IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 18, 2020, accepted March 23, 2020, date of publication March 26, 2020, date of current version April 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2983590

Graph Wavelet-Based Multilevel Graph Coarsening and Its Application in Graph-CNN for Alzheimer's Disease Detection

HIMANSHU PADOLE, (Member, IEEE), SHIV DUTT JOSHI, AND TAPAN K. GANDHI
Department of Electrical Engineering, IIT Delhi, New Delhi 110016, India

Corresponding author: Himanshu Padole (himanshu.pramod.padole@ee.iitd.ac.in)

This work was supported by Indian Institute of Technology Delhi under the Prime Minister's Research Fellowship program, Government of India.

ABSTRACT Along with the classical applications like graph partitioning, graph visualization, etc., graph coarsening has been recently applied in graph convolutional neural network (GCNN) architectures to perform the pooling operation in the graph domain. In this paper, we propose a novel two-stage graph coarsening method rooted on the graph signal processing with its application in the GCNN architecture. In the first stage of coarsening, the graph wavelet transform (GWT) based features are used to obtain a coarsened graph which preserves the topological characteristics of the original graph. In the second stage, the coarsening problem is formulated as an optimization problem where the reduced Laplacian operator at each level is obtained as a restriction of the original Laplacian operator to a specified subspace that also maximizes the topological similarity. The performance of the proposed coarsening algorithm is quantified in the general coarsening context using different graph coarsening quality measures. Its effectiveness as a pooling operator in GCNN is validated by applying it for the graph coarsening operation in the GCNN architecture. This modified GCNN architecture is then used as a graph signal classifier for the early detection of Alzheimer's disease. The results show that the proposed coarsening method outperforms state-of-the-art methods, both in the general coarsening context and as a pooling operator in the GCNN architecture.

INDEX TERMS Graph coarsening, graph signal processing, convolutional neural network, Alzheimer's disease.

I. INTRODUCTION

Recently, a new area of research called graph signal processing (GSP) has emerged in the field of signal processing which deals with the data residing on irregular structure e.g. biological networks [1], sensor networks [2] etc. Although the field of GSP has witnessed a significant growth in the recent years [3]–[5], its application in many real-life problems is limited by the large size of the graphs involved as the computational complexity of most of the algorithms in GSP grow polynomially with the graph size [6]. Traditionally, graph coarsening has been one of the most efficient ways to circumvent this issue wherein the given graph is mapped to a corresponding smaller graph which preserves some of the key properties of the original graph. The required operation or the analysis is first performed on the coarsened graph and the obtained solution is then refined in the original domain which brings down the overall computational complexity.

Graph coarsening has been used extensively in many network analysis tools which includes graph partitioning [7], graph visualization [8] etc. It has been used in GSP to obtain a multiscale representation of a graph signal [9]. Critically sampled graph wavelet filter banks also use the graph coarsening at its core for the downsampling and the reduction of the graph signal [10]. In algebraic multigrid methods, coarsening has been used to find the solution of differential equations and to solve the system of linear equations in almost linear time [11]. Recently, graph coarsening has found its application in the field of machine learning and artificial neural network also where the graph coarsening is used to perform the pooling operation in the graph convolutional neural network (GCNN) [12], [13].

To solve this graph coarsening problem, different approaches have been proposed by the researchers working in different domains ranging from graph theory [6], [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

multigrid methods [11], [14], to a more recently developed graph signal processing (GSP) [9], [10], [15]. Traditionally, most of the coarsening algorithms have been designed such that the coarsened graph preserves the geometric or the topological structure of the original graph [7], [16], although in few papers [15], [17], [18], the authors proposed the coarsening algorithms that preserve the spectral and some related characteristics of the original graph. But most of these existing coarsening methods suffer from two common drawbacks: (i) while designing an algorithm which preserves the topological structure, only the direct connection between nodes is considered, neglecting the local environment and the neighbouring nodes [7], [9], [10]. (ii) The relation between the action of original graph Laplacian operator and the reduced graph Laplacian operator is not taken into account while constructing the reduced graph Laplacian operator [7], [9], [14]. Recently, few attempts [14], [17, [18] have been made to solve one of these two issues, but none of them try to address both the issues simultaneously. As far as the coarsening operation in GCNN is concerned, it comes up with an additional challenge associated with the multilevel coarsening as the pooling operation in multiple layers requires the multilevel coarsening of the initial original graph. In most of the conventional multilevel coarsening methods, at each level of coarsening, the coarsened graph is constructed to approximate the graph obtained in the previous level rather than to approximate the original graph [7], [9], [10]. In this work, we propose a novel two-stage multilevel graph coarsening algorithm, in which we attempt to overcome the above drawbacks using the concepts of graph signal processing and validate the proposed algorithm in the context of GCNN.

Graph signal processing deals with the signals residing on the irregular structure characterized by an associated graph [3], [4]. In the present era of data explosion, the field of GSP has received a lot of attention, due to its ability to efficiently process the data that naturally reside on the irregular structure e.g. social network data, gene expression data etc. [5], [19]. Basic signal processing operations such as shifting, sampling, Fourier transform etc. have been generalized from the regular Euclidean domain to the graph domain, thus forming the basis of the graph signal processing [3]–[5]. Apart from these basic operations, efforts are being made to extend the more involved concepts in signal processing such as multiresolution analysis, wavelet transform, convolutional neural networks etc. from the regular domain to the graph domain [13], [20]–[23]. To circumvent the first problem associated with the graph coarsening algorithms, in the first stage of our proposed algorithm, we use the similarity between the graph wavelets centered around each node to coarsen the graph as the graph wavelet encodes the local information in the graph instead of considering only the direct connections between the nodes [24]. Wavelet transform has been defined for the graph signal using different approaches [21], [22], out of which, we used the spectral graph wavelet proposed in [22]. To overcome the second limitation of the existing coarsening methods along with the first as described earlier, we propose an optimization framework in which the final reduced Laplacian operator is obtained as a restriction of the original graph Laplacian operator to a specified subspace which also matches with Laplacian obtained from the first stage. Finally, we extend the proposed single level coarsening algorithm to a multilevel coarsening algorithm in which at each level of coarsening, we try to preserve the characteristics of the original graph rather than that of the graph obtained by the previous level of coarsening.

Recently, several architectures have been proposed to generalize the convolutional neural network (CNN) to the graph domain [12], [13], [25]. Though, serious efforts are being made to improve the performance of the graph-CNN, most of these architectures try to improve the convolution operation defined on the graph signal while the graph coarsening process representing the pooling operation largely remains neglected. In order to improve the overall performance of the GCNN architectures, we applied our proposed graph coarsening algorithm in the pooling layer of the GCNN and the resulting classification accuracy was compared with that of the existing GCNN classifiers. The improvement in the classification accuracy of the modified GCNN classifier verified the effectiveness of the proposed algorithm in the context of GCNN.

Along with the applications in various domains such as network analysis, image processing, and computer vision [13], the GCNN is also used in the field of biomedical engineering to detect various diseases and abnormalities [26], [27]. Recently, it has been used for the early detection of Alzheimer's disease (AD) which is one of the most common neurodegenerative diseases affecting around 50 million patients worldwide. Although no proper treatment exists to cure this illness, its early detection is helpful in controlling the disease [28]. Various attempts have been made to detect the AD in the early stage called mild cognitive impairment (MCI) using different modalities like graph signal processing [29], machine learning [30], [31] and diffusion model [5] etc. Recently, GCNN has also been applied for this detection purpose [26], [27]. In our previous work [27], we successfully applied the ChebNet architecture of GCNN to detect the AD in the MCI stage using the resting-state fMRI (rs-fMRI) data of an individual. In this work, we use the modified GCNN architecture to improve our previously proposed AD detection model which is then shown to have the higher overall classification accuracy.

In summary, the salient contributions of our present work are as follows.

- A novel two-stage multilevel graph coarsening algorithm is proposed wherein the coarsened graph at each stage preserves both the topological characteristics and the action of the original graph Laplacian operator simultaneously. A comparison study of the proposed method with the existing coarsening methods demonstrated the superiority of the proposed method.

- The existing GCNN architecture is modified by applying the proposed graph coarsening method to perform the pooling operation and its utility is then illustrated in different graph signal classification applications.
- AD detection model is developed for the early detection of AD in which the modified GCNN architecture is used as a graph signal classifier and achieved the state-of-the-art AD detection performance using the proposed AD detection model.

The rest of the paper is organized as follows. Section II explains some basic concepts about the graph wavelet transform and the GCNN along with the existing graph coarsening algorithms that will be required to appreciate the proposed graph coarsening method. In section III, we describe in detail, the proposed two-stage multilevel graph coarsening algorithm. In section IV, the performance of the proposed graph coarsening method is compared with that of the state-of-the-art methods in different contexts and applications. Classification accuracy of the proposed AD detection model using the modified GCNN architecture is also compared with that of the existing methods in this section. Section V concludes the paper along with some possible extensions and future scope.

II. BASIC PRELIMINARIES AND RELATED WORK

In this section, we review some basic theory and concepts about the graph wavelet transform, graph coarsening and the GCNN which form the basis of the proposed GWT-based graph coarsening method and its application in GCNN.

A. GRAPH WAVELET TRANSFORM

In order to define the notion of wavelets in the graph domain, we first need to define the graph signal and the corresponding graph Fourier transform (GFT). Consider the graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ consisting of N nodes where $\mathcal{V}$ denotes the set of nodes $\{v_1, \ldots, v_N\}$, and $\mathbf{A}$ is $N \times N$ adjacency matrix. Each element $\mathbf{A}_{i,j}$ of this adjacency matrix represents the degree of similarity between the nodes v_i and v_j. A graph signal $\mathbf{s}$ is then defined as a mapping:

$$\begin{aligned} \mathbf{s} : \mathcal{V} &\longrightarrow \mathbb{C} \\ v_n &\longmapsto s_n \end{aligned}$$

The graph Laplacian of $\mathcal{G}$ is an $N \times N$ matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is a degree matrix of $\mathcal{G}$ which is diagonal in nature having the sum of the corresponding rows of $\mathbf{A}$ as its diagonal elements. The Fourier transform of the graph signal has been defined in the literature using different approaches e.g. graph Laplacian based approach [4], Jordan decomposition based approach [3] etc. In the present work, we used the graph Laplacian based GFT and the spectral graph wavelets [22] defined using this GFT that are discussed in brief next.

In classical Fourier transform, a given time domain signal is expressed in terms of the exponentials e^{jwt} that can be viewed as the eigenfunctions of the one-dimensional Laplace operator. In the graph Laplacian based GFT, the same concept is extended to the graph domain by defining the eigenvectors of the graph Laplacian $\mathbf{L}$ as the graph Fourier basis. So, for a given graph signal $\mathbf{s}$, its GFT is defined by

$$\hat{\mathbf{s}} = \mathbf{V}^{\mathbf{T}}\mathbf{s}, \tag{1}$$

where $\mathbf{V} = (\mathbf{v}_1|\mathbf{v}_2 \ldots |\mathbf{v}_N)$ is a $N \times N$ matrix consisting of normalized eigenvectors of $\mathbf{L}$.

In [22], the authors defined the wavelet transform on the graph signal by choosing a band pass filter h in the graph Fourier domain which acts as a mother wavelet in the classical wavelet transform. In this GWT approach, the authors first proved that the classical wavelet transform at some fixed scale t can be expressed as a Fourier multiplier operator where the signal frequency components are multiplied by a filter h, scaled at that particular t. Analogously, the GWT at a particular scale t is defined as

$$\mathbf{T^t s} = \mathbf{V H^t V^T s}, \tag{2}$$

where $\mathbf{H^t} = \mathrm{diag}(h(t\lambda_1), \ldots, h(t\lambda_N))$ and $\{\lambda_i\}_{i=1,\ldots,N}$ are the eigenvalues of $\mathbf{L}$.

Having defined the GWT as above, the wavelet $\boldsymbol{\psi_{t,a}}$ at scale t, centred at some particular node a, can be obtained by operating $\mathbf{T^t}$ on $\boldsymbol{\delta_a}$ i.e. by selecting a particular column of the matrix $\mathbf{T^t}$ [22]. The proposed coarsening method makes use of these graph wavelets for the graph downsampling purpose as explained in the subsequent sections.

B. GRAPH COARSENING

In general, graph coarsening can be thought of as a two step process consisting of (i) graph downsampling and (ii) graph reduction which are explained below.

1) GRAPH DOWNSAMPLING

In this step, for a given graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, the set of nodes $\mathcal{V}$ is partitioned into two complementary sets: a set $\mathcal{V}_1$ which is to be retained and the set $\mathcal{V}_1^c$ which is to be discarded. Ideally, in any downsampling process, the set $\mathcal{V}_1$ should be selected such that the maximum of the original graph structure can be embedded in it. This can be achieved by maximizing some sort of correlation between the sets $\mathcal{V}_1$ and $\mathcal{V}_1^c$. The above problem can be mathematically formulated as

$$\underset{\mathcal{V}_1}{\operatorname{argmax}}\ cut(\mathcal{V}_1, \mathcal{V}_1^c), \tag{P1}$$

where $cut(\mathcal{V}_1, \mathcal{V}_1^c) = \sum_{i \in \mathcal{V}_1} \sum_{j \in \mathcal{V}_1^c} \mathbf{A}_{i,j}$.

But it is well-known that the solution of the max cut problem is NP-complete [10]. Therefore, for most of the practical applications, it is required to find the approximate solution of $P1$. In a particular class of graphs called the bipartite graphs or the two-colourable graphs, for a given graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, it is possible to divide the set $\mathcal{V}$ into two disjoint subsets $\mathcal{V}_1$ and $\mathcal{V}_1^c$ having two different colours such that every nonzero edge connects the nodes that have different colours. So, for a bipartite graph, it is easy to downsample the graph by retaining the set $\mathcal{V}_1$ and discarding $\mathcal{V}_1^c$ or vice versa as it maximizes the cut value.

Many different approaches have been proposed in the literature to solve the graph downsampling problem [9], [10], [32]. In [10], the authors approximated a given graph by its maximum spanning tree using the well established algorithms such as Prim's algorithm. Now, as every tree is a bipartite graph, the downsampling is obtained by retaining the nodes of one colour and discarding the others. In [9], the graph was partitioned according to the polarity of the eigenvector corresponding to the highest eigenvalue of the graph Laplacian. It was shown that for a bipartite graph, the suggested algorithm successfully partitions the given graph into two different coloured sets. The authors also explained the connections of their algorithm with spectral clustering and the nodal domain theory. We propose to downsample the graph by using the eigenvectors of the graph adjacency matrix, which will be explained in detail in the next section.

2) GRAPH REDUCTION

Once the subset of nodes, that are to be retained, is selected by applying a downsampling operator on a given graph, the adjacency matrix and thus the corresponding graph Laplacian needs to be determined in order to obtain the coarsened graph. This process of connecting the nodes in the downsampled node set by defining the edge weights between them is termed as graph reduction [9]. Several graph reduction methods such as Kron reduction [9], weighted adjacency matrix based reduction [33], algebraic multigrid based reduction [14] etc. have been proposed in the literature. Some graph coarsening methods like [7] and [11] do not downsample the original graph explicitly and then define its Laplacian but rather combine these two operations together. For example, in popularly used Graclus method [7], the nodes are combined using heavy edge matching algorithm and the edge weight between them is set as the sum of the individual edge weights. In our paper, we use the Kron reduction method for the graph reduction purpose which is explained in brief next.

Consider the given graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ with the graph Laplacian $\mathbf{L}$ and the downsampled node set $\mathcal{V}_1$ on which the reduced graph Laplacian $\tilde{\mathbf{L}}$ is to be defined. Let $\mathbf{L}_{\mathcal{X},\mathcal{Y}}$ denotes the submatrix of size $|\mathcal{X}| \times |\mathcal{Y}|$ with elements having row index in set $\mathcal{X}$ and column index in set $\mathcal{Y}$. The Kron reduced Laplacian $\tilde{\mathbf{L}}$ is defined as the Schur complement of $\mathbf{L}$ w.r.t. $\mathbf{L}_{\mathcal{V}_1^c,\mathcal{V}_1^c}$ [9] i.e.

$$\tilde{\mathbf{L}} = \mathbf{L}_{\mathcal{V}_1,\mathcal{V}_1} - \mathbf{L}_{\mathcal{V}_1,\mathcal{V}_1^c}\mathbf{L}^{-1}_{\mathcal{V}_1^c,\mathcal{V}_1^c}\mathbf{L}_{\mathcal{V}_1^c,\mathcal{V}_1}. \tag{3}$$

The adjacency matrix of the coarsened graph can then be obtained using the relation

$$\tilde{\mathbf{A}} = \mathbf{I}_N \odot \tilde{\mathbf{L}} - \tilde{\mathbf{L}}. \tag{4}$$

The reduced graph thus obtained satisfies many desired conditions like: (i) the reduced Laplacian matrix $\tilde{\mathbf{L}}$ thus obtained is indeed a Laplacian matrix satisfying the necessary constraints, (ii) $\tilde{\mathbf{L}}$ preserves the graph connectivity i.e. if $\mathbf{L}$ is connected, then $\tilde{\mathbf{L}}$ is also connected, (iii) the reduced graph is loopless if the original one is loopless etc. But while reducing the graph, the Kron reduction scheme does not take into the account the action of the reduced Laplacian operator on the graph signal and its relation with the action of the original Laplacian operator on the same signal. This issue is addressed by us in the proposed graph coarsening method by using the restriction of the original Laplacian operator to a smaller subspace.

C. GRAPH CONVOLUTIONAL NEURAL NETWORK

In the last few years, CNN has provided excellent results in almost every signal processing application, possibly due to its ability to learn the nonlinear features from the given data on its own [13]. But the application of conventional CNN has been somewhat limited to the data living on a regular structure e.g images or audio signals owing to the fact that the algorithms for CNN have been developed primarily for the regularly structured data. Recently, several attempts have been made to generalize the CNN to the irregularly structured data which can be modelled by a graph [12], [13]. Conventional CNN architecture consists of two main operations: convolution operation followed by a pooling operation. So, to generalize the CNN to the graph domain, it is required to extend these two operations for the graph signals. For the graph convolution operation, we used the method proposed in the ChebNet architecture [12] wherein the convolution operation is extended to the graph domain by using the graph filters that have a local receptive field. The graph filter is defined as a polynomial of the graph Laplacian $\mathbf{L}$ i.e.

$$\mathbf{y} = h_\theta(\mathbf{L})\mathbf{x}, \tag{5}$$

where $\mathbf{x}$ is the input graph signal and $\mathbf{y}$ is the output graph signal after the convolution. The coefficient vector θ of the polynomial h_θ are learned during the training of the network where the order of the polynomial is decided by the degree of localization required.

The other main operation in CNN is the pooling operation which requires grouping of adjacent data points together. So, the extension of pooling operation to the graph structured data requires to cluster the similar nodes together and replace them with a single node which can be thought of as a graph coarsening operation. While substantial work has been done to improve the convolution part of the graph-CNN [12], [13], [25], the graph coarsening in GCNN has received a little attention and thus making a scope for the improvement in the performance of the existing GCNN architectures. So, in the present work, we propose a new multilevel graph coarsening scheme which is then applied in the ChebNet architecture [12] to improve its overall performance.

III. PROPOSED GRAPH COARSENING METHOD

In this section, we explain the proposed multilevel graph coarsening method in detail wherein the graph coarsening operation is performed by combining two different coarsening approaches. First, the coarsening is performed using the GWT-based features of the original graph in which the coarsened graph tries to preserve the topological characteristics of the original graph. After this initial coarsening step, at each level of coarsening, we construct the reduced graph Laplacian by the restriction of the original Laplacian operator to a dominant subspace which preserves the action of the original Laplacian operator for the graph signals belonging to that subspace. In this second step, a reduced Laplacian is obtained by solving an optimization problem in which we try to make the reduced Laplacian as close as possible to the one obtained in the first step so that the topological characteristics of the original graph are also preserved. The two stages of the proposed graph coarsening are explained in the following.

A. GWT-BASED GRAPH COARSENING

In this initial coarsening stage, we try to maximize the topological similarity between the original graph and the coarsened graph. Let $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ be a given graph which is to be coarsened to a graph $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathbf{A}})$. As discussed in the last section, graph coarsening process can be divided into two operations: 1. graph downsampling and 2. graph reduction. In the graph downsampling operation, most of the existing algorithms e.g. [7] and [9], select the downsampled node set $\tilde{\mathcal{V}}$ using only the direct connection between the nodes. The node selection criterion of maximizing the cut value, mathematically formulated in $P1$, takes into the account only the direct connections between the nodes and ignores the similarity arising between them by virtue of the indirect connections through the common neighbours and local environment around the nodes. To circumvent this issue, we make use of the graph wavelet based similarity measure between the nodes to perform the downsampling operation exploiting the fact that the wavelet centered at each node encodes the local information and the topological environment in the graph [24].

To downsample the graph using the GWT-based features, we first need to define a graph wavelet centered at each node $a \in \mathcal{V}$. Following the notion of graph wavelets introduced in the last section, let $\boldsymbol{\psi}_{t,a}$ be a graph wavelet at scale t, centered at node a. We construct a *correlation matrix* $\mathbf{C} \in \mathbb{R}^{N\times N}$ such that each element

$$\mathbf{C}_{i,j} := 1 + corr(\boldsymbol{\psi}_{t,i}, \boldsymbol{\psi}_{t,j}) = 1 + \frac{\boldsymbol{\psi}_{t,i}^{T}\boldsymbol{\psi}_{t,j}}{||\boldsymbol{\psi}_{t,i}||_2||\boldsymbol{\psi}_{t,j}||_2}, \quad (6)$$

where i, j denote the index of nodes in $\mathcal{V}$ and the offset of 1 ensures the nonnegativity of matrix elements required by some of the downsampling algorithms. The scaling parameter t is set as the mean value of the minimum and the maximum scale values proposed in [24] as it has worked well in most of the graph coarsening applications.

Once this correlation matrix $\mathbf{C}$ is constructed, we need to downsample the graph by using this matrix so that the downsampling operation takes into the account the local environment of the nodes also rather than restricting only to the direct connections. We now treat this matrix $\mathbf{C}$ as a new weight matrix corresponding to the given graph and hence the modified node selection criterion can be formulated as

$$\underset{\tilde{\mathcal{V}}}{\text{argmax}} \sum_{i\in\tilde{\mathcal{V}}} \sum_{j\in\tilde{\mathcal{V}}^c} \mathbf{C}_{i,j}. \quad (7)$$

As discussed earlier, the above maximization problem is NP-complete and hence an approximation is required to solve it. In the present work, we use the approximate coloring algorithm proposed in [32], in which a given graph is partitioned into two complementary sets such that most of the edges satisfy the coloring condition or in other words, the cut value associated with these sets is high. In this approximate coloring method, the graph is partitioned according to polarity of the eigenvector corresponding to the smallest eigenvalue of the weight matrix. So, the downsmpling of the given graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ is performed based on the polarity of the eigenvector corresponding to the smallest eigenvalue of the associated correlation matrix $\mathbf{C}$. The graph downsampling step has to be followed by a graph reduction step. To reduce the obtained downsampled graph, we apply a Kron reduction method [9] which was explained in the last section.

B. GRAPH COARSENING USING LAPLACIAN OPERATOR RESTRICTION

The coarsened graph obtained using the above GWT-based coarsening approach is able to preserve the topological characteristics of the original graph well as the wavelets capture the topological environment in the graph. But this coarsening method does not consider the action of the reduced Laplacian operator on the coarsened graph signal while constructing the reduced Laplacian. Ideally, we want that the reduced Laplacian operator should act similar to the original Laplacian operator when restricted to a specific subspace.

To formulate the idea mathematically, consider a given graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$ consisting of N nodes which is to be coarsened to a graph $\mathcal{G}_c = (\mathcal{V}_c, \mathbf{A}_c)$ having N_c nodes. Let $\mathbf{L}$ be the graph Laplacian of $\mathcal{G}$ and its eigen decomposition be given by

$$\mathbf{L} = \mathbf{V}\Lambda\mathbf{V}^{\mathbf{T}}, \quad (8)$$

where $\mathbf{V} = (\mathbf{v}_1|\mathbf{v}_2|\dots|\mathbf{v}_N)$ is the matrix consisting of eigenvectors of $\mathbf{L}$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ is the diagonal matrix consisting of eigenvalues of $\mathbf{L}$ where $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_N$. The eigenvalues of $\mathbf{L}$ carry the notion of graph frequency which implies that the lower indexed eigenvectors correspond to the lower graph frequency values and vice versa [4]. Let $\mathbf{W}$ be the N_c dimensional subspace spanned by the first N_c eigenvectors of $\mathbf{L}$ corresponding to the lower graph frequencies i.e.

$$\mathbf{W} = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{N_c}\}. \quad (9)$$

In many of the real life applications such as sensor network analysis and brain imaging networks, the signals are relatively smooth on the graph, i.e., graph Fourier coefficients of the signal decay rapidly with the increasing values of graph frequency [29], [34], [35]. Using equation (1), a graph signal $\mathbf{s}$

$$\begin{tikzcd}
\mathbf{W} \arrow[r, "\mathbf{P}"] \arrow[d, "\mathbf{L}"'] & \mathbb{R}^{N_c} \arrow[d, "\mathbf{L}_c"] \\
\mathbf{W} \arrow[r, "\mathbf{P}"] & \mathbb{R}^{N_c}
\end{tikzcd}$$

FIGURE 1. **Isomorphism preserving action of Laplacian operator.**

can be expressed in terms its graph Fourier coefficients and GFT basis as

$$\mathbf{s} = \sum_{i=1}^{N} \hat{s}_i \mathbf{v}_i, \tag{10}$$

where $\hat{s}_i$ denotes the graph Fourier coefficient corresponding to the i^{th} graph frequency. For a smooth signal $\mathbf{s}$, since $\hat{s}_i$ approaches zero as i increases, the projection of signal $\mathbf{s}$ onto the subspace $\mathbf{W}$, given by $\sum_{i=1}^{N_c} \hat{s}_i \mathbf{v}_i$, closely approximates the original signal $\mathbf{s}$ and hence preserves most of its energy content. $\mathbf{W}$ is an N_c dimensional subspace embedded in $\mathbb{R}^N$ and its each individual element is a length N vector. Therefore, a vector space isomorphism

$$\mathbf{P} : \mathbf{W} \longrightarrow \mathbb{R}^{N_c}$$

can be defined which maps a length N vector in $\mathbf{W}$ to a length N_c vector.

Now, we want to define a coarsened Laplacian operator $\mathbf{L}_c \in \mathbb{R}^{N_c \times N_c}$ such that it acts as a restriction of the original Laplacian to N_c dimensional subspace $\mathbf{W}$. This prevents us from choosing $\mathbf{P}$ to be any arbitrary vector space isomorphism. In fact, we want $\mathbf{P}$ to preserve the action of the Laplacian operator, which implies that the diagram in Fig. 1 should commute.

The above condition on $\mathbf{P}$ can be mathematically formulated as

$$\mathbf{P}(\mathbf{L}\mathbf{v}_i) = \mathbf{L}_c(\mathbf{P}\mathbf{v}_i) \quad \text{for} \quad i = 1, \ldots, N_c. \tag{11}$$

As $\mathbf{v}_i$ is an eigenvector of $\mathbf{L}$ with an eigenvalue λ_i,

$$\mathbf{L}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{12}$$

$$\Rightarrow \quad \mathbf{P}(\mathbf{L}\mathbf{v}_i) = \lambda_i(\mathbf{P}\mathbf{v}_i). \tag{13}$$

Substituting (13) in equation (11), we get

$$\mathbf{L}_c(\mathbf{P}\mathbf{v}_i) = \lambda_i(\mathbf{P}\mathbf{v}_i). \tag{14}$$

The above equation implies that $\mathbf{P}\mathbf{v}_i$ is an eigenvector of $\mathbf{L}_c$ with an eigenvalue λ_i. Hence, $\mathbf{L}_c$ can be expressed as

$$\mathbf{L}_c = (\mathbf{P}\hat{\mathbf{V}}_c)\Lambda_c(\mathbf{P}\hat{\mathbf{V}}_c)^T, \tag{15}$$

where $\Lambda_c = \text{diag}(\lambda_1, \ldots, \lambda_{N_c})$ and $\hat{\mathbf{V}}_c = (\mathbf{v}_1|\mathbf{v}_2|\ldots|\mathbf{v}_{N_c})$. Substituting $\mathbf{P}\hat{\mathbf{V}}_c = \mathbf{V}_c$ in (15) reduces it to

$$\mathbf{L}_c = \mathbf{V}_c \Lambda_c \mathbf{V}_c^T. \tag{16}$$

Equation (16) implies that the eigenvalues of the reduced Laplacian are nothing but the initial eigenvalues of the original Laplacian operator $\mathbf{L}$ which in turn implies that this coarsening also preserves the spectral characteristics of the graph.

But matrix $\mathbf{P}$ and hence the matrix $\mathbf{V}_c$ are still unspecified that need to be calculated in order to construct the reduced graph Laplacian $\mathbf{L}_c$. Elements of the matrices $\mathbf{V}_c$ and $\mathbf{L}_c$ cannot assume the arbitrary values and have to satisfy some conditions which are as follows.

1) $\mathbf{V}_c$ is an orthogonal matrix.
2) $\mathbf{L}_c$ is indeed a Laplacian operator implying that all the row sums of $\mathbf{L}_c$ are zero i.e. $\mathbf{L}_c \mathbf{1}_{N_c \times 1} = \mathbf{0}_{N_c \times 1}$
3) The diagonal elements of $\mathbf{L}_c$ are nonnegative and the nondiagonal elements are nonpositive.

An equation similar to (16) was obtained in [15] where the authors formulated it as a constrained optimization problem which was solved using projected subgradient method. So, in our proposed graph coarsening method, the optimization problem can be formulated as

$$\min_{\mathbf{V}_c,\mathbf{t}} \sum_{p \neq q} f(-l_{p,q}) + \sum_{p=1}^{N_c} f(l_{p,p}) + \mu||\mathbf{L}_c - t\tilde{\mathbf{L}}||_F^2$$
$$\text{subject to } \mathbf{V}_c \in \mathcal{O}_{N_c} \cap \mathcal{S}_{N_c}, \tag{17}$$

where $l_{p,q} = \sum_{k=1}^{N_c} \lambda_k \mathbf{V}_{cp,k} \mathbf{V}_{cq,k}$, $f(.)$ is a nonnegative-inducing penalty function penalizing the negative diagonal elements and the positive nondiagonal elements, $\tilde{\mathbf{L}}$ is the reduced graph Laplacian obtained from the first stage of coarsening, $\mathcal{O}_{N_c}$ denotes the set of all orthogonal matrices in $\mathbb{R}^{N_c \times N_c}$ and $\mathcal{S}_{N_c}$ denotes the set of all matrices in $\mathbb{R}^{N_c \times N_c}$ having row sums equal to zero. The above optimization problem is solved using the projected subgradient method as described in [15] to evaluate the coarsened graph Laplacian $\mathbf{L}_c$.

1) MULTILEVEL COARSENING

The single level coarsening method proposed above produces a coarsened Laplacian which preserves the action of original Laplacian operator on the graph signal while the associated graph still preserves the topological structure of the original graph and thus addresses the first two issues involved in the existing graph coarsening schemes. But as discussed earlier, multiple pooling layers in GCNN demands for a multilevel graph coarsening scheme. The multilevel extension of our proposed two-stage coarsening algorithm is presented below as Algorithm 1.

In the proposed multilevel algorithm, at each level of coarsening, the Laplacian of the coarsened graph aims to preserve the action of the original graph Laplacian rather than preserving that of the graph obtained in the previous level. So, with this multilevel extension, our proposed graph coarsening scheme addresses all the three issues associated with the state-of-the-art graph coarsening methods.

To verify the effectiveness of the proposed coarsening method, its performance is first quantified using different coarsening quality measures and then the same is applied in the graph coarsening layer of the ChebNet architecture [12], the effect of which is analysed in the next section.

Algorithm 1 Multilevel Graph Coarsening Algorithm

```
Input: Original graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$,
       Number of coarsening levels $l$.
Set $k \leftarrow 0$, $\mathbf{L}_c^k \leftarrow \mathbf{L}$.
while $k < l$ do
    $k \leftarrow k + 1$.
    Coarsen $\mathbf{L}_c^{k-1}$ using GWT-based first stage of coarsening. Let $\tilde{\mathbf{L}}^k$ be the resulting initial coarsened Laplacian of size $N^k$.
    Construct diagonal matrix $\Lambda_c^k = \mathrm{diag}(\lambda_1, \ldots, \lambda_{N^k})$.
    Calculate a final $k$-level coarsened Laplacian $\mathbf{L}_c^k = \mathbf{V}_c^k \Lambda_c^k (\mathbf{V}_c^k)^T$ by solving an optimization problem:
        $\min_{\mathbf{V}_c^k, t} \sum_{p \neq q} f(-l_{p,q}) + \sum_{p=1}^{N^k} f(l_{p,p}) + \mu ||\mathbf{L}_c^k - t\tilde{\mathbf{L}}^k||_F^2$
        subject to $\mathbf{V}_c^k \in \mathcal{O}_{N^k} \cap \mathcal{S}_{N^k}$.
return $\mathbf{L}_c^k$
```

IV. EXPERIMENTS AND RESULTS

In this section, we first compare our proposed graph coarsening algorithm with state-of-the-art algorithms using three suitable measures. The proposed coarsening method is then applied in the GCNN architecture. This modified GCNN architecture is then used for the classification purpose in three standard datasets and for the early detection of AD using rs-fMRI data. The classification performance obtained using the modified GCNN architecture is compared with that of different graph signal classifiers in these applications to analyse the impact of the proposed coarsening method on the GCNN architecture.

In the graph coarsening process, the first step is graph downsampling. As discussed earlier, the downsampling problem can be mathematically formulated as a maximization problem $P1$ which is NP-complete and hence the approximate solutions must be used. The quality of the approximation can be quantified using the quantity called *cut index* [10] which is defined as a ratio of sum of the edge weights connecting two resulting subsets to the sum of all the edge weights in the graph and can be mathematically expressed as:

$$\textit{cut index} = \frac{\textit{cut}(\mathcal{V}_1, \mathcal{V}_1^c)}{\sum\limits_{i,j \in \mathcal{V}^2} \mathbf{A}_{i,j}}. \tag{18}$$

Obviously, a higher value of cut index implies a better graph downsampling method which in turn implies a better graph coarsening method. So, we compared the cut indices obtained using different existing downsampling methods [9], [10], [36] and used the one leading to the highest cut index in the proposed coarsening scheme. For the illustration purpose, we operated these downsampling methods on the following four different types of graphs [17].

1) Erdös-Rényi random graphs with $N = 500$ nodes and probabilty parameter $p = 0.02$.
2) Airfoil graph consisting of $N = 4000$ nodes.

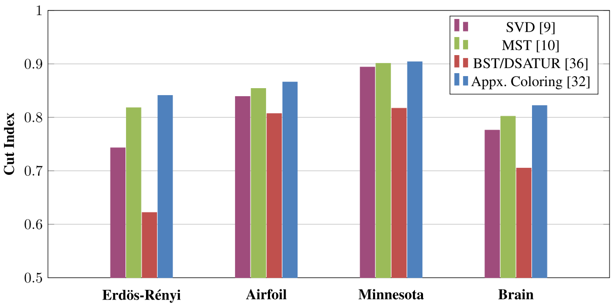


FIGURE 2. Cut index values for different graphs and downsampling methods.

3) Minnesota graph consisting of $N = 2642$ nodes.
4) Brain functional connectivity graphs of subjects from ADNI dataset [37] having $N = 325$ nodes obtained using MIST [38] brain parcellation.

Fig. 2 compares the resulting values of cut indices obtained by different downsampling methods on these graphs wherein the results for the Erdös-Rényi random graphs are averaged over 100 graphs and for the functional connectivity graphs, the results are averaged over the graphs of 100 subjects.

From the plot shown in Fig. 2, it can be concluded that the downsampling based on the eigenvector corresponding to the smallest eigenvalue is best suited for our graph coarsening purpose and hence was used in the proposed coarsening method.

To quantify how closely the reduced Laplacian operator approximates the action of the original Laplacian operator, we propose a new measure called *operator dissimilarity index*. Mathematically, we define the operator dissimilarity index as:

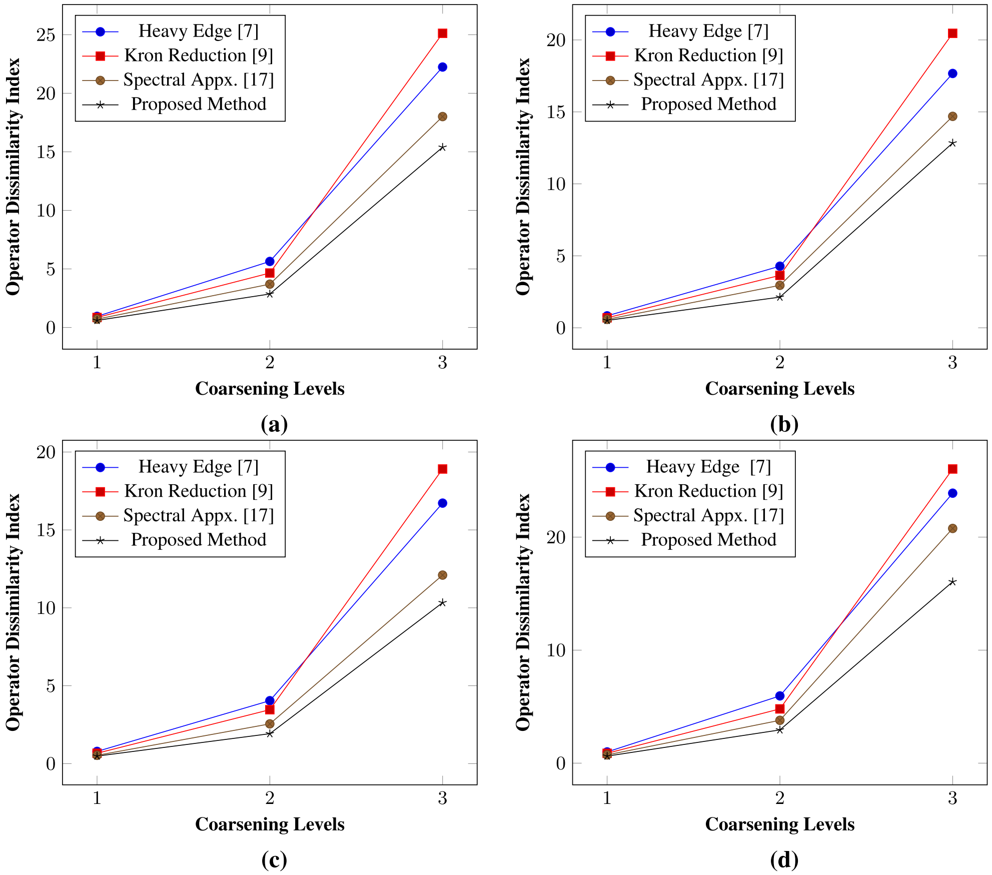
$$\mathbb{E}_{\mathbf{X}}[||(\mathbf{L}\mathbf{x})_c - \mathbf{L}_c \mathbf{x}_c||_2], \tag{19}$$

where $\mathbf{L}$ is the original normalized graph Laplacian, $\mathbf{L}_c$ denotes the coarsened graph Laplacian, $\mathbf{x}$ is the original graph signal having a unit norm and $(.)_c$ denotes the coarsened graph signal. Following the probabilistic graph signal model proposed in [29], the probability distribution of the graph signal is assumed to be given by

$$\mathbb{P}(\mathbf{x}) \propto \exp(-\mathbf{x}^T \mathbf{L}\mathbf{x}), \tag{20}$$

which is a reasonable choice as most of the real life graph signals are predominantly low pass signals and hence the smoother signals occur with the higher probability.

As it can be seen from equation (19), the operator dissimilarity index measures the expected value of the difference between the output signals when the same graph signal is operated by the original Laplacian operator and the coarsened Laplacian operator. So, the lower the value of operator dissimilarity index, the better is the coarsening algorithm. In Fig. 3, we compare the operator dissimilarity index of our proposed coarsening method with that of the state-of-the-art methods using the same graphs which were used to compare the cut index.



(a)

(b)

(c)

(d)

FIGURE 3. Operator dissimilarity index obtained using different coarsening methods on: (a) Erdös-Rényi graph (b) Airfoil graph (c) Minnesota graph (d) Brain connectivity graph.

To evaluate the coarsening performance of our proposed method more comprehensively, we compare the performances of different coarsening methods using another coarsening quality measure called *restricted spectral approximation* [17]. It essentially quantifies how well the coarsened Laplacian operator approximates the action of the original Laplacian operator w.r.t. the subspace $\mathbf{U}_k$ spanned by k lowest eigenvectors of the original Laplacian. Mathematically, it is computed as the smallest value of ϵ for which the following equation holds.

$$||\mathbf{x} - \tilde{\mathbf{x}}||_{\mathbf{L}} \leq \epsilon \; ||\mathbf{x}||_{\mathbf{L}} \quad \text{for all} \quad \mathbf{x} \in \mathbf{U}_k, \tag{21}$$

where $||\mathbf{x}||_{\mathbf{L}} = \sqrt{\mathbf{x}^T \mathbf{L} \mathbf{x}}$, $\mathbf{x}$ is the original graph signal and $\tilde{\mathbf{x}}$ is the lifted approximation of $\mathbf{x}$. In Fig. 4, we plot the values of ϵ for different graphs and graph coarsening methods, where a smaller value of ϵ indicates a better coarsening method.

Results in Fig. 3 and Fig. 4 show that the reduced graph Laplacian obtained using the proposed coarsening method is better able to preserve the action of original Laplacian operator as compared to the existing coarsening methods which in turn marks the superiority of the proposed algorithm.

After validating the performance of the proposed graph coarsening algorithm using different graph coarsening quality measures, we apply it in the coarsening layer of the ChebNet architecture [12] of GCNN. We analyse the performance of this modified GCNN model equipped with the proposed coarsening method in three different applications and compare it with that of the GCNN models constructed using the existing coarsening methods. The illustrative applications considered in this paper are:

1. MNIST handwritten digit classification problem [12].
2. Document classification in Cora and Citeseer datasets [25].
3. Early detection of AD using resting-state fMRI data from ADNI dataset.

In the first application, the objective is to label a given handwritten digit image as one of the digit between 0 to 9. To solve this problem using GCNN classifier, we use the approach presented in [12]. The 8 nearest neighbour graph having $28 \times 28 = 784$ nodes is constructed corresponding to each image from the dataset where the edge weight is calculated as:

$$\mathbf{W}_{i,j} = \exp\left(-\frac{||\mathbf{z}_i - \mathbf{z}_j||_2^2}{\sigma^2}\right), \tag{22}$$

where $\mathbf{z}_i$ denotes the 2-D coordinate of pixel i.

The graphs thus obtained from the training dataset are then used to train the modified GCNN classifier which classifies a given handwritten digit into one of the 10 classes. The GCNN architecture consists of two graph convolution layers of 32 and 64 feature maps respectively, each of which is followed

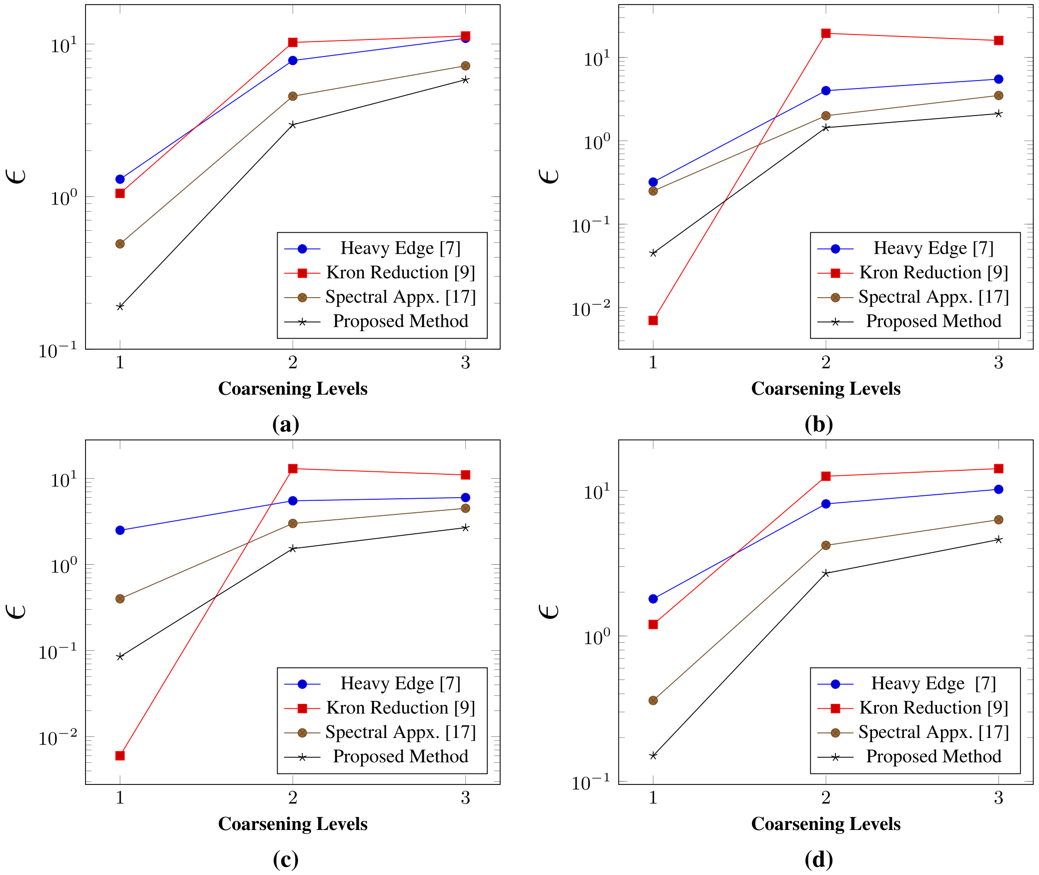


FIGURE 4. Values of ϵ obtained using different coarsening methods on: (a) Erdös-Rényi graph (b) Airfoil graph (c) Minnesota graph (d) Brain connectivity graph. The results are plotted for a representative subspace of size $k = 40$, as the same k-value is used in [17].

TABLE 1. Classification accuracies of GCNN architectures with different graph coarsening methods on MNIST dataset.

Method	Classification Accuracy (%)
Heavy Edge [7]	99.14
Kron Reduction [9]	99.21
Spectral Appx. [17]	98.65
Proposed Method	**99.30**

TABLE 2. Classification accuracies of GCNN architectures with different graph coarsening methods on Cora and Citeseer datasets.

Method	Cora	Citeseer
Heavy Edge [7]	80.68	69.55
Kron Reduction [9]	81.24	70.43
Spectral Appx. [17]	78.90	68.26
Proposed Method	**83.85**	**72.76**

by a graph coarsening layer. The output of these layers is given to a fully connected layer and a softmax regression layer to obtain a class label.

The authors of [12] used Graclus method [7] for the graph coarsening purpose. To compare our proposed graph coarsening algorithm with state-of-the-art algorithms [7], [9], [17], we constructed GCNN classifiers using the same architecture and parameter values but with different graph coarsening algorithms and compared their classification accuracies. Results in table 1 show that our proposed multilevel graph coarsening algorithm outperforms the existing coarsening algorithms when applied in the GCNN architecture.

To validate the performance of the proposed coarsening method in case of graphs constructed from unstructured data as well, we apply our model for the classification of documents in Cora and Citeseer datasets. To solve this document classification problem using GCNN classifier, we again adapt the approach presented in [12], where a document is first represented using the bag-of-words model. The vocabulary words are then represented by the corresponding vectors using the word2vec embedding and the 16 nearest neighbour graph is constructed using these vectors where the edge weight between the nodes is calculated using equation (22), with $\mathbf{z}_i$ denoting the vector associated with word i. Number of nodes of the graph obtained using the above approach depends on the size of the vocabulary of the dataset, thus it generates 1433-node graph for Cora dataset and 3703-node graph for Citeseer dataset.

Having obtained the graphs corresponding to documents in both the datasets, we train the GCNN model described in the MNIST classification application for both of these datasets separately and apply those for the document classification purpose. Table 2 compares the classification accuracy percentages of GCNN architectures constructed using

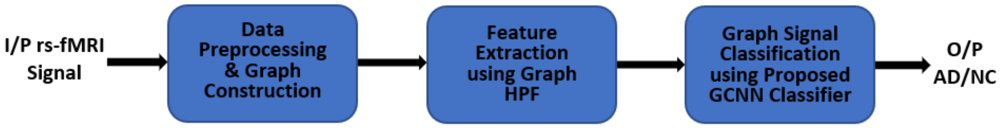

I/P rs-fMRI Signal
Data Preprocessing & Graph Construction
Feature Extraction using Graph HPF
Graph Signal Classification using Proposed GCNN Classifier
O/P AD/NC

FIGURE 5. Block diagram of AD detection model using the proposed GCNN classifier.

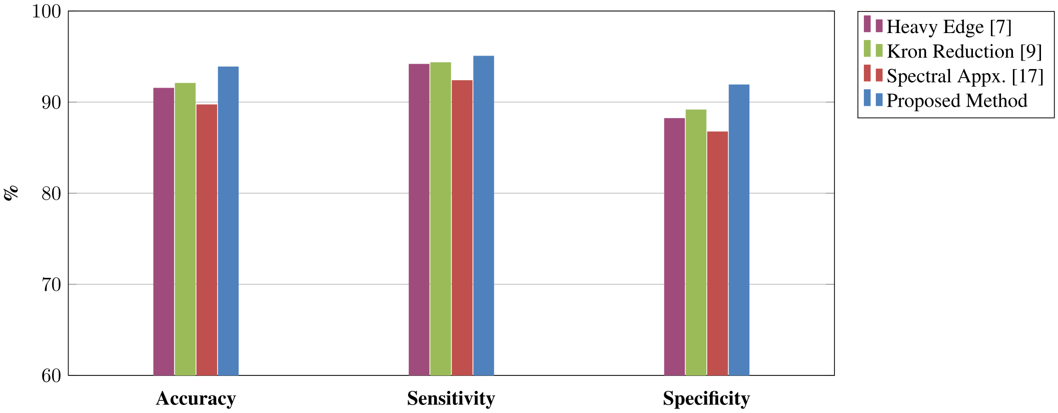


FIGURE 6. Classification performance of GCNN architectures with different graph coarsening methods in AD Detection.

different graph coarsening schemes for Cora and Citeseer datasets.

The classification accuracy values in table 1 and table 2 corroborates the effectiveness of the proposed graph coarsening algorithm as a coarsening method in the GCNN architecture.

After verifying the efficacy of the proposed coarsening scheme in the general classification context, we next apply the modified GCNN architecture in our previously proposed GCNN-based AD detection model [27] which is explained in brief below.

2) AD DETECTION MODEL

Let $\mathbf{X} \in \mathbb{R}^{N \times T}$ be rs-fMRI signal of a subject from ADNI dataset [37] with a known label i.e. normal control (NC) or having MCI where T denotes the number of timepoints in a signal and N denotes the number of brain regions into which the whole brain is parcellated. This fMRI signal $\mathbf{X}$ can be considered as a concatenation of T, N-dimensional graph signals represented by the columns of matrix $\mathbf{X}$. We label each of these individual graph signals with the label of rs-fMRI signal $\mathbf{X}$. As described in detail in our previous work [27], the graph high pass filtered signal acts a discriminating feature for AD detection application. Hence the individual N-dimensional graph signals are first passed through graph HPF that are then provided to a graph signal classifier for the classification purpose. We use the proposed modified GCNN architecture as a graph signal classifier for the early detection of AD. The block diagram of the proposed AD detection model is shown in Fig. 5.

Having discussed the general structure of the proposed AD detection model, we next analyse its performance using the rs-fMRI data of 160 subjects from ADNI dataset. The data consists of rs-fMRI signals of 80 NC subjects and 80 subjects with MCI which is parcellated into 325 regions using MIST parcellation scheme [38]. The graph signal lying on 325-node graph thus obtained is passed throgh graph HPF, which is then used as an input to the proposed GCNN classifier.

The GCNN architecture consists of two graph convolution layers having 32 and 64 feature maps respectively and output of each convolution layer is coarsened by our proposed graph coarsening method. These layers are followed by a fully connected layer having 256 hidden units and a softmax regression layer. The above GCNN model is trained using Adam optimiser for 30 epochs with an initial learning rate of 0.001. 5-fold stratified cross-validation is used to obtain the classification performance of the proposed method which is quantified using three measures viz. accuracy, sensitivity and specificity. To verify the effectiveness of our modified GCNN architecture in AD detection application, we repeated the above classification analysis using GCNN architectures in which the pooling operation is performed using different existing graph coarsening methods. Fig. 6 compares the classification performance of GCNN architectures equipped with different graph coarsening schemes.

To evaluate the performance of the proposed GCNN based AD detection model more comprehensively, in Fig. 7, we compare its classification performance with that of state-of-the-art AD detection methods. The results show an improvement in all the three classification measures, thereby attesting the superiority of the proposed modified GCNN classifier over the existing approaches for early detection of AD.

The improvement in the overall performance can be attributed to the fact that, in the proposed algorithm, the nodes were combined using the correlation between graph wavelets centered at each node which accounted for the similarity

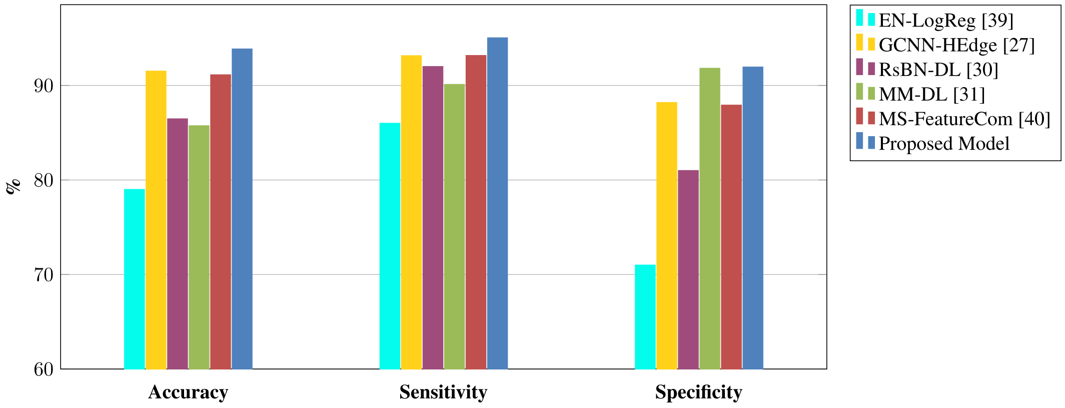


FIGURE 7. Classification performance of different AD detection methods.

between local neighbourhood and environment of the nodes unlike the most of the existing algorithms that consider only the direct connection between them. Apart from that, the constrained optimization problem formulation for the multilevel coarsening ensured that along with topological similarity, the coarsened Laplacian obtained at each coarsening level also preserves the operator characteristics of the original Laplacian which may have reduced the overall information loss in the coarsening process.

V. CONCLUSION AND FUTURE SCOPE

In this paper, a novel two-stage graph coarsening algorithm was presented wherein both the topological characteristics and the action of Laplacian operator of the original graph are simultaneously preserved in the coarsened graph. In the first stage of coarsening, the GWT-based features were used to coarsen graph followed by an optimization based second stage of coarsening in which at each level of coarsening, the restriction of the original graph Laplacian was used to obtain the reduced graph Laplacian. The performance of the proposed algorithm was quantified using graph coarsening quality measures like cut index and operator dissimilarity index. Its effectiveness as a pooling operator in the GCNN was validated by applying it as graph coarsening operator in the ChebNet architecture. The proposed coarsening method outperformed state-of-the-art coarsening methods in terms of coarsening quality measures as well as in the classification performance of corresponding GCNN classifier in different applications including early detection of AD.

In the present work, the proposed coarsening scheme was applied in the ChebNet architecture of the GCNN and its performance was analysed. We can extend its application in the other GCNN architectures [13], [25] and analyse their performance to generalize the proposed coarsening scheme in the context of GCNN. The value of the scaling parameter t in the correlation matrix $\mathbf{C}$ was selected empirically. Although, it worked reasonably well, this parameter selection may not be the optimal and further study about the optimal scaling parameter selection may improve the overall performance.

REFERENCES

[1] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1189–1203, Oct. 2016.

[2] I. Jablonski, "Graph signal processing in applications to sensor networks, smart grids, and smart cities," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7659–7666, Dec. 2017.

[3] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[4] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[5] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[6] D. Gfeller and P. De Los Rios, "Spectral coarse graining of complex networks," *Phys. Rev. Lett.*, vol. 99, no. 3, Jul. 2007, Art. no. 038701.

[7] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1944–1957, Nov. 2007.

[8] C. Walshaw, "A multilevel algorithm for force-directed graph-drawing," *J. Graph Algorithms Appl.*, vol. 7, no. 3, pp. 253–285, 2006.

[9] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "A multiscale pyramid transform for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2119–2134, Apr. 2016.

[10] H. Q. Nguyen and M. N. Do, "Downsampling of signals on graphs via maximum spanning trees," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 182–191, Jan. 2015.

[11] O. E. Livne and A. Brandt, "Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver," *SIAM J. Scientific Comput.*, vol. 34, no. 4, pp. B499–B522, Jan. 2012.

[12] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.

[13] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.

[14] D. Ron, I. Safro, and A. Brandt, "Relaxation-based coarsening and multiscale graph organization," *Multiscale Model. Simul.*, vol. 9, no. 1, pp. 407–423, Jan. 2011.

[15] P. Liu, X. Wang, and Y. Gu, "Coarsening graph signal with spectral invariance," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 1070–1074.

[16] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 359–392, Jan. 1998.

[17] A. Loukas, "Graph reduction with spectral and cut guarantees," *J. Mach. Learn. Res.*, vol. 20, no. 116, pp. 1–42, 2019.

[18] Y. Jin, A. Loukas, and J. F. JaJa, "Graph coarsening with preserved spectral properties," 2018, *arXiv:1802.04447*. [Online]. Available: http://arxiv.org/abs/1802.04447

[19] J. Wang, "Laplacian regularized low-rank representation for cancer samples clustering," *Comput. Biol. Chem.*, vol. 78, pp. 504–509, Feb. 2019.

[20] O. Teke and P. P. Vaidyanathan, "Extending classical multirate signal processing theory to Graphs—Part I: Fundamentals," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 409–422, Jan. 2017.

[21] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, Jul. 2006.

[22] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[23] F. Saberi-Movahed, M. Eftekhari, and M. Mohtashami, "Supervised feature selection by constituting a basis for the original space of features and matrix factorization," *Int. J. Mach. Learn. Cybern.*, vol. 2019, pp. 1–17, Dec. 2019.

[24] N. Tremblay and P. Borgnat, "Graph wavelets for multiscale community mining," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5227–5239, Oct. 2014.

[25] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, "CayleyNets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 97–109, Jan. 2019.

[26] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease," *Med. Image Anal.*, vol. 48, pp. 117–130, Aug. 2018.

[27] H. Padole, S. D. Joshi, and T. K. Gandhi, "Early detection of Alzheimer's disease using graph signal processing on neuroimaging data," in *Proc. 2nd Eur. Conf. Electr. Eng. Comput. Sci. (EECS)*, Dec. 2018, pp. 302–306.

[28] Alzheimer's Association, "2017 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 13, no. 4, pp. 325–373, 2017.

[29] C. Hu, J. Sepulcre, K. A. Johnson, G. E. Fakhri, Y. M. Lu, and Q. Li, "Matched signal detection on graphs: Theory and application to brain imaging data classification," *NeuroImage*, vol. 125, pp. 587–600, Jan. 2016.

[30] R. Ju, "Early diagnosis of alzheimer's disease based on resting-state brain networks and deep learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 1, pp. 244–257, Jan./Feb. 2019.

[31] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang, "Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 361, pp. 185–195, Oct. 2019.

[32] B. Aspvall and J. R. Gilbert, "Graph coloring using eigenvalue decomposition," *SIAM J. Algebr. Discrete Methods*, vol. 5, no. 4, pp. 526–538, Dec. 1984.

[33] S. K. Narang and A. Ortega, "Local two-channel critically sampled filterbanks on graphs," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 333–336.

[34] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[35] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec. 2016.

[36] S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4673–4685, Oct. 2013.

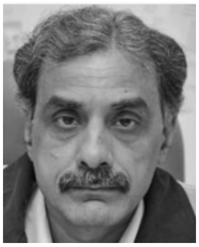
[37] *Alzheimer's Disease Neuroimaging Initiative: ADNI*. Accessed: Apr. 12, 2019. [Online]. Available: http://adni.loni.usc.edu

[38] S. Urchs, J. Armoza, C. Moreau, Y. Benhajali, J. St-Aubin, P. Orban, and P. Bellec, "MIST: A multi-resolution parcellation of functional brain networks," *MNI Open Res.*, vol. 1, p. 3, Mar. 2019.

[39] F. de Vos, "A comprehensive analysis of resting state FMRI measures to classify individual patients with alzheimer's disease," *Neuroimage*, vol. 167, pp. 62–72, Feb. 2018.

[40] Z. Jiao, Z. Xia, X. Ming, C. Cheng, and S.-H. Wang, "Multi-scale feature combination of brain functional network for eMCI classification," *IEEE Access*, vol. 7, pp. 74263–74273, 2019.

HIMANSHU PADOLE (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the Visvesvaraya National Institute of Technology (VNIT), Nagpur, India, in 2014, and the M.Tech. degree in communication engineering from IIT Delhi, New Delhi, India, in 2016, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include graph signal processing, biomedical signal processing, and machine learning. He was a recipient of government of India's prestigious PMRF Fellowship Awarded to excellent doctoral candidates for pursuing their research work.



SHIV DUTT JOSHI received the B.E. degree (Hons.) in electrical and electronics engineering from the Birla Institute of Technology, Pilani, India, in 1981, and the M.Tech. degree in communications and radar engineering and the Ph.D. degree in signal processing from IIT Delhi, New Delhi, India, in 1983 and 1988, respectively. He is currently a Professor with the Electrical Engineering Department, IIT Delhi. His research interests include the development of fast algorithms for stochastic signal processing, speech processing, modeling of stochastic processes, and group theoretical approach to signal processing.

TAPAN K. GANDHI received the B.Sc. degree in physics, the M.Sc. degree in electronics, the M.Tech. degree in bioelectronics, and the Ph.D. degree in biomedical engineering from IIT Delhi, in 2001, 2003, 2006, and 2011, respectively. He is currently an Associate Professor with the Department of Electrical Engineering, IIT Delhi. Prior to joining IIT Delhi, as a Faculty Member, he was a Postdoctoral Fellow with the Massachusetts Institute of Technology (MIT), Cambridge, USA, for three years. His research interests include cognitive computation, artificial intelligence, medical instrumentation, biomedical signal and image processing, and assistive technology.