEZ achieves consistently high EZ localization performance (Accuracy: 0.88 ± 0.03; AUC: 0.73 ± 0.03) that far outstripped any of the baseline methods. This performance is notable given the variability in EZ locations and scanner type across the cohort. *Conclusion:* Our results highlight the promise of using DeepEZ as an accurate and noninvasive therapeutic planning tool for medication refractory epilepsy. *Significance:* While prior work in EZ localization focused on identifying localized aberrant signatures, there is growing evidence that epileptic seizures affect inter-regional connectivity in the brain. DeepEZ allows clinicians to harness this information from noninvasive imaging that can easily be integrated into the existing clinical workflow.

Index Terms— Brain Connectivity, Deep Learning, Epilepsy, Seizure Localization, Resting-State fMRI

I. INTRODUCTION

EPILEPSY is one of the most common neurological disorders, affecting around 50 million people worldwide, and is linked to a fivefold increase in mortality [13]. Epilepsy onset often occurs in childhood, and approximately one third

N. Nandakumar is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA (e-mail: nnandak1@jhu.edu).

D. Hsu is with the Department of Neurology, University of Wisconsin Madison, Madison, WI, USA (e-mail: hsu@neurology.wisc.edu).

R. Ahmed is with the Department of Neurosurgery, University of Wisconsin Madison, Madison, WI, USA (e-mail: raheel.ahmed@neurosurgery.wisc.edu).

A. Venkataraman is with Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA (corresponding author, e-mail: archana.venkataraman@jhu.edu).

of all patients have a medication refractory course that is associated with a disabling cumulative effect on neurocognitive development, lost productivity for the family, and increased societal and healthcare costs [5]. Surgical treatment is a safe and effective therapeutic approach for medication refractory epilepsy, that can provide seizure freedom and improved quality of life [26]. However, surgical candidacy and treatment outcomes are dependent on accurate localization of the epileptogenic zone (EZ) as defined by clinical, radiographic (magnetic resonance imaging, MRI) and physiological (electroencephalography, EEG) features [65]. Long-term treatment failures following surgery most commonly occur due to inaccurate identification and resection of the EZ . Invasive monitoring using implanted intracranial electrodes can provide more accurate EZ localization that can help plan treatment, but is associated with surgical risks [76]. Hence an accurate EZ localization hypothesis is the foundation for effective and safe treatment in epilepsy [45], [52], and is the most important prognostic determinant for long term treatment outcomes.

A. Automated Methods for EZ Localization

Over the past two decades, there has been an increasing focus on automated methods for EZ localization. These methods are most often based on electrographic (EEG) or neuroimaging (structural MRI) modalities and can help reduce interpretative differences and delays in clinical reviews.

Automated methods for EEG localization have largely focused on improving the spatial resolution of the EEG sensors by deconvolving the signals into current dipoles or distributed sources at the millimeter scale [69], [84]. Going one step further, EEG data can be combined with noninvasive magnetoencephalography (MEG) for improved source estimation [58], [60]. Recent studies have demonstrated the translational promise of such methods. However, from a modeling standpoint, these *inverse solvers* require careful annotations of the seizure interval and are sensitive to physiological noise and the underlying head model [19], [20]. More importantly, they rely on high-density recordings of >50 EEG/MEG channels. From a logistical standpoint, the current standard-of-care for long-term EEG monitoring is the 10-20 electrode placement system [42], which contains fewer than 20 EEG channels distributed across the scalp. This resolution is insufficient for accurate and fine-grained inverse source estimation. Moreover, only 27% of epilepsy centers in the United States have

access to and regularly utilize MEG [7]. Thus, while inverse source mapping remains a valuable direction of research with tremendous potential for presurgical evaluation, these methods are not amenable to most clinical workflows. Recently, Temple University Hospital (TUH) released a large public EEG dataset, which has spurred interest in seizure type classification [4], [66], [67], where the goal is to predict the epilepsy subtype from scalp EEG. While this task provides more information than seizure detection and is less reliant on human annotations than inverse source localization, the categories (focal, generalized, complex partial, absence, etc.) are too broad to accurately pinpoint the EZ .

In contrast to EEG, automated methods for MRI localization aim to identify epileptogenic lesions including Focal Cortical Dysplasias (FCDs), that are often difficult to radiographically identify on clinical imaging. Traditionally, these methods were implemented as a two-stage procedure. First, image-based features are extracted from the MRI data, such as cortical thickness, intensity, texture, asymmetry, and voxel-based morphometry [6], [16], [27], [34], [71]. Second, each voxel is classified as normal or FCD using statistical or machine learning algorithms. While these methods work well on large FCD cohorts, they tend to be unreliable for nonlesional patients [3]. In addition, epileptogenic lesions are diverse and can involve cortical, subcortical white matter [23], [63] and vascular abnormalities [24], [31], [78], which are better suited to other data modalities [2], [51], [81].

B. Connectivity as a Biomarker for Epilepsy

Recent neuroimaging studies of epilepsy have implicated global brain network changes in seizure generation and disease progression. Accordingly, epilepsy is increasingly viewed as a network disorder that affects regional and global *connectivity* [10], [17], [38], [63], [68]. Diffusion MRI (d-MRI) assesses white matter properties based on free water diffusion [79] and informs us on the structure of brain networks [8]. In contrast, resting-state functional MRI (rs-fMRI) quantifies the temporal synchrony between brain regions by measuring changes in low frequency BOLD fluctuations. Alterations in structural and functional network properties have been linked to disease onset [25], duration [53], and treatment outcomes [10], [68] in epilepsy. For example, our group has investigated functional topology in subjects with epilepsy [30], demonstrating functional reorganization with a shift of network hubs to the contralateral hemisphere in temporal onset epilepsy [50]. Support Vector Machine-based analysis can discriminate network features in temporal epilepsy and healthy control subjects. In addition, neural connectivity patterns can help predict neuropsychological measures that assess language and memory function. Notably, global connectome changes in epilepsy are associated with a decrement in neurocognitive phenotype [72]. While promising, these connectivity studies are restricted to predefined structural and functional systems and careful patient subtyping (e.g., temporal lobe epilepsy). In addition, the results are correlative and do not quantify how well the biomarkers would generalize to new patients.

Recently, a few seminal studies have explored prospective EZ localization from rs-fMRI connectivity. The earliest work

of [73] computed local and global network theoretic measures from whole-brain rs-fMRI data. The authors used outlying values in these network measures, as compared to a normative cohort, to define the EZ for each epilepsy patient. While promising on a small validation dataset, the performance requires careful tuning of different threshold values. The follow-up study of [75] proposed a hierarchical Bayesian model that inferred patient-specific hubs of abnormal connectivity. The authors demonstrate comparable performance to [73] without any auxiliary parameter tuning. However, both of these studies rely on comparison with a normative cohort, which may not be available in a clinical setting. The work of [85] takes a difference approach by first running independent component analysis (ICA) and the constructing a set of rules (e.g., asymmetry, power spectrum) to select the components associated with the EZ . While the authors demonstrate highly promising localization performance, they rely on visual inspection to select between candidate EZ components. Thus, a careful read of the method suggests that it is not fully automated. Specifically, the authors rely on visual inspection to select between candidate EZ components. Finally, the work of [36] also runs ICA on the rs-fMRI data and extracts a set of hand-crafted features from the components. In this case, the authors employ a support vector machine classifier to automatically learn which components are associated with the EZ . While the authors demonstrate good localization performance on some patients, the reported sensitivity is low for others.

C. Our Contributions

In this paper, we introduce the first deep learning model for EZ localization using interictal rs-fMRI connectivity. The underlying assumption of our work is that the chronic and recurrent seizure activity causes subtle and distributed changes in functional connectivity across the brain. In contrast to rule-based approaches, we used supervised learning to automatically mine and leverage complex relationships in the rs-fMRI data for robust and generalizable EZ identification. Our model, which we call DeepEZ, takes as input a whole-brain connectivity graph, where nodes correspond to regions in our brain parcellation and edges denote the functional connectivity between regions. From here, DeepEZ uses graph convolutional networks (GCNs) [41], [83] to implicitly track information flow along expected anatomical pathways and fully-connected layers to classify each node (i.e., region) as belonging to the EZ or not. We encode anatomical information using d-MRI tractography, which is often viewed as the anatomical substrate for functional signaling in the brain [33], [82]. DeepEZ also incorporates the findings of previous works via an asymmetry term in the loss function to encourage lateralized predictions and a learned subject-specific bias to mitigate environmental confounds. We validate DeepEZ on a heterogeneous dataset of 14 pediatric epilepsy patients collected at the University of Wisconsin (UW) Madison. We demonstrate the DeepEZ outperforms the ICA methods of [36], [85] and ablated versions of the network. We also rigorously evaluate the sensitivity of DeepEZ to parcellation size, the number of network layers, hyperparameter tuning, and data augmentation. Taken together,
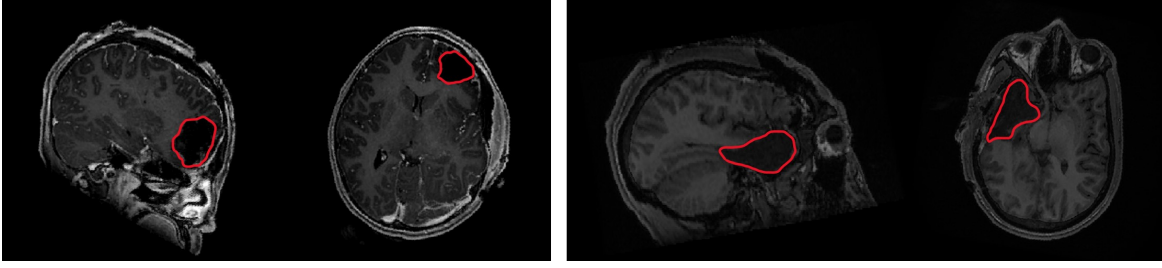
Fig. 1. From (**L-R**), post resection structural MRI scans of two separate patients. We use the resection boundary, delineated via the red lines above, to derive pseudo ground truth labels for the EZ during training and testing.

TABLE I

DEMOGRAPHIC INFORMATION, EZ LOCATION, AND SCANNER TYPE, AND OUTCOME FOR EACH PATIENT IN THE UW MADISON DATASET.

Subject	Age	Gender	EZ Location	Scanner	Seizure Control
1	14	M	left anterior and medial parietal	3T	Engel IA, ILAE 1
2	17	M	left anterior temporal lobe	3T	Engel IA, ILAE 1
3	10	F	left frontal pole	3T	Engel IA, ILAE 1
4	11	M	right temporal lobe	3T	Engel IA, ILAE 1
5	15	F	right anterior frontal/amygdala	3T	Engel IIIA, ILAE 4
6	11	M	left inferior frontal operculum and insula	3T	Engel IA, ILAE 1
7	13	F	right middle frontal region	3T	Engel IIIA, ILAE 2
8	14	M	right posterior temporal and parietal region	1.5T	Engel IA, ILAE 1
9	15	F	right anterior precentral gyrus	1.5T	Engel IIIA, ILAE 4
10	18	M	right temporal	3T	Engel IIB ILAE 2
11	9	M	right middle frontal lobe	1.5T	Engel IA, ILAE 1
12	15	M	right middle parietal lobe	1.5T	Engel IVA, ILAE 5
13	16	M	Left inferior frontal	3T	Engel IIIA, ILAE 4
14	11	M	left inferior frontal gyrus	3T	Engel IA, ILAE 1

our results highlight the promise of using rs-fMRI connectivity as a complementary source of information to localize the EZ in presurgical epilepsy patients.

II. MATERIALS AND METHODS

A. Data and Preprocessing

Our dataset consists of preoperative functional and postoperative structural MRI scans from 14 pediatric subjects with focal epilepsy that underwent a EZ resection procedure at UW Madison. The MRI data was acquired as a part of standard care on either a GE 1.5T or a GE 3T Signa scanner. This study was approved by the University of Wisconsin-Madison Institutional Review Board under protocol 2019-1265 (approved Feb 2020).

Preoperative Rs-MRI (rs-fMRI) data was acquired using an echo planar imaging sequence (EPI, TR = 802 ms, TE = 33.5 ms, flip angle = $50°$, FOV = 20.8 cm, res = 2 mm isotropic). The data was preprocessed using the CPAC pipeline, which includes slice time correction, motion correction, nuisance signal regression, band-pass filtering $(0.01 - 0.1Hz)$, and registration to the MNI template. We use the Brainnetomme atlas [29] to define $N = 246$ cortical and subcortical regions for our analysis. We chose this atlas due to its fine spatial resolution and symmetric region definitions.

Postoperative T1-weighted structural images were acquired using a three-dimensional gradient-echo pulse sequence (MPRAGE, TR = 604 ms, TE = 2.516 ms, flip angle = $8°$, FOV = 25.6 cm, res = 0.8 mm isotropic). After skull stripping, we use affine registration to align the T1 data for each patient to the MNI space. All registrations were visually

inspected for quality assurance. We manually delineate the resection zone and use this boundary to define pseudo ground truth EZ labels for training and evaluation after applying the Brainnetomme atlas. Fig. I-B depicts two examples of the postoperative T1 images, where the resection is marked by red arrows. Finally, Table I reports the age, gender, EZ location, scanner used to acquire the rs-fMRI data for each patient and patient outcome using the Engel [28] and ILAE scale [9] . As seen, our epilepsy cohort is highly heterogeneous and every patient experienced reduced seizures after surgery, with the majority being completely seizure free.

The GCNs in DeepEZ rely on a structural connectivity profile. To ensure consistency across subjects, we derive the graph from d-MRI tractography of 50 subjects from the Human Connectome Project (HCP) dataset [80]. The d-MRI data for each subject was preprocessed using the pipeline of [44] to obtain individual structural connectivity matrices based on the Brainnetome atlas. The steps of the pipeline include linear registration, tensor estimation, tractography, and graph estimation. The graphs are then averaged and thresholded to obtain the template matrix **A** used in the following section. Diffusion MRI (d-MRI) was also acquired for each patient and integrated into one of the baseline algorithms. D-MRI was collected on a 3T GE scanner (TR= $7000ms$, TE= $82.4ms$, res= 1.5 mm isotropic, b-value = 1000).

B. Deep Learning for EZ Localization

DeepEZ is designed under the assumption that there are subtle but widespread connectivity patterns associated with
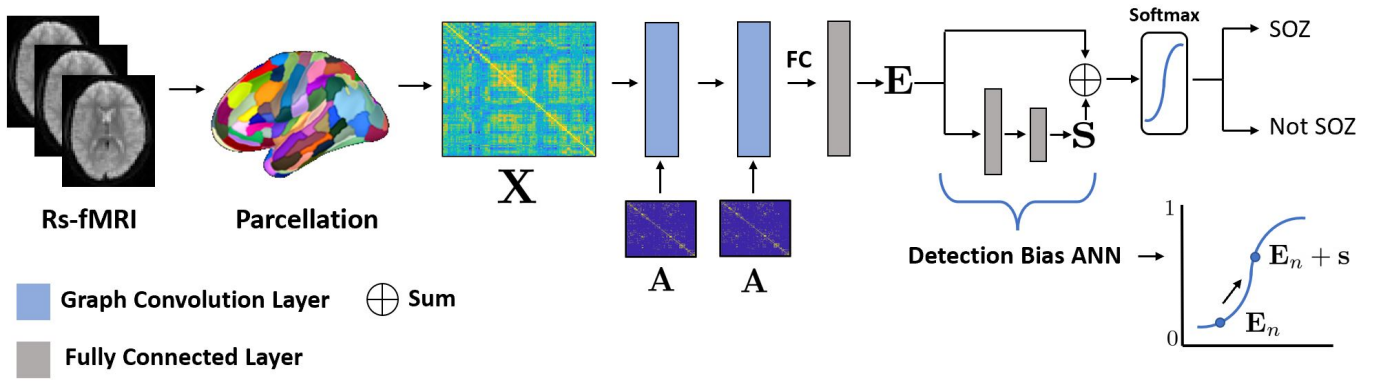
Fig. 2. The overview of our model schematic. First, we apply a parcellation to the rs-fMRI and construct the subject specific functional connectivity information $\mathbf{X}$. Our network contains two graph convolution layers which include the adjacency matrix $\mathbf{A}$. The network uses an artificial neural network (ANN) for node classification. To improve detection of the EZ class, we added a separate ANN to learn a subject-specific bias term $\mathbf{s}$, which is added to the node-wise predictions $\mathbf{E}$. Our model classifies each ROI from the parcellation as either belonging to the EZ or not.

the EZ . Inspired by the rs-fMRI literature, we use a weighted similarity matrix to capture whole-brain rs-fMRI connectivity [46], [47], [54]. Our DeepEZ architecture exploits the topological properties of rs-fMRI connectivity data via a set of graph convolutions. To integrate biological knowledge, we use structural connectivity, derived from d-MRI tractography, to define the underlying graph, thus emphasizing signal propagation along anatomical pathways. DeepEZ includes a subject-specific detection bias to account for patient differences and improve generalizability. Finally, we incorporate a clinically relevant asymmetry term into our loss function to provide crucial lateralization information. Fig. 2 provides a graphical overview of our DeepEZ framework.

1) Graph Convolution Network: Formally, let N be the number of brain regions in our parcellation and T be the number of time points for a rs-fMRI scan. We define $\mathbf{m}_i \in \mathbb{R}^{T \times 1}$ as the average time series extracted from region i, normalized to have zero mean and unit variance. Following the work of [54], we construct the input functional connectivity matrix as follows:

$$\mathbf{X} = \exp\left[\mathbf{M}^T \mathbf{M} - 1\right] \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{T \times N}$ aggregates the region-wise time series $\mathbf{m}_i$ as columns. Eq. (1) ensures non-negative input values, such that anti-correlated regions have connectivity close to zero, and highly correlated regions have connectivity close to one.

As shown in Fig. 2, DeepEZ processes the input data via two spatial graph convolutions. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a binary adjacency matrix used for spatial graph filtering [86]. As described in the previous section, we use d-MRI tractography to construct $\mathbf{A}$. In this case, an entry $\mathbf{A}_{ij} = 1$ denotes an anatomical pathway connecting regions i and j. Each graph convolution produces an activation map $\mathbf{H}_l \in \mathbb{R}^{N \times F_l}$, where $l \in \{1, 2\}$ denotes the layer number. The learnable parameters in each graph convolution are a weight matrix $\mathbf{W}_l \in \mathbb{R}^{F_l \times F_{l+1}}$ and a constant bias $\mathbf{b}_l \in \mathbb{R}^{1 \times F_{l+1}}$. The activation maps are generated via the layer propagation rule:

$$\mathbf{H}_1 = \phi\left(\mathbf{A}\mathbf{X}\mathbf{W}_1 + \mathbf{b}_1\right) \tag{2}$$

$$\mathbf{H}_2 = \phi\left(\mathbf{A}\mathbf{H}_1\mathbf{W}_2 + \mathbf{b}_2\right) \tag{3}$$

The multiplication by $\mathbf{A}$ in Eqs. (2)-(3) aggregates the region-wise representation based on their direct neighborhood [86].

2) Subject-Specific Detection Bias: We treat EZ identification as a two-class classification problem, where each region i classified as either belonging to the EZ or not. Here, the output $\mathbf{H}_2$ of our GCN cascade is fed through a fully-connected layer to obtain $\mathbf{E} \in \mathbb{R}^{N \times 2}$

$$\mathbf{E} = \phi\left(\mathbf{H}_2\mathbf{W}_{fc}\right). \tag{4}$$

Similar to the graph convolutions, the weight matrix $\mathbf{W}_{fc} \in \mathbb{R}^{F_2 \times 2}$ is learned during training.

One challenge with our clinical rs-fMRI dataset is heterogeneity, both in the EZ locations and in the data acquisition procedures (e.g., scanner type). To improve detection of the EZ class, we introduce a novel concept known as subject-specific detection bias (SSDB), which helps to mitigate variation in the input data distributions. The SSDB $\mathbf{s} \in \mathbb{R}^{1 \times 2}$ is learned via a simple 2-layer artificial neural network (ANN) and is added to each row of $\mathbf{E}$ to obtain the final predictions. Mathematically, let $\{\mathbf{G}_l, \mathbf{c}_l\}$ denote the weight matrix and constant offset for each layer $l \in \{1, 2\}$ of the ANN. Our subject-specific bias term $\mathbf{s}$ is computed as follows:

$$\mathbf{s} = \phi\left(\mathbf{G}_2\phi\left(\mathbf{G}_1\mathbf{E} + \mathbf{c}_1\right) + \mathbf{c}_2\right). \tag{5}$$

An illustration of the effect of the SSDB on predicting whether a region n belongs to the EZ class is shown in the bottom right of Fig. 2. Empirically, we observe that the bias improves the sensitivity of detecting the EZ class. Following the SSDB addition, a softmax function is applied and each region is classified as belonging to the EZ or not using a max operator.

3) EZ Classification via Weighted Class Prediction and Contralateral Loss Function: There exists a large class imbalance in our dataset, as on average 7.3 % of the regions lie within the resection boundary that denotes the EZ . Since the GCN layers are designed to operate upon a whole-brain connectivity matrix, traditional data augmentation techniques would not solve our class imbalance problem. Following the work of [55], [56], we train our model with a modified Risk-Sensitive Cross-Entropy loss function [74], which is designed to handle a class membership imbalance. Formally, let δ_i be

TABLE II
HYPERPARAMETERS DETERMINED VIA CROSS VALIDATION ON A
SEPARATE COHORT DRAWN FROM THE HCP DATASET.

| Parameter	Value	Parameter	Value
(δ_1, δ_2)	$(0.29, 1.52)$	λ	0.017
lr	0.005	Epochs	200
ϵ	1×10^{-8}	wd	5×10^{-5}

Fig. 3. We use repeated 7-fold CV for model training and testing. We report the mean and standard deviation of the sensitivity, specificity, AUC, and accuracy across runs. For each baseline, we report the FDR corrected p-value that DeepEZ achieves significantly higher AUC.

the risk associated with class i. If δ_i is large, then we pay a larger penalty for misclassifying samples belonging to class i.

Beyond the class imbalance, it has been shown that contralateral areas of the brain have high rs-fMRI correlation [37], [64], often causing them to be treated similarly in downstream analyses. In contrast, we expect the EZ to be lateralized [14], [36]. We leverage this asymmetry in the second term of our DeepEZ loss function by specifying that regions contralateral to the predicted EZ should be classified as normal.

Let N_e denote the nodes that belong to the EZ (labeled without loss of generality as class #2), and let $c(n)$ denote the contralateral counterpart to region n. Our training loss function consists of the following two terms:

$$\mathcal{L} = -\underbrace{\sum_{n=1}^{N}\sum_{i=1}^{2} \delta_i y_{n,i} \log \hat{y}_{n,i}}_{\text{Weighted Cross Entropy}} - \lambda \underbrace{\frac{1}{N_e}\sum_{n \in N_e}\left(\hat{y}_{n,2} - \hat{y}_{c(n),2}\right)}_{\text{EZ Contralateral Term}}. \quad (6)$$

The quantities $\hat{y}_{n,i}$ in Eq. (6) denote the DeepEZ prediction for the baseline ($i = 1$) and EZ ($i = 2$) classes at each region n. As seen, the first term of Eq. (6) accounts for the class imbalance, and the second term enforces hemispheric asymmetry in the final EZ predictions. Finally, λ balances the contributions of the two loss terms.

4) Implementation Details: We implement DeepEZ in PyTorch [57] using the Adam optimizer with weight decay (wd) and ϵ for regularization. The LeakyReLU (x) $= \max(0, x) + 0.1 \cdot \min(0, x)$ activation function is applied at each hidden layer of the network in Fig. 2. A softmax activation is applied at the final layer for region-wise classification.

To prevent undue bias, we tune the hyperparameters $\delta_1, \delta_2, \lambda$ in Eq. (6) and the Adam optimization routine based on 50 subjects drawn from the HCP dataset. Specifically, we randomly selected a portion of the brain regions in each subject to denote an "artificial EZ". We then use 10-fold cross validation to fix the hyperparameters used in all experiments. For δ_1 and δ_2, we performed a coarse grid search from $0 - 10$ in increments of of 10^{-1} until we found a suitable range of performance. We then performed a finer grid search in increments of 10^{-2}. For the parameter λ, we performed a fine grid search over $0 - 1$ in increments of 10^{-3}. Table II reports the resulting parameter values for our final DeepEZ implementation.

C. Baseline Methods

We compare DeepEZ against three competing methods from the literature and seven ablated versions of our model:

- ICA approach of [36] (ICA1)
- ICA approach of [85] (ICA2)
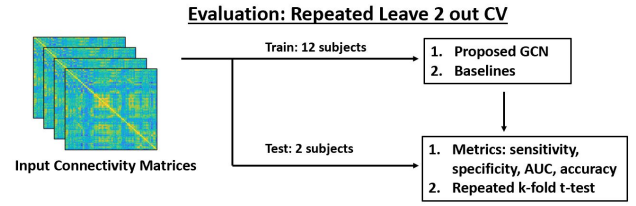
- The BrainNetCNN [43] adopted for region-wise classification (BN-CNN)
- Ablation #1: No SCT and No SSDB (GCN)
- Ablation #2: No SSDB (GCN-SCT)
- Ablation #3: No SCT (GCN-SSDB)
- Ablation #4: DeepEZ with an identity matrix replacing $\mathbf{A}$ (GCN-I)
- Ablation #5: DeepEZ with patient-specific d-MRI matrices replacing $\mathbf{A}$ (GCN-$\mathbf{A}_{subj}$)
- Ablation #6: DeepEZ with a randomly sampled matrix replacing $\mathbf{A}$ (GCN-$\mathbf{A}_{rand}$)
- Ablation #7: DeepEZ with topology preserved matrix replacing $\mathbf{A}$ (GCN-$\mathbf{A}_{top}$)

The first baseline is a traditional machine learning approach for EZ localization described in [36] Specifically, the ICA method extracts features from each independent component (IC) and then classifies each IC as belonging to the EZ or not via a linear support vector machine. The IC farthest from the boundary is selected as the final EZ for that patient. We chose this ICA baseline because it performed the best in the meta-analysis of [15], which compares seven different ICA methods for EZ localization using rs-fMRI. The second baseline described in [85] employs a screening process where ICs are sequentially discarded based on rules such as contralateral correlation and power spectrum density and the remaining ICs are considered belonging to the EZ class. We note that the original method is not fully automated. For example, visual inspection was used on a subject that had multiple independent components (ICs) that survived the rule-based screening process. In an effort to provide fair comparison across methods, we automate the work of [85] by combining the predictions of any ICs that pass the rule-based screening.

The third baseline is a modified version of the BrainNetCNN model developed in [43], which was originally designed to regress cognitive scores from structural connectivity matrices. The BrainNetCNN architecture uses cross-shaped convolutional filters to leverage topological relationships in connectivity data. We have modified the final layers of the original architecture to perform region-wise classification input rather than patient-level phenotypic prediction.

Our next three baselines focus on the EZ contralateral term (SCT) and the subject specific detection bias (SSDB) in our DeepEZ framework. As seen, we systematically ablate the components to determine the performance gain derived from

TABLE III

MEAN PLUS OR MINUS STANDARD DEVIATION FOR SENSITIVITY, SPECIFICITY, PRECISION, F1-SCORE, ACCURACY, AND AUC. THE T-SCORE COMPARES THE AUC OF DEEPEZ WITH EACH BASELINE; WE ALSO NOTE THE CORRESPONDING FDR CORRECTED P-VALUE. WE USE A TWO-SAMPLE T-TEST BASED ON THE REPEATED 7-FOLD CV FOR ALL MODELS, EXCEPT ICA2; HERE, WE USE A ONE-SAMPLE T-TEST.

Method	Sensitivity	Specificity	Precision	F1	Accuracy	AUC	t-score	p-value
ICA1	0.071 ± 0.034	0.78 ± 0.045	0.09 ± 0.035	0.08 ± 0.029	0.57 ± 0.033	0.52 ± 0.023	22.26	$< 10^{-10}$
ICA2	0.25	0.7	0.31	0.28	0.69	0.6	7.13	$< 10^{-10}$
BN-CNN	0.099 ± 0.048	0.72 ± 0.036	0.11 ± 0.046	0.11 ± 0.027	0.78 ± 0.015	0.56 ± 0.035	11.84	$< 10^{-10}$
GCN	0.17 ± 0.051	0.78 ± 0.032	0.26 ± 0.039	0.20 ± 0.038	0.81 ± 0.032	0.62 ± 0.035	7.66	$< 10^{-10}$
GCN-SCT	0.22 ± 0.059	0.81 ± 0.036	0.29 ± 0.042	0.25 ± 0.041	0.83 ± 0.029	0.65 ± 0.038	5.13	8.6×10^{-7}
GCN-SSDB	0.31 ± 0.056	0.83 ± 0.031	0.43 ± 0.048	0.36 ± 0.039	0.85 ± 0.03	0.70 ± 0.034	2.15	0.023
GCN-I	0.27 ± 0.061	0.86 ± 0.034	0.39 ± 0.037	0.31 ± 0.028	0.87 ± 0.041	0.68 ± 0.033	3.69	3.3×10^{-4}
GCN-$\mathbf{A}_{subj}$	0.28 ± 0.046	0.87 ± 0.029	0.37 ± 0.039	0.32 ± 0.034	0.89 ± 0.031	0.70 ± 0.026	2.03	0.031
GCN-$\mathbf{A}_{rand}$	0.33 ± 0.041	0.86 ± 0.031	0.41 ± 0.042	0.37 ± 0.037	0.88 ± 0.034	0.71 ± 0.028	1.91	0.046
GCN-$\mathbf{A}_{top}$	0.35 ± 0.039	0.86 ± 0.035	0.47 ± 0.038	0.4 ± 0.035	0.88 ± 0.036	0.72 ± 0.019	1.63	0.078
DeepEZ	$\mathbf{0.4 \pm 0.044}$	0.85 ± 0.033	$\mathbf{0.52 \pm 0.039}$	$\mathbf{0.45 \pm 0.041}$	0.88 ± 0.034	$\mathbf{0.73 \pm 0.031}$		

each one. The last four baselines use the same architecture and loss function but vary the anatomical connectivity matrix $\mathbf{A}$ in the spatial graph convolutions. The GCN-I baseline replaces $\mathbf{A}$ with the identity matrix, effectively removing the anatomical regularization from our model. The GCN-$\mathbf{A}_{subj}$ baseline replaces $\mathbf{A}$ with personalized graphs computed from the patient-specific d-MRI. Due to variations in the data and tractography outputs, the edges in $\mathbf{A}_{subj}$ vary across patients, i.e., the graphs are slightly mismatched. The GCN-$\mathbf{A}_{rand}$ baseline replaces $\mathbf{A}$ with a randomly-sampled symmetric matrix $\mathbf{A}_{rand}$. Finally, the GCN-$\mathbf{A}_{top}$ baseline replaces $\mathbf{A}$ with a matrix that reflects the same geometric topology as $\mathbf{A}$. To obtain $\mathbf{A}_{top}$, we first bin the edge weights from the unthresholded version of $\mathbf{A}$ and then randomly shuffle edges within each bin [62]. We then threshold to obtain $\mathbf{A}_{top}$. Similar to the proposed model, we kept $\mathbf{A}_{rand}$ and $\mathbf{A}_{top}$ fixed across patients within each CV fold. For robustness, we run a repeated 7-fold CV procedure for all methods, and we sample the matrix $\mathbf{A}_{rand}$ 50 times and report the average statistics.

III. EXPERIMENTS AND RESULTS

A. EZ Detection Performance

We evaluate the performance of each method using a repeated 7-fold CV setup, where each fold contains 2 subjects and each repeated sampling (i.e., run) ensures different fold membership. Fig. 3 shows the evaluation workflow of our experiments. We report the mean and standard deviation performance across 90 unique runs for the following metrics: sensitivity (TPR), specificity (TNR), accuracy, and area under the receiver operating characteristic curve (AUC). To demonstrate a statistically significant performance gain, we perform a t-test on the AUC metric comparing each baseline with DeepEZ. The test statistic corrects for dependencies between the resampled folds, as outlined in [12]. Using this statistic, we compute a p-value and apply FDR correction. Since the ICA2 method is based on deterministic rules and not learned by a classifier, the results are the same across CV folds. Thus, we report only a single average across patients for each metric, as opposed to a mean $\pm$ standard deviation. We use a one-sample t-test to determine statistical significance for ICA2.

Table III summarizes the EZ detection performance for each method. We observe that DeepEZ achieves the highest sensitivity, precision, F1 and AUC. While the specificity is slightly lower than the GCN-$\mathbf{A}_X$ baselines that swap out the HCP matrix $\mathbf{A}$, the performances are within a standard deviation. Minor variations in accuracy can also be attributed to the class imbalance between EZ and non-EZ regions. The performance gains of DeepEZ are underscored by the AUC t-test, where we observe a statistically significant ($p < 0.05$) improvement for DeepEZ over all baseline methods except for GCN-$\mathbf{A}_{top}$ method. This result highlights that the performance gain of DeepEZ can be largely attributed to the graph topology (e.g., small-world connectivity) when fusing the rs-fMRI connectivity information across layers.

We note that the competing ICA1 and BN-CNN methods are not well-suited to the task, possibly due to the heterogeneity of our clinical cohort. While ICA2 performs much better than ICA1, it cannot match the performance of DeepEZ. One issue with the rule-based ICA methods is that the selection criteria on one dataset may not generalize well to another. While an end-to-end model such as DeepEZ can easily be retrained on new data, modifying a rule-based approach is nontrivial. The ablated models perform slightly better than ICA1, ICA2 and BN-CNN, but still not on par with DeepEZ. In fact, we observe a notable performance gain using the SSDB, which suggests that a subject-specific approach may be useful to overcome heterogeneity in clinical prediction tasks. There is a similar performance gain when incorporating the contralateral loss term (SCT), which emphasizes the asymmetry associated with our problem. We observe a marked decline in sensitivity when replacing the d-MRI connectivity matrix $\mathbf{A}$ with identity. This suggests that using information about anatomical pathways is crucial for EZ localization. Interestingly, we also note a performance decline when using the patient-specific d-MRI information encoded in $\mathbf{A}_{subj}$. We hypothesize that this is due to the inconsistency of the edges across patients, particularly in our small dataset. This hypothesis is supported by the results for GCN-$\mathbf{A}_{rand}$ and GCN-$\mathbf{A}_{top}$, which performs slightly better than GCN-$\mathbf{A}_{subj}$. Recall that, while random, the graphs in GCN-$\mathbf{A}_{rand}$ and GCN-$\mathbf{A}_{top}$ are fixed. We find that model training is more stable when the same $\mathbf{A}$ matrix is used for each patient. Finally, GCN-$\mathbf{A}_{top}$ performs the best out of the baselines, implying a benefit to using topological information.
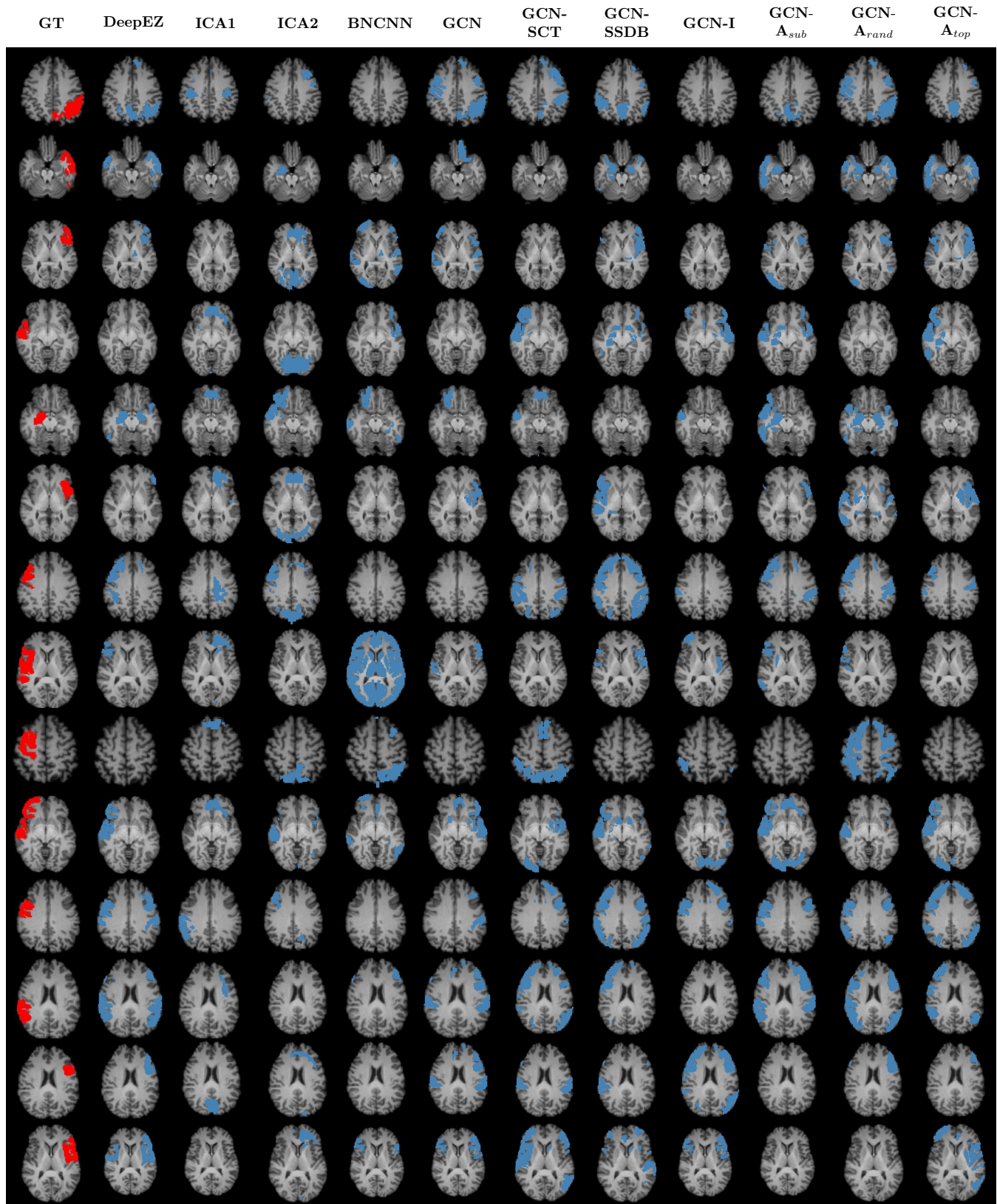
Fig. 4. Axial view of ground truth (red) and model predictions (blue) for all patients in the UW Madison dataset. Each row corresponds to a single patient, organized from 1–14 according to Table I. Model names are displayed at the top of each column.

Fig. 4 illustrates axial views of the ground truth (red) and predicted (blue) labels for all 14 patients across each method considered. Each row represents a patient (numbered 1-14 in Table I) and each column represents one method. As shown, DeepEZ localizes correct regions in most patients while omitting non-EZ regions. An example of this can be see for Patient 1 (first row), where DeepEZ aligns well with the ground truth labels, while avoiding the spurious predictions by the GCN-X methods. Furthermore, DeepEZ is the only method to localize the resections in Patient 3 and Patient 10, while not incurring incorrect contralateral predictions. We observe that the ICA and BNCNN methods are poorly suited for the task and rarely produce correct predictions. Overall, we observe similar predictions made by the DeepEZ and GCN-$\mathbf{A}_{top}$ methods, likely due to the preserved network topology. Overall, we observe that DeepEZ achieves a good balance between correct localization while avoiding false positives.

Finally, we visualize the most commonly learned graph convolutional filters of DeepEZ. Accordingly, we extracted the first-stage graph convolution weight matrices $\mathbf{W}_1 \in \mathbb{R}^{246 \times 120}$ learned during each of our repeated 7-fold CV runs. Each column of $\mathbf{W}_1$ represents a different convolutional filter that can be visualized by plotting the magnitude of the weights back on to the brain. We first aligned the columns of $\mathbf{W}_1$ across the repeated CV folds using a Procrustes algorithm and then masked each column to identify the top ten regions implicated by that filter. We identified two activation patterns that were consistently learned by DeepEZ. Fig. 5 shows these filters projected onto the cortical surface, where the color denotes the average activation across repeated CV folds.

We note that one filter (left) implicates regions within the temporal lobe while the other filter (right) implicates the regions associated with the frontal lobe. These patterns mimic the distribution of EZ labels in our UW Madison dataset. Thus, in a data-driven manner, DeepEZ focuses its analysis on stereotypical patterns associated with the EZ.

B. Assessing Model Robustness

In this section, we assess the robustness of DeepEZ to four aspects of our experimental setup: (1) the choice of brain parcellation, (2) the number of GCN layers , (3) the relative weighting δ_2 for the EZ class in Eq. (6), and (4) the small dataset size used for model training.

1) Varying Parcellation Choice: It has been shown that the choice of parcellation can have a tremendous impact on rs-fMRI analyses [21], [48]. For example, coarse parcellations mitigate the effects of noise but can blur subtle effects, whereas fine parcellations preserve detailed phenomena but can be overwhelmed by environmental confounds. In addition, the Brainnetome atlas (BNA) used in Table III is symmetric where each region has a direct contralateral region, which is not the case with all parcellations. To explore this, we apply DeepEZ using three different scales of the Craddocks functional parcellation [18]. The Craddocks atlas was derived using a spectral clustering algorithm on the rs-fMRI data from healthy subjects. The different scales come from varying the number of clusters. In this work, we use the $N = 178, N =$

TABLE IV
MEAN PLUS OR MINUS STANDARD DEVIATION FOR ACCURACY AND AUC FOR DIFFERENT PARCELLATIONS. THE FINAL COLUMN SHOWS THE FDR CORRECTED P-VALUES WHEN COMPARING THE AUCs OF THE BNA AND CRADDOCKS (CC) PARCELLATIONS.

Atlas	Accuracy	AUC	p-val
BNA-246	0.88 ± 0.034	0.73 ± 0.031	
CC-178	0.85 ± 0.037	0.70 ± 0.032	0.018
CC-318	0.87 ± 0.038	0.72 ± 0.039	0.367
CC-384	0.87 ± 0.035	0.72 ± 0.031	0.312

318, and $N = 384$ scales, which include both coarser and finer parcellations than the BNA $N = 246$ atlas to assess the effect that resolution has on performance. Once again, we use repeated 7-fold CV to quantify performance.

Table IV reports the accuracy and AUC when applying DeepEZ to each of the parcellations defined above. The p-values are computed with respect to the original BNA atlas. To account for the fact that regions in the Craddocks atlases are not symmetrically defined across the hemispheres, our SCT loss function considers the region with centroid closest to the contralateral location as the counterpart $c(n)$ in Eq. (6). Based on a $p < 0.05$ threshold, we only observe a significant performance difference in AUC with the CC-178 atlas. This observation suggests that a finer parcellation is better suited for EZ localization. In contrast, DeepEZ is robust using either the CC-318, CC-384, or BNA atlas, which suggests model stability across different parcellations.

2) Number of GCN Layers: At a high level, the GCN layers of DeepEZ perform a random walk on the brain graph defined by the anatomical connections in d-MRI. Our choice of two GCN layers in DeepEZ can analyze the rs-fMRI connectivity patterns associated with path lengths ≤ 2 but cannot capture higher-order information. To probe this effect, we conduct a robustness experiment in which we vary the number of GCN layers in DeepEZ and use the repeated 7-fold CV strategy in Fig. 3 to quantify the performance of each method.

Table V reports the performance across 1–4 GCN layers. We observe that the proposed architecture (2 GCN layers) achieves the best trade-off between true positive and false positive detections, as quantified via a t-test on the AUC. There are two interpretations for this result. First, it appears that the rs-fMRI connectivity patterns associated with the EZ are the most prominent at a walk of length of two, with diminishing returns beyond this point. Second, increasing the number of GCN layers also increases the number of model parameters, which may lead to overfitting. Taken together, we believe that two GCN layers balances the trade-off between capturing discriminative patterns without overfitting on small datasets.

3) EZ Detection Hyperparameter: One of the key aspects of DeepEZ is the weighted cross-entropy loss to handle class imbalance. To probe this effect, we sweep the EZ detection hyperparameter δ_2 in in increments of 0.1 while keeping the other hyperparameters fixed. Fig. 6 shows the sensitivity (left) and AUC (right) metrics as δ_2 varies over the range $[1.1, 2.0]$. As expected, sensitivity increases with δ_2 due to the higher penalty for incorrectly classifying EZ regions as baseline.
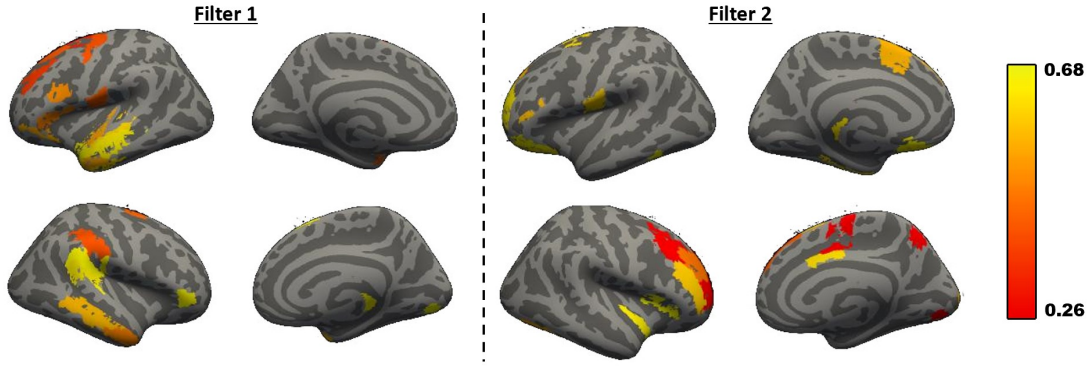
Fig. 5.   Regions implicated by the two most commonly learned convolutional filters in DeepEZ . One filter (**L**) identifies the temporal lobe while the other (**R**) identifies the frontal regions. The distribution of these regions mimic the EZ labels in our dataset.
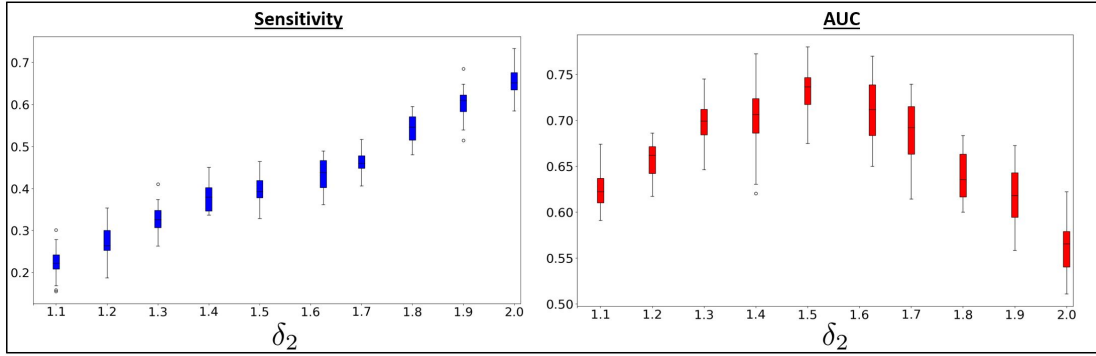

Fig. 6.   Sensitivity (left) and AUC (right) for proposed method while sweeping δ_2 in increments of 0.1. We observe as δ_2 increases, sensitivity increases, but AUC eventually decreases.

TABLE V
LOCALIZATION PERFORMANCE AS THE NUMBER OF GCN LAYERS IS VARIED. THE T-SCORE COMPARES THE AUC OF THE PROPOSED DEEPEZ WITH EACH BASELINE; WE ALSO NOTE THE CORRESPONDING FDR CORRECTED P-VALUE.

Layers	Sensitivity	Specificity	Precision	F1	Accuracy	AUC	t-score	p-value
1	0.09 ± 0.035	0.93 ± 0.021	0.12 ± 0.032	0.1 ± 0.029	0.91 ± 0.039	0.55 ± 0.032	12.04	$< 10^{-10}$
2 (Proposed)	$\mathbf{0.4 \pm 0.044}$	0.85 ± 0.033	$\mathbf{0.52 \pm 0.039}$	$\mathbf{0.45 \pm 0.041}$	0.88 ± 0.034	$\mathbf{0.73 \pm 0.031}$		
3	0.27 ± 0.042	0.87 ± 0.035	0.38 ± 0.039	0.32 ± 0.037	0.88 ± 0.031	0.70 ± 0.027	2.01	0.039
4	0.33 ± 0.039	0.84 ± 0.029	0.45 ± 0.041	0.38 ± 0.029	0.83 ± 0.028	0.68 ± 0.029	3.6	4.9×10^{-4}

However, the AUC metric peaks at 1.5 and steadily decreases, which suggests that DeepEZ incurs too many false positives at larger values of δ_2. We observe relatively stable performance in both metrics over the range $\delta_2 = [1.4-1.6]$. Finally, we note that the weighted cross-entropy loss is useful from a clinical perspective, as it is more important not to miss the EZ regions at this stage of therapeutic planning for epilepsy.

4) Training Data Augmentation: Data augmentation has been shown to improve the performance of deep learning models due to providing more information about the underlying data distribution [59], [61]. Given the small sample size ($N = 14$), we explore whether DeepEZ would benefit from data augmentation. Here, we sub-sampled the time series data using a continuous sliding window to create 10 distinct new training similarity matrices for each subject. Our augmented dataset contains an order of magnitude more data points ($N = 140$). Our evaluation strategy remained the same as depicted in Fig. 3. Table VI reports the EZ detection performance for DeepEZ with and without data augmentation. We observe a

small performance boost in each metric and smaller standard deviations when using data augmentation during training. However, we note that the performance gain in AUC is not statistically significant, as indicated in the last column of Table VI. This result demonstrates that DeepEZ is able to effectively mine the information present in our original dataset for generalizable EZ localization.

IV. DISCUSSION

We have introduced DeepEZ, a novel deep learning approach for EZ localization based on rs-fMRI connectivity. DeepEZ relies on spatial graph convolutions that leverage biologically-inspired anatomical pathways to aggregate neighborhood information during forward propagation. These graph convolutions are complemented with a subject-specific detection bias (SSDB) to mitigate inter-patient differences in connectivity and an asymmetry loss term to encourage lateralized predictions. In comparison to baseline methods,

TABLE VI

MEAN PLUS OR MINUS STANDARD DEVIATION FOR SENSITIVITY, SPECIFICITY, AUC, AND ACCURACY WITH AND WITHOUT DATA AUGMENTATION. THE FINAL COLUMN SHOWS THE FDR CORRECTED P-VALUES FOR THE ASSOCIATED T-SCORE FOR COMPARING AUC.

Augmentation	Sensitivity	Specificity	Precision	F1	Accuracy	AUC	t-score	p-value
Without	0.4 ± 0.044	0.85 ± 0.033	0.52 ± 0.039	0.45 ± 0.041	0.88 ± 0.034	0.73 ± 0.031		
With	0.43 ± 0.033	0.87 ± 0.026			0.89 ± 0.015	0.74 ± 0.019	-1.28	0.76

DeepEZ achieves statistically improved AUC for detecting EZ regions. Via ablation studies, we show that these performance gains are linked to the EZ contralateral term (SCT) and the SSDB. In subsequent anaylses, we demonstrate that DeepEZ is robust to varying the parcellation used for analysis and performs comparably with and without data augmentation. We also show that DeepEZ achieves robust performance within a range of δ_2 in the weighted cross-entropy loss term.

Clinical rs-fMRI studies often lack statistical power due to small sample sizes [1] with logistical constraints making it difficult to acquire additional data for analysis. In our case, the UW Madison dataset contains a specialized cohort of pediatric focal epilepsy patients who underwent surgical resection of the EZ . Currently, rs-fMRI is not a commonly acquired modality for epilepsy patients, which limits our ability to grow the dataset further. Thus, to maximize the sample size for model training and evaluation, we include patients that have their scans taken from two different scanners (1.5T and 3T in Table I) which is common in the literature [49], [70], [77]. Accordingly, we have designed DeepEZ to mine information from smaller heterogeneous datasets. Relevant attributes include a relatively small number of learnable parameters, a biologically informed spatial graph, the SSDB module to improve sensitivity, and a SCT to encourage clinically relevant EZ localization patterns. We demonstrate state-of-the-art in Section 3, along with a robustness to different modeling choices and data augmentation. The fact that DeepEZ harmonizes information across scanners is particularly encouraging, as inter-scanner differences are known to confound deep learning models [39]. Overall, these results increase our confidence in DeepEZ as a prospective clinical tool for epilepsy evaluation.

Perhaps the most interesting result of this work is that using subject-specific anatomical connectivity information does not improve localization performance. In fact, the sensitivity, precision, F1 and AUC are worse than when the graph $\mathbf{A}$ is fixed according to the normative HCP dataset. There are three different facets to this result. From an imaging standpoint, there is more variability in our UW Madison dataset due to scanner differences (e.g., 1.5T versus 3T), patient age (9–18 years), and heterogeneous pathologies. This variability may lead to "incorrect" tractography outputs, as compared to the underlying neurophysiology. In contrast, the HCP acquisition sequences have been carefully validated on a standardized cohort, and the preprocessing pipelines have been optimized for the data. Consequently, the matrices $\mathbf{A}$ may reflect long-range and distributed anatomical pathways more accurately than $\mathbf{A}_{subj}$. From an optimization standpoint, the edges are inconsistent across the patient-specific matrices $\mathbf{A}_{subj}$. This inconsistency can lead to instability during training, particularly given the small sample size ($S = 14$). Future work will

include comparing the influence of $\mathbf{A}$ and $\mathbf{A}_{subj}$ as the dataset grows in size. Finally, from a neurobiological standpoint, our robustness study in Section III-B suggests that two GCN layers is optimal for EZ localization. Thus, DeepEZ only fuses information across two-stage pathways. Given that we operate at the region level, the HCP template may be sufficient for this operation without needing patient-specific connectivity.

In line with the above observation, our experiments demonstrate that replacing the anatomical connectivity matrix with $\mathbf{A}_{top}$ achieves similar performance. This result underscores the importance of *network topology* over individual anatomical connections. It also suggests that DeepEZ is robust to variations in the anatomical connectivity matrix used for $\mathbf{A}$. Thus, we conclude that acquiring rs-fMRI data alone is sufficient for EZ localization, which reduces the logistical burden of integrating DeepEZ into the clinical workflow.

We note that there is considerable prior work that uses ICA for EZ localization in rs-fMRI [11], [32], [36]. The meta-analysis of [15] determined that among these, the machine learning approach of [36] (baseline 1 in this work) achieves the best odds ratio. However, as reported in Table III, this method fails to localize the EZ for our cohort. One possible reason is that the handcrafted features chosen in [36] may not generalize well to different cohorts. Another drawback of this method is that it selects just one independent component as the EZ , when there might exist multiple epileptic sources across overlapping components [35]. In fact, we observed in our experiments that the independent components did not overlap well with the surgical resection boundaries used as the pseudo ground truth EZ . We also note that since this method performs ICA at the subject level, there is substantial variability in the component locations across patients. In contrast, DeepEZ uses a well-defined functional parcellation, which allows for both a fine resolution analysis and group-level concordance in the region definitions. We also demonstrate that DeepEZ is robust to the choice of parcellation, which gives the user more flexibility when applying the framework to clinical data.

Accurate EZ localization through noninvasive imaging can significantly impact epilepsy treatment. Concordance between individual modalities (EEG, structural MRI) improves the prognostic yield of first tier evaluation of patients with epilepsy. In addition, the odds of seizure freedom following surgical treatment in medication refractory epilepsy are associated with correct identification of an epileptogenic lesion [40]. Hence, in non-lesional cases, successful delineation of a EZ based on rs-fMRI connectivity can improve treatment outcomes by better identifying surgical resection targets [22]. One limitation of this paper is that, as our data was acquired by the UW Madison health system during routine preoperative evaluation, we do not have access to healthy controls or

multi-focal epilepsy patients (who would not be considered surgical candidates) for model training. However, acquiring and integrating this data into DeepEZ remains a crucial direction of future work. Finally, we emphasize that our deep learning framework can be adapted to other applications, such as preoperative mapping of the eloquent cortex [54]–[56].

V. CONCLUSION

We have introduced a novel deep learning based method to localize the seizure onset zone in focal epilepsy patients. Our DeepEZ framework combines the representational power of deep neural networks with domain specific modeling choices, such as asymmetry of the EZ and anatomical regularization. We show that DeepEZ outperforms an existing ICA method from the literature and various deep learning baselines in localizing the EZ. We also demonstrate the robustness of our framework to various modeling choices, such as parcellation and hyperparameter choice. Our results provide evidence for EZ identification in epilepsy using noninvasive rs-fMRI based imaging that has the potential to improve the diagnostic work up and therapeutic planning in treatment of epilepsy.

AUTHOR CONTRIBUTIONS

AV and NN conceived and designed the experiments. NN implemented the model and performed the analyses. AV DH and RA contributed data and resources. AV and NN drafted the manuscript. DH and RA edited the manuscript.

REFERENCES

[1] Bhim M Adhikari et al. A resting state fmri analysis pipeline for pooling inference across diverse cohorts: an enigma rs-fmri protocol. *Brain imaging and behavior*, 13(5):1453–1467, 2019.

[2] S. Adler et al. Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *NeuroImage: Clinical*, 14:18–27, 2016.

[3] B. Ahmed et al. Cortical feature analysis and machine learning improves detection of "MRI-negative" focal cortical dysplasia. *Epilepsy & Behavior*, 48:21–28, 2015.

[4] David Ahmedt-Aristizabal et al. 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.

[5] Katharina Allers et al. The economic impact of epilepsy: a systematic review. *BMC neurology*, 15(1):1–16, 2015.

[6] S.B Antel et al. Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. *NeuroImage*, 19(4):1748 – 1759, 2003.

[7] Anto I. Bagic and Richard C. Burgess. Utilization of meg among the us epilepsy centers: A survey-based appraisal. *Journal of Clinical Neurophysiology*, 37(6), 2020.

[8] P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *Journal of Magnetic Resonance*, 111:209–219, 1996.

[9] Colin D Binnie and Charles E Polkey. Commission on neurosurgery of the international league against epilepsy (ilae) 1993–1997: recommended standards. *Epilepsia*, 41(10):1346–1349, 2000.

[10] Ingmar Blümcke et al. The clinicopathologic spectrum of focal cortical dysplasias: A consensus classification proposed by an ad hoc task force of the ILAE diagnostic methods commission. *Epilepsia*, 52(1):158–174, 2011.

[11] Varina L Boerwinkle et al. Correlating resting-state functional magnetic resonance imaging connectivity by independent component analysis-based epileptogenic zones with intracranial electroencephalogram localized seizure onset zones and surgical outcomes in prospective pediatric intractable epilepsy study. *Brain connectivity*, 7(7):424–442, 2017.

[12] Remco R Bouckaert et al. Evaluating the replicability of significance tests for comparing learning algorithms. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 3–12. Springer, 2004.

[13] W.J. Cadarelli and B.J. Smith. The burden of epilepsy to patients and payers. *American Journal of Managed Care*, 16(12):S331–336, 2010.

[14] Maria Centeno and David W Carmichael. Network connectivity in epilepsy: resting state fmri and eeg–fmri contributions. *Frontiers in neurology*, 5:93, 2014.

[15] Arpan R Chakraborty et al. Resting-state functional magnetic resonance imaging with independent component analysis for presurgical seizure onset zone localization: A systematic review and meta-analysis. *Epilepsia*, 61(9):1958–1968, 2020.

[16] O. Colliot et al. Individual voxel-based analysis of gray matter in focal cortical dysplasia. *NeuroImage*, 29(1):162 – 171, 2006.

[17] C.J. Cook et al. Effective connectivity within the default mode network in left temporal lobe epilepsy: Findings from the epilepsy connectome project. *Brain Connectivity*, 9(2):174–183, 2019.

[18] R Cameron Craddock et al. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

[19] A. Crouzeix et al. An evaluation of dipole reconstruction accuracy with spherical and realistic head models in MEG. *Clinical Neurophysiology*, 110:2176–2188, 1999.

[20] B.N. Cuffin. Eeg localization accuracy improvements using realistically shaped head models. *IEEE Trans Biomedical Engineering*, 43:299–393, 1996.

[21] Marcel A de Reus et al. The parcellation-based connectome: limitations and extensions. *Neuroimage*, 80:397–404, 2013.

[22] Jane De Tisi et al. The long-term outcome of adult epilepsy surgery, patterns of seizure remission, and relapse: a cohort study. *The Lancet*, 378(9800):1388–1395, 2011.

[23] F. Deleo et al. Histological and MRI markers of white matter damage in focal epilepsy. *Epilepsy Research*, 140:29 – 38, 2018.

[24] D. Ding et al. Cerebral arteriovenous malformations and epilepsy, part 1: Predictors of seizure presentation. *World Neurosurgery*, 84(3):645 – 652, 2015.

[25] Gaëlle E Doucet et al. Early and late age of seizure onset have a differential impact on brain resting-state organization in temporal lobe epilepsy. *Brain topography*, 28(1):113–126, 2015.

[26] Rekha Dwivedi et al. Surgery for drug-resistant epilepsy in children. *New England Journal of Medicine*, 377(17):1639–1647, 2017.

[27] M. El Azami et al. Detection of lesions underlying intractable epilepsy on t1-weighted mri as an outlier detection problem. *PLOS ONE*, 11(9):1–21, 2016.

[28] JVNP Engel Jr. Outcome with respect to epileptic seizures. *Surgical treatment of the epilepsies*, pages 609–621, 1993.

[29] Lingzhong Fan et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, 26(8):3508–3526, 2016.

[30] C Garcia-Ramos et al. Low functional robustness in mesial temporal lobe epilepsy. *Epilepsy research*, 123:20–28, 2016.

[31] B. Garcin et al. Epileptic seizures at initial presentation in patients with brain arteriovenous malformation. *Neurology*, 78(9):626–631, 2012.

[32] Francisco Gil et al. Beyond the epileptic focus: functional epileptic networks in focal epilepsy. *Cerebral Cortex*, 30(4):2338–2357, 2020.

[33] Michael D Greicius, Kaustubh Supekar, Vinod Menon, and Robert F Dougherty. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex*, 19(1):72–78, 2009.

[34] S.-J. Hong et al. Automated detection of cortical dysplasia type ii in mri-negative epilepsy. *Neurology*, 83(1):48–55, 2014.

[35] Borbála Hunyadi et al. Ica extracts epileptic sources from fmri in eeg-negative patients: a retrospective validation study. *PloS one*, 8(11):e78796, 2013.

[36] Borbála Hunyadi et al. A prospective fmri-based technique for localising the epileptogenic zone in presurgical evaluation of epilepsy. *Neuroimage*, 113:329–339, 2015.

[37] R Matthew Hutchison et al. Functional connectivity of the frontal eye fields in humans and macaque monkeys investigated with resting-state fmri. *Journal of neurophysiology*, 107(9):2463–2474, 2012.

[38] G. Hwang et al. Using low-frequency oscillations to detect temporal lobe epilepsy with machine learning. *Brain Connectivity*, 9(2):184–193, 2019.

[39] Madhura Ingalhalikar et al. Functional connectivity-based prediction of autism on site harmonized abide dataset. *IEEE Transactions on Biomedical Engineering*, 2021.

[40] Barbara C Jobst and Gregory D Cascino. Resective epilepsy surgery for drug-resistant focal epilepsy: a review. *Jama*, 313(3):285–293, 2015.

[41] Eunji Jun et al. Identifying resting-state effective connectivity abnormalities in drug-naïve major depressive disorder diagnosis via graph convolutional networks. *Human Brain Mapping*, 41(17):4997–5014, 2020.

[42] V. Jurcak, D. Tsuzuki, and I. Dan. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, 34:1600–1611, 2007.

[43] Jeremy Kawahara et al. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

[44] Gregory Kiar et al. A comprehensive cloud framework for accurate and reliable human connectome estimation and meganalysis. *bioRxiv*, page 188706, 2017.

[45] M.O. Krucoff et al. Rates and predictors of success and failure in repeat epilepsy surgery: A meta-analysis and systematic review. *Epilepsia*, 58(12):2133–2142, 2017.

[46] Georg Langs et al. Functional geometry alignment and localization of brain areas. In *Advances in neural information processing systems*, pages 1225–1233, 2010.

[47] Georg Langs et al. Decoupling function and anatomy in atlases of functional connectivity patterns: Language mapping in tumor patients. *Neuroimage*, 103:462–475, 2014.

[48] Anton Lord et al. Brain parcellation choice affects disease-related topology differences increasingly from global to local network levels. *Psychiatry Research: Neuroimaging*, 249:12–19, 2016.

[49] Katherine T Martucci et al. The posterior medial cortex in urologic chronic pelvic pain syndrome: detachment from default mode network. a resting-state study from the mapp research network. *Pain*, 156(9):1755, 2015.

[50] Mohsen Mazrooyisebdani et al. Graph theory analysis of functional connectivity combined with machine learning approaches demonstrates widespread network differences and predicts clinical variables in temporal lobe epilepsy. *Brain connectivity*, 10(1):39–50, 2020.

[51] J.-J. Mo et al. Clinical value of machine learning in the automated detection of focal cortical dysplasia using quantitative multimodal surface-based features. *Frontiers in Neuroscience*, 12:1008, 2019.

[52] M. Mohan et al. The long-term outcomes of epilepsy surgery. *PLOS ONE*, 5:1–16, 2018.

[53] Victoria L Morgan et al. Evolution of functional connectivity of brain networks and their dynamic interaction in temporal lobe epilepsy. *Brain connectivity*, 5(1):35–44, 2015.

[54] Naresh Nandakumar et al. A novel graph neural network to localize eloquent cortex in brain tumor patients from resting-state fmri connectivity. In *International Workshop on Connectomics in Neuroimaging*, pages 10–20. Springer, 2019.

[55] Naresh Nandakumar et al. Automated eloquent cortex localization in brain tumor patients using multi-task graph neural networks. *Medical image analysis*, page 102203, 2021.

[56] Naresh Nandakumar et al. A multi-scale spatial and temporal attention network on dynamic connectivity to localize the eloquent cortex in brain tumor patients. In *International Conference on Information Processing in Medical Imaging*, pages 241–252. Springer, 2021.

[57] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019.

[58] Giovanni Pellegrino, Tanguy Hedrich, Rasheda Chowdhury, Jeffery A Hall, Jean-Marc Lina, Francois Dubeau, Eliane Kobayashi, and Christophe Grova. Source localization of the seizure onset zone from ictal eeg/meg data. *Human brain mapping*, 37(7):2528–2546, 2016.

[59] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

[60] Chris Plummer, Simon J Vogrin, William P Woods, Michael A Murphy, Mark J Cook, and David TJ Liley. Interictal and ictal source localization for epilepsy surgery using high-density eeg with meg: a prospective long-term study. *Brain*, 142(4):932–951, 2019.

[61] Khandakar M Rashid and Joseph Louis. Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*, 42:100944, 2019.

[62] James A Roberts, Alistair Perry, Anton R Lord, Gloria Roberts, Philip B Mitchell, Robert E Smith, Fernando Calamante, and Michael Breakspear. The contribution of geometry to the human connectome. *Neuroimage*, 124:379–393, 2016.

[63] R. Rodríguez-Cruces and L. Concha. White matter in temporal lobe epilepsy: Clinico-pathological correlates of water diffusion abnormalities. *Quantitative Imaging in Medicine and Surgery*, 5:264–278, 2015.

[64] Gonzalo M Rojas et al. Study of resting-state functional connectivity networks using eeg electrodes position as seed. *Frontiers in neuroscience*, 12:235, 2018.

[65] F. Rosenow and H. Luders. Presurgical evaluation of epilepsy. *Brain*, 124(9):1683–1700, 2001.

[66] Subhrajit Roy et al. Machine learning for seizure type classification: setting the benchmark. *arXiv preprint arXiv:1902.01012*, 2019.

[67] I.R. Dwi Saputro et al. Seizure type classification on EEG signal using support vector machine. *Journal of Physics: Conference Series*, 1201:012065, 2019.

[68] C. Shepherd et al. A quantitative study of white matter hypomyelination and oligodendroglial maturation in focal cortical dysplasia type ii. *Epilepsia*, 54(5):898–908, 2013.

[69] Abbas Sohrabpour, Zhengxiang Cai, Shuai Ye, Benjamin Brinkmann, Gregory Worrell, and Bin He. Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources. *Nature communications*, 11(1):1–15, 2020.

[70] Chandra Sripada et al. Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder. *Human brain mapping*, 35(9):4693–4705, 2014.

[71] S. Srivastava et al. Feature-based statistical analysis of structural MR data for automatic detection of focal cortical dysplastic lesions. *NeuroImage*, 27:253–266, 2005.

[72] Aaron F Struck et al. Regional and global resting-state functional mr connectivity in temporal lobe epilepsy: Results from the epilepsy connectome project. *Epilepsy & Behavior*, 117:107841, 2021.

[73] S.M. Stufflebeam et al. Localization of Focal Epileptic Discharges using Functional Connectivity Magnetic Resonance Imaging. *Journal of Neurosurgery*, pages 1–5, 2011.

[74] Sundaram Suresh et al. Risk-sensitive loss functions for sparse multi-category classification problems. *Information Sciences*, 178(12):2621–2638, 2008.

[75] A. Sweet et al. Detecting epileptic regions based on global brain connectivity patterns. In *MICCAI: Medical Image Computing and Computer Assisted Intervention*, pages 98–105. LNCS, 2013.

[76] Nitin Tandon et al. Analysis of morbidity and outcomes associated with use of subdural grids vs stereoelectroencephalography in patients with intractable epilepsy. *JAMA neurology*, 76(6):672–681, 2019.

[77] Jewell B Thomas et al. Functional connectivity in autosomal dominant and late-onset alzheimer disease. *JAMA neurology*, 71(9):1111–1122, 2014.

[78] F. Turjman et al. Epilepsy associated with cerebral arteriovenous malformations: a multivariate analysis of angioarchitectural characteristics. *American Journal of Neuroradiology*, 16(2):345–350, 1995.

[79] Eric van Diessen et al. Functional and structural brain networks in epilepsy: what have we learned? *Epilepsia*, 54(11):1855–1865, 2013.

[80] David C Van Essen et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

[81] Y. Wang et al. Voxel-based automated detection of focal cortical dysplasia lesions using diffusion tensor imaging and t2-weighted mri data. *Epilepsy & Behavior*, 84:127–134, 2018.

[82] Zhiqun Wang, Jianli Wang, Han Zhang, Robert Mchugh, Xiaoyu Sun, Kuncheng Li, and Qing X Yang. Interhemispheric functional and structural disconnection in alzheimer's disease: a combined resting-state fmri and dti study. *PLoS One*, 10(5):e0126310, 2015.

[83] Dongren Yao et al. Triplet graph convolutional network for multi-scale analysis of functional connectivity using functional mri. In *International Workshop on Graph Learning in Medical Imaging*, pages 70–78. Springer, 2019.

[84] Shuai Ye, Lin Yang, Yunfeng Lu, Michal T Kucewicz, Benjamin Brinkmann, Cindy Nelson, Abbas Sohrabpour, Gregory A Worrell, and Bin He. Contribution of ictal source imaging for localizing seizure onset zone in patients with focal epilepsy. *Neurology*, 96(3):e366–e375, 2021.

[85] Clara Huishi Zhang et al. Lateralization and localization of epilepsy related hemodynamic foci using presurgical fmri. *Clinical Neurophysiology*, 126(1):27–38, 2015.

[86] Si Zhang et al. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):1–23, 2019.