ATTEND TO CONNECT: GNN FOR PREDICTING BRAIN FUNCTIONAL CONNECTIVITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Functional connectivity (FC) studies have demonstrated the benefits of investigating the brain and its disorders through the undirected weighted graph of fMRI correlation matrix. Most of the work with the FC, however, depends on the way the connectivity is computed and further depends on the manual post-hoc analysis of the FC matrices. In this work, we propose a deep learning architecture (BrainGNN) that learns the connectivity structure while learning to classify subjects. It simultaneously trains a graphical neural network on this graph and learns to select a sparse subset of brain regions important to the prediction task. We demonstrate the model's state-of-the-art classification performance on a schizophrenia fMRI dataset and show how introspection leads to disorder-relevant findings. The graphs learned by the model exhibit strong class discrimination, and the identified sparse subset of relevant regions is consistent with the schizophrenia literature.

1 INTRODUCTION

Functional connectivity which is often computed using cross-correlation among brain regions of interest (ROIs) is a powerful approach which has been shown to be informative for classifying brain disorders and revealing putative bio-markers relevant to the underlying disorder (Liu et al., 2008; Lynall et al., 2010; Yu et al., 2017; Gadgil et al., 2020). Inferring and using functional connectivity through spatio-temporal data, e.g. functional magnetic resonance imaging (fMRI), has been an especially important area of research in recent times. Functional connectivity can improve our understanding of brain dynamics and improve classification accuracy for brain disorders e.g. schizophrenia (Yan et al., 2017; Parisot et al., 2018; Kawahara et al., 2016; Ktena et al., 2017).

Existing studies often heavily depend on the underlying method of functional connectivity estimation (Rashid et al., 2016; Saha et al., 2020; Salman et al., 2019). These studies work very well on classification but do not learn a sparse graph and not helpful for identifying bio-markers in the brain. Other studies Du et al. (2018) on brain disorders utilize ROIs predefined based on anatomical or functional atlases, which are either fixed for all subjects or are based on group differences. These approaches ignore the variations in ROIs due to the underlying disease conditions and discount the possibility that only a small subset of ROIs may be important at a time.

In this work, we address the problems of using a fixed method of learning functional connectivity and using it as a fixed graph to represent brain structure (the standard practices) by utilizing a novel attention based Graph Neural Network (GNN) Li et al. (2016), which we call BrainGNN. We apply it to fMRI data and 1) achieve comparable classification accuracy to existing algorithms, 2) learn dynamic graph functional connectivity, and 3) increase model interpretability by learning which regions from the set of ROIs are relevant for classification.

2 MATERIALS AND METHODS

2.1 MATERIALS

In this study, we worked with the data from Function Biomedical Informatics Research Network (FBIRN) Keator et al. (2016) dataset including schizophrenia (SZ) patients and healthy controls

(HC) for testing our model. Resting fMRI data from the phase III FBIRN were analyzed for this project.

2.1.1 PREPROCESSING

The FBIRN dataset was pre-processed through SPM12 Penny et al. (2011) based on the MATLAB 2019 environment. The slice-timing was first performed on the data, and then subject head motions were corrected by the realignment procedure. After that, the data was warped to MNI space using EPI template and resampled to 3 mm^3 voxels. Finally, the data were smoothed with a 6mm FWHM Gaussian kernel. We selected subjects for further analysis Fu et al. (2021) if the subjects have head motion $\leq 3^\circ$ and ≤ 3 mm, and with functional data providing near full brain successful normalization (Fu et al., 2019). This resulted in a total of 311 subjects with 151 healthy controls and 160 subjects with schizophrenia. Each subject is represented by $\mathbf{X} \in \mathbb{R}^{\mathrm{x}\times\mathrm{y}\times\mathrm{z}\times\mathrm{t}}$, where $\mathrm{x}, \mathrm{y}, \mathrm{z}$ represent the number of voxels in each dimension and t is the number of time points which are 160. To reduce the affect of noise we zscore the time sequence of each voxel independently. To partition the data into regions use automated anatomical labeling (AAL) Tzourio-Mazoyer et al. (2002) which contains 116 brain regions. We take the weighted average of the voxel intensities inside a region. This results in a dataset $D = (S_1, S_2, S_3.....S_n)$ where $S_i \in \mathbb{R}^{r\times\mathrm{t}}$, $n = 311$, $r = 116$, $t = 160$.

2.2 METHOD

We have three distinct parts in our novel attention based GNN architecture: 1) a Convolutional Neural Network (CNN) Lecun et al. (1998), 2) a Self-Attention mechanism Vaswani et al. (2017), and 3) a GNN (Graph Neural Network). In this section we explain the purpose and details of each part separately. Refer to Figure A.1 for the complete architecture diagram of BrainGNN.

2.2.1 CNN ENCODER

We use a CNN Lecun et al. (1998); Kiranyaz et al. (2021) encoder to obtain the representation of individual regions created in the preprocessing step outlined in 2.1.1. Each region vector of dimension $\mathrm{t} = 160$ is passed through multiple layers of one dimensional convolution, and a fully connected layer to get final embedding. We use rectified linear unit (ReLU) as an activation layer between convolution layers. Our one dimensional CNN layer embeds the temporal features of regions and the spatial connections are handled in the attention and GNN parts of the architecture.

2.2.2 SELF ATTENTION

Using the embeddings created by the CNN encoder, we estimate the connectivity between brain regions using multi-head self-attention following (Vaswani et al., 2017) . The self-attention model creates three embeddings namely (key, query, value) for each region, using three simple linear layers. To create weights among regions, the model takes dot product of a region's query with every other region's key embedding to get scores between them. Hence, $score_{ij} = query_i \cdot key_j$. The scores are then converted to weights using softmax, $w_i = Softmax(score_i)$. Weights are multiplied with each region's $value$ embedding and summed together to create new representation for $region_i$.

This process is carried out for all regions, producing new representation of every region and weights between regions. These weights are then used as functional connectivity between brain regions. The self attention layer encodes the spatial axis for each subject and provides with connection between regions. Weights are learned via end to end learning of our model performing classification. This frees us from using predefined models or functions to estimate the connectivity.

2.2.3 GNN

Our graph network is based on a previously published model Li et al. (2016). Each subject is represented by a graph G having V, A, E where $V \in \mathbb{R}^{r\times\mathrm{t}}$ is the matrix of vertices, where each vertex is represented by an embedding acquired by self-attention. $A, E \in \mathbb{R}^{r\times r}$ are the adjacency and edge weight matrices. Since we do not use any existing method of computing edges, we construct a complete directed graph with backward edges, meaning every pair of vertices is joined by two directed edges with weights e_{ij} and $e_{ji} \in E$. For each GNN layer, at every step s, each node/region, sums feature vectors of every other region relative to the weight edge between the nodes and pass

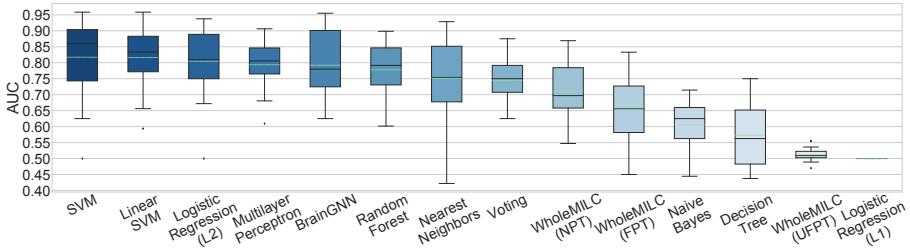


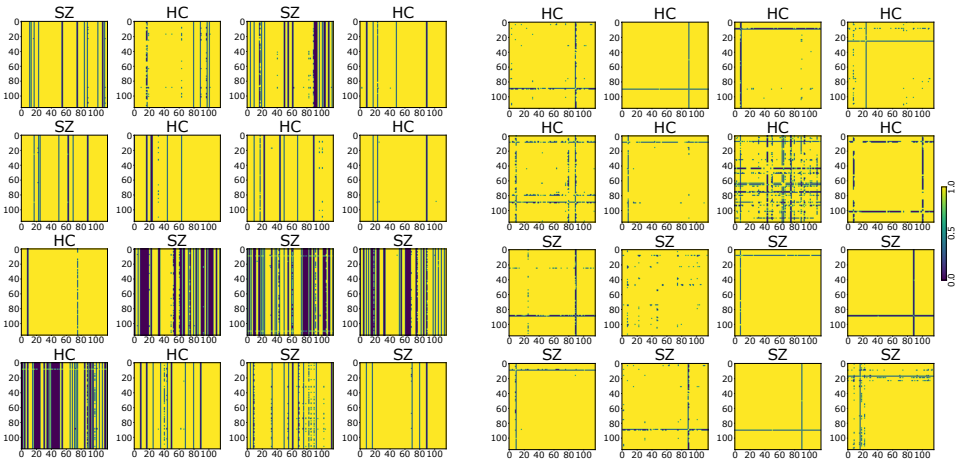
Figure 1: The ROCAUC of various classifiers trained on correlation matrices.

the resultant and it's own feature vector through a gated recurrent unit (GRU) network Cho et al. (2014), to obtain new embedding for itself. $x_s^{n_i} = \mathrm{GRU}(x_{s-1}^{n_i}, \sum_{\forall n_j : n_j -> n_i} e_{ji} x_{s-1}^{n_j})$.

The number of steps is a hyper-parameter which we have set it as 2 based on our experiments. The graph neural network helps nodes to create new embeddings based on the embeddings of other regions in the graph weighted by the edge weights between them. We use 6 GNN layers, with the first 3 followed by a top-k pooling layer Gao & Ji (2019); Knyazev et al. (2019). Pooling is performed to help model focus on the important regions/nodes which are responsible for classification. Since we represent each subject as graph G, in the end we do graph classification by pooling all the feature vectors of the remaining 23 regions/nodes. To get one feature vector from the entire graph we concatenate the output of three different pooling layers. We use graph max pool, graph average pool and attention based pool Vinyals et al. (2016). The resultant feature vector is then passed through a single linear layer for classification.

2.2.4 Training and Testing

To train, validate and test our model we divide the total 311 subjects into three groups of size 215, 80 and 16, for training, validating and testing respectively. We use 19 fold cross validation with 10 trials per fold, resulting in a total of 190 trials. We randomly select 100 subjects per class for each trial. We calculate the area under the ROC (receiver operating characteristic) curve (AUC) for each trial. To optimize our model we train all of our architecture in an end to end fashion, using Cross Entropy to calculate our loss by giving true labels Y as targets. Let θ represent the parameters of the entire architecture. $loss = \mathrm{CrossEntropy}(\hat{Y}, Y)$; $\theta^* = \arg\min_\theta(loss; \theta)$



(a) Regions' Connectivity - BrainGNN (Directed) (b) Regions' Connectivity - sFNC (Undirected)

Figure 2: 1: BrainGNN provides mean AUC as 0.79, which is just (~ 0.02) less than the best performing model (SVM). 2a:Weights of SZ class are sparser than HC. 2b: The matrices are symmetric but are less informative than those produced by BrainGNN.

3 RESULTS

We show three different groups of results in our study. 1) The classification results, 2) Regions' connectivity and 3) Key regions selection. We discuss these in the following sections. The input for the machine learning models is sFNC matrices produced using Pearson product-moment correlation coefficients (PCC).

a) Classification

We use AUC metric to quantify the classification results of our model. The performance is comparable to state of the art classical machine learning algorithms using hand crafted features and existing deep learning approaches such as Mahmood et al. (2020), which performed test on independent component analysis (ICA) components with a hold out dataset. Figure 1 shows that BrainGNN provide state of the art results.

b) Functional Connectivity

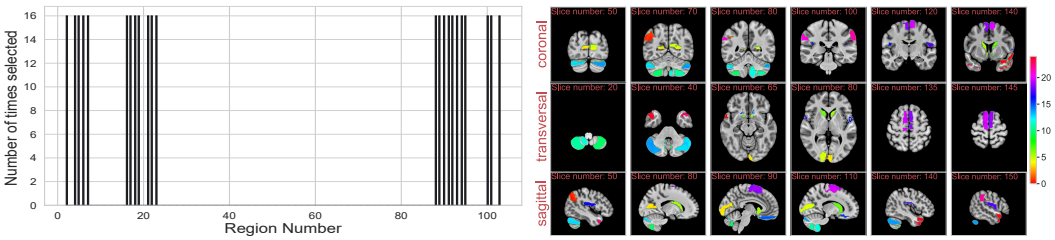
Figure 2a shows weight matrices for the second test set in cross validation. Weight matrices of subjects belonging to SZ class turn out to be much sparser than weights of healthy controls subjects. The result shows that the connectivity is limited to fewer regions, and functional connectivity differs across classes and fewer regions get higher weights in case of SZ subjects. FNC matrices produced using PCC method 2b, do not provide such level of information and almost all regions get unit weight between other regions.

c) Region Selection

As hypothesized, Figure 3a shows that not all regions are equally important/informative for the underlying disease. Figure 3a shows the final 23 regions selected after the last pooling layer in the GNN model (using 2nd test fold) which is just 20 percent of the total brain regions used. Figure 3b shows the location of the selected regions in the MNI brain space, regions are distinguished by color.

4 CONCLUSION

High classification performance shows that the model can accurately classify the subjects and hence can be trusted with the other two interpretative results of the paper. Functional connectivity between regions shown in the paper is of paramount importance as it highlights how brain regions are connected to each other and the variation between classes. The final regions selected by the model strengthens our hypotheses that not all regions are equally important for identifying a particular brain disorder. BrainGNN almost eliminates manual decisions transitioning graph construction and region selection into the data-driven realm. We envision in a future extension of the model that would enable it to work directly from the voxel-level not only connecting and selecting ROIs, but also constructing them.



(a) Region frequency (b) Region maps

Figure 3: 3a: Regions selected after the last pooling layer of GNN. All 23 regions are selected equal number of times (16). 3b: Mapping the 23 regions back on the brain across the three anatomical planes. 100th time point is selected for these brain scans. X axis shows different slices of the plane.

REFERENCES

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

Yuhui Du, Zening Fu, and Vince D. Calhoun. Classification and prediction of brain disorders using functional connectivity: Promising but challenging. *Frontiers in Neuroscience*, 12:525, 2018. ISSN 1662-453X. doi: 10.3389/fnins.2018.00525. URL https://www.frontiersin.org/article/10.3389/fnins.2018.00525.

Zening Fu, Arvind Caprihan, Jiayu Chen, Yuhui Du, John C Adair, Jing Sui, Gary A Rosenberg, and Vince D Calhoun. Altered static and dynamic functional network connectivity in alzheimer's disease and subcortical ischemic vascular disease: shared and specific brain connectivity abnormalities. *Human Brain Mapping*, 2019. doi: 10.1002/hbm.24591.

Zening Fu, Armin Iraji, Jessica A. Turner, Jing Sui, Robyn Miller, Godfrey D. Pearlson, and Vince D. Calhoun. Dynamic state with covarying brain activity-connectivity: On the pathophysiology of schizophrenia. *NeuroImage*, 224:117385, 2021. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2020.117385. URL http://www.sciencedirect.com/science/article/pii/S1053811920308703.

Soham Gadgil, Qingyu Zhao, Adolf Pfefferbaum, Edith V. Sullivan, Ehsan Adeli, and Kilian M. Pohl. Spatio-temporal graph convolution for resting-state fmri analysis. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 528–538, Cham, 2020. Springer International Publishing. ISBN 978-3-030-59728-3. doi: 10.1007/978-3-030-59728-3_52.

Hongyang Gao and Shuiwang Ji. Graph u-nets, 2019.

Jeremy Kawahara, Colin Brown, Steven Miller, Brian Booth, Vann Chau, Ruth Grunau, Jill Zwicker, and Ghassan Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146, 09 2016. doi: 10.1016/j.neuroimage.2016.09.046.

David B Keator, Theo GM van Erp, Jessica A Turner, Gary H Glover, Bryon A Mueller, Thomas T Liu, James T Voyvodic, Jerod Rasmussen, Vince D Calhoun, Hyo Jong Lee, et al. The function biomedical informatics research network data repository. *Neuroimage*, 124:1074–1079, 2016. doi: 10.1016/j.neuroimage.2015.09.003.

Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 1d convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 2021. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2020.107398. URL http://www.sciencedirect.com/science/article/pii/S0888327020307846.

Boris Knyazev, Graham W. Taylor, and Mohamed R. Amer. Understanding attention and generalization in graph neural networks, 2019.

Sofia Ira Ktena, Sarah Parisot, Enzo Ferrante, Martin Rajchl, Matthew Lee, Ben Glocker, and Daniel Rueckert. Distance metric learning using graph convolutional networks: Application to functional brain networks. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne (eds.), *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pp. 469–477, Cham, 2017. Springer International Publishing. ISBN 978-3-319-66182-7. doi: 10.1007/978-3-319-66182-7_54.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Yujia Li, Richard Zemel, Marc Brockschmidt, and Daniel Tarlow. Gated graph sequence neural networks. In *Proceedings of ICLR'16*, April 2016. URL https://www.microsoft.com/en-us/research/publication/gated-graph-sequence-neural-networks/.

Yong Liu, Meng Liang, Yuan Zhou, Yong He, Yihui Hao, Ming Song, Chunshui Yu, Haihong Liu, Zhening Liu, and Tianzi Jiang. Disrupted small-world networks in schizophrenia. *Brain*, 131(4): 945–961, 02 2008. ISSN 0006-8950. doi: 10.1093/brain/awn018. URL https://doi.org/10.1093/brain/awn018.

Mary-Ellen Lynall, Danielle S. Bassett, Robert Kerwin, Peter J. McKenna, Manfred Kitzbichler, Ulrich Muller, and Ed Bullmore. Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, 30(28):9477–9487, 2010. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.0333-10.2010. URL https://www.jneurosci.org/content/30/28/9477.

Usman Mahmood, Md Mahfuzur Rahman, Alex Fedorov, Noah Lewis, Zening Fu, Vince D. Calhoun, and Sergey M. Plis. Whole milc: Generalizing learned dynamics across tasks, datasets, and populations. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 407–417, Cham, 2020. Springer International Publishing. ISBN 978-3-030-59728-3. doi: 10.1007/978-3-030-59728-3_40.

Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease. *Medical Image Analysis*, 48:117 – 130, 2018. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2018.06.001. URL http://www.sciencedirect.com/science/article/pii/S1361841518303554.

William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

Barnaly Rashid, Mohammad R. Arbabshirani, Eswar Damaraju, Mustafa S. Cetin, Robyn Miller, Godfrey D. Pearlson, and Vince D. Calhoun. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fmri brain connectivity. *NeuroImage*, 134:645–657, 2016. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2016.04.051. URL https://www.sciencedirect.com/science/article/pii/S105381191630091X.

Debbrata K. Saha, Eswar Damaraju, Barnaly Rashid, Anees Abrol, Sergey M. Plis, and Vince D. Calhoun. A classification-based approach to estimate the number of resting fmri dynamic functional connectivity states. *bioRxiv*, 2020. doi: 10.1101/2020.06.24.161745. URL https://www.biorxiv.org/content/early/2020/06/25/2020.06.24.161745.

Mustafa S. Salman, Yuhui Du, Dongdong Lin, Zening Fu, Alex Fedorov, Eswar Damaraju, Jing Sui, Jiayu Chen, Andrew R. Mayer, Stefan Posse, Daniel H. Mathalon, Judith M. Ford, Theodorus Van Erp, and Vince D. Calhoun. Group ica for identifying biomarkers in schizophrenia: 'adaptive' networks via spatially constrained ica show more sensitivity to group differences than spatio-temporal regression. *NeuroImage: Clinical*, 22:101747, 2019. ISSN 2213-1582. doi: https://doi.org/10.1016/j.nicl.2019.101747. URL https://www.sciencedirect.com/science/article/pii/S221315821930097X.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, 15(1):273 – 289, 2002. ISSN 1053-8119. doi: https://doi.org/10.1006/nimg.2001.0978. URL http://www.sciencedirect.com/science/article/pii/S1053811901909784.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets, 2016.

Weizheng Yan, Sergey Plis, Vince D Calhoun, Shengfeng Liu, Rongtao Jiang, Tian-Zi Jiang, and Jing Sui. Discriminating schizophrenia from normal controls using resting state functional network connectivity: A deep neural network and layer-wise relevance propagation method. In *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2017. doi: 10.1109/MLSP.2017.8168179.

Renping Yu, Han Zhang, Le An, Xiaobo Chen, Zhihui Wei, and Dinggang Shen. Connectivity strength-weighted sparse group representation-based brain network construction for mci classification. *Human Brain Mapping*, 38(5):2370–2383, 2017. doi: https://doi.org/10.1002/hbm.23524. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23524.

A BRAINGNN ARCHITECTURE

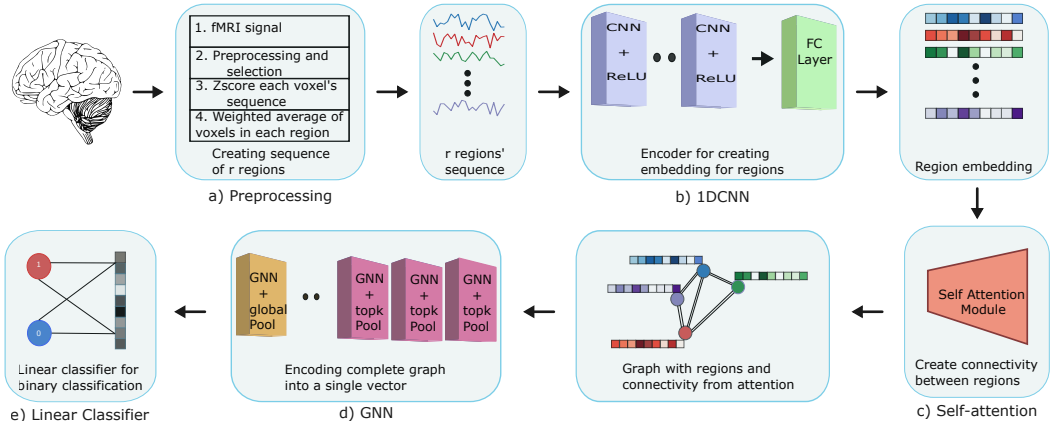


Figure A.1: BrainGNN architecture using a) Preprocessing: To preprocess the raw data with different steps (2.1.1). b) 1DCNN: To create embedding for regions (2.2.1). c) Self-attention: To create connectivity between regions (2.2.2) d) GNN: To obtain a single feature vector for the entire graph (2.2.3) and e) Linear classifier: To obtain the final classification.