

A DEEP LEARNING FRAMEWORK TO CHARACTERIZE NOISY LABELS IN EPILEPTOGENIC ZONE LOCALIZATION USING FUNCTIONAL CONNECTIVITY

Naresh Nandakumar¹, David Hsu², Raheel Ahmed³, and Archana Venkataraman^{1,4}.

¹Department of Electrical and Computer Engineering, Johns Hopkins University, USA

²Department of Neurology, University of Wisconsin School of Medicine, USA

³Department of Neurosurgery, University of Wisconsin School of Medicine, USA

⁴Department of Electrical and Computer Engineering, Boston University, USA

ABSTRACT

Resting-state fMRI (rs-fMRI) has emerged as a viable tool to localize the epileptogenic zone (EZ) in medication refractory focal epilepsy patients. However, due to clinical protocol, datasets with reliable labels for the EZ are scarce. Some studies have used the entire resection area from post-operative structural T1 scans to act as the ground truth EZ labels during training and testing. These labels are subject to noise, as usually the resection area will be larger than the actual EZ tissue. We develop a mathematical framework for characterizing noisy labels in EZ localization. We use a multi-task deep learning framework to identify both the probability of a noisy label as well as the localization prediction for each ROI. We train our framework on a simulated dataset derived from the Human Connectome Project and evaluate it on both the simulated and a clinical epilepsy dataset. We show superior localization performance in our method against published localization networks on both the real and simulated dataset.

Index Terms— Epilepsy, Dynamic Functional Connectivity, Noisy Labels, Semi-supervised Learning

1. INTRODUCTION

The ratio of corrupted to clean labels in real world datasets is reported to be anywhere from 8.0% to 38.5% [1]. The challenge of curating datasets with accurate labels is especially significant in medical imaging. Datasets tend to be small to begin with, and institutional policies or patient privacy can prevent data from being shared. Labeling of medical images is especially resource-intensive and potentially unreliable, as it requires specific domain expertise and there often exists a large degree of inter-observer variability [2].

Epilepsy is one of the most common neurological disorders and is linked to a fivefold increase in mortality [3]. Surgical treatment is a viable approach for medication refractory epilepsy [4]. However, identifying the epileptogenic zone (EZ) is a challenging task that requires many presurgical evaluations. Invasive monitoring using implanted intracranial electrodes can provide accurate EZ localization that can

help plan treatment, but is associated with surgical risks [5]. Resting-state fMRI (rs-fMRI) measures co-activation patterns in the brain, and has shown promise in identifying regions associated with the EZ that could potentially improve presurgical evaluation and localization [6, 7].

Previously established EZ localization models using rs-fMRI connectivity, like those in [6, 8] have used the entire resection area, delineated from the post-operative structural T1 images, to act as the EZ label during train and testing. However, it is well known that these resections usually are larger than the EZ, either as a means of removing secondary tissue that can also be problematic or ensuring the resection is liberal enough to ensure all EZ tissue is removed [9]. Therefore, parts of the EZ label in these models are mislabeled.

We develop a framework to identify noisy labels for EZ localization. Specifically, we model the probability of an incorrect label using the concrete distribution [10], which is a continuous relaxation of the Bernoulli distribution. Leveraging the data augmentation techniques presented in [11] for EZ simulation, we create a simulated dataset that contains noisy labels that reflect the expected pattern of noisy labels. We develop a multi-task neural network architecture to learn the concrete distribution parameters of interest in a strongly pre-trained fashion and perform localization. Our proposed method outperforms localization methods presented in [8, 6] on both datasets and highlight representative examples.

2. METHODS

2.1. Graphical model representation

Fig. 1 shows the graphical model for one subject. Mathematically, let $n \in \{1, \dots, N\}$ index node (brain ROI), and $\mathbf{X} \in \mathbb{R}^{N \times N \times T}$ be the data. Following [8], we construct dynamic connectivity matrices as the input, so $\mathbf{X}_n \in \mathbb{R}^{N \times T}$ is the dynamic connectivity profile associated with node n . Let $\mathbf{Y}_n$ be the observed label and $\tilde{\mathbf{Y}}_n$ be the real, unobserved label for node n and let $\mathbf{Z}_n$ be a latent random variable that captures corruption of label n , which is parameterized by α_n .

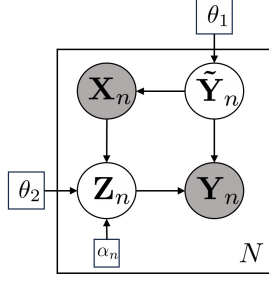


Fig. 1. A graphical model showing the dependencies in our model. Shaded nodes are observed while white are latent.

The joint distribution of $\mathbf{Y}_n, \tilde{\mathbf{Y}}_n, \mathbf{Z}_n$ conditioning on $\mathbf{X}_n$ from the graphical model is

$$P(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n, \mathbf{Z}_n | \mathbf{X}_n) = P(\mathbf{Y}_n | \tilde{\mathbf{Y}}_n, \mathbf{Z}_n) P(\tilde{\mathbf{Y}}_n | \mathbf{X}_n) P(\mathbf{Z}_n | \mathbf{X}_n). \quad (1)$$

We define the helper function $\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n) = \mathbf{Y}_n \tilde{\mathbf{Y}}_n + (1 - \mathbf{Y}_n)(1 - \tilde{\mathbf{Y}}_n)$, which indicates when the observed label is corrupted or not. We define our likelihood term as

$$P(\mathbf{Y}_n | \tilde{\mathbf{Y}}_n, \mathbf{Z}_n) = [\rho^{\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n)} (1 - \rho)^{(1 - \Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n))}]^{(1 - \mathbf{Z}_n)} [\alpha_n^{(1 - \Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n))} (1 - \alpha_n)^{\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n)}] \mathbf{Z}_n \quad (2)$$

where $\rho \approx 0.99$, or essentially is 1. In contrast to existing methods, like the one in [12], our goal is to develop a framework that can learn α_n as opposed setting it *a priori*. The terms $P(\tilde{\mathbf{Y}}_n | \mathbf{X}_n)$ and $P(\mathbf{Z}_n | \mathbf{X}_n)$ describe the class label distribution and noise content in the data respectively. We use two separate deep networks to characterize these distributions, parameterized by θ_1 and $\theta'_2 = \theta_2 \cup \alpha$ respectively.

Concrete distribution Instead of using the discrete Bernoulli distribution, we model the latent random variable $\mathbf{Z}_n$ via the concrete distribution [10]

$$\mathbf{Z}_n = \sigma \left(\frac{1}{t} \left(\log \left(\frac{\alpha_n}{1 - \alpha_n} \right) + \log \left(\frac{u_n}{1 - u_n} \right) \right) \right) \quad (3)$$

which gives a continuous relaxation of the Bernoulli distribution, where u_n is a uniform random variable on the interval of $[0, 1]$. Here, the temperature t is a hyperparameter of the distribution. We observed more stable training from using the concrete Bernoulli compared to the discrete case.

Learning the parameters We use two deep networks to model $P(\tilde{\mathbf{Y}} | \mathbf{X})$ and $P(\mathbf{Z} | \mathbf{X})$ where θ_1 parameterizes $P(\tilde{\mathbf{Y}} | \mathbf{X})$ and θ'_2 parameterizes $P(\mathbf{Z} | \mathbf{X})$. Our goal is to find the optimal $\theta = \theta_1 \cup \theta'_2$ that maximizes the incomplete log-likelihood $P(\mathbf{Y} | \mathbf{X}; \theta)$. We use the EM algorithm to iteratively solve this problem [13]. For an arbitrary distribution $q(\tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{Y}, \mathbf{X})$, we can derive a lower bound of the incomplete log-likelihood

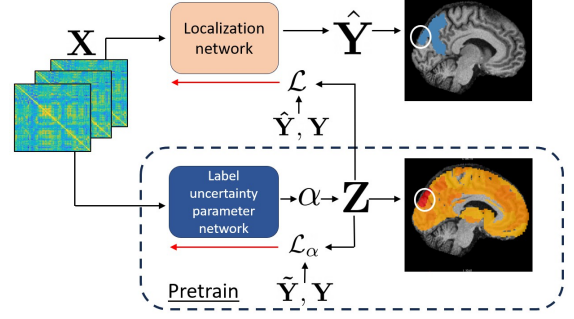


Fig. 2. Bottom: The label uncertainty parameter network is pretrained in a fully supervised manner to learn α from $\mathbf{X}$. **Top:** after pretraining, the localization network is trained only with the observed noisy labels $\mathbf{Y}$.

$$\log P(\mathbf{Y} | \mathbf{X}; \theta) \geq q(\tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{Y}, \mathbf{X}) \log \frac{P(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{X}; \theta)}{q(\tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{Y}, \mathbf{X})}. \quad (4)$$

The E-step involves computing the posterior of the latent variables using the current parameters $\theta^{(t)}$,

$$P(\tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{Y}, \mathbf{X}; \theta^{(t)}) = \frac{P(\mathbf{Y} | \tilde{\mathbf{Y}}, \mathbf{Z}; \theta^{(t)}) P(\tilde{\mathbf{Y}} | \mathbf{X}; \theta^{(t)}) P(\mathbf{Z} | \mathbf{X}; \theta^{(t)})}{\sum_{\tilde{\mathbf{Y}}, \mathbf{Z}} P(\mathbf{Y} | \tilde{\mathbf{Y}}, \mathbf{Z}; \theta^{(t)}) P(\tilde{\mathbf{Y}} | \mathbf{X}; \theta^{(t)}) P(\mathbf{Z} | \mathbf{X}; \theta^{(t)})} \quad (5)$$

where the expected complete log-likelihood can be written as

$$Q(\theta; \theta^{(t)}) = \sum_{\tilde{\mathbf{Y}}, \mathbf{Z}} P(\tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{Y}, \mathbf{X}; \theta^{(t)}) \log P(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{Z} | \mathbf{X}; \theta). \quad (6)$$

For the M-step, we exploit two deep networks to model the probability $P(\tilde{\mathbf{Y}} | \mathbf{X}; \theta_1)$ and $P(\mathbf{Z} | \mathbf{X}; \theta'_2)$. Recall that $\theta'_2 = \theta_2 \cup \alpha$. The gradient of Q with respect to θ can be decoupled into two parts, which are implemented via back-propagation of the two separate neural networks:

$$\frac{\partial Q}{\partial \theta} = \sum_{\tilde{\mathbf{Y}}} P(\tilde{\mathbf{Y}} | \mathbf{Y}, \mathbf{X}; \theta^{(t)}) \frac{\partial}{\partial \theta_1} P(\tilde{\mathbf{Y}} | \mathbf{X}; \theta_1) + \sum_{\mathbf{Z}} P(\mathbf{Z} | \mathbf{Y}, \mathbf{X}; \theta^{(t)}) \frac{\partial}{\partial \theta'_2} P(\mathbf{Z} | \mathbf{X}; \theta'_2) \quad (7)$$

2.2. Deep learning network architecture

Our workflow is shown in Fig. 2. The data is input to both the localization network and the label uncertainty parameter network. The label uncertainty parameter network predicts α from $\mathbf{X}$ in a fully supervised manner (including knowledge of $\tilde{\mathbf{Y}}$) during pretraining. We use the model from [8] as the

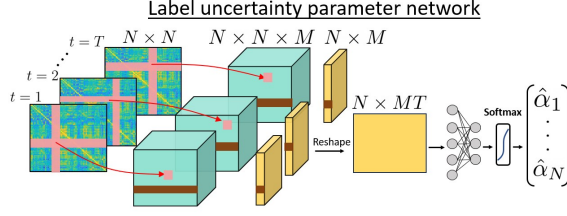


Fig. 3. We use a CNN architecture with an ANN to predict α .

localization network. Once pretrained, we do not include information of $\tilde{\mathbf{Y}}$ when training the localization network.

Label uncertainty parameter network: Fig. 3 shows the label uncertainty parameter network used for pretraining. We use cross-shaped convolutional filters designed in [14] to perform feature extraction from the dynamic connectivity inputs. We then apply a 1D convolution (brown in Fig.3) along the columns of the intermediate features, then reshape the representation into a $N \times MT$ matrix to then feed into a 3 layer ANN and obtain predictions α_n . To ensure $0 \leq \alpha_n \leq 1$, we use a softmax layer. We have access to both $\tilde{\mathbf{Y}}_n, \mathbf{Y}$ using a simulated dataset. We sample $\mathbf{Z}$ and our loss function is

$$\mathcal{L}_\alpha = -\frac{1}{N} \sum_{n=1}^N \epsilon_1 (1 - \Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n)) \log(\mathbf{Z}_n) + \epsilon_2 \Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n) \log(1 - \mathbf{Z}_n). \quad (8)$$

The function $\mathcal{L}_\alpha$ is a weighted cross entropy function that encourages α_n to be high when the labels $\mathbf{Y}_n, \tilde{\mathbf{Y}}_n$ are different and encourages α_n to be low when the labels are the same. We introduce the weights ϵ_1, ϵ_2 to handle the class-imbalance, as only a relatively small subset of the nodes will be mislabeled. This network identifies the connectivity patterns associated with correctly labeled vs. mislabeled nodes.

Combined network training: Once the label uncertainty parameter network is pretrained, we attach the model in [8], as the localization network for combined training. Let $y_n \in \mathbb{R}^{N \times 2}$ be the one-hot encoded version of $\mathbf{Y}_n$. Let $\hat{y}_n \in \mathbb{R}^{N \times 2}$ be the network outputs for node n . We do combined network training with the following loss function:

$$\mathcal{L} = -\frac{1}{N} \left[\underbrace{(1 - \mathbf{Z}_n) [y_{n,1} \delta_1 \log \hat{y}_{n,1} + \delta_2 y_{n,2} \log \hat{y}_{n,2}]}_{\text{Certain term}} + \underbrace{\mathbf{Z}_n \left(\frac{\log \hat{y}_{n,1} + \log \hat{y}_{n,2}}{2} \right)}_{\text{Uncertain term}} \right] + \lambda_1 \underbrace{\left((1 - \mathbf{Z}_n) - \frac{|\sum_{i \in ne(n)} \mathbf{Y}_i - \frac{M}{2}|}{\frac{M}{2}} \right)}_{\text{Neighborhood smoothing term}}. \quad (9)$$

$\mathcal{L}$ is broken down into three main terms. The certain term reflects when we are confident about the observed label. Therefore, we backpropagate the original weighted cross entropy loss as the certain term. The uncertain term reflects the case when we believe label n is mislabeled, and therefore we backpropagate the average of the two prediction terms $\log \hat{y}_{n,1}$ and $\log \hat{y}_{n,2}$. The hyperparameters δ_1, δ_2 are the cross entropy weights which help mitigate class imbalance. The neighborhood smoothing term acts as a biologically inspired regularization term that takes direct spatial neighbors of node n into account, where M is the number of neighbors considered ($M = 6$ in this work). When $\frac{|\sum_{i \in ne(n)} \mathbf{Y}_i - \frac{M}{2}|}{\frac{M}{2}} = 0$, the neighbors of n belong to both classes, and node n is on the boundary of the resection, for which we are more unsure of its label. Given $\lambda_1 > 0$, this case encourages α_n to be close to 1 when we minimize the loss during backpropagation. Conversely, when $\frac{|\sum_{i \in ne(n)} \mathbf{Y}_i - \frac{M}{2}|}{\frac{M}{2}} = 1$, the neighbors of node n all belong to the same class which encourages α_n to be low.

Prediction on test data: We use both $\hat{y}_n$ and α_n for $n \in \{1, \dots, N\}$ during testing. Let $\bar{y} = \frac{1}{N} \sum_{n=1}^N \hat{y}_n$. During testing, we fuse the predictions and alpha parameters for each test subject separately in the following fashion

$$\hat{y}_n^{test} = \hat{y}_n (1 - \alpha_n) + \bar{y} \alpha_n. \quad (10)$$

Eq.10 takes both model outputs to arrive at a prediction $\hat{y}_n^{test}$ for each node n which asserts that the final prediction should be close to the original network prediction when we are confident about the label (α_n low) or be close to the average prediction when we are not confident about the label (α_n high). The way we fuse information in Eq.10 is similar to the weighted combination of uncertain vs. certain terms in $\mathcal{L}$.

Implementation details: We implement our network using the PyTorch [15] deep learning library. We pretrain the label uncertainty parameter network using the Adam optimizer for 250 epochs with a learning rate of 0.001 that decays by a factor of 0.8 every 20 epochs. Once pretrained, we set the initial learning rate for the label parameter network to 0.00005 and train the entire network using the Adam optimizer for 150 epochs with a learning rate of 0.0005 that decays by a factor of 0.8 every 10 epochs. We set the hyperparameters $\epsilon_1 = 1.1$, $\epsilon_2 = 0.15, \delta_1 = 0.16, \delta_2 = 1.2$ and $\lambda_1 = 0.08$ which were determined via cross validation on a separate HCP dataset.

3. EXPERIMENTAL RESULTS

Datasets: We use noise models introduced in [11] to simulate the EZ region to obtain $\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}$ from 400 HCP subjects [16]. To simulate the EZ signal used for $\tilde{\mathbf{Y}}$, we augment the fMRI time series of a spatially continuous region with one of six noise models described in [11]. We create an artificial noisy dataset having the labels implicated in $\tilde{\mathbf{Y}}$ correspond to regions with the noise models (EZ) and $\mathbf{Y}$ include both augmented (EZ) and healthy regions. We create a comprehen-

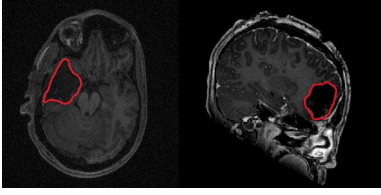


Fig. 4. Resection boundaries for two patients.

Method	Sens	Spec	Acc	AUC	p-value
Proposed	0.5	0.89	0.85	0.71	
Only localization	0.57	0.71	0.76	0.62	< 0.01
DeepEZ	0.36	0.69	0.67	0.57	< 0.01

Table 1. Simulated dataset testing metrics w.r.t $\tilde{Y}$

sive training set by including different types of samples, like where the spatial region implicated by $\tilde{Y}$ is relatively small, large, or on the boundary of Y and the healthy regions.

Our clinical dataset consists of 14 pediatric patients with focal epilepsy from the University of Wisconsin (UW) Madison. Preoperative rs-fMRI data was acquired using an echo planar imaging sequence (EPI, TR = 802 ms, TE = 33.5 ms, flip angle = 50°, res = 2 mm isotropic). The rs-fMRI data is preprocessed using the CPAC pipeline [17]. Postoperative T1-weighted MRI was acquired using a 3D gradient-echo pulse sequence (MPRAGE, TR = 604 ms, TE = 2.516 ms, flip angle = 8°, res = 0.8 mm isotropic). As shown in Fig. 4, we manually segment the resection cavity and consider this area as the pseudo ground truth EZ for each patient. We use the Brainnetomme atlas [18] to define $N = 246$ cortical and subcortical regions for our analysis.

Localization results: We use 10-fold CV and report the average sensitivity, specificity, accuracy and AUC on both the simulated dataset (table 1) and UW clinical dataset (table 2). We use De Long’s test on AUC to show statistical significance in our results. We test using the $\tilde{Y}$ labels for the simulated dataset. We observe the performance of the entire network with the label uncertainty parameter network (proposed) and without (only localization) as well as the DeepEZ method from [6] as a baseline method. We pretrain the label uncertainty parameter network with a different subset of subjects than the entire network training and testing. In both tables, the proposed method outperforms the baselines, as shown by significant difference in the AUC metrics. The proposed method maintains a good sensitivity while having a much higher specificity, so it is not suffering from over predictions where the only localization method is. If trained entirely on a noisy dataset, which is possible in a real world setting, the method from [8] generalizes much worse to real subjects compared to the proposed method that involves learning the label uncertainty parameter and effectively ignoring mislabeled nodes during training.

Method	Sens	Spec	Acc	AUC	p-value
Proposed	0.42	0.94	0.89	0.74	
Only localization	0.51	0.77	0.84	0.68	0.041
DeepEZ	0.25	0.82	0.82	0.59	< 0.01

Table 2. UW dataset testing metrics.

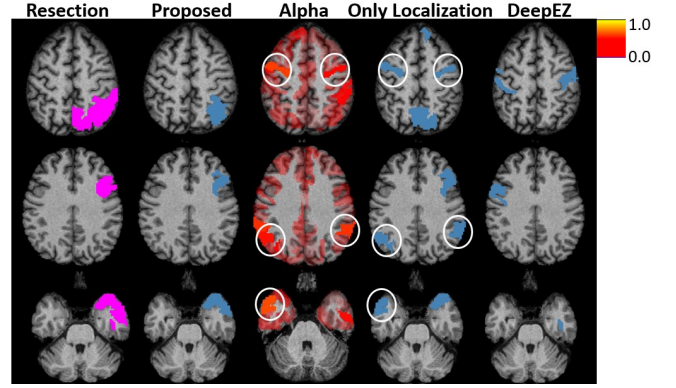


Fig. 5. Resection (pink), predictions (blue) and α (heat map) for three separate epilepsy subjects.

Fig. 5 shows the ground truth (pink) labels, predicted (blue) labels, and recovered alpha values (heat map) for three separate subjects from the UW dataset. The proposed method accurately localizes regions within the ground truth label while having less false positives than the only localization method. Circled in white, we observe regions that correspond to false positives in the baseline and high α values in the proposed. Due to our training procedure and Eq. 10, the proposed method correctly classifies these regions as healthy, showing the localization improvement using our approach.

4. CONCLUSION

In this work, we developed a framework for characterizing noisy labels in EZ localization. We developed a graphical model and derived the EM equations for our setup, and use backpropagation from neural networks to achieve parameter updates. We introduce a deep learning framework that is trained in separate parts to achieve our goal, where we strongly pretrain the label uncertainty parameter network to be able to learn the Bernoulli parameters in a supervised fashion. We create an artificial noisy dataset using EZ simulation methods. We show promising results, where even when trained on the noisy labels, our proposed method outperforms previously established models when testing with the true labels for both the simulated and real dataset.

Acknowledgements: This work is supported by the National Science Foundation CAREER award 1845430 (PI Venkataraman), the National Institutes of Health award R21 CA263804 (PI Venkataraman), and the National Institutes of

Health award R01 EB029977 (PI Caffo).

Compliance with ethical standards: This research study was conducted using human subject data and the study was approved by the institutional review board of the University of Wisconsin School of Medicine.

5. REFERENCES

- [1] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee, “Learning from noisy labels with deep neural networks: A survey,” *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [2] Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour, “Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis,” *Medical image analysis*, vol. 65, pp. 101759, 2020.
- [3] W.J. Cadarelli and B.J. Smith, “The burden of epilepsy to patients and payers,” *American Journal of Managed Care*, vol. 16, no. 12, pp. S331–336, 2010.
- [4] Rekha Dwivedi et al., “Surgery for drug-resistant epilepsy in children,” *New England Journal of Medicine*, vol. 377, no. 17, pp. 1639–1647, 2017.
- [5] Nitin Tandon et al., “Analysis of morbidity and outcomes associated with use of subdural grids vs stereo-electroencephalography in patients with intractable epilepsy,” *JAMA neurology*, vol. 76, no. 6, pp. 672–681, 2019.
- [6] Naresh Nandakumar, David Hsu, Raheel Ahmed, and Archana Venkataraman, “Deepez: A graph convolutional network for automated epileptogenic zone localization from resting-state fmri connectivity,” *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 1, pp. 216–227, 2022.
- [7] Maria Giulia Preti, Nora Leonardi, F Işık Karahanoğlu, Frédéric Grouiller, Mélanie Genetti, Margina Seeck, Serge Vulliemoz, and Dimitri Van De Ville, “Epileptic network activity revealed by dynamic functional connectivity in simultaneous eeg-fmri,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 9–12.
- [8] Naresh Nandakumar, David Hsu, Raheel Ahmed, and Archana Venkataraman, “A deep learning framework to localize the epileptogenic zone from dynamic functional connectivity using a combined graph convolutional and transformer network,” in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.
- [9] Olesya Grinenko, Jian Li, John C Mosher, Irene Z Wang, Juan C Bulacio, Jorge Gonzalez-Martinez, Dileep Nair, Imad Najm, Richard M Leahy, and Patrick Chauvel, “A fingerprint of the epileptogenic zone in human epilepsies,” *Brain*, vol. 141, no. 1, pp. 117–131, 2018.
- [10] Yarin Gal, Jiri Hron, and Alex Kendall, “Concrete dropout,” *Advances in neural information processing systems*, vol. 30, 2017.
- [11] Patrick H Lockett, Luigi Maccotta, John J Lee, Ki Yun Park, Nico UF Dosenbach, Beau M Ances, Robert Edward Hogan, Joshua S Shimony, and Eric C Leuthardt, “Deep learning resting state functional magnetic resonance imaging lateralization of temporal lobe epilepsy,” *Epilepsia*, vol. 63, no. 6, pp. 1542–1552, 2022.
- [12] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang, “Learning from massive noisy labeled data for image classification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2691–2699.
- [13] Shu Kay Ng, Thriyambakam Krishnan, and Geoffrey J McLachlan, “The em algorithm,” *Handbook of computational statistics: concepts and methods*, pp. 139–172, 2012.
- [14] Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh, “Brainnetcn: convolutional neural networks for brain networks; towards predicting neurodevelopment,” *NeuroImage*, vol. 146, pp. 1038–1049, 2017.
- [15] Adam Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.
- [16] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al., “The wu-minn human connectome project: an overview,” *Neuroimage*, vol. 80, pp. 62–79, 2013.
- [17] Cameron Craddock et al., “Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac),” *Front Neuroinform*, vol. 42, pp. 10–3389, 2013.
- [18] Lingzhong Fan et al., “The human brainnetome atlas: a new brain atlas based on connectional architecture,” *Cerebral cortex*, vol. 26, no. 8, pp. 3508–3526, 2016.