HLGSNet: Hierarchical and Lightweight Graph Siamese Network with Triplet Loss for fMRI-based Classification of ADHD

Ranjeet Ranjan Jha, Aditya Nigam, Arnav Bhavsar
School of Computing and Electrical Engineering
Indian Institute of Technology
Mandi, India
d16044@students.iitmandi.ac.in, (aditya, arnav)@iitmandi.ac.in

Gaurav Jaswal
Department of Electrical Engineering
Indian Institute of Technology Delhi
Delhi, India
gauravjaswal@ee.iitd.ac.in

Sudhir K Pathak
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, USA
skpathak@pitt.edu

Rathish Kumar
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Kanpur, India
bvrk@iitk.ac.in

***Abstract*—Attention Deficit Hyperactivity Disorder (ADHD) is a behavior-based disorder that mainly occurs in young children. Resting-state fMRI data have been very popular for diagnosing brain disorders like Autism, ADHD, and schizophrenia, by network-based functional connectivity, since these disorders are associated with both individual brain regions and their connections. Finding patterns among regions of controls' brain and ADHD patients' discriminating brains, is a non-trivial task. For classification of ADHD, we propose an end-to-end lightweight CNN architecture with hierarchical representation learning i.e., HLGSNet. We extract 116 anatomical regions from each subject in both normal and patient conditions, and graphs are built with the help of temporal correlation between different regions, where each region is considered as a node. Following this, a Siamese graph convolution neural network with triplet loss has been trained for finding embeddings so that samples for the same class should have similar embeddings. Finally, along with a fully connected layer, the trained model has been fine-tuned for the classification task. Experiments have been carried out on publicly available ADHD-200 dataset with promising performance.**

***Index Terms*—fMRI, Graph convolution neural networks, ADHD detection, Siamese Classification**

I. Introduction

One of the fastest-growing methods for assessing neural connectivity is the functional magnetic resonance imaging (fMRI). It is broadly categorized into two classes: task-based fMRI [1] and resting-state fMRI [2]. More recently, functional connectivity studies using rs-fMRI has been yielding promising detection results for various diseases, including Alzheimer's, Attention deficit hyperactivity disorder (ADHD), Autism spectrum disorder (ASD), and epilepsy.

ADHD is one of the most severe neuro-developmental diseases impacting 5-10% of young children between the ages of 6 and 12. Since no specific diagnostic approach is reported for ADHD, the diagnosis depends on findings, typically over months, by medical practitioners or parents. Thus, this is a very time consuming and costly process. However, non-invasive brain imaging techniques along with advanced signal processing methods can help in early detection of ADHD [3], [4].

Machine learning methods can play a crucial role in the discovery of the difference in brain connectivity patterns across ADHD and healthy subjects. Many approaches, such as clustering [5], sparse dictionary learning [6], correlation [7], and graph-based techniques [8], have been used for feature extraction and selection, followed by a classifier to predict the actual class label. However, few existing classification techniques use hand-crafted features extractor and traditional machine learning frameworks which do not have automatic feature learning capability. Recently, deep learning methods [9], [10] are being explored to classify subjects based on the functional connectivity of brain regions.

Challenges and Contribution: There have been several scientific attempts to identify and recognize brain diseases, using fMRI activation. However, these methods are usually not generalizable to smaller data sets because traditional networks tend to get over-fitted and do not perform well on testing datasets. Moreover, conventional graph networks fail to learn hierarchical representations. To overcome aforementioned limitations, this paper presents an end-to-end CNN architecture, i.e., Hierarchical and Light-weight Graph Siamese Network (HLGSNet), for classification of neurological disorders, particularly ADHD from fMRI data. Our contribution in this paper is of four folds: (1) Each region has been represented by average signal of that region which helps to get the denoised signal unlike at voxel level. (2) The network is trained in a Siamese framework using a triplet loss function where graph convolution features are learned from the graph

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

representation of the brain. Moreover, training is assisted by the use of adaptive margin, and then a trained network along with a fully connected layer has been fine-tuned for the classification. (3) A differential pooling has been used to build the hierarchical representation of graph structures (4) A thorough experimental study to validate the proposed network on challenging ADHD datasets has been performed, achieving state-of-art performance on test data.

II. Literature Survey

It has been observed that correlation is one of the best methods to compute the functional connectivity network, which has different variants, including Pearson correlation coefficient, Kendall rank correlation, and Spearman rank correlation. Bohland et al. [11] has segmented the brain into 116 regions utilizing Automated Anatomical Labeling (AAL) atlas [12] and functional connectivity has been computed using Patel's Kappa, sparse regularized inverse covariance and Pearson's correlation methods. In [13], the whole brain is parcellating into 190 regions using CC200 atlas [14] and FCN has been constructed. Constructed FCN has been represented by a graph, where each of the regions is represented by a node in a graph; and the correlation between two nodes gives their edge weight. By putting some threshold on edge weight, they have built the unweighted graph.

Conventional Approaches: To measure the similarity between two graph embeddings, four main methods are reported in the literature: (i) graph embedding, (ii) graph edit distance, (iii) graph kernels, and (iv) motif counting. Among these methods, graph embedding [15] works on classification or regression modules to estimate brain graph similarity and have been deployed in various studies. However, this method fails to extract the structural properties of brain graphs [16]. In the case of motif counting, the subgraph patterns are obtained to compare brain graphs, but this a very time-consuming technique. Graph kernels compare brain graphs based on extracting smaller sub-graphs and thus unable to capture the global property of brain patterns.

On the other hand, graph edit distance deploys both structural and semantic feature variations lying on the graphs and also provides significant results to unknown node correspondences. In [17], the authors presented an approach of self-regulation of the right inferior frontal cortex with fMRI-Neurofeedback to enhance functional connectivity between the right inferior frontal cortex and other areas of the cognitive control network. In [18], a mechanism known as general functional connectivity (GFC) is proposed, which offered better test-retest reliability than resting-state fMRI to measures the intrinsic connectivity of the brain. In [19], a high field fMRI and high-density electroencephalogram (EEG) is considered to study the participant-driven activity patterns to retrieve the past episodes. In [20], they applied a classification approach on EEG-fMRI data and achieved IED-related BOLD maps with greater biological meaning compared to the conventional method. In [21], the usages of mutual connectivity analysis with Local Models is investigated to estimate nonlinear measures of interaction between pairs of time-series in terms of cross-predictability. In [22], authors utilized the non-imaging data for classification of an individual as ADHD or non-ADHD resulting from functional connectivity in various regions of the brain. They enabled the clustering algorithm via the method of the density peaks. In [23], efforts were made to classify an fMRI signal of a human as ADHD or non-ADHD based on differences in functional connectivity.

Deep learning Approaches: Deep learning approaches have been become very popular in vision field, and researchers have started to deploy this into medical domain also. In recent literature, several deep learning method could be seen in functional MRI (fMRI), diffusion MRI, and EEG. In addition, few efforts also made in fMRI based diseases classification like ADHD, and autism. In [24], an end-to-end deep learning model is designed that learned the network representations for classification of neurological disorder using fMRI data. Authors have performed experiments in three steps including feature extraction, followed by similarity and classification. Feature extraction and similarity network consist of CNN layers and Fully connected layers, respectively. In third step, classification has been carried out utilizing soft-max classifier. Finally, all the steps have been combined and end to end training would be performed.

In [25], a novel FCNet is proposed to obtain the inherent features of raw time-series signals by getting functional connectivity directly from the raw fMRI time-series signals. There are currently various types of neural networks, such as a convolutional neural network (CNN), built to learn a similarity feature that compares two pictures. In [26], the authors proposed a similarity metric between irregular graphs and achieved 11.9% improvement in accuracy over conventional methods. In specific, the polynomial filters were extended to irregular graphs utilizing a Siamese graph convolutional neural network (CNN). In this way, they have utilized graph convolution in spectral domain. Later, fully connected layer has been used to get the embedding. Finally, classification has been performed for autism dataset.

III. Proposed Hierarchical and Light-weight Graph Siamese Network (HLGSNet)

The proposed framework for ADHD classification based on HLGSNet illustrated in Figure 1. At the pre-processing stage, the first task of the network is to build the brain graphs of the raw fMRI data (as discussed in subsection III-A). Afterward, online triplets of these graphs have been created randomly and given to the Graph CNN network. Next, Graph CNN has been trained based on triplet (anchor, positive and negative) by using the triplet loss function. One can see in Figure 1 that the three legs of graph CNN with shared weights have been used to train the proposed model. Finally, trained graph CNN along with a soft-max classifier, has been used to perform the classification task.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

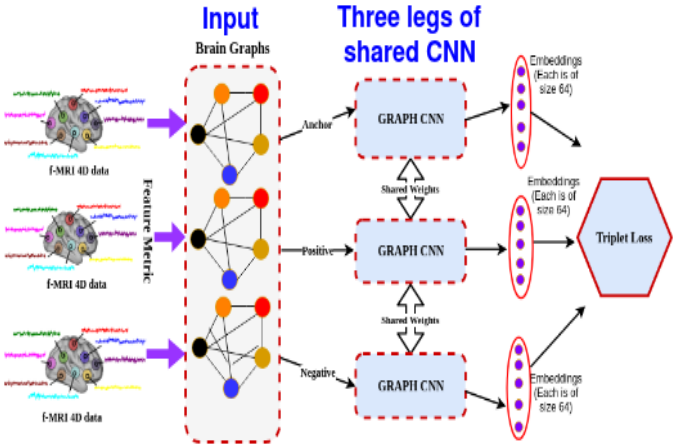

Fig. 1. Proposed Hierarchical and Lightweight Graph Siamese Network

A. Graph Construction for brain classification

A graph is defined by G = (V, E), where V and E are sets of vertices and edges, respectively. In our work, automated anatomical labeling atlas [27] has been used to parcellate the brain into 116 anatomically meaning-full regions for each subject, which represent the nodes. Now for graph construction, first, we compute the average of the signals from all the voxels of each region. In this way, each region would be represented by the average signal of that region. Moreover, the length of each signal is 171 since there are 171 temporal samples at each voxel define as the signal. Hence any brain B, can be represented as a signal X of size 116×171.

Further, the correlation matrix C of size of 116×116 is constructed, wherein each entry is the temporal correlation between these average signals from each pair of regions. We then construct the graph adjacency matrix A by thresholding C using a threshold value α. as defined below:

$$a_{ij} = \begin{cases} 1, & \text{if } c_{ij} > \alpha \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where c_{ij}, a_{ij} are i^{th} row and j^{th} column of C and A, respectively; both of them are of size 116×116. These two matrices (A and X) have been given to the Graph CNN as input for getting embedding.

Fig. 2 exhibits the (thresholded) difference between the mean correlation matrix of normal subjects and the ADHD subjects. Here, we have calculated the absolute difference between the mean correlation matrix of normal and patient to show the difference between normal and ADHD. Afterward, we have put the threshold over absolute difference to get the binary mask (one and zero denote edge and no edge, respectively), as plotted in Fig. 2. Moreover, we have shown the name of every fourth region among the total 116 regions. In the figure, red color indicates one (edge), whereas blue denotes zero (no edge). All the edges in the graph, as mentioned earlier, would be helpful for differentiating between normal and ADHD subjects. In this way, the Fig. 2 highlights connectivity differences between the two cases, which can lead our network to learn discriminative features between the two classes.

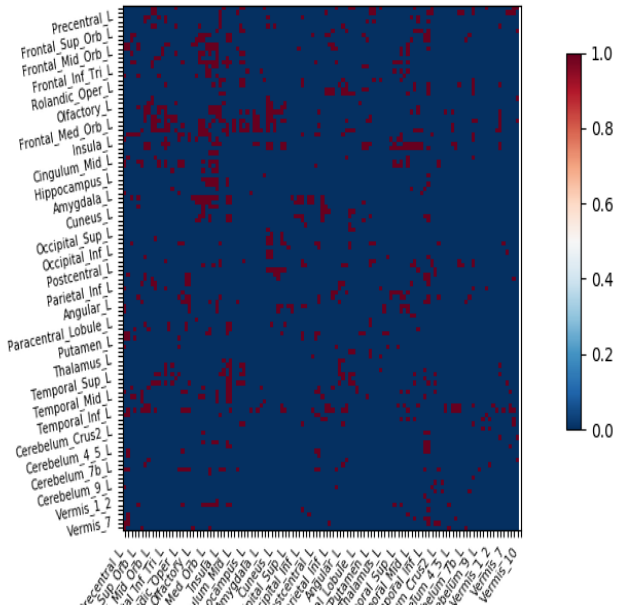

Fig. 2. Absolute difference between mean of normal and patient correlation matrix, followed by thresholding.

B. Graph CNN Network Designing and Architecture

The GCNN design (as shown in Figure 3) consists of the graph convolution layer, differential pooling layer, followed by the fully connected layer, as discussed below:

1) Graph Convolution Layer (GCL): Graph convolution [28] could be defined in the spatial domain or spectral domain by taking advantage of message propagation or filter based graph signal processing. In our case, we have utilized the graph convolution in spectral-domain, which takes the leverage of graph signal processing.

Laplacian: By assuming the graph is undirected, the foremost entity is the normalized Laplacian matrix, which plays a significant role in the realization of spectral graph convolution. The first step in the GCL layer is to calculate normalized Laplacian matrix L_{norm}, which is defined as follows:

$$L_{norm} = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{2}$$

Where A, D, and I_N are adjacency matrix, degree diagonal matrix, and identity matrix, respectively; each of them is of size 116×116.

As L_{norm} matrix is real symmetric positive semi-definite, it can be decomposed into three matrices as defined below (using spectral decomposition theorem):

$$L_{norm} = \mathbf{U}\Lambda\mathbf{U}^{\intercal} \tag{3}$$

where U consists of a set of orthonormal eigenvectors ($U = [ev_1, ev_2, ev_3, ...ev_n]$) , and Λ is a diagonal matrix containing

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

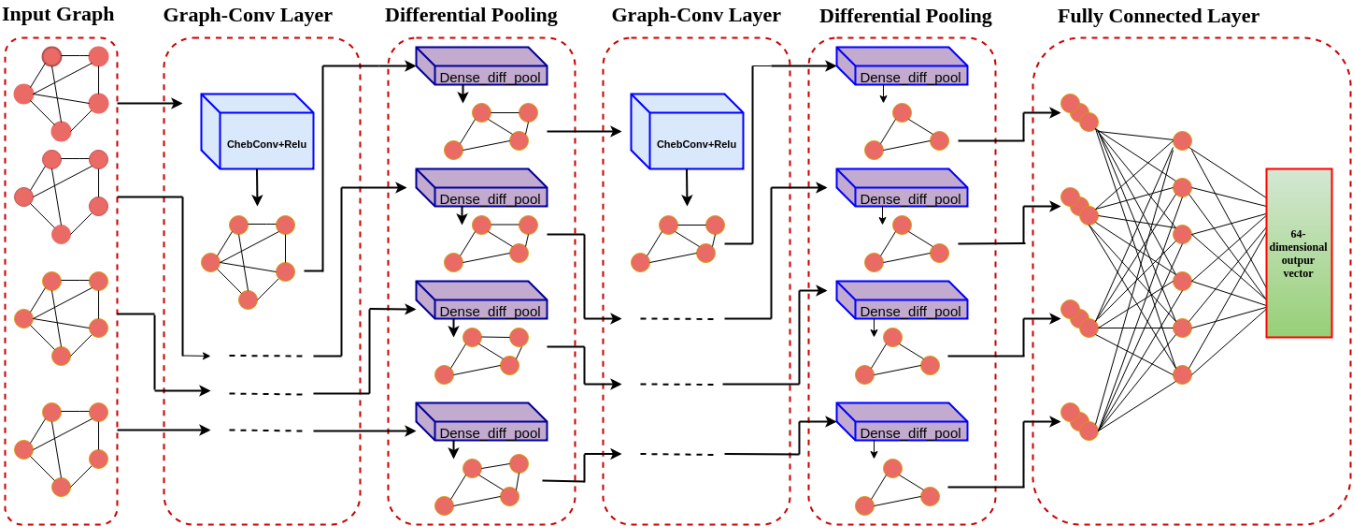

Fig. 3. Proposed Graph CNN Network

all eigenvalues (also called spectrum). Since all the eigenvectors have been defined in orthonormal space, $U^T U = I$.

Graph Fourier Transform: The Graph Fourier Transform (GFT) of X (containing feature vectors of all 116 nodes) is calculated by multiplying eigenvector matrix with graph signal matrix X as follows: $X^{'} = \mathbf{U}^{\intercal} X$, whereas Inverse Graph Fourier transform (IGFT) is defined as $X = \mathbf{U} X^{'}$.

As such, we could see that GFT would transform the input signal to the orthonormal space where the eigenvectors, as mentioned above, represents the basis spanning the space of 116 length vector of brain signals. Thus utilizing the same basis, IGFT would transform the signal defined in orthonormal space to the original space by multiply basis with corresponding $X^{'}$ values.

Defining Graph Convolution: Given two signals Z, W, signal Z can be convolved by signal W as defined below:

$$Z \circledast W = IGFT(GFT(Z) \cdot GFT(W)) = U(U^{\intercal} Z) \cdot (U^{\intercal} W) \tag{4}$$

Where $\circledast$ is convolution operation, and $\cdot$ is the element-wise product of two matrices viz; GFT of Z and GFT of W. Moreover, inverse GFT is applied to get final convolution output in signal space.

In the similar way as discussed above, the spectral convolution of X by a filter p can be defined as:

$$\begin{aligned} X_{conv} = p \circledast X &= IGFT(GFT(p) \cdot GFT(X)) \\ &= U(U^{\intercal} p) \cdot (U^{\intercal} X) \end{aligned} \tag{5}$$

Let us consider the p_θ as a learnable filter, represented by $p_\theta = diag(U^T p)$ and parameterized by θ; where θ is Fourier coefficient vector need to be learned. Hence the simplified spectral convolution can be written as follows.

$$X_{conv} = p \circledast X = \Big(U \big(p_\theta (U^{\intercal} X) \big) \Big) \tag{6}$$

Considering the above mentioned spectral convolution, the input graph signal M^{l-1} with (if^{l-1}) channels at l^{th} layer could be establish as below.

$$M^l_{(:,q)} = \sum_{r=1}^{if^{l-1}} U \Theta^l_{r,q} U^{\intercal} M^{l-1}_{(:,r)} \tag{7}$$

Here, of^l $(q = 1, 2, ..., of^l)$ is number of output channels at layer l. In addition, M^0 = X at layer 1 of Spectral convolution. At last $p_\theta = \Theta^l_{r,q}$ is learnable diagonal matrix at layer l.

Approximation: Above mentioned decomposing of the Laplacian matrix into orthonormal eigen vectors matrix is a time-consuming process (high complexity of order N^3). However, Chebyshev expansion can be employed to approximate eigenvectors matrix decomposition in time order of N. In this way, filter p_θ would be approximated by Chebyshev expansion as follows:

$$p_\theta = \sum_{n=0}^{T} \Omega_n S_n(\Lambda^z) \tag{8}$$

where $\Lambda^z = \frac{2 \times \Lambda}{\lambda_{max}} - I$, so as to normalize between $[-1, 1]$. Ω_n represents n^{th} Chebyshev coefficient, and T denotes Chebyshev polynomial order and $S_n(\Lambda^z)$ is defined as:

$$S_n(\Lambda^z) = \begin{cases} 1, & \text{if } m = 0 \\ \Lambda^z, & \text{if } m = 1 \\ 2S_{n-1}(\Lambda^z) - S_{n-2}(\Lambda^z), & \text{otherwise} \end{cases} \tag{9}$$

In consequence, the convolution of X with filter p_θ, by utilizing the above mentioned formulation, would be given as:

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

$$
\begin{aligned}
X_{conv} =& p \circledast X = \Big(U\big(p_\theta(U^\intercal X)\big)\Big) \\
&= U\Big(\sum_{n=0}^{T} \Omega_n S_n(\Lambda^z)\Big) U^\intercal X \\
&= \sum_{n=0}^{T} \Omega_n S_n(L^z) X
\end{aligned} \tag{10}
$$

here, $S_n(L^z) = US_n(\Lambda^z)U^\intercal$ and $L^z = \frac{2\times L}{\lambda_{max}} - I$. As we have utilized this Chebyshev approximation, this algorithm has reduced the time complexity to order of N from order N^3.

Differential Pooling Layer : In this layer, cluster assignment matrix $C_l(i,j)$ has been learned, which is of size $n_l \times n_{l+1}$. Each row of this matrix is related to one of the n_l clusters (or nodes) at layer l and each column represents one of the n_{l+1} clusters at the next layer $l+1$. Hence, this matrix gives a soft assignment of each node of the present layer to the clusters of the next layer. In other words, we can say that based on this matrix, nodes information are aggregated and combined to form new nodes or cluster for next layer, and accordingly the adjacency and graph signal matrices have been changed as defined below.

$$X_{inp}^{l+1} = C^{l\intercal} X_{out}^{l} \tag{11}$$

$$A_{inp}^{l+1} = C^{l\intercal} A_{out}^{l} C^{l} \tag{12}$$

Where, X_{inp}^{l+1} is input graph signal matrix for layer $l+1$, X_{out}^{l} is output graph signal matrix for layer l. Similarly A_{inp}^{l+1} is input graph adjacency matrix for layer $l+1$, A_{out}^{l} is output graph adjacency matrix for layer l.

Fully Connected Layer: This layer has been utilized to get the final embedding (of size 64) of the input feature map. These embeddings are used for training the triplet-based Siamese network. Once the Siamese network is trained, these embeddings are further used to train the softmax classifier with cross-entropy loss. In this way, one more fully connected layer with activation function soft-max has been added after a trained network (by Siamese). The output would give a two-dimensional probability vector. Finally, the cross-entropy loss function has been utilized to get the classification error.

C. Network Training with Triplet Loss

Training has been done in two steps: initially siamese network has been trained with a series of layers viz; GCL-I, Pooling with embedding, GCL-II, Pooling with embedding and two fully-connected layers. For training of this network triplet loss function has been used, as defined below:

$$Loss = \sum_{i=1}^{N} [\|E_i^{anc} - E_i^{pos}\|_2^2 - \|E_i^{anc} - E_i^{neg}\|_2^2 + \alpha] \tag{13}$$

In this equation, α, E_i^{anc}, E_i^{pos}, and E_i^{neg} are the margin, embedding of anchor, positive and negative example, respectively. This function minimizes the loss of a similar class of embedding by minimizing the distance from the anchor to positive embedding while maximizing the distance from the negative and anchor. These embeddings will help to train the classifier in a proper way without over-fitting as we are now having the network with multiple triplets of samples unlike standard CNN classifier.

In this way, our approach involves a Siamese neural network which contains triplet networks of Graph convolution layers (GCN), importantly with a sophisticated loss, unlike in [26], Moreover, we have also used contemporary GCN architectures for further performance improvement.

TABLE I
METRICS MEASURED FOR DIFFERENT TESTING DATABASES

Dataset	Accuracy	Sensitivity	Specificity
NI	68%	72.73%	64.29%
NYU	68.29%	66.67%	68.96%

IV. PERFORMANCE ANALYSIS AND DISCUSSIONS

This section discusses the testing strategy, dataset, experiments, and robustness of the proposed CNN model w.r.t. various test cases.

A. Datasets Description

We have two datasets to validate our approach, which had also utilized in ADHD-200 Global Competition. Each dataset has provided both training and testing data separately with different apparatus and scan parameters. We have performed our experiments on pre-processed data. FSL [29] and AFNI [30] tools have been utilized to prepossess the above-mentioned dataset, which consists of various steps:

1) First four volumes have been removed.
2) lice time correction.
3) Montreal Neurological Institute (MNI) template with $4\times 4 \times 4$ voxel resolution has been used for registration.
4) The Gaussian filter has been applied for the smoothing of data.

B. Experimental Setup

We have made use of NVIDIA GeForce GTX 1080 Ti graphics card along-with a Linux operating system with 32 GB RAM for the implementation of our proposed network. Initially, the triplet based Siamese network is trained using Adam optimizer together with a learning rate of 0.001. The trained network, accompanied by the soft-max classifier (apply cross-entropy loss), has been further utilized to perform the classification task.

C. Testing Strategy

The proposed HLGSNet has been trained over only 270 subjects (NI and NYU) of both normal and patients and tested individually on NI and NYU with promising results. NI contains 25 subjects for testing in which 14 and 11 belong to normal and patient, respectively. NYU testing data consists of 12 normal and 29 patients subjects, a total of 41 subjects. Since we have less data, we used the siamese-based network to mitigate the over-fitting issue of neural networks. In this way,

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE II
COMPARISONS OF THE PROPOSED METHOD WITH EXISTING METHODS

Data-set	NI			NYU		
Approach	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
Average accuracy [31]	56.9%	N/A	N/A	35.1%	N/A	N/A
Clustering approach [32]	44%	45.4%	42.8%	24.3%	68.9%	60.9%
FCNet [25]	60%	N/A	N/A	58.5%	N/A	N/A
DeepFMRI [24]	67.9%	63.6%	71.4%	73.1%	65.5%	91.6%
Proposed Network	68%	72.73%	64.29%	68.29%	66.67%	68.96%

siamese network would take triplet as input, and huge amount of triplet could be generated using only less amount of data. Finally, several performance parameters have been utilized to check the robustness of our model.

1) Performance parameters: We have used three types of performance parameters including Accuracy, Specificity, and Sensitivity. They have defined via true positive, true negative, false positive and false negative as defined below.

True Negative (TN): # correctly classified normal subjects.

True Positive (TP): # correctly classified ADHD subjects.

False Positive (FP): # normal subjects which is mistakenly classified as ADHD.

False Negative (FN): # ADHD subjects which is mistakenly classified as normal.

The performance parameters are:

- **Sensitivity:** # correctly classified ADHD subjects over total # actual ADHD subjects.

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

- **Specificity:** # correctly classified normal subjects over total # actual normal subjects.

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

- **Accuracy:** # correctly classified subjects over total # subjects.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \tag{16}$$

D. Experimental Analysis

The performance analysis of the test data for the proposed network is presented in Table 1. Among all datasets, our proposed network has achieved the best accuracy on NYU. While in terms of sensitivity, the method yields 72.73 % on the NI dataset, which is highest among all. On the other hand, the highest specificity 68.96% is obtained on the NYU dataset among all the datasets. Hence, these results show that our proposed method yields a promising performance, given the problem's small amount of data and the challenging nature of it.

Comparative Analysis: To show the performance comparison, we have compared the proposed network with four state-of-art methods based on accuracy, as shown in Table 2. One can observe that in the case of NI dataset, our method outperforms all the state-of-the-art approaches. Likewise, in the case of the NYU dataset, our proposed method yields the second-best highest accuracy. If we take the case of sensitivity which is more important performance parameter of any classification model, our proposed technique achieves 72.73% on NI data, that is far better than all the existing approaches in Table 2. Considering specificity for NI dataset, our model has surpassed all the state-of-art methods except [24]. For NYU data, the sensitivity of our proposed model is better than all existing techniques except [32], however in the case of specificity we are better than [32].

While the approach in [24] outperforms the proposed method on the NYU data, in terms of accuracy, we believe that the proposed approach and its paradigm of using GCN variations are encouraging, and can be further explored.

V. CONCLUSIONS

This paper presents an end-to-end hierarchical Siamese classification network (HLGSNet) for classification of ADHD and control subjects. Specifically, HLGSNet is comprised of a graphical CNN feature extractor network that extracts features from the brain graphs and a learnable classification network that calculates embeddings to emphasize similarity and dissimilarity between the graphs corresponding to the class samples via the triplet loss function. Experimental findings of our methodology indicate encouraging success on the ADHD-200 dataset.

REFERENCES

[1] C.-W. Woo, A. Krishnan, and T. D. Wager, "Cluster-extent based thresholding in fmri analyses: pitfalls and recommendations," *Neuroimage*, vol. 91, pp. 412–419, 2014.

[2] B. B. Biswal, "Resting state fmri: a personal history," *Neuroimage*, vol. 62, no. 2, pp. 938–944, 2012.

[3] M. Altınkaynak, N. Dolu, A. Güven, F. Pektaş, S. Özmen, E. Demirci, and M. İzzetoğlu, "Diagnosis of attention deficit hyperactivity disorder with combined time and frequency features," *Biocybernetics and Biomedical Engineering*, 2020.

[4] L. Dubreuil-Vall, G. Ruffini, and J. A. Camprodon, "Deep learning convolutional neural networks discriminate adult adhd from healthy individuals on the basis of event-related spectral eeg," *Frontiers in neuroscience*, vol. 14, 2020.

[5] L. Duan and X. Mai, "Spectral clustering-based resting-state network detection approach for functional near-infrared spectroscopy," *Biomedical Optics Express*, vol. 11, no. 4, pp. 2191–2204, 2020.

[6] M. Ghasemi, M. Kelarestaghi, F. Eshghi, and A. Sharifi, "T2-fdl: A robust sparse representation method using adaptive type-2 fuzzy dictionary learning for medical image classification," *Expert Systems with Applications*, p. 113500, 2020.

[7] M. Rosaleena, W. A. Sethares, V. A. Nair, and P. Vivek, "Rethinking measures of functional connectivity via feature extraction," *Scientific Reports (Nature Publisher Group)*, vol. 10, no. 1, 2020.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

[8] Y. Du, J. Sui, and D. Lin, "Identifying neuroimaging-based markers for distinguishing brain disorders," *Frontiers in Neuroscience*, vol. 14, 2020.

[9] R. R. Jha, G. Jaswal, D. Gupta, S. Saini, and A. Nigam, "Pixisegnet: pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck," *IET Biometrics*, vol. 9, no. 1, pp. 11–24, 2019.

[10] S. Kumari, R. R. Jha, A. Bhavsar, and A. Nigam, "Indoor–outdoor scene classification with residual convolutional neural network," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing*. Springer, 2020, pp. 325–337.

[11] J. W. Bohland, S. Saperstein, F. Pereira, J. Rapin, and L. Grady, "Network, anatomical, and non-imaging measures for the prediction of adhd diagnosis in individual subjects," *Frontiers in systems neuroscience*, vol. 6, p. 78, 2012.

[12] 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*, vol. 15, no. 1, pp. 273–289, 2002.

[13] S. Dey, A. R. Rao, and M. Shah, "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects," *Frontiers in neural circuits*, vol. 8, p. 64, 2014.

[14] R. C. Craddock, G. A. James, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg, "A whole brain fmri atlas generated via spatially constrained spectral clustering," *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.

[15] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. Duncan, "Graph embedding using infomax for asd classification and brain functional difference detection," in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317. International Society for Optics and Photonics, 2020, p. 1131702.

[16] S. H. Hosseini, F. Hoeft, and S. R. Kesler, "Gat: a graph-theoretical analysis toolbox for analyzing between-group differences in large-scale structural and functional brain networks," *PloS one*, vol. 7, no. 7, 2012.

[17] K. Rubia, M. Criaud, M. Wulff, A. Alegria, H. Brinson, G. Barker, D. Stahl, and V. Giampietro, "Functional connectivity changes associated with fmri neurofeedback of right inferior frontal cortex in adolescents with adhd," *NeuroImage*, vol. 188, pp. 43–58, 2019.

[18] M. L. Elliott, A. R. Knodt, M. Cooke, M. J. Kim, T. R. Melzer, R. Keenan, D. Ireland, S. Ramrakha, R. Poulton, A. Caspi *et al.*, "General functional connectivity: shared features of resting-state and task fmri drive reliable and heritable individual differences in functional brain networks," *NeuroImage*, 2019.

[19] L. Bréchet, D. Brunet, G. Birot, R. Gruetter, C. M. Michel, and J. Jorge, "Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with eeg and fmri," *NeuroImage*, 2019.

[20] A. M. DSouza, A. Z. Abidin, and A. Wismüller, "Classification of autism spectrum disorder from resting-state fmri with mutual connectivity analysis," in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10953. International Society for Optics and Photonics, 2019, p. 109531D.

[21] N. K. Sharma, C. Pedreira, U. J. Chaudhary, M. Centeno, D. W. Carmichael, T. Yadee, T. Murta, B. Diehl, and L. Lemieux, "Bold mapping of human epileptic spikes recorded during simultaneous intracranial eeg-fmri: The impact of automated spike classification," *NeuroImage*, vol. 184, pp. 981–992, 2019.

[22] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, "Fusion of fmri and non-imaging data for adhd classification," *Computerized Medical Imaging and Graphics*, vol. 65, pp. 115–128, 2018.

[23] A. Riaz, E. Alonso, and G. Slabaugh, "Phenotypic integrated framework for classification of adhd using fmri," in *International Conference on Image Analysis and Recognition*. Springer, 2016, pp. 217–225.

[24] A. Riaz, M. Asad, S. M. R. Al Arif, E. Alonso, D. Dima, P. Corr, and G. Slabaugh, "Deep fmri: An end-to-end deep network for classification of fmri data," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1419–1422.

[25] A. Riaz, M. Asad, S. M. R. Al-Arif, E. Alonso, D. Dima, P. Corr, and G. Slabaugh, "Fcnet: a convolutional neural network for calculating functional connectivity from functional mri," in *International Workshop on Connectomics in Neuroimaging*. Springer, 2017, pp. 70–78.

[26] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Distance metric learning using graph convolutional networks: Application to functional brain networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 469–477.

[27] "Aal," http://neuro.imm.dtu.dk/wiki/AutomatedAnatomicalLabeling, 2010.

[28] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[29] "Fsl," https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation, 2008.

[30] "Afni," https://afni.nimh.nih.gov/, 1994.

[31] "Adhd-200," http://fcon$_1$000.projects. nitrc.org/indi/adhd200/.

[32] A. Riaz, E. Alonso, and G. Slabaugh, "Phenotypic integrated framework for classification of adhd using fmri," in *International Conference on Image Analysis and Recognition*. Springer, 2016, pp. 217–225.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE