



Diagnosis of Alzheimer's Disease Using Brain Network

Ramesh Kumar Lama and Goo-Rak Kwon*

The Alzheimer's Disease Neuroimaging Initiative, Department of Information and Communication Engineering, Chosun University, Gwangju, South Korea

Recent studies suggest the brain functional connectivity impairment is the early event occurred in case of Alzheimer's disease (AD) as well as mild cognitive impairment (MCI). We model the brain as a graph based network to study these impairment. In this paper, we present a new diagnosis approach using graph theory based features from functional magnetic resonance (fMR) images to discriminate AD, MCI, and healthy control (HC) subjects using different classification techniques. These techniques include linear support vector machine (LSVM), and regularized extreme learning machine (RELM). We used pairwise Pearson's correlation-based functional connectivity to construct the brain network. We compare the classification performance of brain network using Alzheimer's disease neuroimaging initiative (ADNI) datasets. Node2vec graph embedding approach is employed to convert graph features to feature vectors. Experimental results show that the SVM with LASSO feature selection method generates better classification accuracy compared to other classification technique.

OPEN ACCESS

Edited by:

Szilvia Anett Nagy,
University of Pécs, Hungary

Reviewed by:

Kewei Chen,
Banner Alzheimer's Institute,
United States
Alexandra L. Young,
University College London,
United Kingdom

*Correspondence:

Goo-Rak Kwon
grkwon@chosun.ac.kr

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 11 September 2020

Accepted: 06 January 2021

Published: 05 February 2021

Citation:

Lama RK and Kwon G-R (2021)
Diagnosis of Alzheimer's Disease
Using Brain Network.
Front. Neurosci. 15:605115.
doi: 10.3389/fnins.2021.605115

Keywords: Alzheimer's disease, brain network, node2vec, extreme learning machine, support vector machine

INTRODUCTION

Alzheimer's disease (AD), which causes majority of dementia is a progressive neurodegenerative disease (American Psychiatric Association, 1994; Liu F. et al., 2014; Schmitter et al., 2015; Alzheimer's association, 2016). The subtle AD neuropathological process begins years before the visible progressive cognitive impairment, which is trouble to remember and learn new information. Currently there is no cure and treatment to slow or stop its progression. Currently, more research works are focused toward earlier intervention of AD. Thus accurate diagnosis of disease at its early stage makes great significance in such scenario.

With the availability of recent neuroimaging technology, promising result is obtained in the early and accurate detection of AD (Hanyu et al., 2010; Górriz et al., 2011; Gray et al., 2012). The study of progression of disease and early detection is carried out by using different imaging models, such as electroencephalography (EEG) (Pfefferbaum et al., 2000), functional magnetic resonance imaging (fMRI) (Masliah et al., 1993), single-photon emission computed tomography (SPECT) (Chen et al., 2013) and positron emission tomography (PET) (Ly et al., 2014).

Similarly, structural magnetic resonance imaging (MRI) (Hanyu et al., 1999; Canu et al., 2010; Bendlin et al., 2012) is the most commonly used imaging system for study of AD. The feature extracted from MRI is typically gray matter volumes and measured as important biomarker for the study of neurodegeneration, alterations of hippocampal white matter pathways is often observed in AD (Delbeuck et al., 2003; Liu Y. et al., 2014). Several studies reveal the alterations in widely distributed functional and structural connectivity pairs are prevalent in AD and mild cognitive impairment (MCI) (Delbeuck et al., 2007; Acosta-Cabronero et al., 2012). Additionally, in recent studies, the resting-state functional magnetic resonance imaging (rs-fMRI) has been widely used

for the investigations of progression of AD (Stam et al., 2006; Sorg et al., 2007; Supekar et al., 2008; Doan et al., 2017). This imaging system evaluates the impulsive variabilities seen in the blood oxygenation level-dependent (BOLD) indications in various regions of the brain. Several studies are carried out based on aberrant regional spontaneous fluctuation of BOLD, functional connectivity and alteration in functional brain network. These studies are carried out in different networks, such as default mode network, somatomotor network, dorsal attention network, limbic network, and frontoparietal control network (Bullmore and Sporns, 2009). Thus, the graph theory based network analyses of human brain functional connectomes, provides better insights of the network structure to reveal abnormal patterns of organization of functional connectivity in AD infected brain (Rubinov and Sporns, 2010; Sporns, 2011).

Graph theory is a mathematical approach to study complex networks. Network is constructed of vertices which are interconnected by edges. Vertices in our case are brain regions. Graph theory is widely used as tool for identifying anatomically localized subnetworks associated with neuronal alterations in different neurodegenerative diseases (Bajo et al., 2010). In fMRI images, graph represents causal relations or correlations of different nodes in constructed networks. However, the brain network built by graph has non-Euclidian characteristics. Thus, applying machine learning techniques to analyze the brain networks is challenging. We use graph embedding to transform graphs to a vector or set of vectors to overcome this problem. Embedding captures the graph topology, vertex-vertex relationship, and other relevant graph information. In the current study, we used node2vec graph embedding technique to transform vertex and edge of brain network graph to feature vector. With the help of this model we have analyzed and classified the networks of brain from fMRI data into AD, MCI, and HC.

MATERIALS FOR THE STUDY

fMRI Dataset

In our study, we have used the dataset from Alzheimer's disease neuroimaging initiative database (ADNI)¹. The ADNI database was launched in 2004. The database consists of subjects of age ranging from 55–90 years. The goal of ADNI is to study the progression of the disease using different biomarkers. This includes clinical measures and assesses of the structures and functions of brain for the course of different disease states.

All participants were scanned using 3.0-Tesla Philips Achieva scanners at different centers. Same scanning protocol were followed for all participants and the set parameters were ratio of Repetition Time (TR) to Echo Time (TE) i.e., TR/TE = 3000/30 ms, 140 volumes, also voxel thickness as 3.3 mm, acquisition matrix size = 64 × 64, 48 slices, flip angle = 80°. Similarly, 3D T1-weighted images were collected using MPRAGE sense2 sequences with acquisition type 3D, field strength = 3 Tesla, flip angle

9.0 degree, pixel spacing X = 1.0547000169754028 mm; Pixel Spacing Y = 1.0547000169754028 mm, slice thickness = 1.2000000476837158 mm; echo time (TE) 2.859 ms, inversion time (TI) 0.0 ms, repetition time (TR) 6.6764 ms and weighting T1. We selected subjects as specified in Table.

Subjects

We selected 93 subjects from ADNI2 cohort. The purpose of ADNI2 is to examine how brain imaging and other biomarkers can be used to measure the progression of MCI and early AD. The ADNI selects and categorizes participants in specific group based on certain inclusion criteria. The criteria are well defined in². We selected the subjects according to availability of both MRI and fMRI data. Thus, the subjects with following demographic status as shown in **Table 1** with following average age, clinical dementia rating (CDR) and mini-mental state estimation (MMSE) out of all available data in ADNI2 cohort were selected in our study.

1. 31 HC subjects: 14 males, 17 females; age \pm SD = 73.9 \pm 5.4 years with the mini-mental state estimation (MMSE) score of 28.9 \pm 1.65 and the range was 24–30.
2. 31 MCI subjects: 17 males, 14 females; age \pm SD = 74.5 \pm 5.0 with the MMSE score of 27.5 \pm 2.02, and range was 22–30.
3. 31 AD subjects: 13 males, 18 females; age \pm SD = 72.7 \pm 7.0 with MMSE = 20.87 \pm 3.6, and the range was 14–26.

Data Preprocessing

We used data processing subordinate for the resting state fMRI via DPARSF³ (Chao-Gan and Yu-Feng, 2010) and the statistical parametric mapping platform via SPM8⁴ aimed at the preprocessing of rs-fMRI data. All the images initially obtained from scanner were in the format of digital imaging and communications in medicine (DICOM). We converted these images to neuroimaging informatics technology initiative (NIFTI) file format. Signal standardization and participant's adaptation to the noise while scanning each participant are carried out by discarding the first 10 time points for each participant. Next, we preformed preprocessing operation through following steps:

For slice-timing correction last slice was referred reference slice. Friston 24-parameter model with 6 parameters of head motion, 6 parameters of head motion from the previous time point, and 12 corresponding squared items were employed for

²<https://www.nia.nih.gov/alzheimers/clinical-trials/alzheimers-disease-neuroimaging-initiative-2-adni2>

³<http://www.restfmri.net>

⁴<http://www.fil.ion.ucl.ac.uk/spm>

TABLE 1 | Summary of subject's demographic status.

Number of subjects	HC (n = 31)	MCI (n = 31)	AD (n = 31)
	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	73.9 \pm 5.4	74.5 \pm 5.0	SD = 72.7 \pm 7.0
Global CDR	0.04 \pm 0.13	0.5 \pm 0.18	0.95 \pm 0.30
MMSE	28.9 \pm 1.65	27.5 \pm 2.02	20.87 \pm 3.6

¹<http://adni.loni.usc.edu/>

realignment for head movement compensation. Similarly, after the realignment, individual structural images (T1-weighted MPRAGE) were registered to the mean functional image. For the standardization of the rs-fMRI toward the original place was accomplished with the help of diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) as in Ashburner (2007) (resampling voxel size = 3 mm × 3 mm × 3 mm). A 6 mm full width at half-maximum (FWHM) Gaussian kernel spatial smoothing was employed for the smoothing. Next, we performed linear trend exclusion and also the temporal band pass filtering which ranges at (0.01 Hz < f < 0.08 Hz) on the time series of each voxel. Finally, cerebrospinal as well as white matter signals along with six head-motion parameters were regressed out to reduce the effects of nuisance signals.

Proposed Framework

This proposed method consists of the following four major functional steps as shown in **Figure 1**:

1. Construct a brain network using graph theory.
2. Convert graph to feature vector using node2vec graph embedding.
3. Reduce the features.
4. Perform the classification using regularized extreme learning machine (RELM) and linear support vector machine (LSVM).

Construction of Brain Networks

For the construction of network from fMR images, we first preprocessed the raw fMR data as described in data preprocessing section. Next, we used the automated anatomical labeling (AAL) atlas to identify the brain regions of interest (ROI). The whole image was divided in 116 regions with each hemisphere. Next, we calculate the average time series of each ROI for each subject by averaging their time series across the voxels within each ROI. For

each subject, a matrix of 130 rows and 116 columns was obtained. In the matrix, every row denotes the time series conforming to a give ROI, while information of total regions at a definite time point are stored at each column. The mean time series of each brain region were obtained for each individual by averaging the time series within the region. For $L_i = (l_i(1), l_i(2), \dots, l_i(n))$ and $L_j = (l_j(1), l_j(2), \dots, l_j(n))$ are two n length time series of brain region i and j , the Pearson's correlation (PC) between them can be calculated as

$$PC_{ij} = \frac{cov(L_i, L_j)}{\sigma_{L_i} \sigma_{L_j}} \quad (1)$$

Where $cov(L_i, L_j)$ is covariance of variables L_i and L_j . Similarly, σ_{L_i} and σ_{L_j} are standard deviation of variables L_i and L_j . This operation results into 116 × 116 correlation matrix which defines the relation amongst different regions of brain and matches to the functional connectivity network.

Graph-Embedding

Graphs are complex data structures, consisting a finite set of vertices and set of edges which connect a pair of nodes. One of the possible solutions to manipulate prevalent pattern recognition algorithms on graphs is embedding the graph into vector space. Indeed, graph embedding is a bridge between statistical pattern recognition and graph mining. We employ the node2vec (Grover and Leskovec, 2016) algorithm as graph embedding tool in this study. The node2vec algorithm aims to learn a vectorial representation of nodes in a graph by optimizing a neighborhood preserving objective. It extends the previous node embedding algorithm Deepwalk (Canu et al., 2010) and it is inspired from the state of art word embedding algorithm word2vec (Delbeuck et al., 2003).

In word2vec, given a set of sentences also known as corpus, the model learns word embedding by analyzing the context of each word in the body. The word2vec uses the neural network with one hidden layer to transform words into embedding vectors. This neural network is known as Skip-gram. This network is trained to

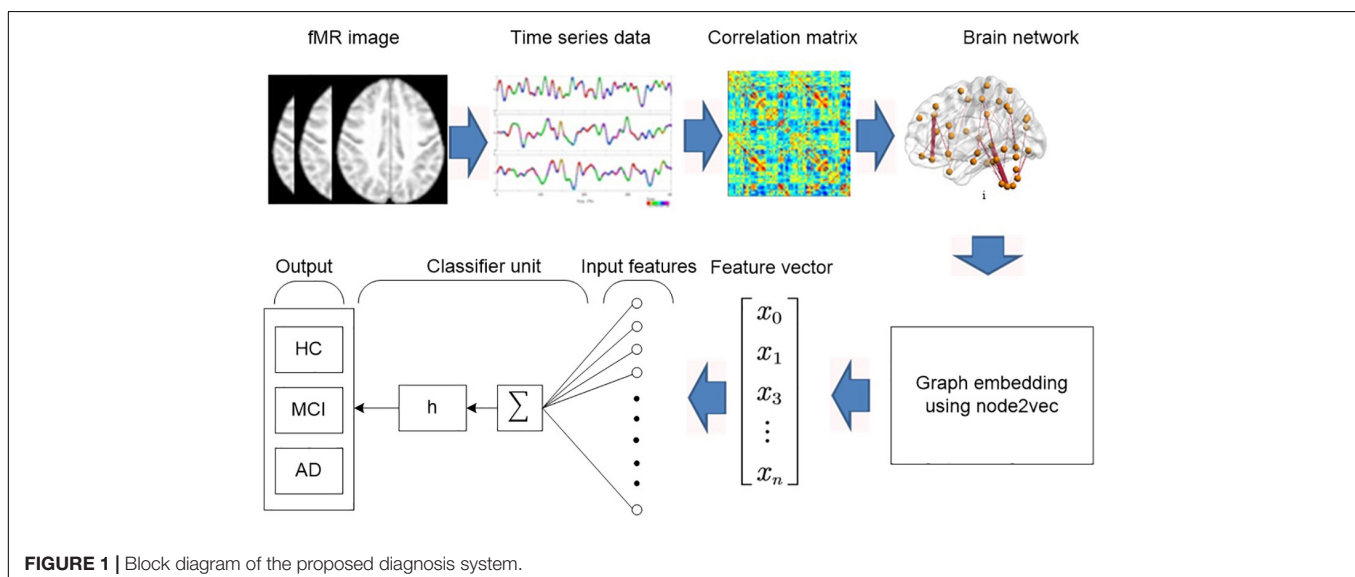


FIGURE 1 | Block diagram of the proposed diagnosis system.

predict the neighboring word in the sentence. It accepts the word at the input and is optimized such that it predicts the neighboring words in a sentence with high probability.

node2vec applies the same embedding approach to train and predict the neighborhood of a node in graph. However, word is replaced by the node and the bag of nodes is used instead of corpus. The sampling is used to generate this bag of nodes from a graph. Generally, the graph embedding consists of three steps:

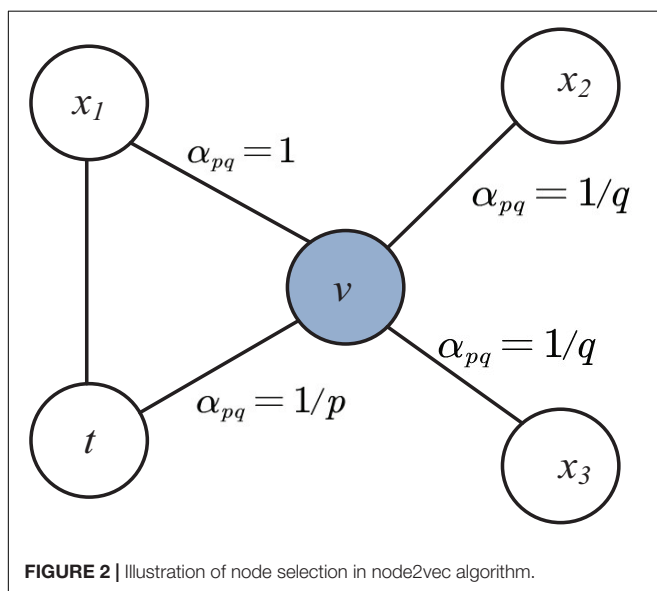
Sampling

A graph is sampled with random walks. This random walk results in bag of nodes of neighborhood from sampling. The bag of nodes acts as a collection of contexts for each node in the network. The innovation of node2vec with respect to Deepwalk is the use of flexible biased random walks on the network. In Deepwalk, random walk is obtained by a uniform random sampling over the linked nodes, while node2vec combine two different strategies for the network exploration: depth-first search (DFS) and breadth-first-search (BFS). For current random walk position at node v and traversed position at previous step at node t and neighboring nodes x_1 , x_2 and x_3 , the sampling of next node x is determined by evaluating the unnormalized transition probabilities π_{vx} on edge (t, v) with the static edge weight w_{vx} as shown in **Figure 2**. This unnormalized transition probability is estimated based on search bias α defined by two parameters p and q such that $\pi_{vx} = \alpha_{pq}(t, x) \cdot w_{vx}$ where.

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p}, & \text{if } d_{tx} = 0 \\ 1, & \text{if } d_{tx} = 1 \\ \frac{1}{q}, & \text{if } d_{tx} = 2 \end{cases} \quad (2)$$

Here d_{tx} denotes the shortest path distance between nodes t and x .

The parameter p determines the likelihood of sampling the node t again during random walk. When the value of p is high



revisit of the node possibility is low. Similarly the parameter q allows to different between local and global nodes. If $q > 1$, the random walk has the likelihood of sampling the nodes around the node v is high.

Training Skip-Gram

The bag of nodes generated from the random walk is fed into the skip-gram network. Each node is represented by a one-hot vector and maximizes the probability for predicting neighbor nodes. The one-hot vector has size same as the size of the set of unique words used in the text corpus. For each node only one dimension is equal to one and remaining are zeros. The position of dimension having one in vector defines the individual node.

Computing Embedding

The output of the hidden layer of the network is taken as the embedding of the graph.

Feature Reduction Techniques

Support Vector Machine-Recursive Feature Elimination (SVM-RFE)

Support vector machine-recursive feature elimination is a multivariate feature reduction algorithm is based on wrapper model. This method is recursive and in each of iteration of the RFE, LSVM model is trained. This method starts by constructing a model on the complete set of features and computing the importance score for each feature. The least important features are removed and the model is rebuilt and the importance scores are again computed. This recursive procedure is continued until all the features are eliminated. Then, the features are ranked according to the order of elimination. A detailed description of SVM-RFE procedure presented in a previous paper (Guyon et al., 2002). In this work, after applying SVM-RFE, the most significant training features that make the most of cross-validated accurateness are kept to train the classifiers.

Least Absolute Shrinkage and Selection Operator (LASSO)

Least absolute shrinkage and selection operator (Tibshirani, 1996) is a powerful method which is used to remove insignificant features. Two major tasks of this method are regularization and feature selection. This method minimizes residual sum of squares of the model using ordinary least square regression (OLS) by placing a constraint on the sum of the absolute values of the model parameters. LASSO computes model coefficients β by minimizing the following function:

$$\begin{aligned} &RSS_{LASSO}(\beta_i, \beta_0) \\ &= \arg \min_{\beta} \left[\sum_{i=1}^n (y_i - (\beta_i x_i + \beta_0))^2 + \alpha \sum_{j=1}^k |\beta_j| \right] \quad (3) \end{aligned}$$

Where, x_i is the graph embedded feature input data, a vector of k values at observation j , and n is the number of observations. y_i is the response at observation i . α is a non-negative user defined regularization parameter. This parameter controls the strength of penalty. When α is sufficiently large then coefficients are forced

to be zero which leads to produce few relevant features. If α approaches 0 the model becomes OLS with more relevant features (Hanyu et al., 2010).

Features Selection With Adaptive Structure Learning (FSASL)

Features selection with adaptive structure learning is an unsupervised method which performs data manifold learning and feature selection. This method first utilizes the adaptive structure of the data to construct the global learning and the local learning. Next, the significant features are selected by integrating both of them with $L_{2,1}$ -norm regularizer. This method utilizes the sparse reconstruction coefficients to extract the global structure of data for global learning. In sparse representation, each data sample x_i can be approximated as a linear combination of all the other samples, and the optimal sparse combination weight matrix.

For local learning, this method directly learns a Euclidean distance induced probabilistic neighborhood matrix (Du and Shen, 2015).

$$\begin{aligned} \min_{W,S,P} & (\|W^T X - W^T X S\|^2 + \alpha \|S\|_1) \\ & + \beta \sum_{i,j} (\|W^T x_i - W^T x_j\|^2 P_{ij} + \mu P_{ij}^2) + \gamma \|W\|_{21} \\ \text{s.t. } & S_{ii} = 0, P_{1n} = 1_n, P \geq 0, W^T X X^T X = I \end{aligned} \quad (4)$$

Where, α is used to balancing the sparsity and the reconstruction error, β and γ are regularization parameters for global and local structure learning in first and second group and the sparsity of feature selection matrix in the third group. Similarly, S is used to guide the search of relevant global feature and P defines the local neighborhood of data sample x_i .

Local Learning and Clustering Based Feature Selection (LLCFS)

LLCFS is clustering based feature selection method. This method learns the adaptive data structure with selected features by constructing the k-nearest neighbor graph in the weighted feature space (Zeng and Cheung, 2011). The joint clustering and feature weight learning is performed by solving the following problem.

$$\begin{aligned} \min_{Y, \{W^i, b^i\}_{i=1}^n, z} & \sum_{i=1}^n \sum_{c'=1}^c \left[\sum_{x_j \in N_{x_i}} \beta (Y_{ic'} - x_j^T W_c^i - b_c^i)^2 \right. \\ & \left. + (W_c^i)^T \text{diag}(z^{-1}) W_c^i \right] \\ \text{s.t. } & 1_d^T z = 1, z \geq 0 \end{aligned} \quad (5)$$

Where z the feature weight vector and N_{x_i} is the k-nearest neighbor of x_i based on z weighted features.

Pairwise Correlation Based Feature Selection (CFS)

CFS selects features based on the ranks attributes according to an empirical evaluation function based on correlations (Hall, 2000).

Subsets made of attribute vectors are evaluated by evaluation function, which are associated with the labels of class, however autonomous among each another. CFS accepts that unrelated structures express a low correspondence with the class and hence they are ought to be overlooked by the procedure. Alternatively, additional features must be studied, as they are typically hugely correlated with one or additional amount of other features.

Classification

Two of the prevalent machine-learning classification algorithms namely, LSVM, and RELM are studied in this article. The results acquired through the experiments of these classifiers show that RELM classifier performs better than others respective of the computation time required and accuracy value. Each of the methods is described in brief in the subsections below.

Support Vector Machine Classifier

Linear support vector machine (Cortes and Vapnik, 1995) is principally a supervised binary classifier that classifies separable and non-separable data. This type of classification is usually used in the field of neuroimaging and is deliberated as one of the finest machine-learning method in the domain of the neuroscience for past decades. It discovers the best hyperplane to split both classes which has optimum boundary from support vectors for the duration of the training. The classifier decides on the basis of the estimated hyperplane to test the new data points. For the patterns that are linearly separable, LSVM can be used. Alternatively, LSVM is not capable of guaranteeing improved performance in the complex circumstances with the patterns that are not separable. In such circumstances, Kernel trick is used to extend the LSVM. The input arrays of linear SVM are plotted to the space dimensions using the kernels. Both the linear as well as non-linear radial basis function (RBF) kernels are extensively trained using SVM kernels.

Extreme Learning Machine

ELM (Extreme Learning Machine) is single layer feedforward neural networks (Huang et al., 2006; Zhang et al., 2015). This neural network is implemented using Moore-Penrose generalized inverse to set its weights (Peng et al., 2013; Cao et al., 2016). Thus, this learning algorithm doesn't require iterative gradient-based backpropagation to tune the artificial hidden nodes. Thus this method is considered as effective solution with extremely reduced complexity (Cambria and Huang, 2013; Qureshi et al., 2016). ELM with L number of hidden nodes and $g(x)$ as activation function is expressed as

$$Y_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x) \beta_i \quad (6)$$

Where x is an input vector. $h_i(x)$ is the input to output node from hidden node output. $\beta = [\beta_1, \dots, \beta_2]^T$ is the weight matrix of i^{th} node. The input weight w_i and the hidden layer biases b_i are generated randomly before the training samples are acquired. Thus iterative back-propagation to tune these parameters is not

needed. Given N training samples $\{(x_j, t_j)\}_{j=1}^N$. The objective function of ELM is expressed as,

$$\|H(w_1, \dots, w_N, b_1, \dots, b_N)\hat{\beta} - T\| = \min_{\beta} \|H\hat{\beta} - T\| \quad (7)$$

with

$$H(w_1, \dots, w_N, b_1, \dots, b_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_L \cdot x_1 + b_L) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix} \quad (8)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix}$$

Here H represents the hidden layer output matrix and T represents output label of training data matrix. The output weight matrix β is calculated as

$$\beta = H^+ T \quad (9)$$

Here, H^+ represents the Moore-Penrose generalized inverse of the matrix H . Since ELM learning approach requires no backpropagation, this method is best suited for the binary and multiclass classification of big data and neuroimaging features. However the decrease in computation time comes with the expense of increase in the error in the output, which ultimately decreases the accuracy. Thus, a regularization term is added to improve generalization performance and make the solution more robust. The output weight of the regularized ELM can be expressed as

$$\beta = \left(\frac{I}{C} + H^T H \right)^{-1} H^T T \quad (10)$$

Performance Evaluation

We evaluated the performance using the SVM and RELM classifiers for each specific test including the binary and multiclass test. Confusion matrix is constructed to visualize the performance of the binary classifier in a form of a as shown in **Table 2**. Correct numbers of prediction of classifier are placed on the diagonal of the matrix. These components are further divided into true positive (TP), true negative (TN), which represent correctly identified controls. Similarly, the false positive (FP) and false negative (FN) represent the number of wrongly classified subjects.

TABLE 2 | Confusion matrix.

Accurate Class	Predicted Class	
	C1	C2
C1	TP	FN
C2	FP	TN

The proportion of subjects which are correctly classified by the classifier is expressed as the accuracy.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

However, for dataset with unbalanced class distribution accuracy may be a good performance metric. Thus two more performance are used. These metrics are known as sensitivity and specificity are used.

$$SEN = \frac{TP}{TP + FN} \quad (12)$$

$$SPE = \frac{TN}{TN + FP} \quad (13)$$

The sensitivity (SEN) measures the rate of true positives (TP) while the specificity (SPE) measures rate of true negatives (TN).

RESULTS

Demographic and Clinical Findings

We did not find a significant group difference in age in AD versus HC, AD versus MCI and MCI versus HC. However significant group difference was found in MMSE ($P < 0.01$) and CDR ($P < 0.01$) in all group combinations. The gender proportion on both AD and HC is male dominant. AD has 54.83% and HC has 45.16% male dominance. **Table 1** shows the detailed descriptions and analysis of these variables.

Classification Results

We have observed the performance of our proposed algorithm and compared it with that of the RELM classifier and LSVM classifier for respective test comprising the binary classification. The performance shown by the binary classifier is envisaged as a confusion matrix as presented in **Table 1**. Elements on the diagonal elements of the matrix specify the accurate estimations by the classifier. These elements are further divided as true positive (TP) and true negative (TN), which signifies appropriately recognized controls. Correspondingly, all the erroneously classified matters can be symbolized by false positive (FP) and false negative (FN). We evaluated the feature selection and classification algorithms on data set using a 10-fold cross validation (CV). First, we divided the subjects into 10 equally sized subsets: each of these subsets (folds), containing 10% of the subjects as test set and remaining 90% for training set. Then feature ranking was performed on the training sets. We used different algorithms to rank the features. Linear SVM and RELM classifiers were trained using these top-ranked features. For each training and test we performed separate feature selection to avoid the feature selection bias during 10-fold cross validation. We calculated cross validated average classification accuracy and standard deviation for specific feature using k -top most ranked features, where k ranges from 1 to 50. We repeatedly tested for 5 iterations and plotted the mean accuracy and standard deviation as shown in **Figure 3** for LASSO feature selection and RELM classifier.

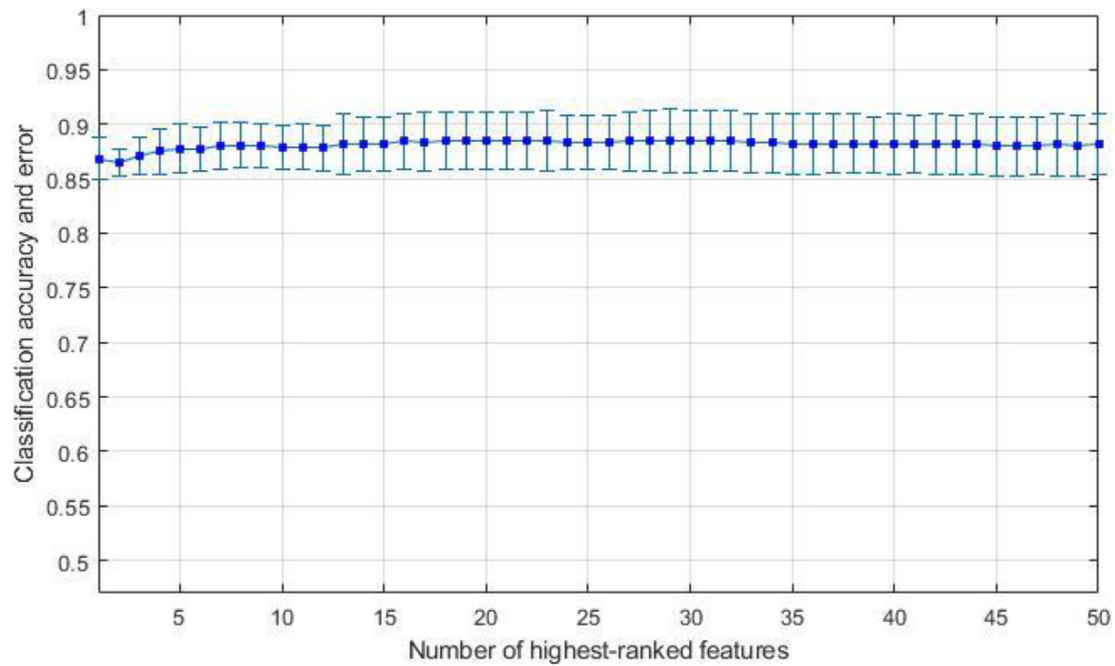


FIGURE 3 | Average accuracy and standard deviation for AD against HC using RELM classification method on reduced datasets using LASSO feature selection.

TABLE 3 | 10-fold cross-validation binary mean classification performance for AD against HC using RELM classifier using different feature selection methods.

Feature selection method	Performance metrics	Mean (%)	Standard deviation	Max (%)	Min (%)
LASSO	Accuracy	87.723	0.468	88.663	85.82
	Sensitivity	90.93	0.341	91.525	89.50
	Specificity	84.52	0.792	85.891	82.14
	F-measure		0.883		
FSASL	Accuracy	76.181	1.069	78.551	73.45
	Sensitivity	76.233	1.255	78.839	72.58
	Specificity	75.664	1.264	77.868	72.20
	F-measure		0.785		
LLCFS	Accuracy	75.737	1.004	78.690	71.64
	Sensitivity	74.205	1.069	77.031	70.11
	Specificity	77.881	1.378	81.036	73.64
	F-measure		0.817		
CFS	Accuracy	80.517	1.737	82.86	74.005
	Sensitivity	80.035	1.813	82.22	73.25
	Specificity	79.16	1.977	81.79	73.084
	F-measure		0.867		
SVM-RFE	Accuracy	68.57	1.186	70.474	65.60
	Sensitivity	75.99	1.676	78.832	71.715
	Specificity	60.92	1.301	63.426	57.34
	F-measure		0.6743		

TABLE 4 | 10-fold cross-validation binary mean classification performance for HC against MCI using RELM classifier using different feature selection methods.

Feature selection method	Performance metrics	Mean (%)	Standard deviation	Max (%)	Min (%)
LASSO	Accuracy	96.11	0.859	96.88	91.33
	Sensitivity	95.03	1.080	95.93	89.84
	Specificity	97.18	0.798	97.84	92.93
	F-measure		0.973		
FSASL	Accuracy	85.85	0.9129	87.503	80.88
	Sensitivity	79.27	0.986	81.484	76.01
	Specificity	92.03	1.4433	93.308	85.40
	F-measure		0.937		
LLCFS	Accuracy	82.29	0.624	83.39	80.77
	Sensitivity	77.54	0.73	78.566	75.408
	Specificity	86.81	1.081	88.41	83.74
	F-measure		0.85		
CFS	Accuracy	80.67	1.68	90.427	74.48
	Sensitivity	88.43	2.328	74.968	80.49
	Specificity	72.38	1.3677	74.486	68.58
	F-measure		0.795		
SVM-RFE	Accuracy	80.20	0.920	82.001	77.89
	Sensitivity	84.00	1.207	85.99	79.86
	Specificity	75.91	1.251	77.806	73.29
	F-measure		0.815		

Finally, we calculated the mean accuracy and standard deviation of highest ranked features for different feature selection and classification methods as depicted in **Tables**

3–8 and the bold values in each table indicate the maximum value of accuracy, sensitivity and specificity. Maximum and minimum value of accuracy, sensitivity

TABLE 5 | 10-fold cross-validation binary mean classification performance for MCI against AD using RELM classifier using different feature selection methods.

Feature selection method	Performance metrics	Mean (%)	Standard deviation	Max (%)	Min (%)
LASSO	Accuracy	93.86	0.766	94.90	89.128
	Sensitivity	91.93	0.757	93.67	89.836
	Specificity	95.92	1.204	96.61	88.580
	F-measure		0.968		
FSASL	Accuracy	85.358	1.030	86.76	80.29
	Sensitivity	85.088	0.951	86.61	80.47
	Specificity	85.201	1.4227	86.829	79.27
	F-measure		0.879		
LLCFS	Accuracy	90.32	1.06316	91.88	86.50
	Sensitivity	93.33	2.0782	95.63	87.43
	Specificity	87.49	0.9471	89.56	85.75
	F-measure		0.895		
CFS	Accuracy	79.13	1.2768	81.595	75.10
	Sensitivity	83.59	1.2281	85.60	78.94
	Specificity	74.83	1.7084	78.927	70.54
	F-measure		0.795		
SVMRFE	Accuracy	77.5974	0.8177	78.93	74.98
	Sensitivity	76.802	1.1299	79.225	73.26
	Specificity	78.182	0.8359	79.289	75.44
	F-measure		0.8169		

TABLE 7 | 10-fold cross-validation binary mean classification performance for HC against MCI using LSVM classifier using different feature selection methods.

Feature selection method	Performance metrics	Mean (%)	Standard deviation	Max (%)	Min (%)
LASSO	Accuracy	98.91	0.456	99.25	95.82
	Sensitivity	99.68	0.56	100.0	95.48
	Specificity	98.11	0.46	98.51	96.00
	F-measure		0.9856		
FSASL	Accuracy	81.28	1.010	83.01	77.64
	Sensitivity	84.61	1.389	86.62	79.81
	Specificity	77.92	1.121	79.81	73.038
	F-measure		0.833		
LLCFS	Accuracy	76.27	0.631	78.31	74.70
	Sensitivity	71.08	1.388	75.37	68.23
	Specificity	81.40	1.005	82.69	76.80
	F-measure		0.800		
CFS	Accuracy	86.16	2.25	88.96	80.47
	Sensitivity	92.28	2.33	95.11	86.14
	Specificity	79.72	2.375	82.75	73.88
	F-measure		0.8517		
SVMRFE	Accuracy	71.92	0.832	74.43	69.93
	Sensitivity	66.90	1.493	71.14	63.53
	Specificity	76.68	1.187	79.61	72.88
	F-measure		0.7762		

TABLE 6 | 10-fold cross-validation binary mean classification performance for AD against HC using LSVM classifier using different feature selection methods.

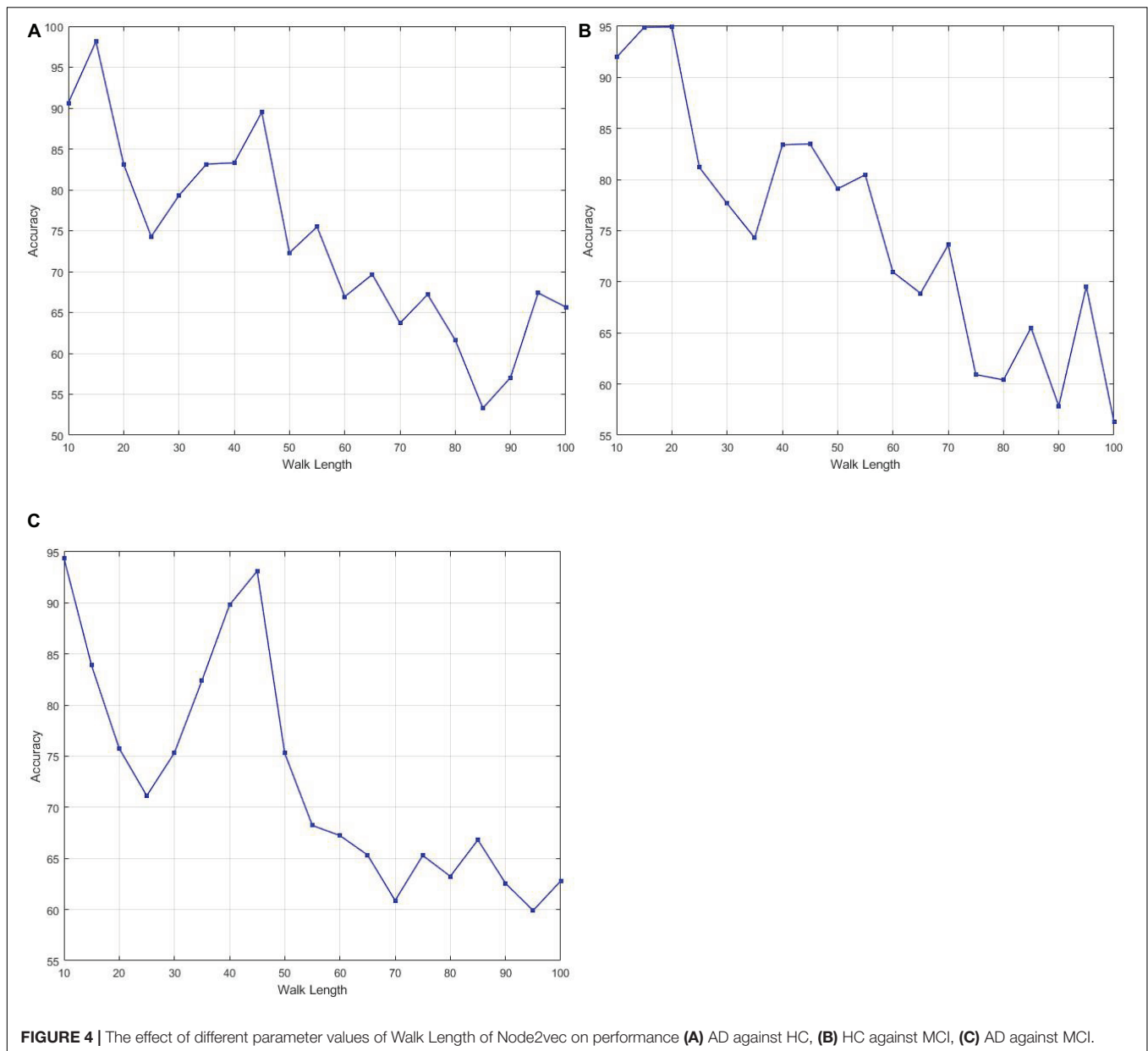
Feature selection method	Performance metrics	Mean (%)	Standard deviation	Max (%)	Min (%)
LASSO	Accuracy	90.63	0.515	91.51	88.52
	Sensitivity	87.044	0.585	88.03	85.44
	Specificity	94.315	0.671	95.35	90.95
	F-measure		0.958		
FSASL	Accuracy	82.895	1.4020	85.60	80.19
	Sensitivity	78.206	1.5118	81.99	75.21
	Specificity	87.712	1.7666	90.02	82.85
	F-measure		0.8360		
LLCFS	Accuracy	81.19	1.438	83.37	77.68
	Sensitivity	85.15	2.087	88.068	80.49
	Specificity	76.39	1.25	78.983	73.65
	F-measure		0.8095		
CFS	Accuracy	88.37	1.78	91.18	83.25
	Sensitivity	87.95	1.72	91.98	84.52
	Specificity	88.79	2.17	90.71	81.28
	F-measure		0.903		
SVMRFE	Accuracy	65.99	1.48	68.51	62.05
	Sensitivity	65.03	1.41	67.73	61.46
	Specificity	67.40	2.327	70.53	61.27
	F-measure		0.671		

TABLE 8 | 10-fold cross-validation binary mean classification performance for MCI against AD using LSVM classifier using different feature selection methods.

Feature selection method	Performance metrics	Mean (%)	Standard deviation	Max (%)	Min (%)
LASSO	Accuracy	97.80	0.9862	98.32	91.99
	Sensitivity	97.62	1.0065	97.92	91.73
	Specificity	97.74	1.0720	98.5	92.07
	F-measure		0.98		
FSASL	Accuracy	83.71	0.90	85.00	78.15
	Sensitivity	90.63	1.57	92.23	84.12
	Specificity	77.10	1.098	79.74	72.55
	F-measure		0.838		
LLCFS	Accuracy	90.04	1.43	92.02	86.07
	Sensitivity	90.23	1.43	91.94	86.74
	Specificity	90.09	1.77	92.52	84.62
	F-measure		0.903		
CFS	Accuracy	84.41	1.95	86.81	79.18
	Sensitivity	90.37	2.14	93.26	83.73
	Specificity	78.10	1.90	80.79	72.84
	F-measure		0.83		
SVMRFE	Accuracy	82.87	0.903	83.96	78.832
	Sensitivity	81.68	1.22	84.34	77.10
	Specificity	84.11	0.94	85.21	80.72
	F-measure		0.854		

and specificity are calculated amongst corresponding values estimated for highest ranked features as shown in Figure 3.

Tables 3–5 show the binary classification results using RELM classifier with five different feature selections. Results obtained through the feature selection methods are compared in



regards to the performance metrics such as accuracy, sensitivity specificity and f-measure. **Table 3** summarizes the AD versus HC classification. The LASSO feature selection method outperforms all other methods consider with the highest mean accuracy of 87.72%, mean specificity of 90.93% and mean sensitivity of 84.52%. Additionally, the standard deviation of LASSO is 0.4 which is less than less than 1. Similarly, the classification results of AD versus MCI and NC versus MCI using RELM are shown in **Tables 4, 5**. As shown in **Table 4**, the highest mean accuracy is 96.11 (± 0.859) for HC against MCI classification and 93.86 (± 0.766) for MCI against AD classification. The standard deviation is less than 1 in both mean classifications. Additionally the F-score is high in all three classifications (0.883 for HC against AD, 0.973 for HC against MCI, 0.968 for AD against MCI)

using LASSO feature selection method compared to other feature selection methods. The value of standard deviation less than one indicates that the data points of accuracy estimated tend to be close to the mean. Hence from the result it is very evident that the less inflated accuracy can be obtained using LASSO. Similarly, the high F-score indicates precision of classification is high compared to other feature selection methods.

Similarly, the comparison of classification of HC, MCI and AD using LSVM classifier with different feature selection methods are shown in **Tables 6–8**. Similar to RELM, the highest performance result in terms of mean accuracy, specificity, sensitivity and F-score was obtained by using LASSO for all three classification tests. As shown in **Table 6**, we obtained the accuracy of 90.63% specificity of 94.315% and sensitivity

of 87.95% and F-score of 0.958 for AD against HC. In **Table 7** the highest mean accuracy, specificity, sensitivity and F-score are obtained as 98.9, 99.68, 98.11, and 0.9856% for HC against MCI classification. Similarly, **Table 8** shows the classification performance of AD against MCI. The highest mean accuracy, specificity, sensitivity and F-score are 97.81, 97.62, 97.74, and 0.98%.

From all these results, it is clearly evident that the use of LASSO as feature selection method is ideal choice for the classification using RELM and LSVM classifiers for the graph embedded data.

From **Tables 3–5** the highest classification accuracies of RELM classifier using LASSO feature selection for AD against HC, HC against MCI and MCI against AD are 87.723% (± 0.468), 96.11% (± 0.859), and 93.86% (± 0.766). Similarly, from **Tables 6–8** the highest classification accuracies of RELM classifier using LASSO feature selection for AD against HC, HC against MCI and MCI against AD are 90.63% (± 0.515), 98.91% (± 0.456), and 97.80% (± 0.9862).

Now, the comparison of performance between two classifiers shows that the SVM can classify the given dataset more accurately with the highest mean accuracy for all three binary classifications. However, the small standard deviation of the classification HC against MCI and MCI against AD suggest that the classification accuracy values are less inflated in RELM as compared to LSVM.

The number of hidden layer nodes influences the performance of the RELM classifier. In our experiment, we found 1000 number of hidden layer generated the best performance in terms of accuracy. Similarly, for SVM we set the default parameter defined for the MATLAB library. We performed the classification by varying different parameters on node2vec graph embedding. **Figure 4** shows the effect of different parameters of node2vec on the performance of RELM classifier. We varied the walk length of node2vec from 10 to 100. In all experiments, increased value of walk length decreases the performance of classifier. For

this experiment, we fixed two other parameters, dimension and number of walks to 32 and 200. Similarly, we set the parameters p and q to correspond localized random walks. With the smaller value of p and larger value of q , the random walk is easy to sample to the high-order -order proximity. Thus, we selected p and q randomly and performed graph embedding with $p = 0.1$ and $q = 1.6$.

DISCUSSION

Several studies based on rs-fMRI have been carried out for the classification of AD and MCI from HC subjects. Binary classification in combination of different classifier with different feature measure reported the accuracy ranging from 85 to 95% for AD against HC and 62.90 to 72.58% to and MCI against HC as shown in **Tables 9, 10**. These studies used the same MCI and HC subjects from the ADNI2 cohort. One can clearly notice that the number of subjects directly influences the accuracy. As the number of subjects increase the accuracy is decreased. As reported in previous section the highest accuracy for the classification of AD from is obtained in proposed work is 90.63% using the combination of LASSO and LSVM. If we compare the results for MCI against HC, the results obtained in current study outperform all the state of art methods. However, it is not fair to compare performance with other studies directly because each work employ different datasets, preprocessing pipelines, feature measures, and classifiers. Majority of works including (Eavani et al., 2013; Leonardi et al., 2013; Wu et al., 2013; Wee et al., 2014; Suk et al., 2016) have used subjects less than or nearly equal to 30 in each subject class. The main reason behind small number of dataset is the availability of fMRI data in ADNI2 cohort. All of these studies performed classification and made conclusion. Likewise, we also conducted our study using ADNI2 cohort with nearly equal number of subjects with

TABLE 9 | Comparison of performance of binary classification AD against HC with state of the art methods using rs-fMRI.

Dataset	Feature measures	Classifier	Accuracy (%)	Reference
AD:77, HC: 173	Combination of FC matrices, FC dynamics, ALFF	AUC	85	de Vos et al., 2018
AD: 12, HC: 12	Difference between DMN and SN map	LDA	92	Zhou et al., 2010
AD: 34, HC: 45	Graph measures	Naïve Bayes	93.3	Khazaei et al., 2017
AD: 15, HC: 16	Averaged voxel intensities of core regions in resting state networks: DMN, DAN, VAN	Multivariate ROC	95	Wu et al., 2013

TABLE 10 | Comparison of performance of binary classification MCI against HC with state of the art methods using rs-fMRI.

Dataset	Feature measures	Classifier	Accuracy (%)	References
MCI: 31, HC: 31	Functional activity co-variations of ROIs	SVM	62.90	Eavani et al., 2013
MCI: 31, HC: 31	Group sparse representation	SVM	66.13	Wee et al., 2014
MCI: 31, HC: 31	SDFN	SVM	70.97	Leonardi et al., 2013
MCI: 31, HC: 31	Deep auto encoder and HMM	SVM	72.58	Suk et al., 2016

FC, functional connectivity; ALFF, amplitude of low-frequency fluctuation; AUC, area under the curve; DMN, default mode network; SN, salience network; LDA, linear discriminant analysis; ROC, receiver operating characteristic; ROI, region of interest; AAL, automated anatomical labeling; DMN, default mode network; SDFN, sliding window-based dynamic functional network; HMM, hidden markov model.

previous studies and the cross validation was also done using these dataset.

Mild cognitive impairment is a transitional stage between the healthy non dementia and dementia stage². This stage is further divided into early MCI (EMCI) and late MCI (LMCI), according to extent of episodic memory impairment. The risk conversion from MCI to AD is higher in LMCI than in EMCI. In this study, we included only EMCI subjects in MCI group. The MCI converted and non-converted to is classified according to CDR and MMSE score. MCI subjects whose CDR undergoes change from 0.5 to 1 and MMSE score goes below 26 in subsequent visits are considered to have fulfilled the criteria to be MCI converted. In our study majority of subjects fulfill to be non-converted MCI. Only few subjects either have changed CDR score or MMSE score during the visits in the interval of 3, 6, 12, and 18 months. Additionally, none of the MCI subjects are recorded in the list of AD subjects.

Limitations

While this study is focused on the stage diagnosis of AD progression using fMRI alone using ADNI2 cohort, the major limitation of this study is the limited sample size of ADNI2 (31 AD, 31 MCI, and 31 HC). In this context, the entire population is not represented adequately with the dataset we used. Thus, we cannot guarantee the generalization of our results to other groups.

CONCLUSION

It is widely accepted that the early diagnosis of AD and MCI plays an import role to take preventive action and to delay the future progression of AD. Thus the accurate classification task of different stages of AD progression is essential. In this study, we demonstrated graph based features from functional magnetic resonance (fMR) images can be used for the classification of AD and MCI from HC. Additionally, we used multiple feature selection techniques to cope with the smaller number of subjects with larger number of feature representations. The appropriate amount of features is extracted from standard Alzheimer's disease Neuroimaging Initiative cohort that lead to maximal classification accuracies as compared to all other recent researches. Among different feature section methods LASSO together with LSVM on graph based features significantly improved the classification accuracy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Alzheimer's Disease Neuroimaging Initiative

(ADNI). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RL has generated the idea and conducted the experiments. G-RK has reviewed idea and final verification of results. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. NRF-2019R1A4A1029769).

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Correspondence should be addressed to GR-K, grkwon@chosun.ac.kr.

REFERENCES

- Acosta-Cabrero, J., Alley, S., Williams, G. B., Pengas, G., and Nestor, P. J. (2012). Diffusion tensor metrics as biomarkers in Alzheimer's disease. *PLoS One* 7:e49072. doi: 10.1371/journal.pone.0049072
- Alzheimer's association (2016). 2016 Alzheimer's disease facts and figures. *Alzheimers Dement.* 12, 459–509. doi: 10.1016/j.jalz.2016.03.001
- American Psychiatric Association (1994). "Task force on DSM-IV," in *Diagnostic and Statistical Manual of Mental Disorders, DSM-IV*, 4th Edn, Vol. xxv, (Washington, DC: American Psychiatric Association).
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113. doi: 10.1016/j.neuroimage.2007.07.007
- Bajo, R., Maestú, F., Nevado, A., Sancho, M., Gutiérrez, R., Campo, P., et al. (2010). Functional connectivity in mild cognitive impairment during a memory task: implications for the disconnection hypothesis. *J. Alzheimers Dis.* 22, 183–193. doi: 10.3233/jad-2010-100177
- Bendlin, B. B., Carlsson, C. M., Johnson, S. C., Zetterberg, H., Blennow, K., and Willette, A. (2012). CSF T-Tau/A β 42 predicts white matter microstructure in healthy adults at risk for Alzheimer's disease. *PLoS One* 7:e37720. doi: 10.1371/journal.pone.0037720
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575
- Cambria, E., and Huang, G.-B. (2013). Extreme learning machines. *IEEE Intell. Syst.* 28, 2–31.
- Canu, E., McLaren, D. G., Fitzgerald, M. E., Bendlin, B. B., Zoccali, G., and Alessandrini, F. (2010). Microstructural diffusion changes are independent of macrostructural volume loss in moderate to severe Alzheimer's disease. *J. Alzheimers Dis.* 19, 963–976. doi: 10.3233/jad-2010-1295
- Cao, J., Zhang, K., Luo, M., Yin, C., and Lai, X. (2016). Extreme learning machine and adaptive sparse representation for image classification. *Neural Netw.* 81, 91–102. doi: 10.1016/j.neunet.2016.06.001
- Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: a MATLAB toolbox for "Pipeline" data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013
- Chen, Y. J., Deutsch, G., Satya, R., Liu, H. G., and Mountz, J. M. (2013). A semi-quantitative method for correlating brain disease groups with normal controls using SPECT: Alzheimer's disease versus vascular dementia. *Comput. Med. Imaging Graph.* 37, 40–47. doi: 10.1016/j.compmedimag.2012.11.001
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297.
- de Vos, F., Koini, M., Schouten, T. M., Seiler, S., van der Grond, J., Lechner, A., et al. (2018). A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease. *Neuroimage* 167, 62–72. doi: 10.1016/j.neuroimage.2017.11.025
- Delbeuck, X., Collette, F., and Van der Linden, M. (2007). Is Alzheimer's disease a disconnection syndrome? Evidence from a crossmodal audio-visual illusory experiment. *Neuropsychologia* 45, 3315–3323.
- Delbeuck, X., Van der Linden, M., and Collette, F. (2003). Alzheimer's disease as a disconnection syndrome? *Neuropsychol. Rev.* 13, 79–92.
- Doan, N. T., Engvig, A., Persson, K., Alnæs, D., Kaufmann, T., Rokicki, J., et al. (2017). Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum. *Sci. Rep.* 7:45131.
- Du, L., and Shen, Y. D. (2015). "Unsupervised feature selection with adaptive structure learning," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 209–218.
- Eavani, H., Satterthwaite, T. D., Gur, T. E., Gur, R. C., and Davatzikos, C. (2013). Unsupervised learning of functional network dynamics in resting state fMRI. *Inf. Process. Med. Imaging* 23, 426–437. doi: 10.1007/978-3-642-38868-2_36
- Górriz, J. M., Segovia, F., Ramírez, J., Lassl, A., and Gonzalez, D. S. (2011). GMM based SPECT image classification for the diagnosis of Alzheimer's disease. *Appl. Soft. Comput.* 11, 2313–2325. doi: 10.1016/j.asoc.2010.08.012
- Gray, K. R., Wolz, R., Heckemann, R. A., Aljabar, P., Hammers, A., and Rueckert, D. (2012). Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *Neuroimage* 60, 221–229. doi: 10.1016/j.neuroimage.2011.12.071
- Grover, A., and Leskovec, J. (2016). "Node2Vec: scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, NY: ACM).
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422.
- Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. *Proc. 17th Int'l Conf. Machine Learning* 359–366.
- Hanyu, H., Asano, T., Sakurai, H., Imon, Y., Iwamoto, T., Takasaki, M., et al. (1999). Diffusion-weighted and magnetization transfer imaging of the corpus callosum in Alzheimer's disease. *J. Neurol. Sci.* 167, 37–44. doi: 10.1016/s0022-510x(99)00135-5
- Hanyu, H., Sato, T., Hirao, K., Kanetaka, H., Iwamoto, T., and Koizumi, K. (2010). The progression of cognitive deterioration and regional cerebral blood flow patterns in Alzheimer's disease: a longitudinal SPECT study. *J. Neurol. Sci.* 290, 96–101. doi: 10.1016/j.jns.2009.10.022
- Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70, 489–501. doi: 10.1016/j.neucom.2005.12.126
- Khazaei, A., Ebrahimzadeh, A., and Babajani-Feremi, A. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behav. Brain Res.* 322, 339–350. doi: 10.1016/j.bbr.2016.06.043
- Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.-M., Schlupe, M., et al. (2013). Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *Neuroimage* 83, 937–950. doi: 10.1016/j.neuroimage.2013.07.019
- Liu, F., Zhou, L., Shen, C., and Yin, J. (2014). Multiple kernel learning in the primal for multi-modal Alzheimer's disease classification. *IEEE J. Biomed. Health Inform.* 18, 984–990. doi: 10.1109/jbhi.2013.2285378
- Liu, Y., Yu, C., Zhang, X., Liu, J., Duan, Y., Alexander-Bloch, A. F., et al. (2014). Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. *Cereb. Cortex* 24, 1422–1435. doi: 10.1093/cercor/bhs410
- Ly, M., Canu, E., Xu, G., Oh, J., McLaren, D. G., Dowling, N. M., et al. (2014). Midlife measurements of white matter microstructure predict subsequent regional white matter atrophy in healthy adults. *Hum. Brain Mapp.* 35, 2044–2054. doi: 10.1002/hbm.22311
- Masliah, E., Mallory, M., Hansen, L., DeTeresa, D., and Terry, R. (1993). Quantitative synaptic alterations in the human neocortex during normal aging. *Neurology* 43, 192–197. doi: 10.1212/wnl.43.1_part_1.192
- Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput. Biol.* 4:e1000100. doi: 10.1371/journal.pcbi.1000100
- Peng, X., Lin, P., Zhang, T., and Wang, J. (2013). Extreme learning machine-based classification of ADHD using brain structural MRI data. *PLoS One* 8:e79476. doi: 10.1371/journal.pone.0079476
- Pfefferbaum, A., Sullivan, E., Hedehus, M., Lim, K., Adalsteinsson, E., and Moseley, M. (2000). Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. *Magn. Reson. Med.* 44, 259–268. doi: 10.1002/1522-2594(200008)44:2<259::aid-mrm13>3.0.co;2-6
- Qureshi, M. N. I., Min, B., Jo, H. J., and Lee, B. (2016). Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI Study. *PLoS One* 11:e0160697. doi: 10.1371/journal.pone.0160697
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Schmitter, D., Roche, A., Maréchal, B., Ribes, D., Abdulkadir, A., Bach-Cuadra, M., et al. (2015). An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *NeuroImage* 7, 7–17.
- Sorg, C., Riedel, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läger, L., et al. (2007). Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *Proc. Nat. Acad. Sci. U.S.A.* 104, 18760–18765. doi: 10.1073/pnas.0708803104

- Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125. doi: 10.1111/j.1749-6632.2010.05888.x
- Stam, C. J., Jones, B. F., Manshandena, I., van Cappellen van Walsum, A. M., Montez, T., Verbunt, J. P. A., et al. (2006). Magnetoencephalographic evaluation of resting-state functional connectivity in Alzheimer's disease. *Neuroimage* 32, 1335–1344. doi: 10.1016/j.neuroimage.2006.05.033
- Suk, H. I., Wee, C. Y., Lee, S. W., and Shen, D. (2016). State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *Neuroimage* 129, 292–307. doi: 10.1016/j.neuroimage.2016.01.005
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Wee, C. Y., Yap, P. T., Zhang, D., Wang, L., and Shen, D. (2014). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Struct. Funct.* 219, 641–656. doi: 10.1007/s00429-013-0524-8
- Wu, X., Li, J., Ayutyanont, N., Protas, H., Jagust, W., Fleisher, A., et al. (2013). The receiver operational characteristic for binary classification with multiple indices and its application to the neuroimaging study of Alzheimer's disease. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10, 173–180. doi: 10.1109/tcbb.2012.141
- Zeng, H., and Cheung, Y. (2011). Feature selection and kernel learning for local learning-based clustering. *IEEE Transactions on PAMI* 33, 1532–1547. doi: 10.1109/TPAMI.2010.215
- Zhang, W., Shen, H., Ji, Z., Meng, G., and Wang, B. (2015). “Identification of mild cognitive impairment using extreme learning machines model,” intelligent computing theories and methodologies,” in *Proceedings of the 11th International Conference, ICIC 2015*, Fuzhou.
- Zhou, J., Greicius, M. D., Gennatas, E. D., Growdon, M. E., Jang, J. Y., Rabinovici, G. D., et al. (2010). Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain* 133, 1352–1367. doi: 10.1093/brain/awq075

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Lama and Kwon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.