A DEEP LEARNING FRAMEWORK TO LOCALIZE THE EPILEPTOGENIC ZONE FROM DYNAMIC FUNCTIONAL CONNECTIVITY USING A COMBINED GRAPH CONVOLUTIONAL AND TRANSFORMER NETWORK

Naresh Nandakumar[1], *David Hsu*[2], *Raheel Ahmed*[3], *and Archana Venkataraman*[1,4].

[1]Department of Electrical and Computer Engineering, Johns Hopkins University, USA
[2]Department of Neurology, University of Wisconsin School of Medicine, USA
[3]Department of Neurosurgery, University of Wisconsin School of Medicine, USA
[4]Department of Electrical and Computer Engineering, Boston University, USA

ABSTRACT

Localizing the epileptogenic zone (EZ) is a critical step in the treatment of medically refractory epilepsy. Resting-state fMRI (rs-fMRI) offers a new window into this task by capturing dynamically evolving co-activation patterns, also known as connectivity, in the brain. In this work, we present the first automated framework that uses dynamic functional connectivity from rs-fMRI to localize the EZ across a heterogeneous epilepsy cohort. Our framework uses a graph convolutional network for feature extraction, followed by a transformer network, whose attention mechanism learns which time points of the rs-fMRI scan are important for EZ localization. We train our framework on augmented data derived from the Human Connectome Project and evaluate it on a clinical epilepsy dataset. Our results demonstrate the clear advantages of our convolutional + transformer combination and data augmentation procedure over ablated and comparison models.

Index Terms— Functional Connectivity, Epilepsy, Deep Learning, Graph Convolutions, Transformer Models

1. INTRODUCTION

Epilepsy is a debilitating neurological disorder characterized by recurrent and unprovoked seizures. While anti-epileptic drugs are effective in many cases, roughly a third of patients have a *medication refractory* course [1]. Surgery is an effective therapeutic approach for medication refractory epilepsy, provided that seizures have focal onset and can be accurately localized to a discrete epileptogenic zone (EZ) [2]. At present, identification of EZ, relies on a localization hypothesis based on concordance between noninvasive electroencephalogram (EEG) and structural MRI for coarse localization and invasive EEG monitoring over the presumed target zone using surgically implanted intracranial electrodes. Inaccurate noninvasive EZ localization can reduce the diagnostic efficacy of intracranial EEG monitoring and compound surgical risks [3]. Resting-state fMRI (rs-fMRI) measures co-activation patterns in the brain, and may elucidate network interactions associated with the EZ that could potentially inform presurgical evaluation and localization [4, 5].

However, localization of the EZ using rs-fMRI is a challenging problem with limited success described in literature. The earliest work used graph theoretic measures derived from rs-fMRI connectivity for EZ localization [6]. The follow-up study by the same authors extended this method to identify patient-specific hubs of abnormal functional connectivity [7]. While these works showed good correlation between identified regions and the EZ, there is limited quantitative evaluation for this approach based on model accuracy and clinical outcomes. The authors of [8] use independent component analysis to generate feature maps, followed by EZ classification of these components. While promising, this method relies on visual inspection of the independent components prior to classification. The automated method proposed by [9] trains a 3D convolutional neural network on the rs-fMRI data to determine the hemisphere of the EZ. This approach uses a data augmentation technique, by which artificial lesions are inserted into the connectivity data from healthy controls. Finally, the work of [10] introduces a graph convolutional network for rs-fMRI connectivity matrices that performs *region-wise identification* of the EZ. While this method achieves high AUC, the sensitivity is low, and it occasionally misses the EZ.

Recent work in rs-fMRI literature has increasingly leveraged the dynamic evolution of connectivity information to improve predictive performance [11]. For example, the work of [12] uses an LSTM network to capture temporal dependencies within the rs-fMRI scan to classify subjects with autism from controls. Likewise, the authors of [13] use an LSTM network as a temporal attention module to improve localization of eloquent cortex in brain tumor patients. The model is applied to dynamic functional connectivity (dFC) matrices computed using a sliding window. Finally, the method of [14] uses a transformer network applied to dFC to predict the brain age of patients with Alzheimer's disease.

We present the first deep learning model to localize the EZ in focal epilepsy patients based on dFC. Our deep network

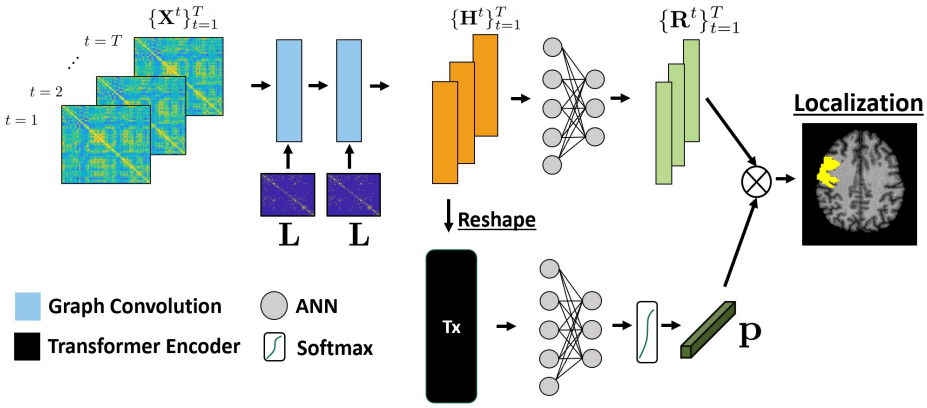

Fig. 1. Network overview. **Top**: We use a multi-modal GCN and fully-connected layers to obtain node-wise predictions of the EZ over time $\mathbf{R}$. **Bottom**: Our transformer and fully-connected layers network extract a temporal attention vector $\mathbf{p}$ that selects specific windows of the dFC input. The attention $\mathbf{p}$ is combined with $\mathbf{R}$ to obtain the final EZ predictions.

architecture uses an anatomically-regularized graph convolutional network (GCN) for feature extraction. From here, a transformer network learns a temporal attention vector, which identifies relevant time windows of the rs-fMRI scan that aid in localization. Following [9], we train our network entirely on simulated data derived from the Human Connectome Project (HCP), and we test it on a clinical epilepsy dataset from the University of Wisconsin Madison. We demonstrate the significantly improved performance of our framework, as compared to both ablated versions of our model and the method of [10]. Our results highlight the promise of using rs-fMRI connectivity for preoperative EZ localization.

2. METHODS

Fig.1 shows an overview of our framework. Our method uses a two layer GCN to obtain intermediate node-wise features over time. These intermediate features are fed into both a temporal attention network as well as a node-classifier network for EZ localization. We use the sliding window technique to generate the dFC inputs to our framework. Formally, let N be the number of brain regions in our parcellation, T be the number of sliding windows, and $\{\mathbf{X}^t\}_{t=1}^T$ be the dFC matrices. $\mathbf{X}^t \in \mathbb{R}^{N\times N}$ is computed from a segment $\mathbf{Z}^t \in \mathbb{R}^{N\times d}$ of the rs-fMRI time series, where d is the sliding window size.

GCN for feature extraction: The first stage of our model is an anatomically-regularized GCN for feature extraction. We use diffusion MRI (d-MRI) tractography to construct the binary adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ used for graph filtering [15]. In this case, an entry $\mathbf{A}_{ij} = 1$ denotes an anatomical connection between regions i and j. Let $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ be the normalized graph Laplacian of $\mathbf{A}$, where $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$.

Let $\mathbf{A} \in \mathbb{R}^{N\times N}$ denote the binary adjacency matrix used for graph filtering [15]. We use d-MRI tractography to construct $\mathbf{A}$. In this case, an entry $\mathbf{A}_{ij} = 1$ denotes an anatomical connection between regions i and j. Let $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ be the normalized graph Laplacian of $\mathbf{A}$, where $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. Each layer produces an activation map $\mathbf{H}_l \in \mathbb{R}^{N\times G_l}$, where $l \in \{1, 2\}$ denotes the layer number. The learnable parameters in each graph convolution are a weight matrix $\mathbf{W}_l \in \mathbb{R}^{G_l\times G_{l+1}}$ and a constant bias $\mathbf{b}_l \in \mathbb{R}^{1\times G_{l+1}}$. The intermediate activation $\mathbf{H}_1$ is generated via the propagation rule:

$$\mathbf{H}_1^t = \phi\Big(\mathbf{L}\mathbf{X}^t\mathbf{W}_1 + \mathbf{b}_1\Big), \tag{1}$$

with the activation $\mathbf{H}_2$ generated likewise from $\mathbf{H}_1$.

Transformer-Based Temporal Attention: The outputs $\{\mathbf{H}_2^t\}_{t=1}^T$ of the GCN correspond to intermediate node-level features per time point. From here, the temporal attention module leverages the encoder stage of a transformer network [16], followed by a fully-connected artificial neural network (ANN). Our transformer employs multi-headed self-attention (MHA) and feed-forward networks with residual connections to process sequential data. Formally, let $\mathbf{H}' \in \mathbb{R}^{T\times NG_2}$ be a flattened version of $\{\mathbf{H}_2^t\}_{t=1}^T$. A single encoder layer in the transformer is computed as follows:

$$\mathbf{C}_1 = \text{MHA}(\mathbf{H}') + \mathbf{H}' \qquad \mathbf{C}_2 = \text{FF}(\mathbf{C}_1) + \mathbf{C}_1, \tag{2}$$

where the $\text{FF}(\cdot)$ operation denotes a feed-forward network.

The $\text{MHA}(\cdot)$ function in Eq. (2) consists of multiple self-attention (SA) operations, where each SA_i for $i \in \{1 \cdots I\}$ is computed as $\text{SA}_i(\mathbf{V}_i) = \mathbf{M}_i\mathbf{V}_i$. As introduced in [16], the attention mask $\mathbf{M}_i \in \mathbb{R}^{T\times T}$ captures the similarity between a query matrix $\mathbf{Q}_i = \mathbf{W}_i^q\mathbf{H}'$ and a key matrix $\mathbf{K}_i = \mathbf{W}_i^k\mathbf{H}'$, both of which are linear functions of the input data:

$$\mathbf{M}_i = \text{Softmax}\left(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{NG_2}}\right) \tag{3}$$

Likewise, the value matrix $\mathbf{V}_i = \mathbf{W}_i^v \mathbf{H}'$ is also obtained via a linear layer. The matrices $\mathbf{W}_i^q$, $\mathbf{W}_i^k$, and $\mathbf{W}_i^v$ in the above expressions denote the learned weights for head i.

The self-attention outputs are concatenated across heads and fed through a linear layer to obtain one MHA operation. The transformer combines the MHA$(\cdot)$ operation with a residual connection. The subsequent FF$(\cdot)$ operation consists of two fully-connected layers plus another residual connection. The encoding procedure in Eq. (2) optimizes the mixing across the sequential input features for the downstream task.

The output of the transformer is fed through two fully-connected layers and a softmax function to obtain our temporal attention vector $\mathbf{p} \in \mathbb{R}^{T\times 1}$. The attention $\mathbf{p}$ is designed to identify which time points are more relevant for downstream node classification. Both the intermediate features $\{\mathbf{H}\}_{t=1}^T$ and attention vector $\mathbf{p}$ appear in the EZ classification stage.

Classification and Loss Function: We treat the problem of EZ localization as a region-wise classification problem, where each region is identified as either belonging to the EZ class or to the "normal" class. Formally, the intermediate features $\{\mathbf{H}\}_{t=1}^T$ are fed into a two-layer ANN to obtain node-wise predictions over time $\{\mathbf{R}\}_{t=1}^T$, where $\mathbf{R}^t \in \mathbb{R}^{N\times 2}$. Our temporal attention vector $\mathbf{p}$ is combined with $\{\mathbf{R}\}_{t=1}^T$ via an inner product to obtain a single prediction for each region.

We adopt a modified version of the loss function presented in [10], which uses a weighted cross-entropy loss and a regularization term to suppress activations in regions contralateral to the EZ. Let $\mathbf{Y} \in \mathbb{R}^{N\times 2}$ be the one-hot encoded labels, N_e denote the nodes that belong to the EZ class and let $c(n)$ denote the contralateral counterpart to region n. Our training loss function consists of the following two terms:

$$\mathcal{L}(\{\mathbf{X}^t\}_{t=1}^T, \mathbf{Y}) = \underbrace{-\sum_{n=1}^{N}\sum_{i=1}^{2} \delta_i \log\Big(\sigma\Big(\sum_{t=1}^{T} \mathbf{R}_{n,c}^t \cdot \mathbf{p}^t\Big)\Big)\mathbf{Y}_{n,c}}_{\text{Weighted Cross Entropy}}$$
$$\underbrace{-\lambda \frac{1}{N_e}\sum_{n\in N_e}\Big(\sigma\Big(\sum_{t=1}^{T} \mathbf{R}_{n,2}^t \cdot \mathbf{p}^t\Big) - \sigma\Big(\sum_{t=1}^{T} \mathbf{R}_{c(n),2}^t \cdot \mathbf{p}^t\Big)\Big)}_{\text{EZ Contralateral Term}}. \quad (4)$$

Data Augmentation for Training: Rs-fMRI studies of focal epilepsy patients are often limited in size. Therefore, following [9], we train our deep network entirely on augmented data derived from a neurotypical control dataset. For each training sample, we augment the healthy rs-fMRI data by first randomly selecting a spatially continuous neighborhood of voxels to form the EZ and then modifying the time series at those voxels via one of six noise models: (1) adding normally distributed noise, (2) adding uniformly distributed noise, (3) adding power-law noise, (4) adding Brownian noise, (5) adding noise generated by a Levy walk process, and (6) randomly permuting the time series. Since there is no established ground truth for how the EZ affects rs-fMRI, the combination of these six noise models exposes our network to a broad range of data abnormalities during training [9]. To our knowledge, our work is the first to use data augmentation for EZ *localization* based on rs-fMRI connectivity.

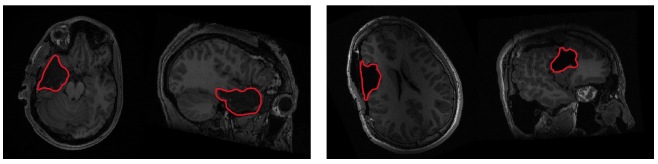

Fig. 2. Resection boundaries (red) for two epilepsy patients.

Implementation Details: We implement our network in Pytorch [17] using the ADAM optimizer and a leaky-ReLU activation function with slope = -0.1 between each layer. To prevent data leakage, all hyperparameters of our network are set using cross-validation on 100 EZ-augmented subjects from the Human Connectome Project (HCP) dataset.

Baselines: We compare our proposed framework against competing methods from the literature (first two below) and ablated versions of our framework (last four below):

- **BN-CNN:** A modified version of the BrainNetCNN architecture developed in [18] that performs region-wise, rather than subject-level, prediction.
- **DeepEZ:** The model developed by [10] to localize the EZ based on static rs-fMRI connectivity.
- **NoAttn:** Ablation #1 that removes the temporal attention mechanism. Final predictions are averaged over time.
- **ANNattn:** Ablation #2 that uses a fully-connected ANN rather than a transformer as the temporal attention model.
- **LSTM:** Ablation #3: that uses an LSTM rather than a transformer as the temporal attention model.
- **NoAugment:** Ablation #4 that trains the deep network directly on the clinical data with no augmentation.

3. EXPERIMENTAL RESULTS

Datasets: Our training data consists of 300 HCP subjects [19]. We generate training three samples per subject ($S = 900$ total) by varying the EZ location and/or noise model used for data augmentation. We use the Brainnetomme atlas [20] to define $N = 246$ cortical and subcortical regions for our analysis. We construct the adjacency matrix $\mathbf{A}$ used in our GCN from d-MRI tractography of 50 additional HCP subjects. Individual structural connectivity matrices are generated according to [21]. We average and threshold these matrices to compute $\mathbf{A}$, used for both training and testing.

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

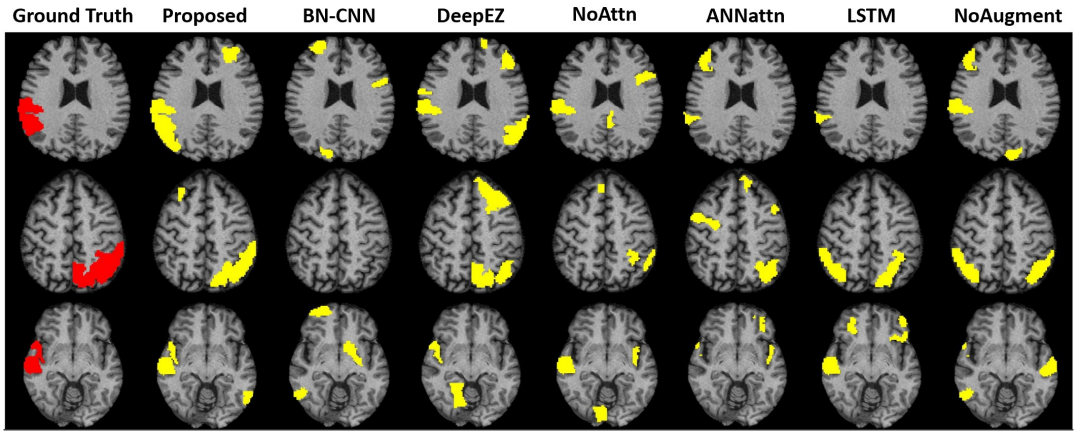

Fig. 3. Ground truth (red) and model predictions (yellow) for three test subjects.

Method	Sens	Spec	Acc	AUC	p-value
BN-CNN	0.17	0.81	0.69	0.58	$< 10^{-8}$
DeepEZ	0.34	0.88	0.87	0.7	0.011
NoAttn	0.35	0.86	0.86	0.69	< 0.01
ANNattn	0.41	0.86	0.88	0.71	0.019
LSTM	0.41	0.88	0.88	0.73	0.052
NoAugment	0.28	0.90	0.90	0.68	< 0.01
Proposed	**0.51**	0.89	**0.92**	**0.77**	

Table 1. Performance metrics for EZ classification.

data is preprocessed using the CPAC pipeline [22]. 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. 2, we manually segment the resection cavity and consider this area as the pseudo ground truth EZ for each patient.

Quantitative Performance: Table 1 reports the performance of each model. We use a De Long's test on the AUC metric [23] to determine statistically significant improvement between our proposed framework and each baseline. We note an improvement in sensitivity and AUC when using the transformer to extract the temporal attention weights. Likewise, we note an improvement when using data augmentation for training. This is likely because our clinical dataset is too small to extract information from using our dynamic model. Finally, Fig. 3 shows model the ground truth (red) and predicted (yellow) labels for three Epilepsy patients among all models. The observed trends reflect the metrics in Table 1.

Temporal Attention: Fig. 4 shows the temporal attention weights recovered from each method that uses attention (proposed, ANNattn, and LSTM) for each of the 14 epilepsy patients during the testing phase. We observe a larger dynamic range in the weights recovered from the proposed framework, as compared to the ablated models. We conjecture that the transformer learns more nuances in the dFC data that improve

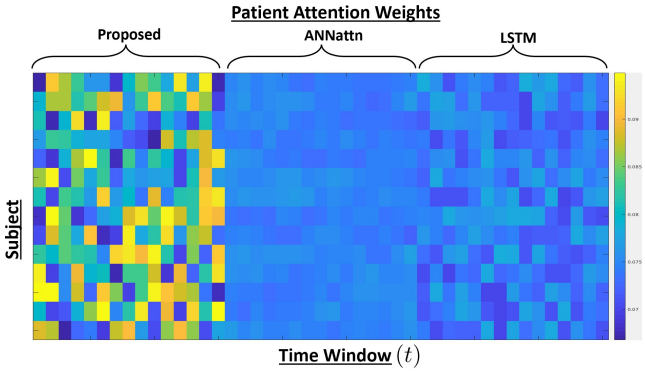

Fig. 4. Temporal attention weights recovered for **Left:** the proposed framework, **Middle:** the ANNattn ablation model, and **Right:** the LSTM ablation model for all epilepsy patients.

the region-wise EZ classification. We hypothesize that the MHA operation, which inherently captures similarities and differences between time-points, is responsible for better honing in on the relevant intervals for prediction.

4. CONCLUSION

We propose a novel end-to-end model based on dynamic functional connectivity to localize the EZ in focal epilepsy patients. Our model leverages a combined GCN + transformer architecture for feature extraction and temporal tracking. In parallel, we leverage a simple yet effective data augmentation strategy for robust training. We show statistically significant improvements over the baseline methods, and hypothesize that the performance gain is directly related to using the transformer-based attention module, which hones in on relevant intervals of the dFC time series for EZ prediction. Our work shows increased promise in using rs-fMRI as a preoperative protocol for noninvasive EZ localization.

Acknowledgements: This work is supported by the Na-

tional 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] Katharina Allers et al., "The economic impact of epilepsy: a systematic review," *BMC neurology*, vol. 15, no. 1, pp. 1–16, 2015.

[2] Rekha Dwivedi et al., "Surgery for drug-resistant epilepsy in children," *New England Journal of Medicine*, vol. 377, no. 17, pp. 1639–1647, 2017.

[3] N. Tandon et al., "Analysis of morbidity and outcomes associated with use of subdural grids vs stereoelectroencephalography in patients with intractable epilepsy," *JAMA neurology*, vol. 76, no. 6, pp. 672–681, 2019.

[4] 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*, vol. 5, pp. 264–278, 2015.

[5] 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.

[6] S.M. Stufflebeam et al., "Localization of Focal Epileptic Discharges using Functional Connectivity Magnetic Resonance Imaging," *Journal of Neurosurgery*, 2011.

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

[8] Clara Huishi Zhang et al., "Lateralization and localization of epilepsy related hemodynamic foci using presurgical fmri," *Clinical Neurophysiology*, vol. 126, no. 1, pp. 27–38, 2015.

[9] Patrick H Luckett et al., "Deep learning resting state functional magnetic resonance imaging lateralization of temporal lobe epilepsy," *Epilepsia*, 2022.

[10] Naresh Nandakumar et al., "Deepez: A graph convolutional network for automated epileptogenic zone localization from resting-state fmri connectivity," *IEEE Transactions on Biomedical Engineering*, 2022.

[11] R Matthew Hutchison et al., "Dynamic functional connectivity: promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013.

[12] Nicha C Dvornek et al., "Identifying autism from resting-state fmri using long short-term memory networks," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2017, pp. 362–370.

[13] 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*. Springer, 2021, pp. 241–252.

[14] Y. Minowa et al., "Comparing deep learning models for age prediction based on the resting state fmri dataset from the" brain dock" service in japan," 2022.

[15] Si Zhang et al., "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019.

[16] A. Vaswani et al., "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[17] Adam Paszke et al., "Automatic differentiation in pytorch," 2017.

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

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

[20] 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.

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

[22] 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.

[23] Elizabeth R DeLong et al., "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, pp. 837–845, 1988.