Int J Syst Assur Eng Manag (March 2024) 15(3):931–949
https://doi.org/10.1007/s13198-023-02180-z

ORIGINAL ARTICLE

Agglomeration of deep learning networks for classifying binary and multiclass classifications using 3D MRI images for early diagnosis of Alzheimer's disease: a feature-node approach

Rashmi Kumari[1] · Subhranil Das[2] · Raghwendra Kishore Singh[3]

Received: 2 June 2023 / Revised: 21 July 2023 / Accepted: 20 September 2023 / Published online: 10 October 2023

Abstract Alzheimer's disease is a degenerative brain condition causing memory loss in the elderly. Existing machine learning methods often yield low classification accuracy due to evaluating single modality features. This paper presents a novel approach that combines Graph Attention Networks and Deep Convolutional Graph Neural Networks to leverage 3D 1.5 T and 3 T T1-weighted MRI images as nodes, enabling faster feature extraction. Three Graph Convolutional Network layers are introduced to improve the classification accuracy for three binary classifications (AD vs. CN, MCI vs. CN, and MCI vs. AD) and multiclass classification (AD vs. CN vs. MCI). The model is optimized for weight updates using the Adaptive Stochastic Gradient Descent technique. Comparative analysis with efficient 3DNET, Squeeze3D-NET, and GoogLENET demonstrates superior performance of the proposed DCGNN network. Furthermore, evaluations against four state-of-the-art techniques for binary and multiclass classifications show its potential in diagnosing the early stages of Alzheimer's disease. The developed model exhibits promise as an effective tool for diagnosing Alzheimer's disease at its early stages.

Keywords Alzheimer's disease · Deep neural networks · Graph attention · Magnetic resonance imaging · Mild cognitive impairment

✉ Rashmi Kumari
Rashmi.kumari@bennett.edu.in

[1] School of Computer Science Engineering and Technology, Bennett University, Greater Noida, Uttar Pradesh 201310, India

[2] Department of Computer Science Engineering, Parul University, Gujrat 391760, India

[3] Department of Electronics and Communication Engineering, National Institute of Technology, Jamshedpur 831014, India

1 Introduction

Most researchers concur that the brain is one of the body's most crucial organs. All actions and reactions that enable us to believe and think are regulated by and maintained by the brain. It also assists in maintaining audio memories and feelings (Menéndez 2017). Alzheimer's disease is a progressive and degenerative brain disorder that is the most common cause of dementia, which is a decline in cognitive function that affects memory, thinking, and behavior. The disease was first identified by Dr. Alois Alzheimer in 1906 and is named after him (Goedert and Ghetti 2007). Alzheimer's disease is a complex and multifactorial condition characterized by the accumulation of amyloid plaques and neurofibrillary tangles in the brain, disrupting the communication between neurons and leading to their death. This results in gradually losing cognitive abilities, such as memory, language, decision-making, and personality changes.

Research on Alzheimer's disease (AD) is driven by the urgent need to address the profound impact this neurodegenerative condition has on individuals, families, and society. AD is a progressive and irreversible brain disorder that primarily affects memory, cognition, and daily functioning, leading to significant cognitive decline and loss of independence in affected individuals. As the global population ages, the prevalence of AD is increasing, making it a pressing public health concern. The primary motivations for researching AD include seeking a deeper understanding of its underlying causes, risk factors, and disease mechanisms. This knowledge is crucial for developing effective preventive measures, diagnostic tools, and therapeutic interventions. Research also focuses on identifying early biomarkers to enable early detection, improving disease management, and ultimately finding a cure. By unraveling the complexities of AD, researchers aim to alleviate its burden on affected

individuals and caregivers, enhance the quality of life for patients, and contribute to the broader mission of advancing neuroscience and geriatric medicine. Aging, skull fractures, and lifestyle habits are just a few of the warning factors that might lead to AD. Although age is the main risk factor for dementia, studies show that a person beyond 65 has a 1–2% chance of developing the condition. This risk can escalate by 30% by age 85 (Cui et al. 2019; Demirhan 2016). By 2050, AD may impact 1 in every 85 persons, according to recent research studied by Alzheimer's Disease Research (Farooq et al. 2017).

Early detection and treatment of AD patients are essential to have good living conditions in the future. There are many ways to diagnose and forecast the disease by considering the different modalities such as MRI, Position Emission Tomography (PET), and Computed Tomography (CT) scans. However, MRI is the neuroimaging modality most frequently used to diagnose AD patients (Demirhan 2016; Farooq et al. 2017). Previous studies have combined Machine Learning (ML) and Deep Learning (DL) algorithms with MRI-based classification approaches (Fritsch et al. 2019). Some of the studies are explained in the subsequent paragraphs.

In a few years, it has been discovered that Machine Learning (ML) approaches are beneficial for diagnosing Alzheimer's (Chen et al. 2022; Cui et al. 2023a, 2023b). Kloppel et al. (2008) claimed that SVM can aid in diagnosing AD by using it to classify pathologically proven instances of AD with CN. According to references (Javeed et al. 2023, 2023; Zhao et al.), linear kernel outperforms polynomial or RBF kernel for high dimensional data classification performance. Nonetheless, scientists also make use of polynomial kernels. The polynomial kernel was utilized by Lahmiri et al. (2014) for the multiclass categorization of CN, MCI, and AD. A polynomial kernel helps classify CN vs. AD using PCA features, according to Zhang et al. (2015). In 2018, Lahmiri et al. (2019) classified CN vs. AD using a polynomial kernel, volumetric characteristics, and cognitive test results. Moreover, several researchers combined many kernels. Alam et al. (2017) categorized CN, MCI, and AD using multiple kernel SVM. Kamathe et al. (2018) used linear, polynomial, and RBF kernels to categorize CN vs. AD. Zhu et al. (2016) proposed the application of a temporally structured SVM (TS-SVM) for classifying longitudinal MR images of MCI converters and non-converters, allowing for early identification of Alzheimer's. A random forest robust SVM (RF-RSVM) was proposed by Lu et al. (2017) for the classification of CN vs. MCI using FDG-PET images. Using dual-tree complex wavelet transform (DTCWT), LDA, and PCA features, TWSVM is employed to classify CN vs. AD (Sharma et al. 2022). From the methods mentioned above, ML methods require extra computational time in processing the features, resulting in more significant resource costs. Therefore, Deep Learning (DL) techniques are more valuable by incorporating the feature extraction step in the model, resulting in higher classification accuracy. Some of the recent Deep Learning techniques have been explained in the next paragraph.

A classification method for AD developed by Cui et al. (2023) uses convolutional and recurrent neural networks (RNNs). First, CNN was created to categorize MR pictures by learning spatial components. The RNN and cascaded Bidirectional Gated Recurrent Unit (BGRU) layers were created using CNN outputs at various time intervals to extract longitudinal characteristics for AD classification. The approach suggested jointly learning spatial and longitudinal qualities to improve performance rather than isolating elements individually. The recommended method would also use an RNN to create a longitudinal study from imaging data gathered over time. According to the results, the recommended method had a greater accuracy of about 91.33% when comparing AD to NC and 71.71% when contrasting Progressive Mild Cognitive Impairment (pMCI) to Stable Mild Cognitive Impairment (sMCI). Another innovative Voxel-based Hierarchical Feature Extraction (VHFE) technique was put forth by Yue et al. (2019) to detect AD early. With the help of a template, it divides the entire brain into 90 separate ROIs (AAL). Informative voxels were used for each ROI using a baseline of values and organized into a single vector to separate informative from uninformative data. Convolutional Neural Networks were fed the brain feature maps for each subject utilizing voxels, with the first stage of these features being selected (CNN). Finally, the authors examined the approach on a portion of an ADNI database to confirm its viability. The testing outcomes demonstrated the method's robustness in comparison to other methods.

The proposed methodology addresses critical technical gaps observed in existing works for Alzheimer's disease (AD) classification. Unlike isolated spatial or longitudinal analysis in traditional methods, the proposed approach integrates both aspects using convolutional and recurrent neural networks (CNNs and BGRUs). This joint analysis allows for a more comprehensive understanding of AD progression, simultaneously capturing spatial and temporal patterns. Leveraging longitudinal data helps identify dynamic changes over time, enhancing the model's ability to distinguish between Mild Cognitive Impairment (MCI) and AD stages. Furthermore, the innovative Voxel-based Hierarchical Feature Extraction (VHFE) technique overcomes limitations in traditional voxel-based approaches, improving the discriminative power of brain imaging data. The proposed methodology exhibits greater accuracy, robustness, and potential for early-stage AD detection, promising more effective and reliable diagnostic tools for AD and related cognitive disorders.

After considering the advantages of all the models mentioned above, we have proposed a new technique where the features are stored as nodes that act as input to GAT

networks in which the single layer attention networks have been considered. Then, the optimized nodes generated from GAT networks are given as inputs to the proposed Deep Convolutional Graph Neural Networks (DCGNN) for classifying AD, MCI, and CN subjects. The main contributions of the paper are:

1. For early detection of Alzheimer's Disease, 3D T1 weighted MRI images of Magnetic Field Strength 1.5T, and 3T have been considered for our study from three different ADNI datasets.
2. For extracting different features, such as in the hippocampus area, left cerebellum cortex, and structural brain volume, FreeSurfer 6.0.1 software installed in the High Performance Computing (HPC) has been utilized in 3D MRI images.
3. Graph Attention Networks (GAT) have been proposed to acquire information about the extracted features as nodes. Then, Single Graph Attention Layer (SGAT) is constructed for getting optimized nodes which are given as inputs to the proposed Deep Convolutional Graph Neural Networks (DCGNN) for three binary classifications and one multi class. Also, Adaptive Stochastic Gradient Descent (ASGD) optimization technique is applied for updating the weights of three Graph Convolutional Network (GCN) layers in the proposed networks.
4. To compare the effectiveness of the proposed DCGNN network with three other existing networks, 857 participants from three separate ADNI databases have been used. Additionally, the proposed methodology has been compared to four cutting-edge techniques for each of the three binary classifications and multiclass classification, and simulation results indicate that it generates the best classification outcomes.

The organization of the paper has been as follows: The detailed description of the dataset and the specifications of MRI protocol as well as FreeSurfer software, has been mentioned in Sect. 2. Sect. 3 depicts the proposed network's working methodology and the optimization technique for updating the weights. Simulation findings have been elaborated on in Sect. 4. Finally, the conclusion and future of the research directions have been explained in Sect. 5.

2 Materials and methods

2.1 Description of the dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (2010) is a large-scale, multicenter study that aims to identify and track the progression of Alzheimer's disease (AD) through various clinical, imaging, genetic, and cognitive assessments. The dataset consists of data from three cohorts: ADNI-1, ADNI-2, and ADNI-3, which were collected between 2004 and 2016. The ADNI dataset contains a wide range of data types, including clinical assessments, cognitive tests, magnetic resonance imaging (MRI) scans, positron emission tomography (PET) scans, and genetic data (Petersen et al. 2010).

Real-time data from the baseline 3D T1-weighted MRI images of magnetic strength 3 T from the three ADNI-1, ADNI-2, and ADNI-3 datasets have been acquired. Here, 273, 326, and 258 subjects have been considered for the study, comprising AD, MCI, and CN classes. Moreover, the subjects have been selected based on four criteria, such as Functional Activities Questionnaire (FAQ), Mini-Mental State Examination (MMSE), Global Clinical Dementia Rating (Global CDR), and Geriatric Depression Scale (GDSCALE). The magnetic field strength of 3 T has been taken into consideration for 3D T1 MRI images consisting of 1726 scans in the ADNI-1 dataset, 1342 scans for 1.5 T magnetic field strength in ADNI-2, and 1023 scans of 3 T magnetic field strength for ADNI-3 dataset. All MR images are normalized for N3 and B1 nonuniformity before further processing. The Neuroimaging Informatics Technology Initiative (NIfTI) format has been employed for saving and storing MR images. The FMRIB Software Library (FSL) toolset registers MR images (McKhann et al. 2011). Table 1 provides a detailed description of the three ADNI subjects and highlights the subjects' demographic characteristics.

2.2 Procedure of the study

In this study, there are two phases of operation. In its 1st phase, real-time data is extracted through Freesurfer Software for all classes, such as AD, CN, and MCI, where the different optimized features are removed on the 3D T1-weighted MRI images based on age, sex, MMSE, FAQ, MMSE, Global CDR, and GD Scale. In the 2nd phase, these optimized features are stored as nodes of the proposed Graph Attention Network (GAT), where the Single Graph Attention Layer (SGAL) has been added to reduce the redundant features. The output of the layer is the optimized nodes generated from GAT networks, which act as inputs to the proposed Deep Convolutional Graph Neural Networks (DCGNN), where the patients are classified with AD, CN, and MCI. The flow diagram of the total procedure of the study is shown in Fig. 1.

2.3 MRI protocol

The specifications of 1.5 T and 3 T T1-weighted MRI protocol are as follows: SAGITTAL Acquisition Plane, 3D Acquisition Type, 8HRBRAIN Coil, 1.5 Tesla Magnetic Strength, 8° Flip Angle, GE MEDICAL SYSTEMS Manufacturer, 256

Table 1 Summary of demographic characteristics for ADNI-1, ADNI-2, and ADNI-3 subjects

Dataset	Demographic characteristics	AD	MCI	CN
ADNI-1	Total number of subjects (Male/Female)	62/42	48/45	41/35
	Total number of MRI Images	647	467	714
	Age (Minimum/Maximum)	68.6/92.7	56.8/91.1	71.4/88.9
	FAQ (Minimum/Maximum)	(1/27)	(2/25)	(2/7.5)
	MMSE (Minimum/Maximum)	(11/27)	(17/34)	(22/29)
	Global CDR (Minimum/Maximum)	(0.2/3)	(1/4)	(−0.7/1.8
	GD Scale (Minimum/Maximum)	(1/7.3)	(1/10)	(0/3)
ADNI-2	Total number of subjects (Male/Female)	57/36	68/33	71/51
	Total number of MRI Images	522	411	409
	Age (Minimum/Maximum)	67.7/87.2	59.1/90.8	68.6/90.4
	FAQ (Minimum/Maximum)	(2/25)	(2.5/24.5)	(2.5/10.5)
	MMSE (Minimum/Maximum)	(102/28.3)	(14.3/32.5)	(20.8/29.8)
	Global CDR (Minimum/Maximum)	(0.4/2.8)	(1.2/4.7)	(0.8/2.7)
	GD Scale (Minimum/Maximum)	(1.2/6.8)	(1/9)	(0/2.7)
ADNI-3	Total number of subjects (Male/Female)	52/41	60/32	43/30
	Total number of MRI Images	412	303	308
	Age (Minimum/Maximum)	70.7/90.3	55.7/89.2	69.6/93.5
	FAQ (Minimum/Maximum)	(2/25.5)	(1.3/23.8)	(2.5/6.5)
	MMSE (Minimum/Maximum)	(9/26.5)	(15.5/32.8)	(20.7/28.4)
	Global CDR (Minimum/Maximum)	(0/2.8)	(1.3/4.7)	(0.3/1.9)
	GD Scale (Minimum/Maximum)	(1.8/6.8)	(1.6/9.6)	(0/2.7)

pixels of Matrix X and Matrix Y, 166 Pixels of Matrix Z, SIGNA HDx Manufacturer Model, 0.9 mm of Pixel Spacing X and Y, RM Pulse Sequence, 1.2 mm of Slice thickness, 3.8 TE ms, 1000 TI ms, 8.6 TR ms (Kumari et al. 2022).

2.4 Specifications of high performance computing (HPC)

The configurations of HPC are Master node (2) and Computer Node (1) of E5-2630 v3 Intel Xeon 2.4 GHz processors with 8-core, Hard Disk Capacity of 500 GB; Compute Node (2) of Nvidia K20 GPU and 64 GB memory; 1 Cloud node of E5-2620 V3@2.4, 6 core processors E5-2620 and 64 GB Memory, and Hard Disk Drive of 1 TB (Kumari et al. 2022).

2.5 Freesurfer

FreeSurfer (2012) is a software package for analysing and visualizing structural and functional brain imaging data. It has been widely used in analysing the early stages of Alzheimer's Disease as it is developed as an open-source and cross-platform program. Also, this software has been primarily used in analysing data by including structural T1-weighted images and functional imaging data.

The correction of intracranial volume produces better outcomes in FreeSurfer. This software provides all the various measurements for brain volume and white matter segmentations on the equivalent MRI scanner. The benefit of utilizing software is that when low intensities of white matter are computed where any MS abnormalities unalter the output. Also, this software provides a complete pipeline of B1 bias field correction, skull stripping, and localization of grey and white matter for sMRI data acquisition. The gray-white and pial surfaces can also be considered when reconstructing cortical surface models. Steps for evaluating the processing of global volume in FreeSurfer software:

1. *Creation of directory file* Build a directory consisting of two files, of which the first one contains a compilation of slices from each structure MRI scan. The statistics of the second structural scan statistics are specified in the second file.
2. *Conversion of the data* The following command is utilized by FreeSurfer to convert the raw data into.mgz format. "Recon-all-i < in volume > -s < subject.name > "
3. *Simultaneous acquisitions* The same command can be translated into a different directory if several acquisitions exist for a single subject. Additionally, a single representation is created by combining and averaging these acquisitions.
4. *Motion averaging and correction* Using the first scan analysis, several acquisitions by each subject are first registered. Using the command "recon-all-s subject name > -motion.cor", where the output from each subject is generated from a motion-corrected volume.

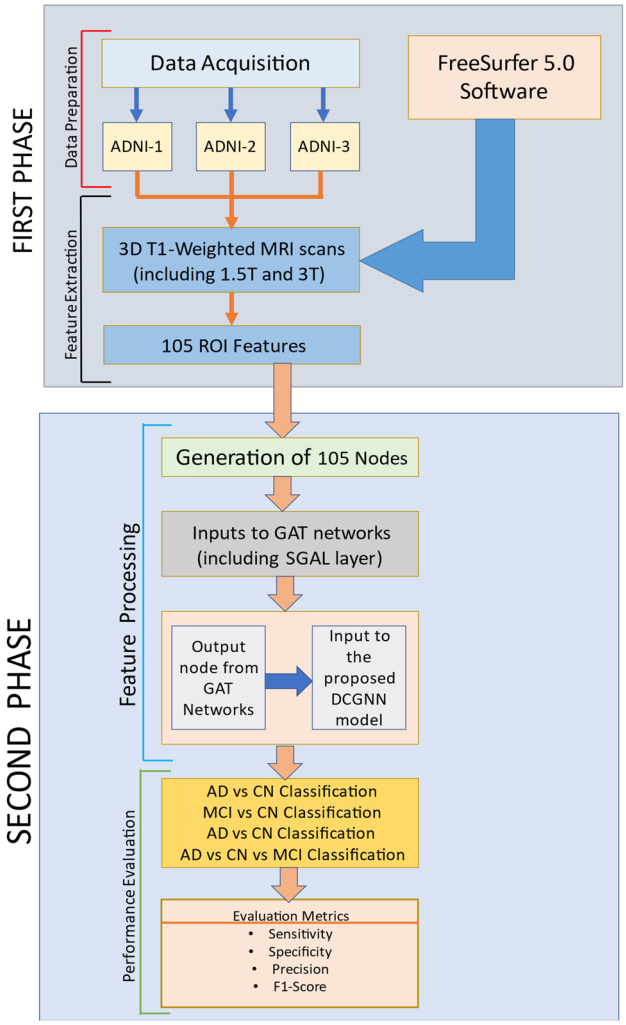

Fig. 1 Flow diagram of the total procedure of the study

5. *Quantization and Adjustment of Intensities* The original orig.mgz file has been converted into a series of steps of histogram equalizations of a series of Talarich space.
6. *Skull stripping* A watershed technique that strips away background noise from the skull section and produces a BrainMask volume that utilizes the T1-weighted intensity corrected data as input**.**
7. *Segmentation of Cortical and Subcortical regions* Following the generation of the BrainMask volume, subcortical characteristics are processed, which could take up to 16 h. An aseg.mgz file is created during the preprocessing stage, where the statistics file is produced.
8. *Segmentation of White Matter* In the final processing stage, the input volume has been standardized and then segmented to produce the white matter volume (wm.seg.mgz). The labels provided from the ASEG file are then employed for filling the ventricles. After that, mri fill distinguishes the two hemispheres, creating a filled binary mask labeled as mgz.

The values of age, estimated intracranial volume (eTIV), and Magnetic Field Strength are measured to develop Linear Regression (LR) models to compute the cortical regions of the whole brain region. Here, 12-fold cross-validation (CV) has been employed to acquire the normalized white and grey matter values by using the DKT atlas to prevent overfitting issues.

3 Proposed graph attention network (GAT)

Graph Attention Networks (GATs) offer several advantages over graph-based and non-graph-based models. Here are some of the key benefits of GAT, such as adaptive feature aggregation, can efficiently handle large graphs due to their localized attention mechanism, flexibility in handling both homogeneous and heterogeneous graphs, robustness to graph irregularity, provide interpretable attention weights, and adaptation to graph generation tasks. Classifying the graph node has become essential for learning the new feature representation for each node over multiple layers. In the proposed network, we have introduced the single graph attention layer for constructing the GAT by aggregating the information by a central node from its neighbors.

Step 1: (Single Graph Attention Layer) It is the primary layer while formulating the different complex architecture of GAT. The reference (Bahdanau et al. 1409) shows that the attention mechanism has been closely related to the designed framework. In this layer, the input layer is defined as $\mathbf{H} = [\overrightarrow{F_1}, \overrightarrow{F_2}, \overrightarrow{F_3}, \ldots, \overrightarrow{F_n}]$, $\overrightarrow{H_K} \in R^f$, where n denotes the number of nodes and f signifies features present in each node. The output of the layer generates a different set of features (f') present in the node, $\mathbf{H}' = [\overrightarrow{F_1'}, \overrightarrow{F_2'}, \overrightarrow{F_3'}, \ldots, \overrightarrow{F_n'}]$.

A Linear Transformation of at least one learnable parameter is required to obtain expressive power to transform the input feature vector into a high-level feature vector. A weight Matrix is defined as $W_i \in R^{f' \times f}$ has been applied to each node in the input layer. The self-attention mechanism has been used in the next step for more accurate results for particular nodes. This method is known as the local shared attention mechanism, which is represented as a linear transformation weight matrix for calculating attenuation coefficients as given in Eq. 1.

$$\gamma_{Ki} = A\left(W\overrightarrow{H_k}, W\overrightarrow{H_i}\right) \text{where A} : R^{f' \times f} \tag{1}$$

Step 2 (Determination of the value of the nodes) The attenuation coefficients determine the level of importance given at features present in node i to node k. Next, the graph

structure is inserted into the local attention mechanism by dropping the insignificance values, or the values are normalized of attenuation coefficients where the optimal value γ_{Ki} Is computed for the nodes i Ɛ P_j. The j indicates the 8 neighbourhood values of node i in the graph, as shown in Fig. 2.

$$n_i = \frac{\sum_{p=1}^{8} n_{ip}}{8} \tag{2}$$

The values obtained after computing the γ_{Ki}, the first order attention coefficients of the nodes are obtained. Therefore, the first order nodes are normalized across the Regions of Interest (P_j) by applying the softmax function as shown in Eq. 3. The softmax function is a crucial component for obtaining meaningful class probabilities, enabling model interpretation, training, and performance evaluation in multiclass classification tasks.

$$\gamma_{Ki} = \text{softmax}(\gamma_{Ki}) = \frac{\exp(\gamma_{Ki})}{\sum_{l \in P_j} \exp(\gamma_{Ki})} \tag{3}$$

Step 3: (Normalization of Attenuation coefficients) In the experiments conducted above, the local shared attention mechanism $\vec{A}$ consists of single layer feedforward neural network by assigning the different values of the weight matrix W_i Ɛ $R^{2\times f}$. Since the variations in the pixel value are less when we consider ROI at different locations in the image, the feedforward neural network produces satisfactory results. In the last step, for each ROI (P_j), the average computed node $\overrightarrow{F'_{1avg}}$ is calculated by applying the Leaky RELU function. The attenuation coefficients obtained by the first order nodes are further computed using a nonlinearity function, i.e., Leaky RELU with a negative slope of 0.3, expressed as in Eq. 4.

$$\gamma_{Ki}^{N} = \frac{\exp(LRELU(\overrightarrow{A^T}[W\overrightarrow{H_K}||W\overrightarrow{H_i}]))}{\sum \exp(LRELU(\overrightarrow{A^T}[W\overrightarrow{H_K}||W\overrightarrow{H_i}]))} \tag{4}$$

where || represents the concatenation operation and $.^T$ indicates transposition. The normalized γ_{Ki}^{N} attention coefficients are given in Fig. 3. The threshold value for obtaining attenuation coefficients is 0.8.

Step 4: (Multi-Headed Mechanism) According to Vaswani et al. (2017), a multi-headed mechanism has been applied for the learning process of self attenuation with K = 4 heads which is given in Fig. 4.

The features present in each node are concentrated, resulting in the output features mathematically given in the Eqn.

$$\overrightarrow{F'_1} = \prod_{K=1}^{4} \sigma\left(\sum_{j=2}^{8} \gamma_{kj}^{K} W^K \overrightarrow{F_j}\right) \tag{5}$$

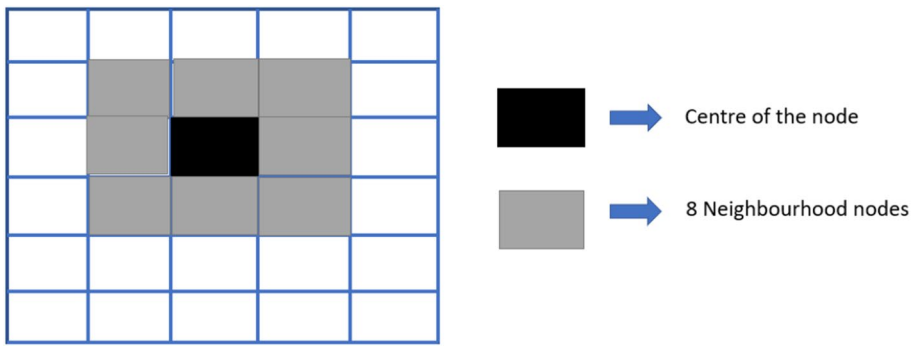

Fig. 2 Representation of the nodes in the GAT network

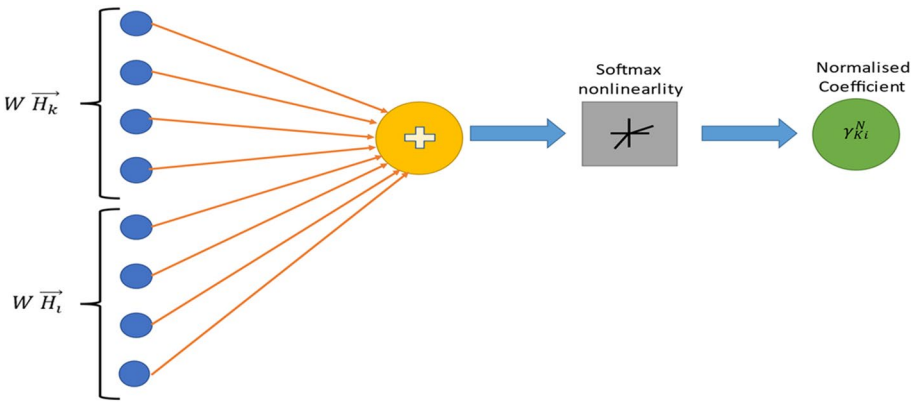

Fig. 3 Normalised coefficients are obtained through GAT networks

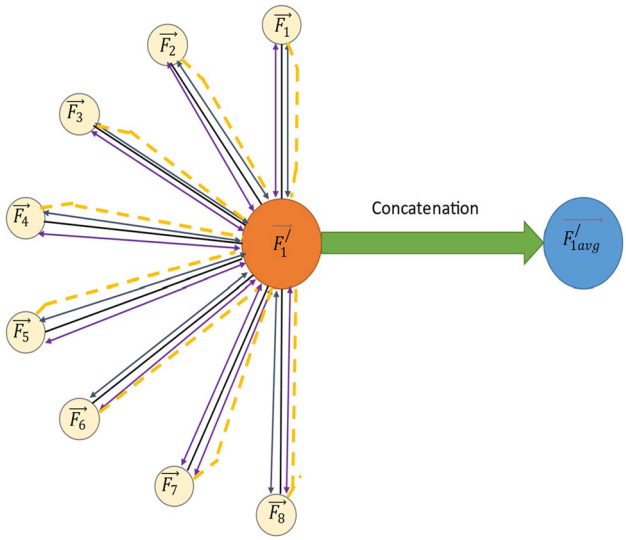

Fig. 4 Multi-headed mechanism of self attenuation with K = 4 heads

where γ_{kj}^{K} are the normalized attenuation coefficients, W^{K} represents the weight matrix and $\prod$ denotes the concatenation. After calculation of $\overrightarrow{F_1^{/}}$ node, other nodes such as $\overrightarrow{F_2^{/}}$, $\overrightarrow{F_3^{/}}$, ….., $\overrightarrow{F_n^{/}}$ are computed similarly.

3.1 Structure of the proposed deep convolutional graph neural networks (DCGNN)

In this section, the network of DCGNN consists of 3 GCN layers, a Leaky RELU activation function, and a drop out of 0.5, described in Fig. 5.

3.1.1 1st GNN layer

Since the outputs nodes generated from GAT networks, i.e., $\overrightarrow{F_1^{/}}$, $\overrightarrow{F_2^{/}}$, $\overrightarrow{F_3^{/}}$, ….., $\overrightarrow{F_n^{/}}$ has been divided into blocks of two, three, and four up to n/2 to process the complexity generated by the nodes. Five convolutional layers, three pooling layers, and four dropout layers of rate 0.5 have been applied in this layer. In addition, three dropout layers of rate 0.5 have been added in intermediate steps for extracting the features from the surrounding Regions of Interest (ROI). A 3 × 3 kernel size has been kept fixed for the 1st and 2nd convolutional layer, which has been applied to 32 kernels, whereas for other convolutional layers, a 3 × 3 kernel size has been kept standard, which is used for each of the 64 kernels. In addition, max-pooling layers are incorporated after the 2nd convolutional layer to enhance filtered node processing. The model architecture consists of 64 nodes in the 1st convolutional layer, 32 in the 2nd layer, and 16 in the 3rd layer, as illustrated in Table 2.

3.1.2 2nd GCN layer

In this layer, the outputs generated from the previous layer have been divided into three groups as input to the 1st convolutional layer. For extracting minute features present in each node, a kernel size of 5 × 5 has been applied to each of the 48 kernels. After that, the kernel size of 3 × 3 has kept fixed for the 2nd layer consisting of 32 kernels following a max pooling layer. For the 3rd layer, 2 × 2 kernel size has been applied to 20 kernels. A dropout of 0.5 is utilized between the layers so that extra irreverent features can be eliminated for faster node processing. Here, 32 nodes are used in the 1st convolutional layer, then 16 nodes in the 2nd layer, then 8 nodes in the 3rd layer. Lastly, a fully connected layer has been added with 700 nodes with one node as output, shown in Table 3.

3.1.3 3rd GCN layer

Single node are generated from each of the blocks of the previous layer and are fed as input to 1st convolutional layer, where 64 nodes are present. Since the image complexities are less, only two convolutional layers are sufficient for processing each feature in the nodes. Here, 2 × 2 kernel size has been utilized for 20 and 10 kernels, respectively. Lastly, a fully convolutional layer has been added in the end with 500 nodes and 3 nodes as output with a Leaky RELU activation function shown in Table 4.

3.2 Adaptive stochastic gradient descent (ASGD) optimization technique

The Adaptive Stochastic Gradient Descent (ASGD) optimization technique offers several advantages in training machine learning models. ASGD dynamically adapts the learning rates based on the historical gradients, resulting in faster convergence and improved robustness to noisy gradients. It enhances the model's generalization capabilities by preventing overfitting and efficiently handles sparse data, leading to more efficient updates and reduced sensitivity to hyperparameters. ASGD's adaptability to distributed computing environments makes it well-suited for parallelization, enabling faster training of large-scale models. Moreover, ASGD is a widely-used and well-studied optimization approach in the literature, providing researchers with a reliable and established choice for training a wide range of machine learning algorithms. For updating the weights of the proposed Deep Convolutional Graph Neural Networks,

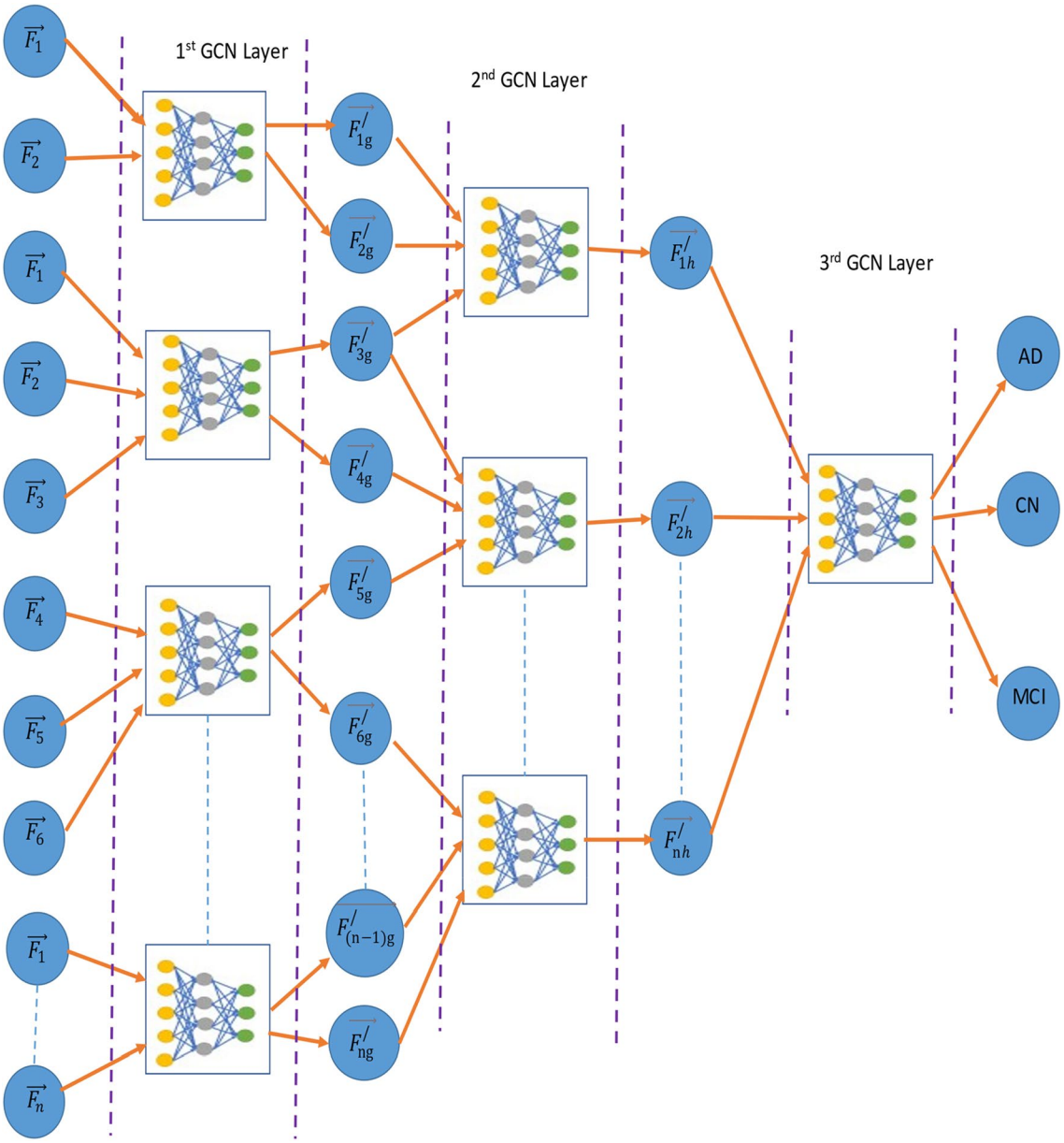

Fig. 5 Overall structure for the proposed DCGNN Network

the optimization technique named Adaptive Stochastic Gradient Descent has been applied to generate two optimized nodes at the end of blocks created by 1st GCN Layer and one optimized node by the end of 2nd GCN layer. Here, the SGD optimization technique has been applied to 1st and 2nd GCN layers, shown in Fig. 6.

From Fig. 6a, it has been deduced that the $\overrightarrow{F_n^{/}}$ nodes are formed from GAT networks in n-dimensional form and given to the adaptive algorithm as inputs. From each of the blocks, two nodes are created, which are combined into one node, i.e., $\overrightarrow{F^{(p)}_{i\frac{n/}{2}g}}$ as the predicted node. The value of this node is compared with the desired value, $\overrightarrow{F^{(d)}_{i\frac{n/}{2}g}}$ which gives the error rate that is used for updating the weights of the nodes, i.e., $\overrightarrow{W}_{(1)} = [\overrightarrow{w_1}, \overrightarrow{w_2}, \ldots\ldots\ldots\ldots, \overrightarrow{w_g}]$. In Fig. 6b, the generated nodes formed the first layer are given in the form of $\overrightarrow{F_h} = [\overrightarrow{F^{/}_{1g}}, \overrightarrow{F^{/}_{2g}}, \ldots\ldots\ldots, \overrightarrow{F^{/}_{\frac{n}{2}g}}]$ that are given to the inputs to the adaptive algorithm. Here, the predicted output node of the 2nd layer is shown as $\overrightarrow{F^{/(h)}_{ih}}$ whereas the

Table 2 Structure of 1st GCN layer along with its parameters

Layer_name	Shape of the output	Parameters
Convolutional layer_1	(32, 32, 36)	936
Convulational_layer_2	(28, 28, 32)	28,832
Maxpooling_1	(14, 14, 32)	0
Dropout_1	(14, 14, 32)	0
Convolutional layer_3	(12, 12, 64)	5824
Convolutional layer_4	(10, 10, 64)	11,584
Maxpooling_2	(5, 5, 64)	0
Dropout_2	(5, 5.64)	0
Convolutional layer_5	(2, 2, 64)	7744
Maxpooling_3	(1, 1, 64)	0
Dropout_3	(1, 1, 64)	0
Flatten_1	(64)	0
Dense_1	(256)	16,640
Dropout_4	(256)	0
Dense_2	(2)	1024

Table 3 Structure of 2nd GCN layer along with its parameters

Layer_name	Shape of the output	Parameters
Convolutional layer_6	(28, 28, 32)	1232
Convulational_layer_7	(24, 24, 32)	9248
Maxpooling_4	(12, 12, 32)	0
Dropout_5	(12, 12, 32)	0
Convolutional layer_8	(10, 10, 64)	5184
Maxpooling_5	(5, 5, 64)	0
Flatten_2	(64)	0
Dense_3	(700)	44,800
Dropout_5	(700)	0
Dense_4	(1)	708

Table 4 Structure of 3rd GCN layer along with its parameters

Layer_name	Shape of the output	Parameters
Convolutional layer_9	(14, 14, 30)	110
Maxpooling_4	(7, 7, 30)	0
Dropout_5	(7, 7, 30)	0
Convolutional layer_8	(4, 4, 15)	605
Maxpooling_5	(2, 2, 15)	0
Flatten_2	(15)	0
Dense_3	(500)	45,015
Dropout_5	(500)	0
Dense_4	(3)	1503

desired output node as $\overrightarrow{F_{ih}^{/(d)}}$ gives an error difference between desired and predicted nodes. The SGD algorithm uses the error difference to update the values of the weights of the nodes, which are given as $\overrightarrow{W}_{(2)} = [\overrightarrow{w_1}, \overrightarrow{w_2}, \ldots\ldots\ldots\ldots, \overrightarrow{w_h}]$.

4 Simulation results and discussion

Different K-features are optimized from the T1-weighted 3D MRI images, which are stored in each node from axial, sagittal, and coronal planes. These features act as inputs to the nodes of the proposed Deep Convolutional Graph Neural Networks (DCGNN) where three binary classifications such as AD vs MCI, MCI vs CN, and AD vs MCI, and one multiclass classification have been proposed. Here, 715 subjects have been considered for our study, which has been acquired from the ADNI-1, ADNI-2, and ADNI-3 datasets. Real-time experiments have been carried out on High Performance (HPC) computing with the specifications of 3 Nos. of Nvidia K20 GPU processor with 1.5 TB memory disk, 12 Intel CORE i10 processors, and 16 core of E5-2630 Intel Xeon 3.7 GHz processors. Here, the Adaptive Stochastic Gradient Descent (ASGD) optimization technique has been applied at the 1st and 2nd GCN layers with different learning rates. The batch size has kept at an optimum value so that differentiation of gradients becomes constant after executing several epochs. For all experiments, the learning rate has been varied so that overfitting does not occur. The dataset has been randomly distributed so that 70% of total images has been simulated for training, 10% of total images for validation and 20% for testing purposes. A confusion matrix has been created by calculating the values of Sensitivity, Specificity, and Classification Accuracy from the testing samples for three binary classifications, i.e., AD vs CN, MCI vs CN, and AD vs MCI, and multiclass classification, AD vs CN vs MCI as shown in Fig. 7.

4.1 Performance of the evaluation metrics

Diagonal elements of the Confusion Matrix (CM) represent the different measures where the classifier correctly predicts to measure the proposed algorithm's performance. The elements are further separated into two groups, consisting of one group containing the elements with the True Positive (TP) and True Negative (TN) and another group consisting of False Positive (FP) and False Negative (FN) as non-diagonal elements. There are other parameters for measuring the efficiency of the proposed algorithm, such as Classification Accuracy (CA), Sensitivity (SEN), Specificity (SP), Precision (PR), and F1-Score, which is given by mathematical expressions as shown below.

$$\mathrm{CA} = \frac{TP + TN}{FP + FN + TP + TN} \tag{6}$$

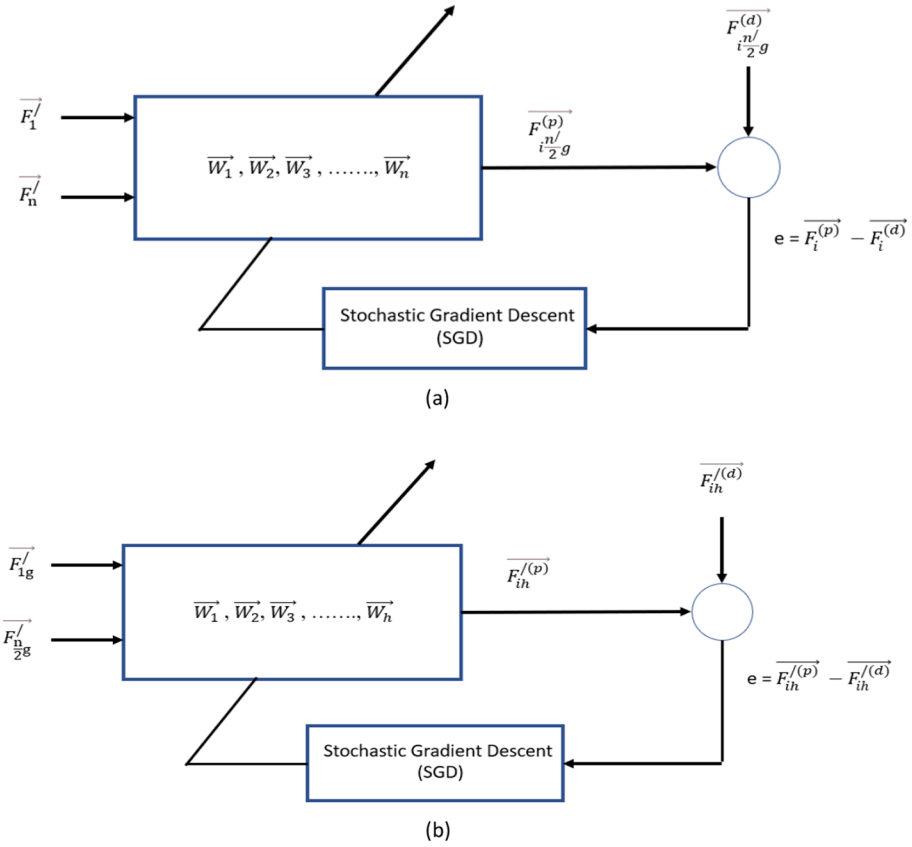

Fig. 6 ASSS **a**, **b** Updation of weights in the GCN layer through the proposed DCGNN network

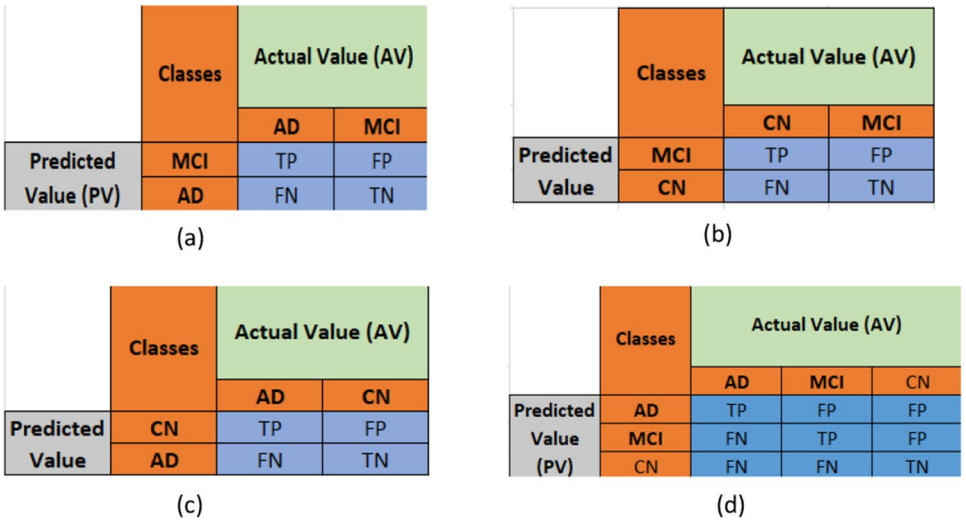

Fig. 7 Confusion matrix for **a** AD vs CN **b** MCI vs CN **c** AD vs MCI **d** AD vs MCI vs CN classifications

$$\text{SP} = \frac{TN}{TN + FN} \quad (7)$$

$$\text{SEN} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{PR} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{F1} - \text{Score} = \frac{2TP}{2TP + FP + FN} \quad (10)$$

Different optimal features from the 3D T1-weighted MRI images are extracted from Freesurfer software, which act as separate nodes in inputs to the proposed DCGNN networks. Tensor Flow libraries are installed on the PC where the specifications of Windows11 Operating system, 1 TB installed harddisk memory, IntelCore 9th Generation with a clock frequency of 2.60 GHz processor by including 6-core and RAM memory of 16 GB RAM where real-time experiments are executed on it. The summary of three binary classifications and one multi class classification report are shown in Table 5.

Table 5 Classification report for three binary and multiclass classification

Dataset	Parameters	AD vs CN	CN vs MCI	MCI vs AD	CN vs MCI vs AD
ADNI-1	CA	89.24	80.68	69.14	94.35
	SP	69.47	73.37	82.23	86.67
	SENS	80.55	75.28	81.14	83.34
	PR	79.94	69.81	77.65	75.77
	F1-SCORE	73.34	75.05	78.28	82.26
ADNI-2	CA	91.18	82.23	72.28	95.68
	SP	72.26	77.51	80.77	88.37
	SENS	83.33	73.86	80.35	81.28
	PR	81.77	72.23	75.55	74.73
	F1-SCORE	70.24	74.11	80.18	83.76
ADNI-3	CA	93.34	80.67	75.1	94.37
	SP	70.65	73.38	78.66	90.34
	SENS	81.17	71.14	75.39	77.61
	PR	80.23	68.82	78.36	70.39
	F1-SCORE	69.95	77.68	81.17	82.21

From Table 5, it is deduced that the classification accuracy percentage of the ADNI-1 dataset is the lowest as compared to ADNI-2 and ADNI-3 datasets. In the ADNI-1 dataset, there are non-uniformities such as T1 correction and irregularities in the T1W1 MRI images' pixels, resulting in more significant complexities while processing. As a result, the other parameters, such as Specificity, Sensitivity, Precision, and F1Score values, are affected. In the ADNI-2 dataset, less uniformities are present due to pre-processing of the images, resulting in higher classification accuracy and other parameters of all three binary and multiclass classifications. However, the highest classification accuracy has been achieved due to less distortion of the images acquired from the ADNI-3 dataset. Also, the T1-weighted images have good clarity in the pixels, resulting in smooth image processing.

In evaluating the efficiency of the proposed DCGNN model, Region of Operating (ROC) curves are drawn for each of the binary classifications, i.e., AD vs CN, AD vs

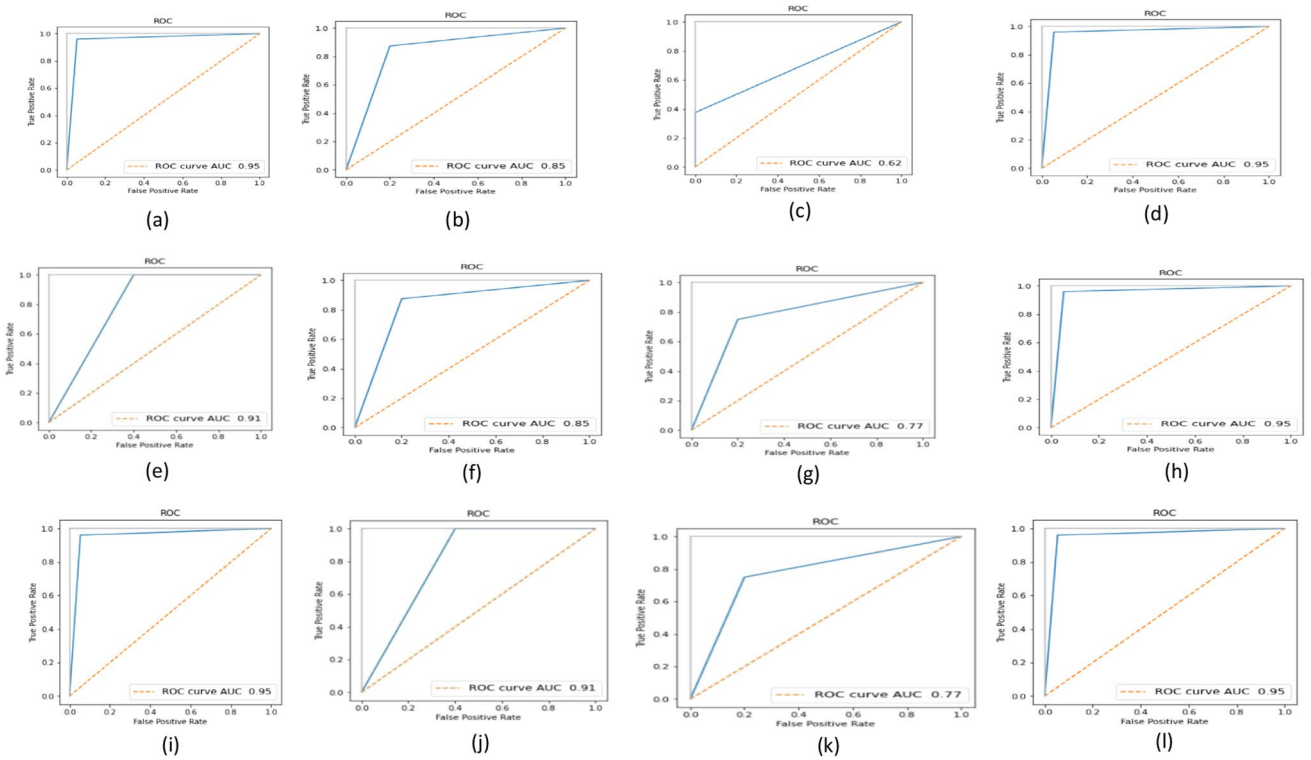

Fig. 8 ROC curves for three binary classifications and multiclass classification for **a–d** ADNI-1 **e–h** ADNI-2 (i)–(l) ADNI-3 datasets

MCI, MCI vs CN, and multiclass classification, MCI vs CN vs AD, which has been considered as one of the vital parameters. From each of the ROC curves, the accuracy of the ROC curve (AUC-ROC) has been determined as shown in Fig. 8. The X-axis of the ROC curve represents the True positive rate (TPR), whereas Y-axis denotes False Positive Rate (FPR) as shown in Fig. 8.

From Fig. 8, the AUC-ROC accuracy percentage is around 91% for AD vs CN for ADNI-2 dataset, whereas for ADNI-1 and 3 datasets the AUC-ROC accuracy has been calculated as 95%. For AD vs MCI binary classification, AUC-ROC accuracy for ADNI-1 and ADNI-2 is around 85% and 91% for ADNI-3 dataset. It is observed from the graphs of ROC that 95% AUC-ROC is obtained for multi class in all of the three ADNI datasets.

4.2 Performance of the proposed DCGNN model

The proposed DCGNN model has been constructed based on GAT networks, where the ASGD optimization technique has been applied to have updated weights for each output node in each of the three GCN layers. The nodes are formed in such a way that the information about the optimal features is stored to classify AD, CN, and MCI subjects. Here, 107 optimal features are extracted from the different Region of Interest (ROI) in 3D T1-weighted MRI scans considered for training 2863 images, validating 409 images, and testing 819 images. Tensor flow 2.0 libraries have been applied in this model to get higher training and validation accuracy while computing the loss score, as shown in the Fig. 9.

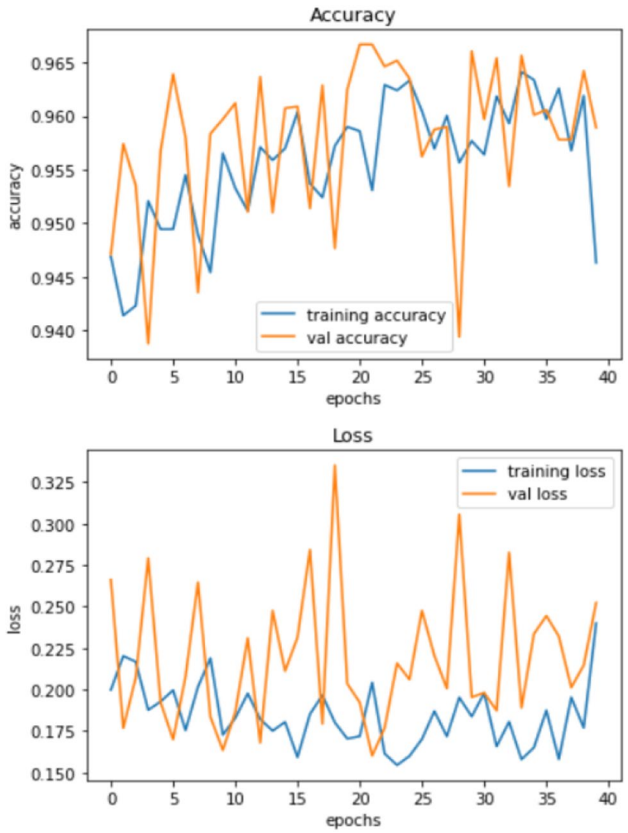

Fig. 9 Training and Validation accuracy and loss graphs

From the Fig. 9, it can be inferred that training and validation accuracy, and loss score, are calculated at each epoch where the simulation has been carried out for 40 epochs. Moreover, the proposed model does not overfit because a slight difference has been observed during training and validation accuracy. The values of the training, validation accuracy, and losses are computed where the variations of ± (1–3) % occurred, as shown in Fig. 9. For each of the layers of GCN, the training and validation accuracy has been computed in Table 6.

From Table 6, it has been observed that the number of layers in the GCN increases from Layer-I to Layer-III, both training accuracy and validation accuracy generally improve. This indicates that deeper GCN architectures can learn more complex and discriminative representations, leading to better performance on the training and validation datasets. With deeper layers, the gap between training and validation accuracy tends to decrease. In Layer-I, the difference between training accuracy (88%) and validation accuracy (79%) is significant, suggesting some overfitting level due to irregularities in the pixels of 3D T1 Weighted MRI images. However, as we move to Layer III, the training accuracy (97%) is still higher. Still, the validation accuracy (91%) is closer at the end of 10 epochs, indicating that the deeper model is generalizing better to unseen data. Also, it can be observed that in Layer-III, the validation loss (3%) is smaller than the training loss (10%), which indicates that the model generalizes well to unseen data during the training process. With this value, three binary classification tasks, such as AD vs CN, MCI vs CN, MCI vs CN, and multiclass classification, AD vs CN vs MCI, have been carried out, described in the subsequent sections.

4.2.1 Performance of AD vs CN subjects

From the two classes, various 3D T1 weighted scans have been taken from three planes for all three datasets, where

Table 6 Summary of all GCN layers with different scores of training and validation datasets

S. no	GCN layer	Training accuracy (%)	Validation accuracy (%)	Training loss (%)	Validation loss (%)
1	Layer-I	88	79	31	11
2	Layer-II	90	85	21	7
3	Layer-III	97	91	10	3

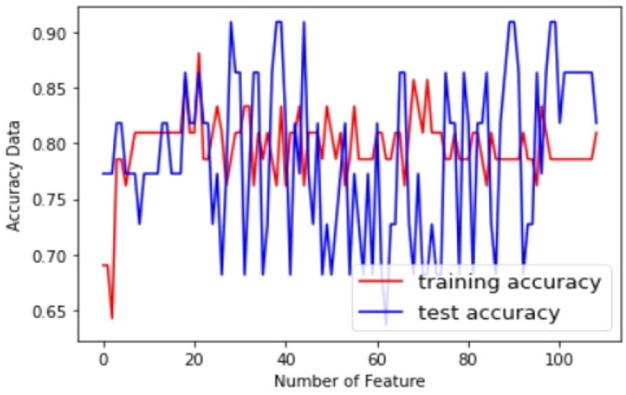

Fig. 10 Training and Testing graphs for AD vs NC subjects

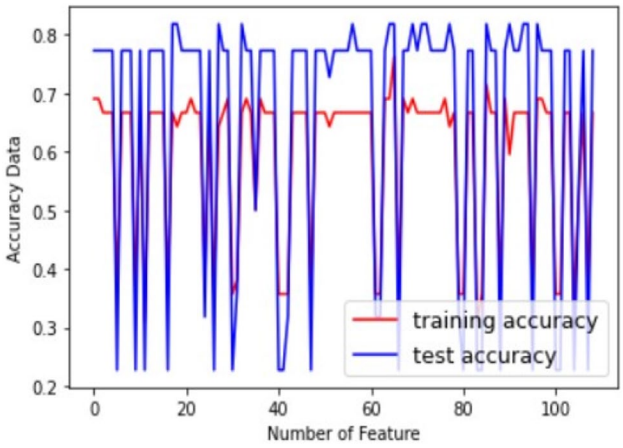

Fig. 11 Training and Testing Accuracy graphs for AD vs MCI dataset

70% of the total images have been trained for training purpose, 10% for validating the images, and the rest, 20% has been kept for testing purpose. In this dataset, all 107 optimal features are extracted from the hippocampal area, left caudate, right caudate, volume of right and left Cerebellum White Matter, and Surface Area of Inferior Temporal has been considered. By applying MMSE, FAQ, and Global CDR values, all the features are normalized in the FreeSurfer software, where each feature has been trained and tested in the form of the graph as shown in Fig. 10.

Based on the information presented in Fig. 10, the average Classification Accuracy (CA) is 91.25%. The highest accuracy is achieved using 24, 38, 39, 44, 90, 91, 97, and 98 features. These selected features are then fed into the proposed DCGNN deep learning model to achieve the best classification accuracy results while considering only relevant features and eliminating redundancy.

To reduce the high-dimensional feature vector, a statistical test, specifically the t-student test, has been performed. This test helps to efficiently decrease the feature vector dimensions, making the computation more manageable and maintaining the model's performance while reducing computational complexity.

4.2.2 Performance of AD vs MCI subjects

The MRI scans from MCI patients have less in number in all three ADNI datasets, resulting in classifying MCI affected patients from AD. The fewer MRI scans are due to the similar symptoms experienced by AD patients. However, large differences are observed in the frontal lobes portion of the MCI brain, making it more challenging to derive the solution. To overcome the above limitations, the proposed model is first trained on CN vs AD dataset and then tested on AD vs MCI datasets. For hypertunning the proposed model, we have set the dropout of 0.3 at the initial layers and then retrained the model by adding some additional features from the MCI dataset. The training and testing accuracy are calculated at each feature, as shown in Fig. 11.

The analysis of Fig. 11 reveals that the testing accuracy exhibits significant fluctuations compared to the training accuracy. This discrepancy can be attributed to the limited availability of MRI scans and the uneven distribution of pixels in T1-weighted MRI images. Due to the scarcity of testing data and skewed representation of pixel values, the testing accuracy experiences more significant variability during evaluation.

Among the different feature sets, the highest testing accuracy is observed when utilizing 18, 19, 31, 35, 56, 62, 63, 70, 72, 73, 78, 83, 90, 93, 94, and 99 features. These optimized feature sets are stored as nodes, which undergo further optimization through Graph Attention Networks (GAT) networks.

The output of these optimized nodes serves as input to the proposed Deep Convolutional Graph Neural Network (DCGNN) model, which aims to achieve the desired Classification Accuracy (CA). By leveraging GAT and DCGNN, the model seeks to improve the overall accuracy of the classification task despite the challenges posed by limited MRI scans and pixel distribution irregularities.

4.2.3 Performance of CN vs MCI subjects

Since the obtained images from MCI subjects are less clear in pixels, it is difficult to process these images to get the the required classification accuracy. Therefore, these images are further pre-processed by making appropriate slice corrections particularly in the hippocampus, left thalamus, right cerebellum cortex, and volume of Ventral DC to avoid overfitting. As a result, approximate equal distribution of the data is constructed for both classes to have a smooth operation. Here, the proposed model has been trained for 1432 images,

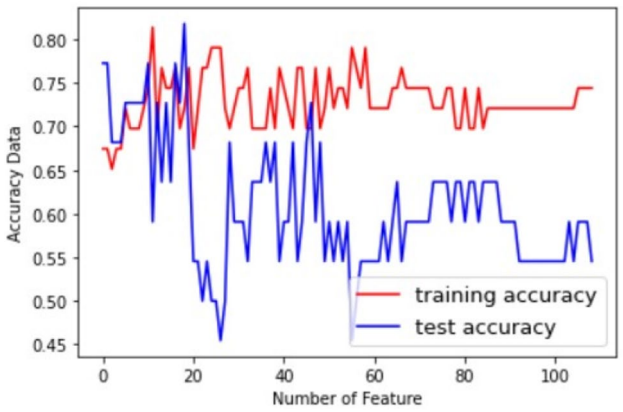

Fig. 12 Training and testing accuracy graphs for CN vs MCI dataset

validating 205 images and 410 images for testing purposes. For each of the 107 features obtained, training and testing accuracy graphs have been plotted in the Fig. 12.

Based on the findings in Fig. 12, it is evident that the classification accuracy tends to be slightly higher when using a lower number of features, whereas it decreases as the number of features increases. Additionally, significant variations are observed in the training and test datasets, indicating potential challenges in generalization.

To achieve the desired level of classification accuracy, an average testing accuracy of 68.32% is obtained by selecting a specific set of features, namely 31, 38, 42, 49, and 55, and storing them as nodes. Subsequently, these nodes undergo further processing using Graph Attention Networks (GAT). The application of GAT allows for removing redundant features, streamlining the feature representation and potentially improving the overall classification accuracy.

4.3 Comparison of training, validation, and testing simulation time

For all the ADNI datasets, 70% of the total data has been used for training, 10% of the total data for validating the samples whereas remaining 20% for samples testing. Training is the total time needed to complete one epoch of the proposed DCGNN model. However, training on more powerful hardware, such as GPUs or TPUs, can reduce training time significantly. Here, the GPU is turned on while processing the images, which shortens the time needed to achieve the desired accuracy compared to when the GPU is switched off. A shorter training time can result in optimal use of computational resources.

Validation time is an essential part of the machine learning process, as it helps ensure that the model can generalize well to new data. The validation time for a machine learning model is the time required to evaluate the model's performance on a validation dataset. During training, a portion of the dataset is typically set aside as a validation dataset, which is used to evaluate the DCGNN model's performance during training. This is important to prevent overfitting, which occurs when a model is too complex and learns to fit the training data too closely, resulting in poor performance on new data.

It's important to note that testing time is a critical part of the machine learning process, as it provides a final evaluation of the proposed model's performance before it's deployed in a real-world setting. It's crucial to ensure that the testing dataset represents the real-world data that the DCGNN model will encounter to ensure that the model will perform well in practice. The testing time for a machine learning model is the time required to evaluate the model's performance on a testing dataset. A summary of the training, validation, and testing times for all three ADNI datasets has been shown in Table 7.

Table 7 shows that GCN-1 consistently has the shortest training time across all ADNI datasets, followed by GCN-2, GCN-3, and DCGNN. This suggests that GCN-1 is the most computationally efficient during the training phase. As the dataset size increases from ADNI-1 to ADNI-3, the training time for all models also increases. This is expected since larger datasets require more computational resources and time for model training. GCN-1 and GCN-2 have similar validation times, while GCN-3 and DCGNN generally need more time for validation. As

Table 7 Training, validation, and testing time for all three ADNI datasets

ADNI datasets	Performance measure (sec)	GCN-1	GCN-2	GCN-3	DCGNN
ADNI-1	Training time	566	413	217	1196
	Validation time	412	392	256	1060
	Testing time	433	362	307	1102
ADNI-2	Training time	728	617	432	1777
	Validation time	527	418	391	1336
	Testing time	623	475	337	1435
ADNI-3	Training time	835	783	706	2324
	Validation time	647	568	483	1698
	Testing time	606	503	483	1592

with training time, the validation time increases with the dataset size, with ADNI-3 having the longest validation time. Similar to validation time, GCN-3 and DCGNN have longer testing times than GCN-1 and GCN-2. While there is some variation, the testing time remains relatively consistent across different datasets. GCN-1 consistently outperforms the other models in terms of both training and validation times. It is the most efficient model across all datasets. GCN-2 generally performs better than GCN-3 and DCGNN regarding training and validation times. DCGNN tends to have the longest training and validation times among all models.

4.4 Comparison of performance metrics with other networks

Different real-time experiments have been conducted by considering all the three axes of T1-weighted 3D MRI images from three other ADNI datasets. Here, the three standard Deep Neural Networks, such as Efficient 3DNET (Cai et al. 2021), Squeeze 3DNET (He et al. 2022), and GoogLeNET (Ak et al. 2022), have been applied to ADNI-1, ADNI-2, and ADNI-3 datasets. Here, the parameters of the Confusion Matrix, such as Sensitivity, Specificity, Precision, and F1-score, are used in comparing the DCGNN model with four cutting-edge techniques. These metrics provide valuable insights into the model's performance in correctly identifying positive and negative instances and its overall accuracy and balance between precision and recall. By evaluating these parameters, all three binary classifications, i.e., AD vs CN, CN vs MCI, MCI vs AD, and one multiclass classification, AD vs MCI vs CN, in all three ADNI datasets have been shown in Table 8. The activation function is prominent in determining the performance of the proposed DCGNN network, Efficient 3DNet, Squeeze3DNET, and GoogLENET. Here, three activation functions, such as RELU, LeakyRELU, and sin, have been applied, shown in Table 8.

Table 8 shows that the models using the sin activation function tend to perform better in terms of CA, SP, SENS, PR, and F1-Score across most disease groups compared to models using the RELU and LeakyRELU activation functions. The sin activation function appears to provide a good balance between sensitivity and specificity, leading to high F1-Scores, which are beneficial for imbalanced classification tasks. All models achieve relatively high CA, SENS, PR, and F1-Scores for the AD vs CN classification task, suggesting good discrimination between Alzheimer's disease (AD) and control (CN) groups. The MCI vs AD classification task generally yields lower accuracy and F1-Score compared to the AD vs CN task, which can be expected as distinguishing between mild cognitive impairment (MCI) and AD might be more challenging. The CN vs MCI and AD vs MCI vs CN classification tasks show varying performance, with some models performing better than others depending on the disease groups involved. The proposed DCGNN consistently achieves competitive performance across all disease groups and activation functions, with high CA, SP, SENS, PR, and F1-Score values. Efficient 3DNET, Squeeze3DNET, and GoogLENET models also demonstrate reasonable performance, but the proposed DCGNN consistently outperforms them.

4.5 Comparison with other existing techniques

One of the biggest challenges in diagnosing AD is catching it at the onset form. Most of the previous research has resulted in reliable classifications to determine if an individual is suffering from AD or its the normal state (Gupta et al. 2019). However, the binary classification of AD vs MCI and MCI vs NC is still an issue that needs addressing in the community of Alzheimer's Disease. The outcomes from the proposed work for classifying three binary categories of AD vs MCI, MCI vs CN, and NC vs AD, and multi-classification, i.e., AD vs MCI vs CN, are shown in Table 9. The simulation results obtained have been compared with other state-of-the-art techniques by considering 3D T1 weighted MRI biomarker and cognitive assessment values. It seems intuitive that discriminating between NC and AD should be simpler to do, and this is supported by our results, which indicate that the binary classification of AD vs NC proved 100% accurate. The other two binary classifications, AD vs MCI and NC vs MCI have accuracy rates of 95% and 91%, respectively. This reduction in accuracy is due to the features obtained after the segmentation of cortical and subcortical regions and the SUVR being less distinct for MCI and AD subjects when compared to the detection of AD vs NC.

The proposed DCGNN model offers several key advantages over existing methods for graph-based learning tasks. This model excels at learning expressive representations of graph-structured data using deep convolutional layers, allowing it to capture complex patterns and dependencies among nodes and edges. It efficiently incorporates spatial information and considers local neighborhood interactions, making it highly effective in extracting meaningful features from varying-sized graphs. DCGNN's adaptability to different tasks, robustness to graph irregularities, and reduced computational complexity further enhance its versatility. Additionally, DCGNN has demonstrated state-of-the-art performance on benchmark datasets, making it a powerful and promising model for graph-based learning applications, such as node classification, graph classification, and dynamic graph generation.

Table 8 Comparison of performance metrics using three standard models in three ADNI datasets

Activation function	Models	Classification	CA	SP	SENS	PR	F1-score
RELU	Efficient 3DNET	AD vs CN	88.13	69.75	80.24	80.33	70.88
		MCI vs AD	70.25	80.51	79.33	78.75	82.86
		CN vs MCI	81.25	75.22	81.75	69.33	88.91
		AD vs MCI vs CN	94.67	85.25	80.67	75.75	75.25
	Squeeze3DNET	AD vs CN	90.25	70.28	78.25	78.75	76.28
		MCI vs AD	73.75	82.37	77.93	80.24	81.33
		CN vs MCI	80.75	77.68	82.17	72.45	77.47
		AD vs MCI vs CN	93.34	80.29	80.38	77.39	81.35
	GoogLENET	AD vs CN	88.03	67.27	80.25	78.85	74.25
		MCI vs AD	68.25	79.71	79.94	81.24	81.82
		CN vs MCI	80.38	74.78	77.35	75.26	77.63
		AD vs MCI vs CN	94.13	85.27	80.84	73.82	79.87
	Proposed DCGNN	AD vs CN	89.24	68.23	79.26	81.13	75.03
		MCI vs AD	71.48	80.64	78.86	79.67	80.25
		CN vs MCI	82.67	73.11	79.91	70.45	76.33
		AD vs MCI vs CN	95.25	88.33	81.27	76.75	80.75
LeakyRELU	Efficient 3DNET	AD vs CN	86.27	68.82	78.52	82.24	73.67
		MCI vs AD	73.71	80.24	77.48	77.67	70.41
		CN vs MCI	80.22	70.51	76.37	68.29	74.47
		AD vs MCI vs CN	94.64	86.84	80.31	75.13	79.93
	Squeeze3DNET	AD vs CN	85.61	66.73	77.48	79.78	70.28
		MCI vs AD	72.35	78.25	80.67	75.27	77.61
		CN vs MCI	81.43	71.38	75.56	65.28	76.79
		AD vs MCI vs CN	92.68	85.27	81.74	77.74	80.56
	GoogLENET	AD vs CN	86.24	64.35	79.64	80.27	73.35
		MCI vs AD	71.25	77.45	78.81	76.32	78.53
		CN vs MCI	80.67	68.18	72.25	68.81	77.74
		AD vs MCI vs CN	94.25	85.29	81.37	76.25	81.82
	Proposed DCGNN	AD vs CN	88.28	67.25	80.38	83.39	74.25
		MCI vs AD	73.37	81.37	81.75	78.27	79.75
		CN vs MCI	83.78	72.23	78.82	69.25	78.35
		AD vs MCI vs CN	95.75	87.17	86.75	78.34	82.81
sin	Efficient 3DNET	AD vs CN	88.37	67.31	80.35	80.17	74.78
		MCI vs AD	74.82	77.57	78.81	78.67	80.39
		CN vs MCI	78.81	72.25	81.76	69.75	78.28
		AD vs MCI vs CN	90.18	86.94	83.69	80.11	81.17
	Squeeze3DNET	AD vs CN	87.75	67.52	78.57	80.37	73.38
		MCI vs AD	74.67	79.18	79.17	78.25	78.44
		CN vs MCI	78.82	72.42	80.44	68.48	77.71
		AD vs MCI vs CN	91.17	84.67	83.78	81.71	80.62
	GoogLENET	AD vs CN	88.58	66.58	78.87	78.49	72.21
		MCI vs AD	73.37	78.12	79.69	77.23	78.37
		CN vs MCI	78.45	71.32	80.53	67.97	77.44
		AD vs MCI vs CN	93.67	86.83	83.82	80.83	78.82
	Proposed DCGNN	AD vs CN	90.75	68.28	81.27	82.33	75.27
		MCI vs AD	75.71	80.47	80.34	79.78	81.71
		CN vs MCI	80.64	73.38	82.27	70.71	79.86
		AD vs MCI vs CN	94.13	88.27	84.11	82.75	83.34

Table 9 Comparing the proposed model with other existing methods by considering 3D T1 MRI biomarker

Classification categories	References	Model	Dataset	CA
AD vs CN	Lian et al. (2018)	FCN	ADNI-1, ADNI-2	90.3
	Liu et.al. (2020)	3D Dense Net	ADNI	88.9
	Bi et al. (2020)	RF	ADNI	90
	Liu et al.(2021)	Deep Separable CNN	ADNI	78.02
	Proposed model	**DCGNN**	**ADNI-1, ADNI-2, ADNI-3**	**93.34**
CN vs MCI	Liu et.al. (2021)	3D Dense Net	ADNI	76.2
	Abral et.al. (2020)	CNN + 3D ResNET	ADNI	83.01
	Feng et al. (2020)	3DCNN + SVM	ADNI	76.82
	Xiu et al. (2021)	LSTM	ADNI-1, ADNI-2	67.5
	Proposed model	**DCGNN**	**ADNI-1, ADNI-2, ADNI-3**	**83.67**
AD vs MCI	Janani et al. (2021)	DL + RF	ADNI	74
	Gupta et al. (2013)	NIBR-Net	ADNI	78.2
	Khan et.al. (2022)	DJMAD-Net	ADNI	62.9
	Alizadeh et al. (2022)	1DCNN	ADNI	76.6
	Proposed model	**DCGNN**	**ADNI-1, ADNI-2, ADNI-3**	**79.1**
AD vs CN vs MCI	Hosseni et. al. (2018)	CNN-3D	ADNI-1	94.8
	payan et al. (2015)	3D-CNN-PAD	ADNI-1	94.8
	Murugun et al. (2021)	DEMNET	ADNI-1	95.2
	Jian et al. (2019)	CNN-AD (VGG16)	ADNI-1	95.13
	Proposed model	**DCGNN**	**ADNI-1, ADNI-2, ADNI-3**	**95.68**

Bold indicates a model that is suggested and evaluated against various techniques using a 3D T1 MRI biomarker

5 Conclusion and future scope

In this paper, Graph Attention Networks (GAT) have been proposed for processing the extracted features from 3D 1.5 T and 3 T magnetic field strength MRI images from three different ADNI subjects as nodes which can given as inputs. A Single Graph Attention Layer has been applied for faster computation of extracted features present in the nodes to have optimized nodes for giving inputs to the proposed DCGNN networks. This proposed work has a CA of 93.34% for AD vs CN, 83.67% CA for CN vs MCI, 79.1% Classification Accuracy for AD vs MCI, and 95.68% CA for AD vs CN vs MCI. Here, the proposed network has been applied to 857 participants comprising all classes from ADNI-1, ADNI-2, and ADNI-3 database. The simulation results indicate that the performance has been regarded as the best among the four recent state-of-the-art techniques.

Other learning strategies, including DL algorithms like Long Short-Term Memory (LSTM), Deep Autoencoders (DAE), and additional Artificial Neural Networks (ANN) variants, may be used in upcoming research to achieve significantly improved classification accuracy for all different classes of Alzheimer Disease. Additionally, a blend of feature selection approaches and deep learning models would be applied to single modalities as well as multi-modalities, which can be tackled as a future topic with a primary focus on the interpretability of clinical diagnosis. Also, research is mainly focused on feature-based methodologies rather than differentiating Alzheimer's disease from standard control. The enormous dataset can be used for classification in the future. Other machine learning technologies are also available for use. As a result, these models can also be used to detect other disorders, including blood cancer, Parkinson's disease, brain cancer, eplissey, and Parkinson's disease. Based on this concept, various ensemble models can

be created to improve the classification accuracy of other datasets, such as text mining and text classification, which classify text written in various languages and formats.

Funding The authors have not disclosed any funding.

Declarations

Conflict of interest All the authors declare that they have no conflict of interest related to this review. This article does not contain any studies with human or animal subjects performed by any authors.

Ethical approval As there are no animals involved in this study, ethical approval is not required.

References

Abrol A, Bhattarai M, Fedorov A, Yuhui Du, Plis S, Calhoun V, Initiative ADN (2020) Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease. J Neurosci Methods 339:108701

Ak A, Topuz V, Midi I (2022) Motor imagery EEG signal classification using image processing technique over GoogLeNet deep learning algorithm for controlling the robot manipulator. Biomed Signal Process Control 72:103295

Alam S, Kwon G-R, Initiative ADN (2017) Alzheimer disease classification using KPCA, LDA, and multi-kernel learning SVM. Int J Imaging Syst Technol 27(2):133–143

Alizadeh F, Homayoun H, hosseinBatouli A, Noroozian M, Sodaie F, Salary HM, Rad HS (2022) Differential diagnosis among alzheimer's disease, mild cognitive impairment, and normal subjects using resting-state fMRI data extracted from multi subject dictionary learning atlas. Front Biomed Technol 9(4):297–306

Asl EH, Ghazal M, Mahmoud A, Aslantas A, Shalaby A, Casanova M, Barnes G, Gimel'farb G, Keynton R, El Baz A (2018) Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. Front Biosci Landmark 23:584–596

Bahdanau D, Kyunghyun C, Yoshua B (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473

Bi X-a, Xi Hu, Hao Wu, Wang Y (2020) Multimodal data analysis of Alzheimer's disease based on clustering evolutionary random forest. IEEE J Biomed Health Inform 24(10):2973–2983

Cai B, Cheng E, Liang P, Xiong C, Sun Z, Zhang Q, Song B (2021) Ghost-light-3dnet: efficient network for heart segmentation. In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI) (pp 1704–1708). IEEE

Chen Z, Wang Z, Zhao M, Zhao Q, Liang X, Li J, Song X (2022) A new classification network for diagnosing Alzheimer's disease in class-imbalance MRI datasets. Front Neurosci 16:807085

Cui R, Liu M, Initiative ADN (2019) RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. Comput Med Imaging Graph 73:1–10

Cui J, Zhaoxia W, Seng-Beng H, Erik C (2023) Survey on sentiment analysis: evolution of research methods and topics. Artif Intell Rev 1–42

Cui C, Haichun Y, Yaohong W, Shilin Z, Zuhayr A, Lori AC, Keith TW, Bennett L, Yuankai H (2023) Deep multi-modal fusion of image and non-image data in disease diagnosis and prognosis: a review. Progress Biomed Eng

Cui Z, Nianzhi Tu, Itoh M (2023a) Effects of brand and brand trust on initial trust in fully automated driving system. PLoS ONE 18(5):e0284654

Demirhan A (2016) Classification of structural MRI for detecting Alzheimer's disease. Int J Intell Syst Appl Eng 4(1):195–198

Farooq A, Anwar S, Awais M, Alnowami M (2017) Artificial intelligence based smart diagnosis of Alzheimer's disease and mild cognitive impairment. In: 2017 international smart cities conference (ISC2) (pp 1-4), IEEE

Feng W, Halm-Lutterodt NV, Tang H, Mecum A, Mesregah MK, Ma Y, Guo X (2020) Automated MRI-based deep learning model for detection of Alzheimer's disease process. Int J Neural Syst 30(06):2050032

Fischl B (2012) FreeSurfer. Neuroimage 62(2):774–781

Fritsch J, Wankerl S, Nöth E (2019) Automatic diagnosis of Alzheimer's disease using neural network language models. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp 5841-5845), IEEE

Goedert M, Ghetti B (2007) Alois Alzheimer: his life and times. Brain Pathol 17(1):57–62

Gupta A, Ayhan M, Maida A (2013) Natural image bases to represent neuroimaging data. In: International conference on machine learning (pp 987–994). PMLR

Gupta Y, Kun HL, Kyu YC, Jang JL, Byeong CK, Goo-Rak K (2019) Alzheimer's disease diagnosis based on cortical and subcortical features. J Healthcare Eng

He R, Liu Y, Xiao Y, Xingyu Lu, Zhang S (2022) Deep spatio-temporal 3D densenet with multiscale ConvLSTM-Resnet network for citywide traffic flow forecasting. Knowl-Based Syst 250:109054

Jain R, Jain N, Aggarwal A, Jude Hemanth D (2019) Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cognit Syst Res 57:147–159

Javeed A, Dallora AL, Berglund JS, Idrisoglu A, Ali L, Rauf HT, Anderberg P (2023) Early prediction of dementia using feature extraction battery (FEB) and optimized support vector machine (SVM) for classification. Biomedicines 11(2):439

Kamathe RS, Joshi KR (2018) A novel method based on independent component analysis for brain MR image tissue classification into CSF, WM, and GM for atrophy detection in Alzheimer's disease. Biomed Signal Process Control 40:41–48

Khan R, Qaisar ZH, Mehmood A, Ali G, Alkhalifah T, Alturise F, Wang L (2022) A practical multiclass classification network for the diagnosis of Alzheimer's disease. Appl Sci 12(13):6507

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Frackowiak RS (2008) Automatic classification of MR scans in Alzheimer's disease. Brain 131(3):681–689

Kumari R, Nigam A, Pushkar S (2022) An efficient combination of quadruple biomarkers in binary classification using ensemble machine learning technique for early onset of Alzheimer disease. Neural Comput Appl 34(14):11865–11884

Lahmiri S, Boukadoum M (2014) New approach for automatic classification of Alzheimer's disease, mild cognitive impairment and healthy brain magnetic resonance images. Healthcare Technol Lett 1(1):32–36

Lahmiri S, Shmuel A (2019) Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease. Biomed Signal Process Control 52:414–419

Lian C, Liu M, Zhang J, Shen D (2018) Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Trans Pattern Anal Mach Intell 42(4):880–893

Liu M, Li F, Yan H, Wang K, Ma Y, Shen Li, Mingqing Xu, Initiative ADN (2020) A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. Neuroimage 208:116459

Liu J, Li M, Yuling Luo Su, Yang WL, Bi Y (2021) Alzheimer's disease detection using depthwise separable convolutional neural networks. Comput Methods Programs Biomed 203:106032

Lu S, Xia Y, Cai W, Fulham M, Feng DD, Alzheimer's Disease Neuroimaging Initiative (2017) Early identification of mild cognitive impairment using incomplete random forest-robust support vector machine and FDG-PET imaging. Comput Med Imaging Graph 60:35–41

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement 7(3):263–269

Menéndez G (2017) La revolución de la longevidad: cambio tecnológico, envejecimiento poblacional y transformación cultural. Revista De Ciencias Sociales 30(41):159–178

Murugan S, ChandranVenkatesan MG, Sumithra X-ZG, Elakkiya B, Akila M, Manoharan S (2021) DEMNET: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. IEEE Access 9:90319–90329

Payan A, Montana G (2015) Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv preprint

Petersen RC, Paul SA, Laurel AB, Michael CD, Anthony CG, Danielle JH, Clifford RJ et al (2010) Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology 74(3):201–209

Sharma R, Goel T, Tanveer M, Murugan R (2022) FDN-ADNet: fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans. Appl Soft Comput 115:108099

Vaswani A, Noam S, Niki P, Jakob U, Llion J, Aidan NG, Łukasz K, Illia P (2017) Attention is all you need. Adv Neural Inform Process Syst 30

Venugopalan J, Tong L, Hassanzadeh HR, Wang MD (2021) Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci Rep 11(1):3254

Xue C, Karjadi C, Paschalidis IC, Au R, Kolachalama VB (2021) Detection of dementia on voice recordings using deep learning: a Framingham Heart Study. Alzheimer's Res Ther 13:1–15

Yue L, Gong X, Li J, Ji H, Li M, Nandi AK (2019) Hierarchical feature extraction for early Alzheimer's disease diagnosis. IEEE Access 7:93752–93760

Zhang Y, Dong Z, Phillips P, Wang S, Ji G, Yang J, Yuan T-F (2015) Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. Front Comput Neurosci 9:66

Zhao Z, Chuah JH, Lai KW, Chow CO, Gochoo M, Dhanalakshmi S, Wu X (2023) Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: a review. Front Comput Neurosci 17:1038636

Zhu Y, Zhu X, Kim M, Shen D, Wu G (2016) Early diagnosis of Alzheimer's disease by joint feature selection and classification on temporally structured support vector machine. In: Medical image computing and computer-assisted intervention–MICCAI 2016: 19th international conference, athens, greece, 2016 Proceedings, Part I 19 (pp 264–272). Springer International Publishing