Received: 28 February 2023 | Revised: 15 February 2024 | Accepted: 24 February 2024

DOI: 10.1002/ima.23054

RESEARCH ARTICLE

WILEY

A joint autoencoder and classifier deep neural network for AD and MCI classification

Reema Ganotra[1] | **Shailender Gupta**[1] | **Shirin Dora**[2]

[1]Department of Electronics Engineering, J. C. Bose University of Science and Technology, YMCA, Faridabad, India

[2]Department of Computer Science, School of Science, Loughborough University, Loughborough, UK

Correspondence
Reema Ganotra, Department of Electronics Engineering, J. C. Bose University of Science and Technology, YMCA, Haryana, India.
Email: reemaganotra01@gmail.com

Abstract

In this article, we present a new approach to distinguish progressive mild cognitively impaired (pMCI) subjects, who eventually develop Alzheimer's disease (AD) from stable MCI (sMCI) subjects whose situation does not deteriorate into AD. The proposed approach combines the discriminating capabilities of classifiers and representation learning capacities of autoencoders into a unified architecture, and is hence termed as joint autoencoder and classifier deep neural network (JACDNN). JACDNN employs a single classifier and multiple autoencoders that are trained together to perform pattern classification. The classifier in JACDNN is trained using standard approaches to distinguish between subject from different classes using the binary cross entropy loss. The autoencoders in JACDNN, regularizes individual layers in the network used for classification to learn representations useful for reconstructing a given input. The performance of JACDNN has been evaluated on several machine learning problems pertaining to dementia, namely AD versus cognitively normal (CN) subjects, AD versus sMCI, CN versus pMCI, and pMCI versus sMCI. These problems are targeted using two datasets. The first dataset consist of gray matter (GM) features of subjects and the second dataset consist of combination of GM and white matter (WM) features. It is observed that better classification results are obtained when the classifier is built on GM and WM as compared with GM features alone. Performance comparison of JACDNN with other existing approaches has been conducted for these problems. The results clearly indicate that JACDNN performs better than other existing approaches for these problems.

KEYWORDS

Alzheimer's disease, deep neural networks, magnetic resonance imaging, mild cognitive impairment, regularization

1 | INTRODUCTION

About 50 million people in the world are living with Alzheimer's disease (AD) and this number is likely to triple by 2050.[1] This poses an immense challenge for healthcare systems around the world. At present, there is no cure for AD, but it is possible to delay the progression of disease upon diagnosis through early intervention.[2] Clinical diagnosis of AD employs neuro-psychological and behavioral assessments to identify the cognitive deficits that are associated with AD. However, these tests are prone to subjective interpretations and do not take into

consideration an individual's personal and social circumstances, which impacts their effectiveness, especially during the early stages of the disease.[3]

It is known that neuropathological effects of AD start affecting the brain years before clinical symptoms appear.[4] These effects can be observed using noninvasive neuroimaging techniques like structural magnetic resonance imaging (sMRI) and positron emission tomography (PET) much before clinical diagnosis of AD.[5] This prompted research into clinically identifiable stages on the pathway to AD. One of the most studied preclinical stage of AD is mild cognitive impairment (MCI) in which a person exhibits cognitive decline, which is less severe than AD. Patients with MCI are far more susceptible to develop AD than CN subjects.[6–8] However, in many cases, MCI patients stay stable for several years or advance towards a form of dementia other than AD.[9] This heterogeneity has given rise to an interest in the development of machine learning methods that can discriminate between MCI patients with different prognosis. Much of the research in this direction has focused on the problem of using MRI or PET to distinguish between patients with progressive MCI (pMCI) that turns into AD and those with stable MCI (sMCI).

Many existing approaches for this problem employ a preprocessing pipeline to identify informative features, which are then used to build a pattern classifier that distinguishes between pMCI and sMCI subjects. The purpose of preprocessing pipeline is to improve the quality of the obtained features through application of statistical techniques like feature selection[10,11] and principal component analysis.[12,13] In the work by Moradi et al.,[11] the preprocessing stage involved processes for removal of effects related to normal aging and feature selection using regularized logistic regression. The selected features are then used to compute a biomarker for classification of subjects into pMCI and sMCI using low-density separation. The preprocessing pipeline in the work presented by Beheshti et al.[10] estimates the regions of interest (RoI) based on gray matter atrophy and then genetic algorithm is used to select informative voxels from each RoI. The values of selected voxels are presented to a support vector machine (SVM) for the classification task. The multiple steps in the preprocessing pipeline render these approaches time-consuming and hard to reproduce[14] as each stage needs to be optimized independently.

Recently, the success of deep learning in tasks like image classification,[15,16] natural language processing,[17] and image set based classification has led to an increased interest in its use for prediction of neurodegenerative disorders like AD and autism spectrum disorder[18] based on neuroimaging data. The image set based classification approach[19,20] holds promise for the exploration of high-dimensional neuroimaging data for differentiating AD patients from those with normal cognitive function. These approaches harness the richness of both spatial and temporal information embedded within image sequences. This strategic utilization of image sequences contributes to an enhanced accuracy in classifying samples with neurodegenerative diseases as these techniques[21–23] automatically estimate the relevant features for a given problem.

Deep learning approaches normally require large number of samples for training due to the large number of parameters associated with these networks. But, barriers in acquiring standardized neuroimaging data for large number of individuals with neurodegenerative disorders make it difficult to obtain large amount of training data. To overcome this problem, two main approaches have been adopted in the literature. First, several researchers have made use of stronger regularization techniques to prevent overfitting on the training data.[24] Spasov et al.[24] have used deep neural networks with separable and grouped convolutions to develop a network with fewer learnable parameters. Second, several approaches have used transfer learning techniques, which involve pretraining a network on large number of samples and then fine-tuning the pretrained network to perform a specific classification task.[25–28] In the approach used by Suk et al.,[27] a stacked autoencoders is pretrained using data from different classes irrespective of their labels. After pretraining, a single output layer is added to the network for classification and weights in the whole network are fine-tuned using back-propagation. Oh et al.[26] proposed a convolutional autoencoder to learn representations for distinguishing between AD and CN subjects and then used the same network for pMCI versus sMCI classification. In a similar approach, Wee et al.[28] trained a graph-convolutional neural network using samples of different classes from ADNI-2 cohort and then used this pretrained network to classify subjects from ADNI-1 cohort and an Asian cohort.

Most transfer learning approaches employ autoencoders for pretraining, which utilize an objective function that involves reconstructing the network input. It has been argued that transfer learning approaches provide the network with a better starting point[27] prior to fine-tuning the network parameters for a particular classification task. But, the lack of a reconstruction-related objective during fine-tuning may allow the network to drift away from the starting point, thereby reducing the network to a simple feedforward classification network.

In this article, we developed a novel network architecture, termed as joint autoencoder and classifier deep neural network (JACDNN), that together uses a classification and reconstruction objective to impose a stronger regularization and force the network to learn representations suitable for classification.

JACDNN employs a network architecture that consists of a single classification sub-network and multiple reconstruction sub-networks. The classification sub-network is a feedforward network that consists of multiple convolutional layers. The final layer in the classification sub-network consists of as many neurons as the number of classes and is used to determine the predicted class for a given input. Each convolution layer in the classification sub-network is also connected to multiple convolutional transpose layers, which form a single reconstruction sub-network associated with a particular convolutional layer. The output of each reconstruction sub-network has the same shape as the network input. During training, JACDNN is simultaneously trained to achieve high classification performance based on the output of the classification sub-network and high reconstruction performance based on the outputs of all reconstruction sub-networks.

The performance of JACDNN has been evaluated using data from Alzheimer's Disease Neuroimaging Initiative (ADNI) on four different classification problems involving AD and pMCI subjects. The study is conducted on two different datasets. First dataset comprises of features extracted from gray matter (GM) images of the subjects and the second dataset comprises of the features extracted from GM and white matter (WM) images of the subjects. The results of performance evaluation have been compared with the performance of recently proposed approaches for all of these problems. It can be clearly observed that JACDNN performs better than existing approaches for all of the problems. Specifically, the pMCI versus sMCI problem achieved 75.62% accuracy, 72.15% recall, and 71.44% precision. Similarly, for classification of AD versus CN 91.08% accuracy, 92.98% recall and 90.03% precision are obtained. For AD versus sMCI, 82.17% accuracy, 73.72% recall, and 75.31% precision and for CN versus pMCI, 81.93% accuracy, 76.17% recall, and 79.35% precision are achieved.

The rest of the article is organized into following sections: Section 2 describes the details of the dataset and data-augmentation techniques used in this study. Section 3 presents the architecture of JACDNN and its training procedure. Section 4 presents the results of performance comparison between JACDNN and other existing approaches. Section 5 summarizes the conclusions from this study.

2 | MATERIALS

2.1 | Participants and data

All data used in this work is obtained from the database for Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu). Only T1-weighted sMRI images for baseline assessment of all subjects are used in this study. This includes records for 580 male subjects and 464 female subjects. In total, there are 321 AD subjects, 324 CN subjects, 228 pMCI subjects and 171 sMCI subjects with ages between 55 and 91 years. Table 1 provides details about demographics, neuro-psychological testing, Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and ApoE genotyping testing for all subjects in the four classes. All T1 weighted sMRI images were subjected to standard preprocessing steps decided by ADNI[29] to remove artifacts that occurred during the image acquisition process. These steps involve sequentially applying adjustments of 3D gradwrap correction, B1 nonuniformity correction, and N3 bias field correction.

2.2 | Data preprocessing

All T1 weighted sMRI images are preprocessed using Statistical Parametric Mapping v12 (SPM 12) toolbox. First, the unified segmentation algorithm[30] is used to estimate tissue probability maps pertaining to GM, WM, and CerebroSpinal Fluid (CSF) tissue in the sMRI images. Only preprocessed images pertaining to GM and WM regions are used for developing the models presented in this article. This unified segmentation algorithm performs segmentation, bias correction, and normalization simultaneously. For the purpose of segmentation, the algorithm conducts volumetric comparison between the sMRI image of a given subject and a reference template to estimate GM, WM, and CSF tissue probability maps. This procedure also allows removal of artifacts like skull, scalp, and air. These tissue probability maps are bias corrected for removal of nonuniform intensities. The bias-corrected images are normalized with respect to the International Consortium for Brain Mapping (ICBM) templates. Next, DARTEL toolbox[31] is used to estimate a local template and register all images to this template. The local template is generated by an iterative process that involves averaging across all segmented images to compute an average image and then warping all images according to this average image. All images are manually checked and corrected for registration defects and misalignment. The registered images are normalized to the

Characteristics	AD	CN	pMCI	sMCI
Total subjects	321	324	228	171
Age	75.23 (±7.8)	75.35 (±5.49)	73.91 (±7.15)	73.75 (±7.33)
Weight	73.11 (±14.70)	76.05 (±15.21)	75.25 (±14.59)	77.36 (±13.01)
Gender (M/F)	174/147	165/159	132/96	109/62
Education	15 (±2.9)	16.3 (±2.7)	15.75 (±2.81)	15.69 (±2.91)
CDR score	0.75 (±0.24)	0 (±0)	0.50 (±0.03)	0.49 (±0.07)
GDS score	1.655 (±1.42)	0.83 (±1.15)	1.65 (±1.34)	1.65 (±1.47)
MMSE score	23.29 (±2.04)	29.13 (±0.99)	26.62 (±1.70)	27.85 (±1.63)
ApoE4 Score1	3.13 (±0.47)	2.88 (±0.41)	3.10 (±0.46)	2.94 (±0.37)
ApoE4 Score2	3.66 (±0.47)	3.26 (±0.45)	3.69 (±0.45)	3.38 (±0.48)

TABLE 1 Demographic, neuropsychological, and genotype details of the participants. In the table, age and education of participants are specified in years and weight of participants is specified in kilograms.

Montreal Neurological Institute (MNI) template and smoothed using a 10 mm full width half maximum isotropic Gaussian kernel. In the final preprocessing step, images are modulated using the deformation fields of their own Jacobian determinants. To reduce the dimensionality of the preprocessed images, voxel size is downsampled to a resolution of 3 mm, resulting in images with a size of $61 \times 73 \times 61$ voxels.

2.3 | Data augmentation

Large number of parameters in a deep neural network necessitate the need for a large number of training samples. For this purpose, data augmentation techniques are used with preprocessed images pertaining to GM and WM tissues in the brain. An established approach used for augmentation of brain images is to flip its left and right side across the central axis. This doubles the number of samples available for training. In this paper, this technique is used separately for both GM and WM images. A single sample for the network is obtained by concatenating images corresponding to GM and WM tissues for a single subject along the channel axis. Different combinations of the original and flipped images of both tissues are used to generate four different training samples for a given subject. These include samples obtained by concatenating original GM and original WM images, original GM and flipped WM images, flipped GM and original WM images, flipped GM and flipped WM images. This results in a fivefold increase in the number of training samples. After this, training dataset consists of 1125, 1135, 800, and 600 samples from AD, CN, pMCI, and sMCI classes. It may be noted that the dataset is divided into training and testing sets by a ratio of 70:30 and data augmentation is only used for training samples. The samples obtained after flipping the tissue maps are further augmented by randomly rotating the two tissue maps by (±)10 degrees.

3 | JOINT AUTOENCODER AND CLASSIFIER DEEP NEURAL NETWORK

In this section, the architecture and the training procedure of JACDNN are provided.

3.1 | Architecture

Figure 1 shows the architecture of the JACDNN. The input layer is used to present samples $(x_1,y_1),...,(x_i,y_i),...$ to the network where x_i is the preprocessed MRI data and $y_i \in [0,1]$ is the associated class. Each input sample has a shape of $h \times w \times c$. Here, h and w represent the height and width of the input sample along the sagittal plane. Similarly, c correspond to the channels of the input MRI image along coronal plane.

The network consists of a single classification sub-network and multiple reconstruction sub-networks (Figure 1). The output of the classification sub-network is used to determine the predicted class and the output of the reconstruction sub-network is used to reconstruct a given input sample. All sub-networks comprise of three types of layers, namely, three-dimensional convolution layer, three-dimensional convolution transpose layer, and Gaussian noise layer. The functionality and the purpose of each of these layers are described below.

3.1.1 | 3-D convolution layer

The sub-network for classification consists of three 3-D convolution layers followed by a single fully connected

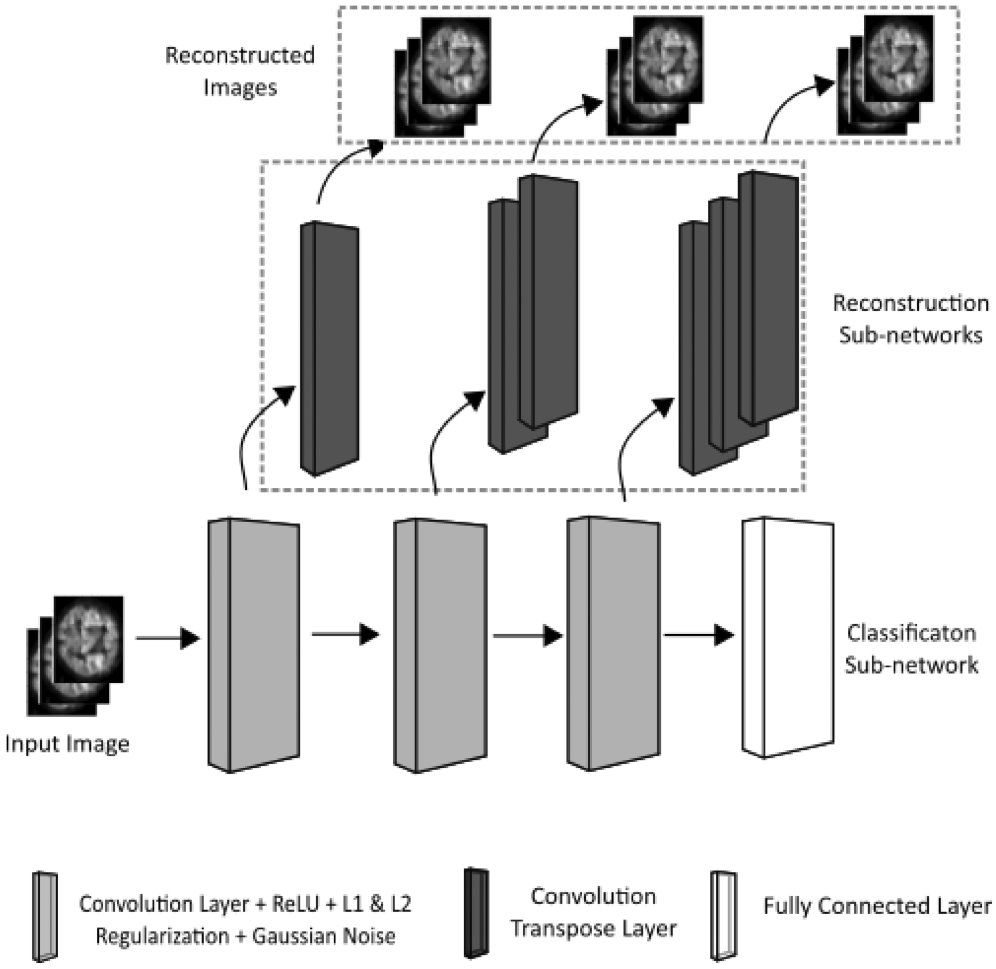

FIGURE 1 Architecture of JACDNN showing classification sub-network and reconstruction sub-networks.

layer with two neurons whose responses are used to determine the predicted class for a given input sample. A 3-D convolution layer applies convolution across all three dimensions of the input. The output (h^l) of the l^{th} convolution layer is obtained by convolving multiple filters with the output of the previous layer. The use of multiple filters allows extraction of multiple feature maps that may be useful for discriminating between the input samples from different classes. Each feature map is a 3-D tensor that is obtained by applying the convolution operation to the output of the previous layer using a single filter. The k^{th} feature map (h_k^l) in the l^{th} convolution layer is given by

$$h_k^l = f\left(w_k^l * h^{l-1} + b_k^l\right), \qquad k \in [1, ...K_l]. \tag{1}$$

where "*" represents the convolution operation. w_k^l and b_k^l represent the weights and biases associated with the k^{th} filter of the l^{th} convolution layer. K_l is the number of filters used in l^{th} layer and f is the activation function, which is chosen to be rectified linear unit (ReLU) in this article. Based on the estimated feature maps, the output of the convolution layer is a four-dimensional tensor given by

$$h^l = \left[h_1^l, ..., h_k^l ..., h_{K_l}^l\right] \tag{2}$$

Each 3D convolution layer of network shown in Figure 1 consists of kernel size 3 and stride 2. The first, second, and third convolution layers employ 16, 32, and 64 filters, respectively.

3.1.2 | 3-D convolution transpose layer

Each convolution layer in the network has an associated reconstruction sub-network (Figure 1). These sub-networks in JACDNN are trained to reconstruct the current input (x_i) for the network based on the feature map of the l^{th} convolution layer. In each of these sub-networks, a given convolution layer is connected to multiple 3-D convolution transpose layers, which are used to upsample the output of the particular convolution layer. The parameters of these convolution transpose layers are

chosen such that the output ($\widehat{x}_i^l$) of the last layer in the l^{th} reconstruction sub-network has the same dimensions as that of the input sample. A convolution transpose layer also applies a convolution operation to its input. Thus, its output is determined using Equation 2.

As shown in Figure 1, the first reconstruction sub-network consists of single 3D convolution transpose layers employing only one filter. The second reconstruction sub-network consist of two 3D convolution transpose layers composed of 16 and one filter, respectively. The third reconstruction sub-network consist of three 3D convolution transpose layers composed of 16, 16, and 1 filter, respectively. Each convolution transpose layer employs kernel size 3 and stride 2.

3.1.3 | Gaussian noise layer

Each convolution layer in the classification sub-network is followed by a Gaussian noise (GN) layer. A GN layer adds a random noise sampled from the normal distribution $N \in [0,\sigma]$ with a mean of zero and a standard deviation of σ to the output of a convolution layer.[32] For all results reported in this article, a value of 0.01 is used for σ. This augments the input data used for training the network and helps in avoiding overfitting. Note that all GN layers are only used during training.

3.2 | Objective function

The objective function for JACDNN constitutes two kinds of losses, namely a classification loss and a reconstruction loss. The classification loss (L_c) employs binary cross entropy to improve the classification accuracy of the network and is given by

$$L_c = \sum_{i=1}^{n} y_i \log \widehat{y}_i + (1-y_i)\log(1-\widehat{y}_i) \tag{3}$$

where $\widehat{y}_i$ denotes the output of the classification network for the i^{th} input sample.

The reconstruction loss is used to generate better reconstructions of a given input based on all the reconstruction sub-networks and utilizes mean squared error to observe the difference between input image x_i and reconstructed image $\widehat{x}_i$. The reconstruction loss (L_r) associated with the l^{th} sub-network in JACDNN is given by

$$L_r = \frac{1}{n}\sum_{i=1}^{n} \log (x_i - \widehat{x}_i)^2 \tag{4}$$

Based on the loss functions described in Equations (3) and (4), the total loss ($\mathscr{L}$) for JACDNN is given by

$$\mathscr{L} = L_c + \beta\left(\sum_{i=1}^{l}(L_{rl})\right) \tag{5}$$

where β is termed as the reconstruction coefficient. A high value of β forces the reconstruction sub-networks in JACDNN to generate better reconstructions of a given input, which results in strong regularization and may lead to lower classification performance. A lower value of β results in weak regularization due to the reconstruction objective in which case JACDNN behaves like a simple deep neural network for classification. In addition, L1 (Lasso) and L2 (Ridge) regularization is employed in each convolution layer of the network. The coefficients for both L1 and L2 regularization is set to 0.01 for all layers. JACDNN is trained to minimize the objective function in Equation 5 using stochastic gradient descent.

4 | RESULTS AND DISCUSSION

In this section, the performance of JACDNN is evaluated using the preprocessed GM and WM features from the ADNI dataset for four different classification problems. These problems include AD versus CN, pMCI versus sMCI, AD versus sMCI, and CN versus pMCI. The problem of AD versus CN involves classification of subjects into AD patients and cognitively normal individuals. pMCI versus sMCI problem focuses on distinguishing between patients with progressive MCI that can deteriorate into AD and subjects with a stable form of MCI. The problem of AD versus sMCI involves discriminating between subjects with AD and a stable form of MCI. The CN versus pMCI problem focuses on the task of discriminating between cognitively normal subjects and patients with progressive MCI that can potentially turn into AD. The performance of the network on these experiments have been calculated using metrics of accuracy (η_a), recall (η_r), precision $\left(\eta_p\right)$, and specificity (η_s). Accuracy is defined as the percentage of samples that are correctly classified by the network, given by

$$\eta_a = \frac{\mathrm{TP}+\mathrm{TN}}{\text{Total number of samples}} \tag{6}$$

where TP and TN stand for true positive and true negative samples. TP and TN represent the number of respective AD or pMCI and CN or sMCI subjects that are correctly classified by the network. Recall is defined as

TABLE 2 Parametric study for Gaussian noise layer for AD and CN classification.

Study	Accuracy (%)	Recall (%)	Precision (%)
JACDNN without GN	89.11	85.56	92.22
JACDNN with GN	91.08	92.08	90.03

TABLE 3 Parametric study for reconstruction loss for AD and CN classification.

Study	Accuracy (%)	Recall (%)	Precision (%)
JACDNN without β	86.52	83.50	89.01
JACDNN with β	91.08	92.08	90.03

the percentage of the total number of subjects with pMCI or AD that are correctly classified by the network, given by

$$\eta_r = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{7}$$

where FN stands for false negative samples. It denotes the number of pMCI or AD subjects that are misclassified by the network. Precision is defined as the percentage of actual pMCI or AD subjects among all subjects classified as pMCI or AD, respectively, and is given by

$$\eta_p = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{8}$$

where FP stands for false positive samples. It denotes the number of CN or sMCI subjects that are incorrectly classified as AD or pMCI. Specificity is defined as the percentage of the total number of subjects with CN or sMCI that are correctly classified by the network, given by

$$\eta_s = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{9}$$

All experiments have been conducted using Python 3.6.9 and Tensorflow 2.0 in a Google Colab notebook with GPUs having 25 GB of memory. The results presented in this section have been obtained using fivefold cross validation. Each network presented here has been trained for 70 epochs with a batch size of 5. A learning rate of 0.01 with no momentum is used for stochastic gradient descent.

4.1 | Parametric study for Gaussian noise layer

In this section, a study is conducted to understand the impact of GN on the classification performance of JACDNN. Table 2 shows the performance results for networks with and without GN for classification of AD and CN subjects. It can be clearly observed that the network with GN performs better than the network without GN. The accuracy of the network is improved by approximately 2% while the sensitivity of the network is improved by more than 6%. The improvement in sensitivity indicates the better generalization of network on test samples. Therefore, it is observed that JACDNN with GN layer provides a superior performance with subjects with the disease as compared with JACDNN without GN layer.

4.2 | Parametric study for reconstruction loss

In this section, a study is conducted to understand the impact of reconstruction coefficient β on the classification performance of JACDNN. For this purpose, the performance of the network is observed with reconstruction loss and without reconstruction loss for classification of AD and CN subjects, as shown in Table 3. It can be clearly observed that the network with reconstruction loss performs better than the network without reconstruction loss. The accuracy of the classifier is improved by 4.5%, with regularization provided by reconstruction coefficient β, with approximately 9% improvement in the sensitivity of the network. Similar to other forms of regularization, β is set to a low value of 0.1.

It may be noted that for low values of β, JACDNN behaves like a feedforward neural network trained to perform classification. For high values of β, JACDNN focuses only on generating accurate reconstructions of the input samples based on each of the reconstruction sub-networks, thereby resulting in lower classification performance. Figure 2 shows the reconstructions obtained from each of the sub-networks in JACDNN trained for an AD subject. It can be observed that reconstructions obtained using earlier convolution layers are closer to the input MRI than the reconstructions obtained using deeper layers in the network. This is due to the fact that β determines the strength of regularization provided by the reconstruction loss, as shown in Equation 5. Due to this, representations associated with deeper layers in the network are less suitable for reconstruction and are more suitable for classification. This results in less effective reconstructions based on deeper layers in the network. The value of β is fixed at 0.1 for all experiments.

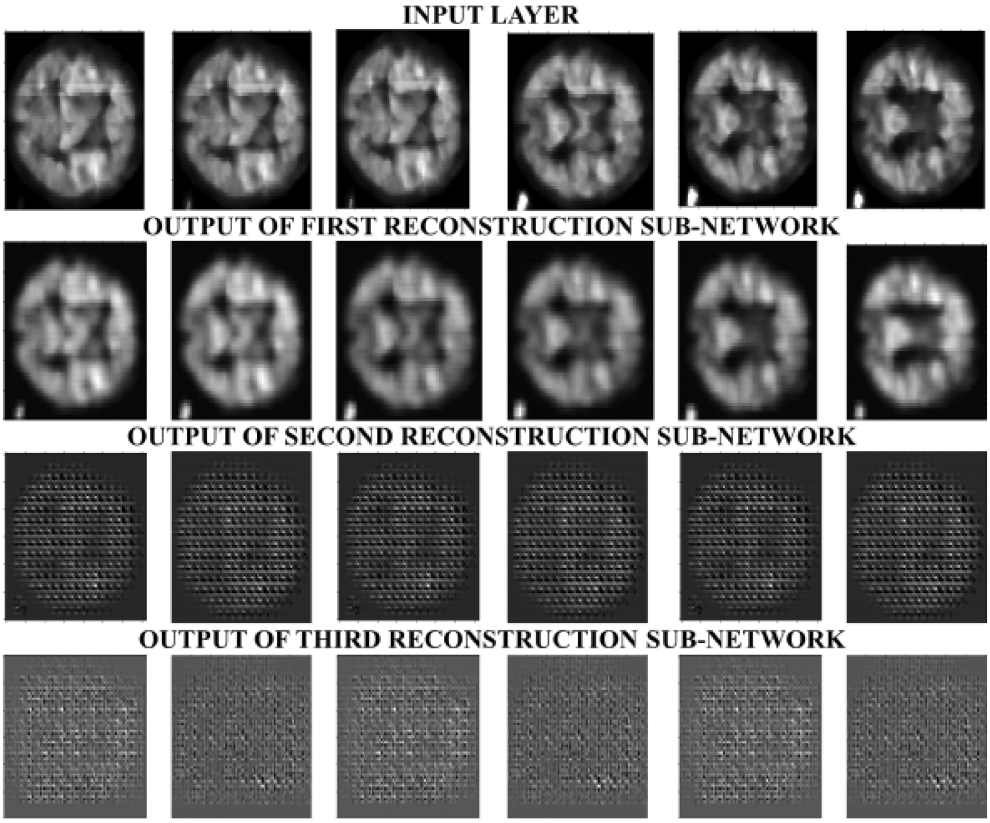

FIGURE 2 Output images obtained from each reconstruction sub-network for subject of AD.

TABLE 4 Classification results of all the experiments.

	AD versus CN	pMCI versus sMCI	AD versus sMCI	CN versus pMCI
GM features				
Accuracy	84.76 (±1.52)	72.77 (±1.8)	77.68 (±1.57)	77.69 (±1.59)
Recall	84.94 (±3.1)	63.13 (±6.9)	66.27 (±2.88)	68.23 (±4.79)
Precision	84.69 (±2.34)	70.73 (±3.6)	68.51 (±2.77)	75.60 (±3.59)
Specificity	84.55 (±3.16)	79.99 (±5.06)	83.74 (±2.14)	82.98 (±2.77)
GM + WM features				
Accuracy	91.08 (±0.38)	75.62 (±0.75)	82.17 (±0.79)	81.93 (±0.59)
Recall	92.98 (±2.63)	72.15 (±5.17)	73.72 (±7.70)	76.17 (±4.3)
Precision	90.03 (±1.04)	71.44 (±1.62)	75.31 (±4.25)	79.35 (±1.95)
Specificity	88.84 (±2.71)	78.23 (±2.99)	86.65 (±4.90)	85.97 (±2.49)

4.3 | Performance comparison using GM and WM features

In this section, the performance of JACDNN is evaluated for the four classification problems using two types of datasets obtained from preprocessed sMRI images. The first dataset includes only those features that are obtained from GM regions in the brain. The second dataset includes features that are obtained from both GM and WM regions in the brain, and hence will be denoted by GM + WM. Table 4 shows the results of performance evaluation of the proposed network for the four classification problems using two types of datasets. The table provides average and standard deviation for accuracy, recall, and precision across fivefold cross validation.

For the problem of AD versus CN, the accuracy of the network trained using GM + WM features is 7% better than the network trained using only GM features. For the classification problems involving pMCI subjects, the accuracy of the networks trained using GM + WM features is 3%–4% better than the networks that utilized only GM features. Similarly, the accuracy of the networks

TABLE 5 Performance comparison of JACDNN with recent approaches for the problems of AD versus CN, pMCI versus sMCI, AD versus sMCI, and CN versus pMCI.

Article	Dataset size (Class I/Class II)	ACC (%)	SEN (%)	SPE (%)
AD versus CN				
Cui and Liu[36]	198/229	91.33	86.87	95.20
Ganotra et al.[37]	137/162	89.26	85.29	92.59
Tong et al.[38]	198/231	89	84.9	92.6
Oh et al.[26]	198/230	86.60 (±3.66)	88.55	84.54
Sun et al.[39]	137/162	89.3	93.8	83.8
Proposed work	**321/324**	**91.08 (±0.38)**	**92.98 (±2.63)**	**88.84 (±2.71)**
pMCI versus sMCI				
Spasov et al.[24]	181/228	72	63	81
Beheshti et al.[10]	71/65	75	76.92	73.23
Cui and Liu[36]	167/236	71.71	65.27	76.27
Tong et al.[38]	167/238	70.4	66.5	73.1
Oh et al.[26]	166/101	73.95 (±4.82)	77.46	70.71
Sun et al.[39]	76/134	65.4	64.2	67.6
Proposed work	**228/171**	**75.62 (±0.75)**	**72.15 (±5.17)**	78.23 (±2.99)
AD versus sMCI				
Oh et al.[26]	198/101	75.06 (±3.86)	76.55	73.39
Proposed work	**324/171**	**82.17 (±0.79)**	**73.72**	**86.65 (±4.90)**
CN versus pMCI				
Oh et al.[26]	230/166	77.35	81.03	74.07
Proposed work	**324/228**	**81.93 (±0.59)**	**76.17 (±4.3)**	**85.97 (±2.49)**

Abbreviations: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity.

trained using GM + WM features is 5% better than the networks that employed only GM features for the problem of AD versus sMCI.

Using GM and WM features resulted in an improvement of 6%–9% in recall compared with using GM features alone for the four classification problems. For the problems of AD versus CN and pMCI versus sMCI, using GM + WM features allowed JACDNN to achieve improvements of 7%–8% and 8%–9%, respectively. Similarly, improvements of 1%–7% are observed in the precision of the classifiers that are trained using both GM and WM features.

To statistically validate the results of performance comparison, the performance of the networks trained using GM + WM and GM datasets have been statistically compared using one-way ANOVA test. For this purpose, the null hypothesis was that accuracy of the networks trained using GM + WM and GM features are not statistically different from each other. The null hypothesis is rejected with a p-value of 0.019, which is lower than 0.05. Therefore, the null hypothesis was rejected with a confidence interval of 95%. These results clearly show that the classifiers built using both GM and WM features perform better than the classifiers that utilize only GM features. This is in line with recent clinical studies that have highlighted a pathological relationship between atrophy in GM and WM regions.[33–35] Further, improvements in classification performance on pMCI versus sMCI and CN versus pMCI problems indicate that atrophy in WM regions might start in early stages of the disease.

4.4 | Performance comparison with recent approaches

In this section, the performance of JACDNN has been evaluated for the four classification problems of AD versus CN, pMCI versus sMCI, AD versus sMCI, and CN versus pMCI using preprocessed GM and WM data from the ADNI database. The results were compared with the performance of other recent approaches on these problems. All the approaches used for comparison have reported performance using fivefold cross validation except Spasov et al.[24] and Behesti et al.,[10] which reported results using 10-fold cross validation. For a fair comparison of results in different approaches, performance based only on sMRI data is included.

Table 5 shows the results of performance comparison for the four problems mentioned above. For the first experiment, that is, AD versus CN, the performance of JACDNN is compared with the performance of classifiers employed in recent works.[36,38,39] Cui et al.[36] reported an accuracy of 91.33%, which is almost equal to the proposed work; however, their work is concentrated on features extracted from longitudinal studies, which are although tedious but can provide with additional features. On the other hand, the proposed work is focused on baseline studies that are comparatively accurate and convenient to conduct and maintain. In terms of sensitivity, the performance of JACDNN is 6% higher than that of Cui et al.[36] Similarly, on comparing the performance of JACDNN with other works, the proposed architecture provides increased accuracy of 91.08%, which is 2% higher than Tong et al.,[38] Ganotra et al.,[37] and Sun et al.,[39] and 5% higher than Oh et al.[26] In terms of sensitivity, the proposed work offers 92.98%, which is 7%, 8%, and 4% higher than Ganotra et al.,[37] Tong et al.[38] and Oh et al., respectively.[26] Although Sun et al.[39] show 1% higher sensitivity in their results, this may be attributed by the novel penalty introduced by them for wrong classification.

For classification of pMCI and sMCI subjects, the performance of JACDNN is compared with sMRI-based results of various recent approaches.[10,24,26,36,38,39] Beheshti et al.[10] employed morphometry and SVM accompanied by t-test to report the accuracy of 75% that lie closest to accuracy provided in this work. However, their work employs two independent and mutually exclusive models for feature selection and classification compared with single architecture proposed by this work. Spasov et al.[24] have reported an accuracy of 72% on sMRI samples using separable and grouped convolutions. For the same classification task Sun et al.,[39] Oh et al.,[26] Cui et al.,[36] and Tong et al.[38] have reported accuracies of 65.4%, 73.95%, 71.71%, and 70.4%, respectively. In comparison with the abovementioned approaches, JACDNN provides an accuracy of 75.62%, which is 3%, 4%, 5%, 2%, and 10% higher than Spasov et al.,[24] Cui et al.,[36] Tong et al.,[38] Oh et al.,[26] and Sun et al.[39] Similarly, in terms of sensitivity, JACDNN obtained 9%, 6%, 7%, and 8% higher results than Spasov et al.,[24] Cui et al.,[36] Tong et al.,[38] and Sun et al.,[39] respectively. Beheshti et al.[10] reported higher sensitivity than the proposed work; however, their work is focused on extracting features from volume of interest in GM atrophy regions. These features are ranked as per their t-test score and used for classification. This method may show high affinity to the subject affected with the disease, and therefore attribute more towards sensitivity. Additionally, Beheshti et al.[10] used only seven samples for evaluation.

The performance of JACDNN on classification of AD versus sMCI and CN versus pMCI is compared with the performance of convolutional autoencoder proposed by Oh. et al.[26] Their work has reported the accuracy of 75.06% for AD and sMCI classification and 77.35% for CN and pMCI classification, respectively. The JACDNN provides improvement of 6%–7% and reports accuracy of 82.17% and 81.93% in respective cases. The joint classifier and autoencoder approach of the JACDNN helps it in this improvement in accuracy. In terms of sensitivity, the performance of Oh et al.[26] is 3%–5% better than the proposed model. This may contribute to the fact that Oh et al.[26] employed a multistage approach, which involved training an autoencoder and a classifier separately. The autoencoder is used to extract representations, and then these representations are used to train the classifier. Multistage approaches can be difficult to train as the two models, autoencoder and classifier, are trained separately, but the performance of the classifier strongly depends on the quality of representations obtained from the encoder. On the other hand, the proposed method is a simpler approach as it involves directly training a neural network using a loss function that combines reconstruction- and classification-related losses. Thus, the proposed method provides a simpler training approach with a small loss in performance.

4.4.1 | Parameter comparison with recent approaches

In this section, a study is conducted to compare the total parameters of neural networks reported in recent studies. Cui and Liu[36] reported a total of 273 292 parameters for their approach, which combines convolutional and recurrent neural network (RNN) architectures for longitudinal AD studies. They utilized convolutional neural networks (CNN) to capture spatial structural features and RNNs to extract longitudinal features from different time points. The incorporation of this two-stage feature learning process substantially increases the total number of parameters in the network. Similarly, Oh et al.[26] employed a two-stage process for AD detection, involving a convolutional autoencoder to learn encoded visual features in the first stage, followed by a CNN for classification in the second stage, resulting in a total of 340 000 parameters. Spasov et al.[24] reported the usage of 550 000 network parameters for classification tasks, combining imaging data with demographic, neuro-psychological, and genetic data.

Notably, not all methods in the field are solely based on neural networks. For instance, Tong et al.,[38] Beheshti et al.,[10] Sun et al.,[39] and Ganotra et al.[37] have presented

TABLE 6 Parameter comparison with recent studies.

Article	Parameters
Cui and Liu[36]	273 292
Oh et al.[26]	340 000
Spasov et al.[24]	550 000
Proposed work	162 404

state-of-the-art methods employing machine learning techniques such as support vector machines for classification. Therefore, a direct comparison of parameters may not be suitable in these cases.

Table 6 presents the number of parameters for neural networks based methods used for comparison in this article. The proposed method is based on a three-layer neural network architecture, where each layer is connected to a transpose convolution network for input reconstruction with 162 404 network parameters. The proposed approach distinguishes itself by not involving separate stages for feature selection and classification and operates as a cohesive learning-based system with automatic feature selection, eliminating the need for a multistage pipeline.

5 | CONCLUSION

In this work, we presented an approach to combine discriminating capabilities of a classifier with representation learning capacities of autoencoders in an integrated architecture to segregate pMCI subjects from sMCI subjects. The JACDNN architecture consists of single classifier sub-network with multiple autoencoder sub-networks. Each convolution layer of the classifier is equipped with an autoencoder, which trains to learn the representations for reconstruction of the input of the layer. These representations are learned by autoencoder while it regularizes the same layer. The classifier sub-network and autoencoder sub-networks are trained simultaneously on various classification problems associated with AD namely AD versus CN, pMCI versus sMCI, AD versus sMCI, and CN versus pMCI. The performance of the architecture is evaluated on two datasets based on GM and combination of GM and WM. It is observed that better classifier results are obtained on GM and WM features as compared with GM features alone. The performance of the JACDNN is compared with other existing approaches, which show that the proposed work provides better results. Furthermore, this work can be replicated for identification and classification of other neurodegenerative diseases and imaging modalities.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Alzheimer's Disease Neuroimaging Initiative at https://adni.loni.usc.edu/. These data were derived from the following resources available in the public domain:—ADNI, https://adni.loni.usc.edu/.

ORCID

Reema Ganotra https://orcid.org/0000-0001-9700-8065

REFERENCES

1. Patterson C. The state of the art of dementia research: new frontiers. *World Alzheimer Rep*. 2018;15:1473. doi:10.1016/j.jalz.2019.06.4115
2. Weller J, Budson A. Current understanding of alzheimer's disease diagnosis and treatment. *F1000Res*. 2018;7:1161. doi:10.12688/f1000research.14506.1
3. Parra MA. Overcoming barriers in cognitive assessment of alzheimer's disease. *Dement Neuropsychol*. 2014;8:95-98. doi:10.1590/S1980-57642014DN82000002
4. Association, A. 2020 alzheimer's disease facts and figures. *Alzheimers Dement*. 2020;16:391-460. doi:10.1002/alz.12068
5. Dubois B, Albert ML. Amnestic mci or prodromal alzheimer's disease? *Lancet Neurol*. 2004;3:246-248. doi:10.1016/S1474-4422(04)00710-0
6. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia: meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*. 2009;119:252-265. doi:10.1111/j.1600-0447.2008.01326.x
7. Petersen RC, Lopez O, Armstrong MJ, et al. Practice guideline update summary: mild cognitive impairment. *Neurology*. 2018;90:126-135. doi:10.1212/WNL.0000000000004826
8. Ward A, Tardiff S, Dye C, Arrighi HM. Rate of conversion from prodromal alzheimer's disease to alzheimer's dementia: a systematic review of the literature. *Dement Geriatr Cogn Dis Extra*. 2013;3:320-332. doi:10.1159/000354370
9. Tarawneh R, Holtzman DM. The clinical problem of symptomatic alzheimer disease and mild cognitive impairment. *Cold Spring Harb Perspect Med*. 2012;2:a006148. doi:10.1101/cshperspect.a006148
10. Beheshti I, Demirel H, Matsuda H. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput Biol Med*. 2017;83:109-119. doi:10.1016/j.compbiomed.2017.02.011
11. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage*. 2015;104:398-412. doi:10.1016/j.neuroimage.2014.10.002
12. Ahmed OB, Benois-Pineau J, Allard M, Catheline G, Amar CB. Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple

kernel learning. *Neurocomputing*. 2017;220:98-110. doi:10.1016/j.neucom.2016.08.041
13. Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*. 2015;151:139-150. doi:10.1016/j.neucom.2014.09.072
14. Samper-Gonzalez J, Burgos N, Bottani S, et al. Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. *NeuroImage*. 2018;183:504-521. doi:10.1016/j.neuroimage.2018.08.042
15. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Proces Syst*. 2012;60:1097-1105. doi:10.1145/3065386
16. Liu M, Zhang J, Adeli E, Shen D. Joint classification and regression via deep multi-task multi-channel learning for alzheimer's disease diagnosis. *IEEE Trans Biomed Eng*. 2018;66:1195-1206. doi:10.1109/TBME.2018.2869989
17. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. *OpenAI Blog*. 2019;1:9.
18. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the abide dataset. *Neuroimage Clin*. 2018;17:16-23. doi:10.1016/j.nicl.2017.08.017
19. Gao X, Niu S, Wei D, et al. Joint metric learning-based class-specific representation for image set. *IEEE Trans Neural Netw Learn Syst*. 2022;1-15. doi:10.1109/TNNLS.2022.3212703
20. Gao X, Sun Q, Xu H, Gao J. Sparse and collaborative representation based kernel pairwise linear regression for image set classification. *Expert Syst Appl*. 2020;140:112886. doi:10.1016/j.eswa.2019.112886
21. Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin*. 2019;21:1. doi:10.1016/j.nicl.2018.101645
22. Khvostikov A, Aderghal K, Benois-Pineau J, Krylov A, Catheline G. 3d cnn-based classification using smri and md-dti images for Alzheimer disease studies. 2018 arXiv preprint arXiv:1801.05968 doi:arXiv:1801.05968[cs. CV].
23. Wang SH, Phillips P, Sui Y, Liu B, Yang M, Cheng H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J Med Syst*. 2018;42:85. doi:10.1007/s10916-018-0932-7
24. Spasov S, Passamonti L, Duggento A, Li'o P, Toschi N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *NeuroImage*. 2019;189:276-287. doi:10.1016/j.neuroimage.2019.01.031
25. Bae J, Stocks J, Heywood A, et al. Transfer learning for predicting conversion from mild cognitive impairment to dementia of alzheimer's type based on a three-dimensional convolutional neural network. *Neurobiol Aging*. 2021;99:53-64. doi:10.1016/j.neurobiolaging.2020.12.005
26. Oh K, Chung YC, Kim K, Kim WS, Oh IS. Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. *Sci Rep*. 2019;9:1. doi:10.1038/s41598-019-54548-6
27. Suk HI, Lee SW, Shen D, Initiative ADN, et al. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med Image Anal*. 2017;37:101-113. doi:10.1016/j.media.2017.01.008
28. Wee CY, Liu C, Lee A, et al. Cortical graph neural network for ad and mci diagnosis and transfer learning across populations. *Neuroimage Clin*. 2019;23:101929. doi:10.1016/j.nicl.2019.101929
29. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685-691. doi:10.1002/jmri.21049
30. Ashburner J, Friston KJ. Unified segmentation. *NeuroImage*. 2005;26(3):839-851. doi:10.1016/j.neuroimage.2005.02.018
31. Ashburner J. A fast diffeomorphic image registration algorithm. *NeuroImage*. 2007;38(1):95-113. doi:10.1016/j.neuroimage.2007.07.007
32. Li Y, Xu R, Liu F. Whiteout: Gaussian adaptive regularization noise in deep neural networks. *Stat*. 2016;5:1050-1069. arXiv:1612.01490.
33. Acosta-Cabronero J, Williams G, Pengas G, Nestor P. Absolute diffusivities define the landscape of white matter degeneration in Alzheimer's disease. *Brain*. 2010;133:529-539. doi:10.1093/brain/awp257
34. Agosta F, Pievani M, Sala S, et al. White matter damage in Alzheimer disease and its relationship to gray matter atrophy. *Radiology*. 2011;258:853-863. doi:10.1148/radiol.10101284
35. Bozzali M, Falini A, Franceschi M, et al. White matter damage in Alzheimer's disease assessed in vivo using diffusion tensor magnetic resonance imaging. *J Neurol Neurosurg Psychiatry*. 2002;72:742-746. doi:10.1136/jnnp.72.6.742
36. Cui R, Liu M. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput Med Imaging Graph*. 2019;73:1-10. doi:10.1016/j.compmedimag.2019.01.005
37. Ganotra R, Dora S, Gupta S. Identifying brain regions contributing to alzheimer's disease using self regulating particle swarm optimization. *Int J Imaging Syst Technol*. 2020;1–12:106-117. doi:10.1002/ima.22458
38. Tong T, Wolz R, Gao Q, Guerrero R, Hajnal J, Rueckert D. Multiple instance learning for classification of dementia in brain MRI. *Med Image Anal*. 2014;18:808-818. doi:10.1016/j.media.2014.04.006
39. Sun Z, Qiao Y, Lelieveldt B, Staring M. Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification. *NeuroImage*. 2018;178:445-460. doi:10.1016/j.neuroimage.2018.05.051

How to cite this article: Ganotra R, Gupta S, Dora S. A joint autoencoder and classifier deep neural network for AD and MCI classification. *Int J Imaging Syst Technol*. 2024;34(2):e23054. doi:10.1002/ima.23054