Received: 6 May 2019 | Revised: 24 March 2020 | Accepted: 1 April 2020

DOI: 10.1111/exsy.12566

SPECIAL ISSUE PAPER

Expert Systems WILEY

A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging

Riyaj Uddin Khan[1] | **Mohammad Tanveer**[1] | **Ram Bilas Pachori**[2] | **Alzheimer's Disease Neuroimaging Initiative (ADNI)**

[1]Discipline of Mathematics, Indian Institute of Technology Indore, Indore, India

[2]Discipline of Electrical Engineering, Indian Institute of Technology Indore, Indore, India

Correspondence
Mohammad Tanveer, Discipline of Mathematics, Indian Institute of Technology Indore, Simrol, Indore 453552, India.
Email: mtanveer@iiti.ac.in

Funding information
Alzheimer's Disease Neuroimaging Initiative (ADNI), Grant/Award Number: U01 AG024904; Science and Engineering Research Board (SERB), Grant/Award Number: SB/S2/RJN-001/2016; Council of Scientific and Industrial Research (CSIR), Grant/Award Number: 22 (0751)/17/EMR-II

Abstract

Alzheimer's disease (AD) is the most prevalent form of dementia. Although fewer people, who suffer from AD are correctly and promptly diagnosed, due to a lack of knowledge of its cause and unavailability of treatment, AD is more manageable if the symptoms of mild cognitive impairment (MCI) are in an early stage. In recent years, computer-aided diagnosis has been widely used for the diagnosis of AD. The main motive of this paper is to improve the classification and prediction accuracy of AD. In this paper, a novel approach is developed to classify MCI, normal control (NC), and AD using structural magnetic resonance imaging (sMRI) from the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset (50 AD, 50 NC, 50 MCI subjects). FreeSurfer is used to process these MRI data and obtain cortical features such as volume, surface area, thickness, white matter (WM), and intrinsic curvature of the brain regions. These features are modified by normalizing each cortical region's features using the absolute maximum value of that region's features from all subjects in each group of MCI, NC, and AD independently. A total of 420 features are obtained. To address the curse of dimensionality, the obtained features are reduced to 30 features using a sequential feature selection technique. Three classifiers, namely the twin support vector machine (TSVM), least squares TSVM (LSTSVM), and robust energy-based least squares TSVM (RELS-TSVM), are used to evaluate the classification accuracy from the obtained features. Five-fold and 10-fold cross-validation are used to validate the proposed method. Experimental results show an accuracy of 100% for the studied database. The proposed approach is innovative due to its higher classification accuracy compared to methods in the existing literature.

KEYWORDS

Alzheimer's disease (AD), computer-aided diagnosis (CAD), FreeSurfer, LSTSVM, mild cognitive impairment (MCI), RELS-TSVM, TSVM

Abbreviations: ANN, artificial neural network; LVQ, learning vector quantization; ODC, optimally differentiating clusters; PCA, principal component analysis; PLS, partial least squares; RLR, regularized logistic regression; ROI, region of interest; SPARE-AD, spatial pattern of abnormalities for recognition of early AD; SVM, support vector machine; VBM, voxel-based morphology

1 | INTRODUCTION

Alzheimer's disease (AD) affects the functional and structural parts of the human brain, which results in cognitive decline and, ultimately, death. It is currently a growing threat to individuals, particularly those above 65 years of age, and it is dangerous because its odds of getting AD double every five years after 65 years of age. Environmental and hereditary influences play a critical role in the advancement and onset of the disease. There has been a reduction in deaths due to heart disease, prostate tumours, breast cancer, human immunodeficiency virus (HIV), and strokes by 16%, 8%, 2%, 42%, and 23%, respectively. On the other hand, deaths due to AD have increased by 68% (Alzheimer's, 2015). There are about 36 million people with AD globally, and this figure may rise to 115 million by the end of 2050 (Brookmeyer, Johnson, Ziegler-Graham, & Arrighi, 2007). Recently, Tanveer et al. (2020) presented an updated review of the diagnosis of AD using machine learning techniques.

The diagnosis of AD includes investigation of an individual's history, clinical perceptions, and incident history from associates. The intellectual decline of individuals can be measured using the mini-mental state examination (MMSE) and clinical dementia rate (CDR). However, these methods are tedious and cannot determine past changes in the brain that can be markers for the progression of AD. For the diagnosis of early AD, it is crucial to develop automated methods that can help experts suggest treatments that can slow the progression of the disease.

In recent years, there has been significant research focus on determining the early stages of AD. Machine learning algorithms based on single or multiple biomarkers for the classification task are widely used for the prediction of AD and mild cognitive impairment (MCI), Neuroimaging is the latest biomarker added in this field and gives reliable and efficient results for the identification of AD and MCI. Recent studies have shown that multiple modalities of the same subject can be used. However, the collection of different modalities for the same subject is not easily feasible, which results in a reduced number of subjects for study.

Electroencephalogram (EEG) data were also used in Kulkarni and Bairagi (2017) for the classification of normal control (NC) versus AD subjects using a support vector machine (SVM). Mazaheri et al. (2018) used EEG recordings of word comprehension by subjects to classify MCI converter (MCIc) from MCI non-converter (MCInc), and NC. Some recent EEG decomposition techniques such as empirical mode decomposition (EMD) based filtering in Gaur, Pachori, Wang, and Prasad (2015), multivariate EMD based filtering in Gaur et al. (Gaur, Pachori, Wang, & Prasad, 2016b; Gaur, Pachori, Wang, & Prasad, 2018), single and multi-channel EMD-based filtering in Gaur et al. (Gaur, Kaushik, Pachori, Wang, & Prasad, 2019; Gaur, Pachori, Wang, & Prasad, 2016a), and intrinsic mode function selection in Gaur, Pachori, Wang, and Prasad (2019), which enhances the classification of two class EEG signals, motivated us to develop algorithms for better classification of AD.

Despite remarkable research on the automatic classification of MCI or AD from NC, these classification techniques are not accurate. Earlier studies were based on either the analysis of the voxel-based morphometry (VBM) or region of interest (ROI) using classification algorithms. The limitations of ROI- and VBM-based techniques can be overcome by high-dimensional pattern classification techniques that use a large number of features. These techniques are effective but require high computational time as these techniques involve the extraction of a large number of features. The normalization of features and selection of the most distinct features may enhance the results of the proposed approach and will eventually reduce computational time. Tanveer and Pachori (2019) explained various techniques for the diagnosis of diseases, such as time-frequency analysis, feature extraction, and machine learning applications for the classification of different diseases.

The proposed approach in this paper performs classification for three groups of subjects (i.e., AD vs. NC, AD vs. MCI, and NC vs. MCI) using structural MRI (sMRI). The main focus is determining whether the given subject has AD or MCI. Once pre-processing by FreeSurfer is done on sMRI, features based on the surface area, cortical thickness, intrinsic curvatures, volume, and white matter (WM) are extracted. Surface area, cortical thickness, intrinsic curvature, and volume features play a major role in prior identification of AD. The surface area, cortical thickness, and curvature of brain regions change, and densities of white and grey matter are reduced due to changes in brain cells, as many of these cells die due to AD. A total of 420 features were extracted from the pre-processed MRI data. The obtained features were normalized then sequential feature selection technique was used to reduce the feature set to 30.

The main task in the proposed approach is classification. For this, we use classifiers such as the twin support vector machine (TSVM) (Jayadeva, Khemchandani, & Chandra, 2007), least squares TSVM (LSTSVM) (Kumar & Gopal, 2009), and robust energy-based least squares TSVM (RELS-TSVM) (Tanveer, Khan, & Ho, 2016) for the classification of AD, MCI, and NC.

In recent years, concepts based on constructing a different variation of non-parallel hyperplanes as in Tanveer (2015c), Tanveer and Shubham (2017), and Tanveer (2015a)) have emerged. TSVM is an efficient classification method, and, as mentioned in Jayadeva et al. (2007), the two non-parallel proximal hyperplanes are designed in such a way that each hyperplane is as close as possible to one class and as far as possible from the other class. Tanveer (2015b) proposed an implicit Lagrangian TSVM classifier in which a pair of unconstrained minimization problems (UMPs) in dual variables is formulated, whose solutions are obtained using the finite Newton method. Due to modified UMPs, it is necessary to solve only two systems of linear equations, compared to two quadratic programming problems (QPPs) in TSVM and twin bounded SVM (TBSVM), which makes the algorithm fast and simple. Shao, Zhang, Wang, and Deng (2011) proposed a TBSVM in which two small QPPs are solved by adding a regularized term for empirical risk minimization to construct two non-parallel hyperplanes, compared to TSVM. Tian, Qi, Ju, Shi, and Liu (2014) proposed a non-parallel SVM (NPSVM) in which the kernel trick can be directly applied without solving another two primal problems for the non-linear case, compared to existing non-linear TSVMs. High computational speed has been obtained by substituting convex QPP with a convex linear system (Kumar & Gopal, 2009). Recently, Tanveer, Sharma, and Suganthan (2019) proposed a novel general TSVM with pinball loss

(Pin-GTSVM) that is insensitive to noise and performs better for noise corrupted datasets. To retain the sparsity in the Pin-GTSVM, Tanveer, Tiwari, Choudhary, and Jalan (2019) proposed a novel sparse pinball twin SVM (SPTSVM) that is insensitive to outliers and retail sparsity. Richhariya and Tanveer (2020) used universum learning for the first time to solve the class imbalance problem in which reduced kernel was incorporated for reducing storage and computation cost. Nasiri, Charkari, and Mozafari (2014) proposed an energy-based least squares TSVM (ELS-TSVM) that utilizes different energy for each class and can also handle unbalanced dataset problems. Gautam et al. (2020) introduced deep kernel regularized least squares method for one-class classification by embedding minimum variance information, which improves the classifier capability by reducing the intra-class variance and can also work effectively with small size datasets. Tanveer et al. (2016) proposed RELS-TSVM, which maximizes the margin with a positive definite matrix formulation to overcome the positive semi-definite matrices formulation in the cases of TSVM, LSTSVM, and ELS-TSVM. Additionally, the RELS-TSVM algorithm does not need a special optimizer but uses energy parameters to reduce the effect of outliers and noise, due to which the classification becomes not only stable but also robust to outliers and noise (Tanveer et al., 2016). Recently, Tanveer et al. (2019a) provided an exhaustive analysis of 08 variants of TSVM-based classifiers along with 179 classifiers of 17 families. The results reveal that RELS-TSVM (Tanveer et al., 2016) performed the best among all the TSVM variants.

Hence, this inspired us to use these classifiers for the classification of AD. The experiment is individually performed for each type of feature set using TSVM, LSTSVM, and RELS-TSVM. The reduced feature set performs better than the complete feature set, which reduces the computational cost.

The paper is organized as follows: Section 2 discusses the background and related works on the diagnosis of AD using existing machine learning techniques. Section 3 explains the dataset and demographic characteristics of all subjects. Section 4 presents the proposed approach. Section 5 outlines the selection of features and the details of machine learning algorithms for classification problems. Section 6 presents the results, analysis, discussion, and a comparison of the proposed approach using the existing methods. Section 7 draws conclusions about the work.

2 | BACKGROUND AND RELATED WORK

There is growing interest in medical imaging that employs machine learning and computer-aided techniques for the diagnosis of AD. In machine learning algorithms, features such as voxel intensity, tissue density, and shape are used to train the classifier for classifying subjects such as those with AD and MCI as well as NC. These techniques are either whole brain-based or ROI-based. Prior knowledge is required for accessing the brain region in ROI-based methods as they do not include all of the information available from brain MRI.

Modern neuroimaging techniques help researchers evaluate and estimate several brain functionalities and structures that are beneficial in the diagnosis of AD and MCI. Recognized and risk-free techniques for brain imaging exist, and these are useful for the evaluation of physiological, anatomical, and pathological brain features with acceptable outcomes. Table 1 presents related work that has used machine learning algorithms for the classification of AD, along with limitations and possible advantages. The table provides the dataset used in each study, types of pre-processing operations, types of classifiers, and extracted features used for classification of the groups, namely AD, MCI, and NC. It also provides performance measures such as accuracy (acc), specificity (spe), and sensitivity (sen), which are used to compare the results of the proposed approach, which shows obvious improvement.

Misra et al. (2009) predicted the conversion of MCI to AD using a method based on pattern recognition that employs a baseline and longitudinal scan of the brain to measure atrophic spatial patterns. The aim is to predict whether MCI will transform into AD using baseline and longitudinal scans of region-based brain tissues. Fan, Shen, and Davatzikos (2005) utilized deformation-based morphology and machine learning techniques for the classification of medical images. High-dimensional template wrapping was utilized to acquire a morphological ROI, and the watershed division was used to separate the features for classification. The separated features were then ranked using SVM recursive feature elimination (RFE) strategy. Finally, SVM was utilized for the classification of subjects using the best set of features.

Fan, Resnick, et al. (2008) used features from positron emission tomography (PET) and MRI and applied high-dimensional pattern classification (PC) to classify AD from MCI. Wang et al. (2007) used functional connectivity from resting-state functional magnetic resonance imaging (rs-fMRI) to discriminate AD from NC using logistic regression. An increase in positive correlation was observed between the parietal, prefrontal, and occipital lobes while a decrease was observed for the parietal and prefrontal lobes. Davatzikos, Resnick, et al. (2008) used voxel-based analysis for the classification of AD and frontotemporal dementia (FTD) from NC using a high-dimensional PC method. Better diagnostic accuracy was obtained using high-dimensional multivariate discriminant analysis than using conventional measurement. Hinrichs et al. (2009) used linear programming for the classification of AD with sMRI (T1) and fluorodeoxyglucose PET (FDG-PET) with regularization using spatial smoothness. López et al. (2009) used PCA for dimension reduction and Bayesian classifiers for classifying NC versus AD. Horn et al. (2009) used ROI-based features, which were reduced by partial least squares (PLS) regression. The best accuracy was obtained using k-nearest neighbors (KNN) for classifying NC versus AD. Richhariya, Tanveer, Rashid, and Alzheimer's Disease Neuroimaging Initiative (2020) used universum support vector machine-based recursive feature elimination (USVM-RFE) technique for the diagnosis of AD. More recent techniques for the classification of AD, such as SVM, deep learning, transfer learning, and ensemble methods, are reviewed in Tanveer et al. (2020).

TABLE 1 Comparison of recent studies

	Classifier			Preprocessing					Features			Results		
References	Database	Classifier used	Groups	Alignment	Image Wrapping	Registration	Segmentation	Normalization	Extraction techniques	Feature type	Reduction	Accuracy	Specificity	Sensitivity
Klöppel et al. (2008)	—	SVM (linear)	AD-NC		✓	✓	✓	✓	—	—	—	—	95.0	95.0
			Mild AD-NC									—	91.2	75.8
			AD-FTLD									—	83.3	94.7
Fan, Resnick, Wu, and Davatzikos (2008)	BLSA	SVM (linear)	AD-MCI	✓	✓	✓	✓	✓	Adaptive regional	Regional features	ODC	90	—	—
Davatzikos, Bhatt, Shaw, Batmanghelich, and Trojanowski (2011)	BLSA	SVM (linear)	MCInc-MCIc	✓	✓	✓	✓	✓	—	SPARE-AD Cerebro-spinal fluid	—	—	—	—
Zhang et al. (2011)	ADNI	LDS, SVM	AD-NC						RLR	Voxel base	—	90.06	—	—
Davatzikos, Resnick, Wu, Parmpi, and Clark (2008)	ADNI	SVM (non linear)	AD-FTD	✓	✓	✓	✓	✓	ROI	Region by region	PCA	82	70	70.9
			AD-NC									94.3		
Misra, Fan, and Davatzikos (2009)	ADNI	SVM	MCIc-MCInc	✓	✓	✓	✓	✓	*t*-test	Brain regions	—	81.5	—	—
Fan et al. (2008)	ADNI	SVM	AD-NC	✓	✓	✓	✓	✓	Atlas wrapping	VBM	—	82	—	—
			MCI-NC									76.0		
			AD-MCI									58.3		
Davatzikos, Fan, Wu, Shen, and Resnick (2008)	BLSA	SVM (non linear)	AD-MCI	✓	✓	✓	✓	✓	—	Brain regions	RFE	90	—	—
Gerardin et al. (2009)	—	SVM	AD-NC							SPHARM	Bagging strategy	94	92	96
			MCI-NC									83	84	83
		FCM										83.3	93.3	73.3
		SVM	AD-NC									90.0	93.3	86.7
Farzan, Mashohor, Ramli, and Mahmud (2015)	ADNI	k-mean				✓	✓	✓	PCA	PBVC	DA	83.3	93.3	73.3
Khedher et al. (2015)	ADNI	SVM (linear)	AD-NC			✓	✓	✓	PLS	Volumetric	PLS	88.49	85.11	91.27
			MCI-NC									81.89	81.62	82.16
			MCI-AD									85.41	83.78	87.03
		SVM (RBF)	MCI-AD									88.49	86.17	90.39
			MCI-AD									80.27	82.7	73.51
			MCI-AD									85.41	84.84	85.95
		PCA-SVM										88	66	66
		VAF-SVM	AD-NC									90	76	76

TABLE 1 (Continued)

	Classifier			Preprocessing					Features			Results		
Ortiz et al. (2013)	ADNI	LVQ-SVM				✓	✓	✓	LVQ	Histogram	PCA	91	81	71
Ye, Pohl, and Davatzikos (2011)	ADNI	Embedding+ LapSVM	MCIc-MCInc			✓	✓	✓	RAVENS	Morphological features	ISOMAR algorithms	56.1	40.8	94.1
		Embedding+ SVM										55.3	42	88.2
		COMPARE+ SVM										52.3	37	89.8
Filipovych et al. (2011)	ADNI	SVM	MCIc-MCInc	✓	✓	✓	✓	✓	–	–	SVM-RFE	82.9	85.71	79.63
Zhang, Wang, and Dong (2014)	OASIS	kSVM-DT	AD-MCI-NC		✓	✓		✓	ALS-PCA	Genetically features	PCA	80	–	–
Hu, Wang, Chen, Hou, and Zhang (2016)	ADNI	SVM	AD-NC	✓		✓		✓	AAL	Wavelet ROI	Deep sparse Learning	–	–	–
			MCIc-MCInc											
Kamathe and Joshi (2018)	–	SVM	AD-NC							ROI	ICA	100	–	–
Kar and Majumder (2019)	ADNI	ANN	AD-NC							GM and WM	–	100	100	100

The fundamental principle behind AD classification approaches based on machine learning algorithms is the use of different brain imaging techniques for feature extraction. A large number of features are available for the classification of AD and MCI from NC. We use features that include WM, surface area, cortical thickness, volume, and curvature of different regions of the brain for the classification of AD and MCI. These are presented in Tables S1 and S2.

3 | MATERIALS AND METHODS

3.1 | ADNI database

For the presented work, we obtained data from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the Food and Drug Administration (FDA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), non-profit private pharmaceutical companies, and other organizations, with funding of $60 millions for the five-year private-public partnership.

The main goal of ADNI is to examine whether PET, serial MRI, clinical, and neuropsychological assessment, and other biological markers can be combined to measure early AD and the progression of MCI. Determination of specific and sensitive markers of very early AD progression is intended to aid clinicians and researchers develop new treatments and monitor their effects and as well as minimize the time and cost of clinical trials.

Michael W, Weiner, MD, VA Medical Center, and the University of California, San Francisco, is the principal investigator of this initiative.

3.2 | MRI acquisition

All the participants in this study were scanned with a GE Medical Systems MRI scanner with 1.5 Tesla field strength. The 3D T1-weighted MRI scans were captured with the following information: sagittal acquisition plane, 3D acquisition type, 1.5 Tesla field strength, 8.0° flip angle, =1.2-mm slice thickness, TE=3.96 ms, TI=1000.0 ms, TR=9.12 ms, and weighting=T1.

3.3 | Subjects

The ADNI dataset comprises more than 6,000 subjects with ages ranging from 18 to 96 years. From this database, we selected 150 subjects aged between 58 and 91 years. The selected participants met the criteria defined in the ADNI protocol. The balance dataset of 150 subjects was constructed as follows:

1 50 NC subjects: 25 males, 25 females; average age ± standard deviation (SD) =80.36 ± 4.98 years, range =70 – 91 years, mini-mental state estimation (MMSE) score =29.02 ± 1.16, range =25 – 30.
2 50 MCI subjects: 30 males, 20 females; average age ± SD =76.09 ± 7.85 years, range =58 – 91 years, MMSE score =25.66 ± 3.20, range =18 – 30.
3 50 AD subjects: 25 males, 25 females; average age ± SD =75.76 ± 7.40 years, range =58 – 90 years, MMSE =20.68 ± 4.70, range =5 – 27.

The complete demographic characteristics of all the selected subjects are listed in Table 2.

TABLE 2 Details of dataset

Subjects	NC	MCI	AD
Number of subjects	50	50	50
Sex (male/female)	25/25	30/20	25/25
Age (mean ± standard deviation)	80.36 ± 4.98	76.09 ± 7.85	75.76 ± 7.40
MMSE (mean ± standard deviation)	29.02 ± 1.16	25.66 ± 3.20	20.68 ± 4.70
Clinical dementia rating (CDR) (mean ± standard deviation)	0.03 ± 0.21	0.65 ± 0.44	1.05 ± 0.49

4 | PROPOSED ALGORITHM

Figure 1 shows the overall scheme of the proposed method. First, the cerebral image features of both hemispheres were measured by processing T1-weighted MRI brain images using FreeSurfer. Then, features such as volume, area, thickness, curvature, and WM volume of different brain regions were measured. A detailed explanation of how FreeSurfer works is provided below.

To calculate area, curvature, and cortical thickness, all the T1-weighted MRI images were processed using the freely available FreeSurfer software package v6.0.0 (http://surfer.nmr.mgh.harvard.edu). To obtain images with a relatively high contrast to noise ratio (CNR), normalization of intensity was applied to the MRI. First, the boundary between white and grey matter was found in an image. Then, a triangular mesh consisting of 160,000 vertices in each hemisphere was formed around the WM. The grey matter surface was formed by deforming the mesh outward so that it closely tracks the boundary between the cerebral spinal fluid (CSF) and grey matter. To measure cortical thickness, the distance between the grey matter and WM surfaces was calculated for each vertex. Using the cortical folding pattern, the image was then registered to FreeSurfer's common template. The neocortex of the Desikan et al. (2006) atlas was then parcellated into 68 neocortical regions comprising 34 regions from each

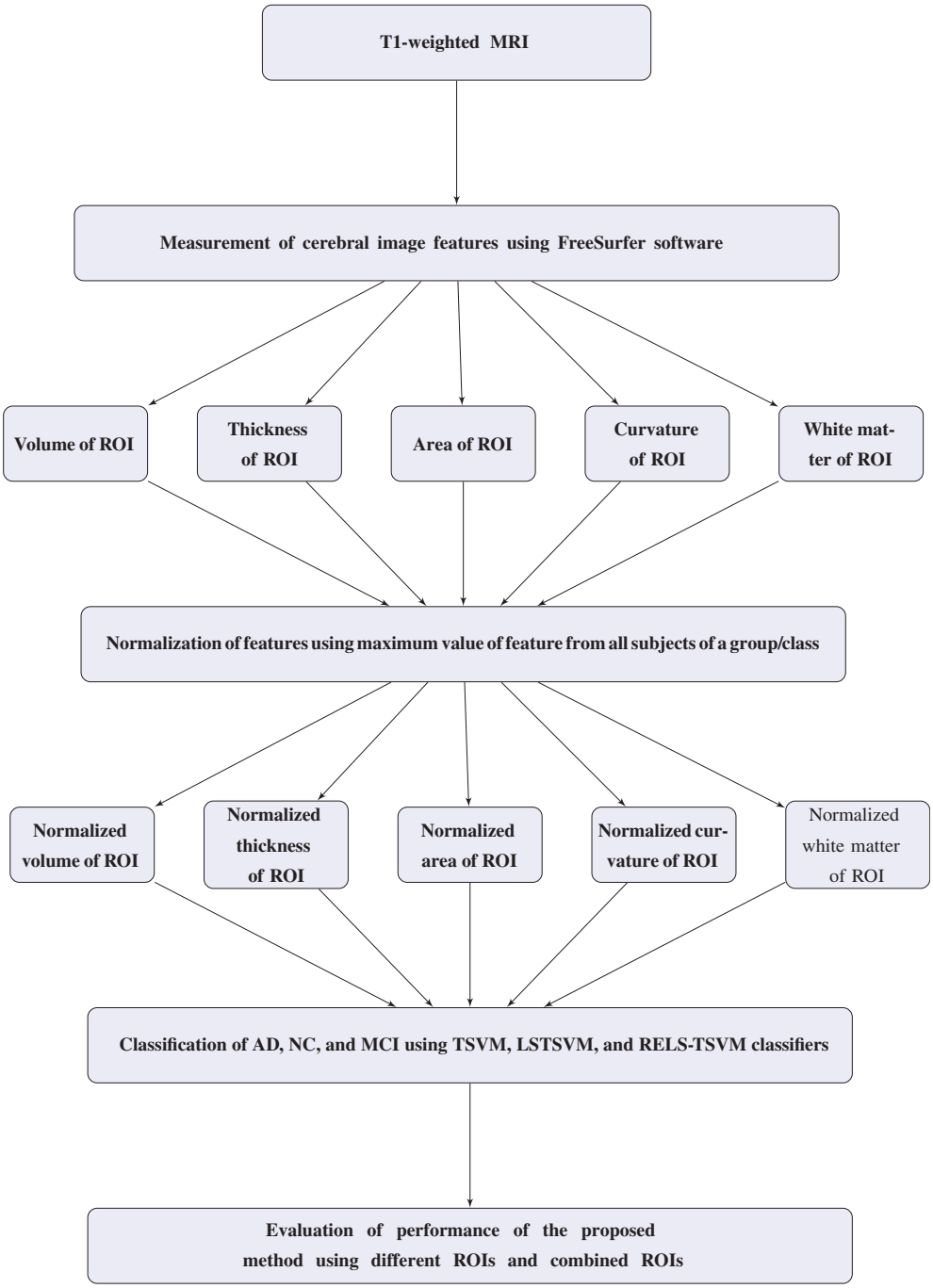

FIGURE 1 Overall scheme of the proposed method

hemisphere. In each parcellation unit, the thicknesses of all the vertices were calculated and then averaged. The obtained thickness is the thickness of that parcellation unit. This yielded 68 features of cortical thickness for each subject. To calculate the cortical surface area, all areas of the mesh triangles formed on the grey matter were summed, which yielded 68 features of the cortical area for each subject. To calculate the cortical curvature of the given surface, the curvature of vertices in two principle directions was calculated, which is the inverse of the radius of the osculating circles in the two principal directions and the average of all these curvatures, known as the cortical curvature of the parcellation unit. On average 68, cortical curvature features were obtained per subject (Figure 2).

Next, the cortical features such as WM intensity, volume, thickness, area, and curvature of different regions of the brain were extracted from stat files produced by FreeSurfer for all the MRIs. First, normalization of these features using the maximum absolute value of each cortical region of the subjects of each group (i.e., AD, MCI, and NC) was done in such a way that the value of each feature lies in the range of 0 – 1. This was done by dividing each feature of each cortical region by the maximum absolute value of features from that region of all subjects from the same group. With this operation, the features that are less linearly separable become more linearly separable. Figures 3, 4, 5, 6, 7, and 8 show the distribution of the first two features selected by sequential forward feature selection (SFFS) before normalization and after normalization for curvature and all combined features. From these figures, it is clear that after normalization, features become more separable (Figures 9 and 10).

The performance of the proposed method was evaluated using TSVM, LSTSVM, and RELS-TSVM classifiers. The performance of three groups (i.e., AD vs. NC, NC vs. MCI, and AD vs. MCI) was evaluated in terms of accuracy, sensitivity, and specificity. To validate our obtained results, we used fivefold and 10-fold cross-validation methods.

The proposed approach can be summarized as follows:

1 First, process all the MRI images using FreeSurfer.
2 Obtain features of the desired cortical regions of the brain.
3 Normalize the features as explained in Section 4
4 Obtain the relevant features using the SFFS method.

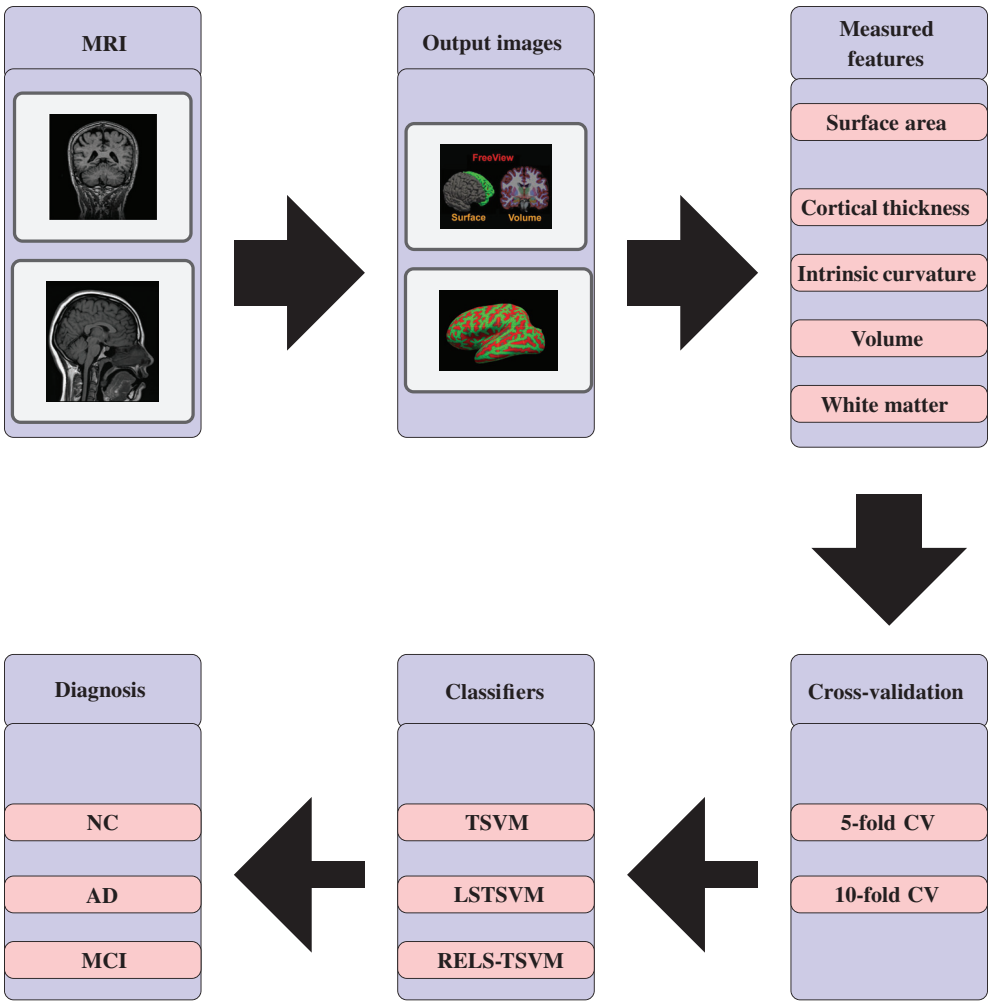

FIGURE 2 Systematic block diagram for automatic detection of AD

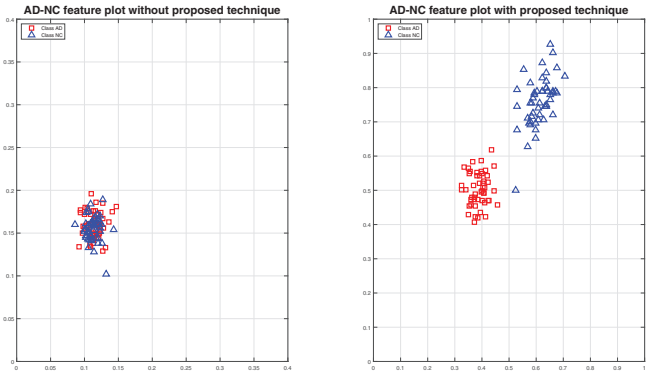

FIGURE 3 Normalization of AD and NC using curvature

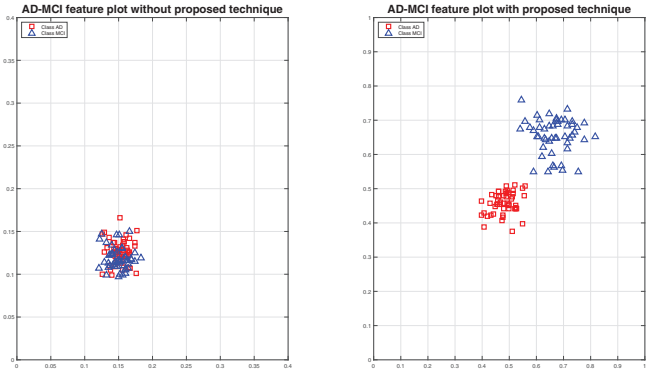

FIGURE 4 Normalization of AD and MCI using curvature

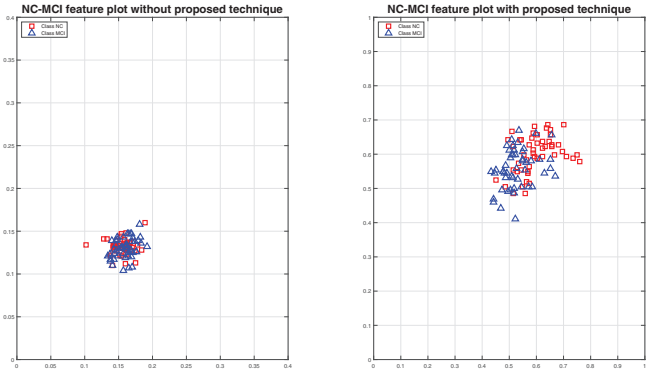

FIGURE 5 Normalization of NC and MCI using curvature

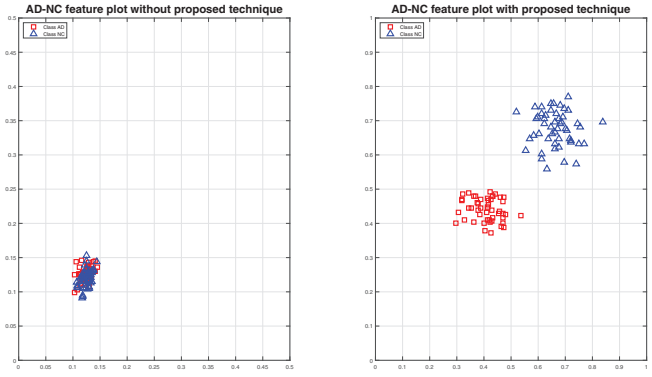

FIGURE 6 Normalization of AD and NC using all features

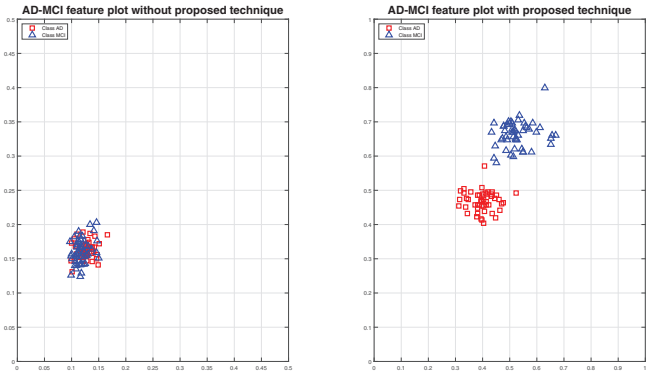

FIGURE 7 Normalization of AD and MCI using all features

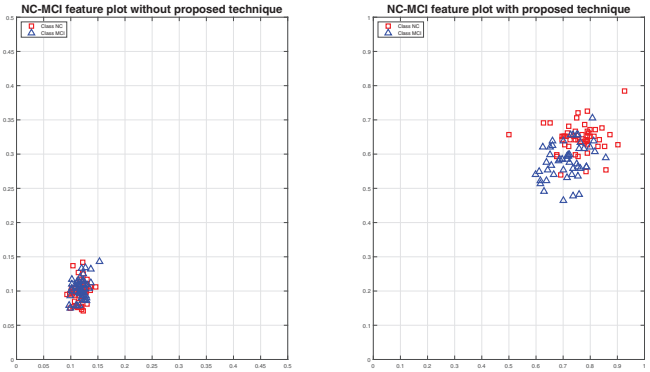

FIGURE 8 Normalization of NC and MCI using all features

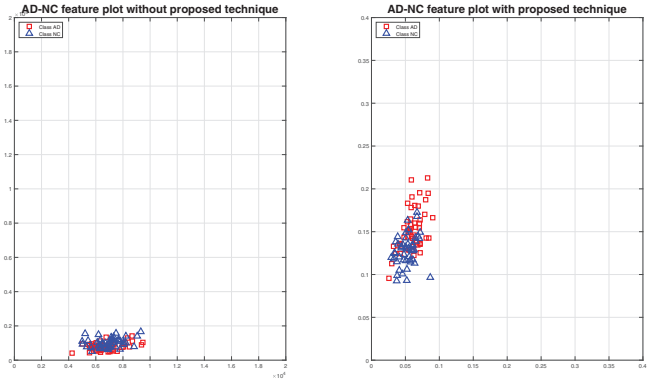

FIGURE 9 Normalization of AD and NC using white matter

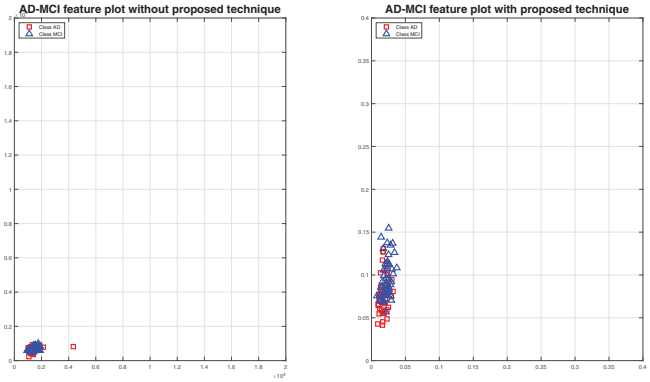

FIGURE 10 Normalization of AD and MCI using white matter

5 Perform classification.
6 Validate the classification results.

5 | PROPOSED APPROACH FOR THE CLASSIFICATION OF AD

We used TSVM, LSTSVM, and RELS-TSVM for the classification of AD. Figure 2 is a systematic block diagram of stepwise classification stages of AD prediction for the proposed approach. First, the obtained raw MRI from ADNI was processed by the freely available FreeSurfer software, and then features such as cortical thickness, surface area, curvature, volume, and WM were extracted. After extracting the individual features, all the ROI were normalized within the groups. The normalization of features was followed by SFFS to obtain the most relevant features. Finally, these relevant features were used for the classification of AD.

5.1 | Effective feature selection

In the analysis of neuroimages, a large number of features is involved compared to the number of subjects, which gives rise to the curse of dimensionality. To combat this effect, we applied an efficient and widely used SFFS method in the proposed approach. Hence, we get low-dimensional data from high-dimensional neuroimaging data.

The SFFS algorithm begins with an empty set, and the features are added sequentially in such a way that minimizes the error criteria. The 10-fold cross-validation is used internally, in which training data is divided into 10 partitions, and the subset of the selected is added feature in each iteration based on the criteria calculated for the internal test dataset. The algorithm stops when the criteria for improvement are exhausted or the number of predefined features is selected. In our study, the residual error is used in SFFS as the optimization function to be minimized. Acceptable is determined by a selected subset of small features using SFFS when a suitable classifier is used. The SFFS algorithm can be described as follows:

1 Start with a null subset $S = \phi$,
2 select the next higher rank feature h=arg min $(Er(x \cup S))$, $x \in (X - S)$,
3 update the feature subset $S = S \cup h$,
4 and stop if the criteria are satisfied; otherwise go to step 2.

where S is a null set, X denotes the set of all features, and Er denotes the classification error.

5.2 | Twin support vector machine (TSVM)

TSVM belongs to a class of binary classification that uses two non-parallel hyperplanes for the classification of data, instead of one hyperplane as with conventional SVMs (Jayadeva et al. (2007)). In conventional SVMs, a large QPP is solved to obtain the hyperplane. However, in TSVM, two small QPPs are solved to obtain two non-parallel hyperplanes. Consider a problem of binary classification in n-dimensional real space R^n, in which m_1 data points belong to class +1 and m_2 data points belong to class −1. Let A and B be the matrix in $R^{m_1 \times n}$ and $R^{m_2 \times n}$ representing the data points of one class (class +1) and the other class (class −1), respectively. For the given binary classification problem, the two non-parallel hyperplanes in R^n of linear TSVM can be expressed as follows (Figure 11):

$$\begin{aligned} \min_{w^{(1)},b^{(1)}} \quad & \frac{1}{2}\left(Aw^{(1)} + e_2 b^{(1)}\right)^T \left(Aw^{(1)} + e_2 b^{(1)}\right) + a_1 \|\xi_1\| \\ \text{subject to} \quad & -\left(Bw^{(1)} + e_1 b^{(1)}\right) + \xi_1 \geq e_1, \quad \xi_1 \geq 0, \end{aligned} \tag{1}$$

$$\begin{aligned} \min_{w^{(2)},b^{(2)}} \quad & \frac{1}{2}\left(Bw^{(2)} + e_1 b^{(2)}\right)^T \left(Bw^{(2)} + e_1 b^{(2)}\right) + a_2 \|\xi_2\| \\ \text{subject to} \quad & \left(Aw^{(2)} + e_2 b^{(2)}\right) + \xi_2 \geq e_2, \quad \xi_2 \geq 0, \end{aligned} \tag{2}$$

where $w^{(1)}$ and $w^{(2)}$ are weight vectors, e_1 and e_2 are vectors of suitable dimensions, whose all elements are of unit magnitude, a_1 and a_2 are positive penalty parameters, $b^{(1)}$ and $b^{(2)}$ are biased values, and ξ_1 and ξ_2 are slack variables.

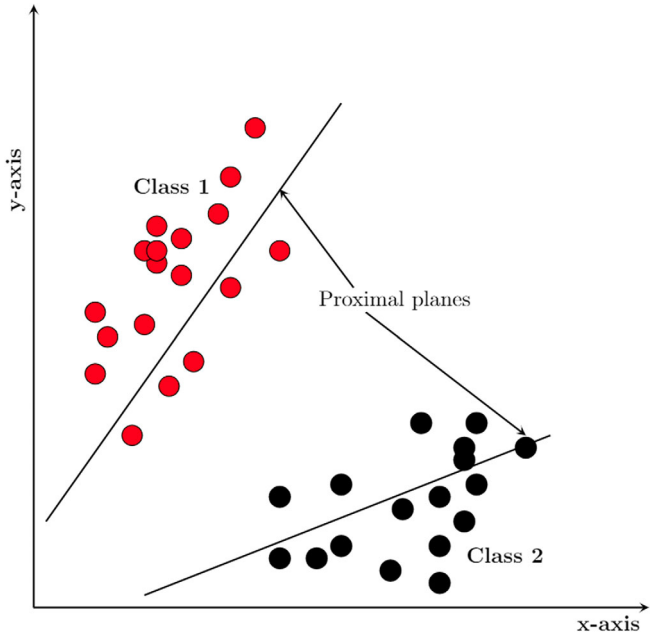

FIGURE 11 Geometrical interpretation of TSVM

The Wolfe duals of QPPs (1) and (2) were solved in Jayadeva et al. (2007). The QPPs (1) and (2) in terms of Lagrangian multipliers $\alpha \in R^{m_1}$ and $\beta \in R^{m_2}$ are given by (3) and QPPs by (4), respectively.

$$\begin{aligned} &\max_{\alpha} \quad e_1^T\alpha - \frac{1}{2}\alpha^T Q\left(P^T P\right)^{-1} Q^T\alpha \\ &\text{subject to} \quad 0 \le \alpha \le a_1, \end{aligned} \tag{3}$$

and

$$\begin{aligned} &\max_{\beta} \quad e_2^T\beta - \frac{1}{2}\beta^T P\left(Q^T Q\right)^{-1} P^T\beta \\ &\text{subject to} \quad 0 \le \beta \le a_2, \end{aligned} \tag{4}$$

respectively.

Here $P = [A \quad e_2]$ and $Q = [B \quad e_1]$. The solution of QPPs shown in (3) and (4) provides the non-parallel hyperplanes of (d), which are given by (5) and (6), respectively.

$$v_1 = -\left(P^T P\right)^{-1} Q^T\alpha, \quad \text{where} \quad v_1 = \left[w^{(1)} \quad b^{(1)}\right]^T \tag{5}$$

$$v_2 = \left(Q^T Q\right)^{-1} P^T\beta, \quad \text{where} \quad v_2 = \left[w^{(2)} \quad b^{(2)}\right]^T \tag{6}$$

It is advantageous for bounded constraints to solve two QPPs as this reduces the number of parameters of QPPs (3) and QPP (4), which are m_1 and m_2, respectively, when compared with parameters $l = m_1 + m_2$ of QPPs of SVM. However, it should be noted that TSVM requires inversion of the matrix twice of size $(n + 1) \times (n + 1)$ in addition to solving dual QPPs (3) and QPPs (4), where $n \ll l$. Support vectors that are important in determining the hyperplanes (1) are defined by data points for which $0 < \alpha_i < a_1 (i = 1, 2, \ldots, m_2)$ or $0 < \beta_j < a_2 (j = 1, 2, \ldots, m_1)$, and the support vectors lie on the corresponding hyperplane.

Here, P^{TP} or Q^{TQ} is always positive semidefinite, so it is possible that in some situations it is not well-conditioned. To avoid this condition, the inverse matrices $(P^T P)^{-1}$ of (5) are replaced by $(P^T P + \delta I)^{-1}$ and that of (6) is replaced by $(Q^T Q + \delta I)^{-1}$, where I represents the identity matrix and δ represents a scalar with a very small positive value. Thus, the modified dual problems (3) and (4) can be expressed as follows:

$$\begin{aligned} &\max_{\alpha} \quad e_1^T\alpha - \frac{1}{2}\alpha^T Q\left(P^T P + \delta I\right)^{-1} Q^T\alpha \\ &\text{subject to} \quad 0 \le \alpha \le a_1 \end{aligned} \tag{7}$$

and

$$\begin{aligned} \max_{\beta} \quad & e_2^T\beta - \frac{1}{2}\beta^T P\left(Q^T Q + \delta I\right)^{-1} P^T\beta \\ \text{subject to} \quad & 0 \le \beta \le a_2, \end{aligned} \tag{8}$$

respectively.

Thus, the solution of α and β QPPs (7) and (8) gives the two non-parallel hyperplanes as follows:

$$\begin{bmatrix} w^{(1)} \\ b^{(1)} \end{bmatrix} = -\left[P^T P + \delta I\right]^{-1} Q^T \alpha, \tag{9}$$

$$\begin{bmatrix} w^{(2)} \\ b^{(2)} \end{bmatrix} = \left[Q^T Q + \delta I\right]^{-1} P^T \beta. \tag{10}$$

After obtaining the non-parallel hyperplanes. A new data point $x \in R^n$ is assigned a label of class +1 or −1 on the basis of its shortest perpendicular distance from two hyperplanes, and the class label value is given as follows:

$$\text{Class i} = \arg \min_{i=1,2} \frac{|x^T w_i + b_i|}{\|w_i\|}. \tag{11}$$

For non-linear classification problems, TSVM can be used by replacing two linear kernels with two non-linear kernels generated by surfaces as follows:

$$K\left(x^T, C^T\right)u^{(1)} + \gamma^{(1)} = 0 \quad \text{and} \quad K\left(x^T, C^T\right)u^{(2)} + \gamma^{(2} = 0, \tag{12}$$

where $C = \begin{bmatrix} A \\ B \end{bmatrix}$ and K is an arbitrary kernel. Corresponding to the surface (12), the primal QPPs of non-linear TSVM can be expressed as follows:

$$\begin{aligned} \min_{u^{(1)},\gamma^{(1)}} \quad & \frac{1}{2}\left\|\left(K\left(A, C^T\right)u^{(1)} + e_2\gamma^{(1)}\right)\right\|^2 + a_1\|\xi_1\| \\ \text{subject to} \quad & -K\left(B, C^T\right)u^{(1)} + e_1\gamma^{(1)}) + \xi_1 \ge e_1, \quad \xi_1 \ge 0, \end{aligned} \tag{13}$$

$$\begin{aligned} \min_{u^{(2)},\gamma^{(2)}} \quad & \frac{1}{2}\left\|\left(K\left(B, C^T\right)u^{(2)} + e_1\gamma^{(2)}\right)\right\|^2 + a_2\|\xi_2\| \\ \text{subject to} \quad & K\left(A, C^T\right)u^{(2)} + e_2\gamma^{(2)}) + \xi_2 \ge e_2, \quad \xi_2 \ge 0. \end{aligned} \tag{14}$$

The duals of QPPs (13) and (14) can also be solved to obtain the hyperplanes (12). It should be noted that the solution of non-linear TSVM also requires the inversion of two matrices of order $m_1 \times m_1$ and $m_2 \times m_2$ along with the solution of two QPPs. In addition, it should be noted that TSVM may fail in some cases where the dataset is perfectly symmetric. This problem can be reduced by either using non-linear kernels or by slightly shifting a point, which disturbs the symmetry. Jayadeva et al. (2007) showed experimentally using University of California Irvine (UCI) machine learning datasets that TSVM performance is better than that of a generalized eigenvalue proximal support vector machine (GEPSVM) and conventional SVM in the case of linear and non-linear kernels. The TSVM can be trained four times faster using a linear kernel compared to conventional SVM (Jayadeva et al., 2007).

5.3 | Least squares TSVM (LSTSVM)

Kumar and Gopal (2009) also explored LSTSVM for which two non-parallel hyperplanes are solved, similarly to TSVM (??). In LSTSVM, the training data points are assigned close to one of the two non-parallel proximal hyperplanes and far from the other hyperplane. LSTSVM is a simple and fast algorithm that only requires a solution to a system of linear equations to achieve both linear and non-linear classifiers. The QPP (15) uses the square of the two-norm of slack variables a_1 with weight $\frac{a_1}{2}$ instead of the one-norm of ξ_1 with weight a_1, as in (??), which makes the constraint $\xi_1 \gg 0e_1$ redundant. The optimization problems of LSTSVM can be expressed as follows:

$$\min_{w^{(1)},b^{(1)}} \frac{1}{2}\left(Aw^{(1)}+e_2b^{(1)}\right)^T\left(Aw^{(1)}+e_2b^{(1)}\right)+\frac{a_1}{2}\|\xi_1\|^2$$
$$\text{subject to} \quad -\left(Bw^{(1)}+e_1b^{(1)}\right)+\xi_1=e_1, \tag{15}$$

$$\min_{w^{(2)},b^{(2)}} \frac{1}{2}\left(Bw^{(2)}+e_1b^{(2)}\right)^T\left(Bw^{(2)}+e_1b^{(2)}\right)+\frac{a_2}{2}\|\xi_2\|^2$$
$$\text{subject to} \quad \left(Aw^{(2)}+e_2b^{(2)}\right)+\xi_2=e_2. \tag{16}$$

The linear LSTSVM fully solves the classification problems by simply finding the inverse of two matrices with much smaller dimensions of order $(n+1)\times(n+1)$, where $n \ll l$ (Kumar & Gopal, 2009). After solving the QPPs (15) and (16), the following two systems of linear equations can be solved to obtain the two non-parallel hyperplanes:

$$\begin{bmatrix} w^{(1)} \\ b^{(1)} \end{bmatrix} = -\left[a_1Q^TQ+P^TP\right]^{-1}a_1Q^Te_1, \tag{17}$$

$$\begin{bmatrix} w^{(2)} \\ b^{(2)} \end{bmatrix} = \left[a_2P^TP+Q^TQ\right]^{-1}a_2P^Te_2, \tag{18}$$

where $P=[A \quad e_2]$, $Q=[B \quad e_1]$, and a_1 and a_2 are some positive penalty parameters.

A new datapoint $x \in R^n$ is assigned a class label of +1 or −1 based on its perpendicular distance from two hyperplanes.

5.4 | Robust energy-based least squares TSVM (RELS-TSVM)

Consider a binary classification problem in which matrix $A \in R^{m_1 \times n}$ represents data points of class +1 and matrix $B \in R^{m_2 \times n}$ represents data points of class −1.

The linear RELS-TSVM (Tanveer et al., 2016) includes a pair of minimization problems as follows:

$$\min_{w^{(1)},b^{(1)},\xi^{(1)}} \frac{1}{2}\left(Aw^{(1)}+e_2b^{(1)}\right)^T\left(Aw^{(1)}+e_2b^{(1)}\right)+\frac{a_1}{2}\xi_1^T\xi_1+\frac{a_3}{2}\left\|\begin{bmatrix} w^{(1)} \\ b^{(1)} \end{bmatrix}\right\|^2$$
$$\text{subject to} \quad -\left(Bw^{(1)}+e_1b^{(1)}\right)+\xi_1=E_1, \tag{19}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(2)}} \frac{1}{2}\left(Aw^{(1)}+e_1b^{(1)}\right)^T\left(Aw^{(1)}+e_1b^{(1)}\right)+\frac{a_2}{2}\xi_2^T\xi_2+\frac{a_4}{2}\left\|\begin{bmatrix} w^{(2)} \\ b^{(2)} \end{bmatrix}\right\|^2$$
$$\text{subject to} \quad \left(Aw^{(2)}+e_2b^{(2)}\right)+\xi_2=E_2, \tag{20}$$

where e_1 and e_2 are the vectors of ones of suitable dimensions, $a_1,\ a_2,\ a_3,\ a_4 > 0$ are penalty parameters, and E_1 and E_2 represent the energy parameters of the hyperplanes.

The QPP (19) can be solved as follows:

$$z_1 = -\left(a_1Q^TQ+P^TP+a_3I\right)^{-1}a_1Q^TE_1, \tag{21}$$

where $z_1=\begin{bmatrix} w^{(1)} \\ b^{(1)} \end{bmatrix}$. In a similar way, the solution of QPP (20) is given by following equation:

$$z_2 = \left(a_2P^TP+Q^TQ+a_4I\right)^{-1}a_2P^TE_2, \tag{22}$$

where $z_2 = \begin{bmatrix} w^{(2)} \\ b^{(2)} \end{bmatrix}$, $P = [A\ e_2]$, $Q = [Be_1]$. The following decision function is used to assign the class label $i = \{+1, -1\}$ to unknown data point $x_i \in R^n$:

$$f(x_i) = \begin{cases} +1, & \text{if } \left| \dfrac{x_i w^{(1)} + eb^{(1)}}{x_i w^{(2)} + eb^{(2)}} \right| \leq 1 \\ -1, & \text{if } \left| \dfrac{x_i w^{(1)} + eb^{(1)}}{x_i w^{(2)} + eb^{(2)}} \right| > 1 \end{cases} \tag{23}$$

6 | EXPERIMENTAL RESULTS AND ANALYSIS

6.1 | Permutation testing

The statistical importance of the classifier can be calculated using permutation tests. In addition, the performance of the classifier can be evaluated using the test error as a statistic, which gives the dissimilarity between two classes as evaluated in Golland and Fischl (2003). The evaluation process begins with the choice of the test statistics of the classifier and the allocation of arbitrary labels to the classifier by permuting the class labels for the training dataset. Cross-validation (CV) is involved in permutation testing, in which a randomly permuted class label is given to the diagnostic dataset. This leads to the misprediction of clinical data labels due to the distribution of classification results under the null hypothesis. The significance of a classifier can be indicated by p value, which is defined as the permuted estimation rate against the estimation rate with the original data labels. In our work, we use fivefold CV and 10-fold CV methods and carry out binary classification of AD using TSVM, LSTSVM, and RELS-TSVM.

6.2 | Performance evaluation methods

The performance of the proposed algorithm was evaluated using the TSVM, LSTSVM, and RELS-TSVM classifiers. Table 3 depicts the confusion matrix, which can be used to evaluate the binary classification performance of classes A and B. The number of correctly predicted values by the classifier is indicated by the diagonal elements of the confusion matrix. The correctly identified controls can be represented by further dividing the elements into true negative (TN) and true positive (TP). Similarly, the classes that are wrongly classified by the classifier can be represented by false negative (FN) and false positive (FP).

The portion of the subjects correctly labeled by the classifier can be measured by accuracy (acc), which is defined as follows:

$$\text{acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{24}$$

If the dataset has a highly unbalanced class distribution, then the accuracy as defined in Equation (24) may not be the correct performance measure. Therefore, two other performance measures were used, namely specificity (spe) and sensitivity (sen), which are defined below as follows:

$$\text{spe} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{25}$$

TABLE 3 Confusion matrix

	Predicted class	
True class	A	B
A	TP	FP
B	FP	TN

$$\text{sen} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{26}$$

The rate of TNs is measured by the spe as defined in Equation (25) while the rate of TPs is measured by the sensitivity as defined in Equation (26).

6.3 | Classification results and analysis

In this paper, the features extracted by the FreeSurfer are used by three classifiers to evaluate the performance of the proposed algorithm. The main aim is to develop a method that can distinguish healthy (normal) subjects from AD and MCI patients. In the first case, we trained all three classifiers with obtained features from AD and NC MRI scans (group 1 analysis).

6.3.1 | Classification results for NC and AD

In this section, the classification results for distinguishing AD and NC subjects are presented. For this group analysis, we used 50 AD subjects and 50 NC subjects, of which 40 subjects from each group were used to train the classifiers, and the remaining 10 subjects of each group were tested using three classifiers (i.e., TSVM, LSTSVM, and RELS-TSVM). The acc, spe, and sen for the various features are given in Table 4. The lowest acc was obtained for volume using TSVM. The highest acc of 100% was obtained for all the classifiers using cortical curvature and for all the features

TABLE 4 Classification results in percentage (%) for NC versus AD

Classifier	Performance in (%)	White matter	Volume	Area	Thickness	Curvature	All combined
TSVM	acc	75.5	54	99.5	94.5	100	100
	sen	69	80	100	96	100	100
	spe	82	28	99	93	100	100
LSTSVM	acc	79.5	72.5	100	98.5	100	100
	sen	86	74	100	98	100	100
	spe	73	71	100	99	100	100
RELS-TSVM	acc	81.33	54.67	100	95.33	100	100
	sen	73.33	67.33	100	96	100	100
	spe	75.72	64.36	100	96.18	100	100

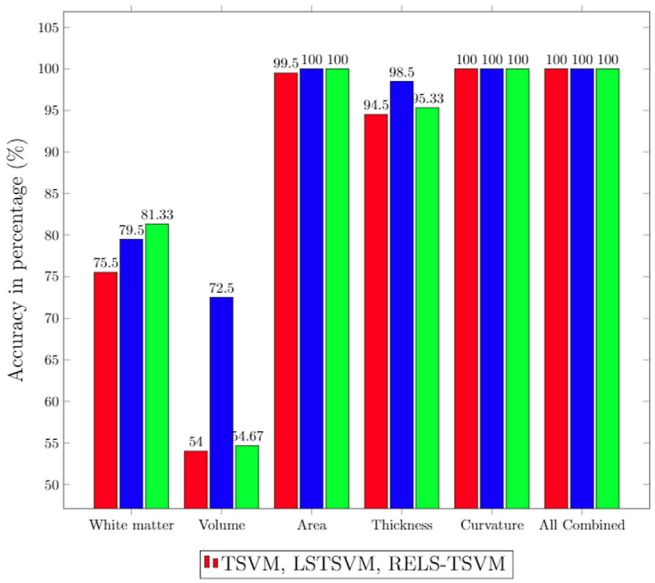

FIGURE 12 Classification accuracy for AD versus NC

combined. Hence, combining all the features increases the acc, spe, and sen. For WM analysis, the lowest acc of 75.5% was obtained by TSVM, and the highest acc of 81.33% was obtained by linear RELS-LSTSVM. For volumetric analysis, the lowest acc of 54% was obtained by TSVM, and the highest acc of 72.5% was obtained by LSTSVM. For cortical area analysis, the lowest acc of 90.5% was obtained by TSVM, and the highest acc of 100% was obtained by LSTSVM and RELS-TSVM. For cortical curvature analysis, all the three classifiers had acc of 100%. Figure 12 shows all the classification acc.

TABLE 5 Classification results in percentage (%) for NC versus MCI

1	Performance (%)	White matter	Volume	Area	Thickness	Curvature	All combined
TSVM	acc	89.5	50.5	90	84	99	99.5
	sen	93	78	93	82	98	99
	spe	86	23	87	86	100	100
LSTSVM	acc	82	63.5	98.5	90.5	99	100
	sen	82	66	99	86	98	100
	spe	82	61	98	95	100	100
RELS-TSVM	acc	79.33	76	96	87.33	98	100
	sen	92	46.67	96	85.33	100	100
	spe	91.64	58.82	96.25	86.47	100	100

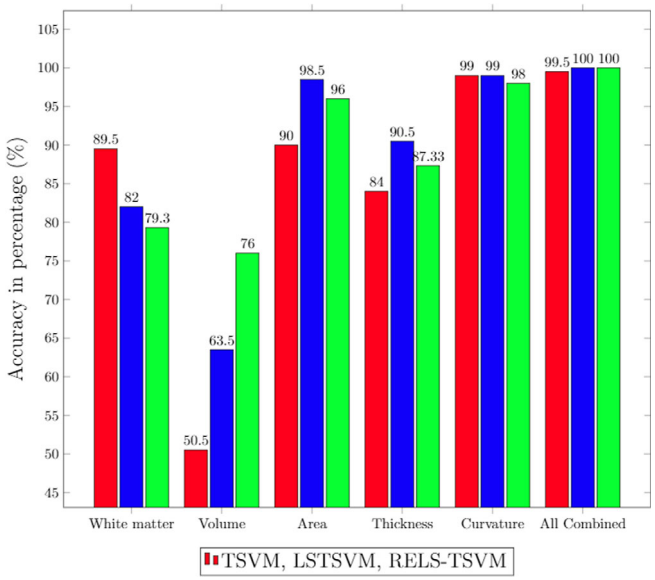

FIGURE 13 Classification accuracy for NC versus MCI

TABLE 6 Classification results in percentage (%) for AD versus MCI

Classifier	Performance (%)	White matter	Volume	Area	Thickness	Curvature	All combined
TSVM	acc	71.5	62.5	94.5	79	99.6	100
	sen	80	62	90	79	100	100
	spe	63	63	99	79	100	100
LSTSVM	acc	76	78	100	91	100	100
	sen	74	80	100	94	100	100
	spe	78	76	100	88	100	100
RELS-TSVM	acc	71.33	54.67	100	68.67	100	100
	sen	78	54.67	99.33	84	100	100
	spe	77.52	55.1	99.38	81.35	100	100

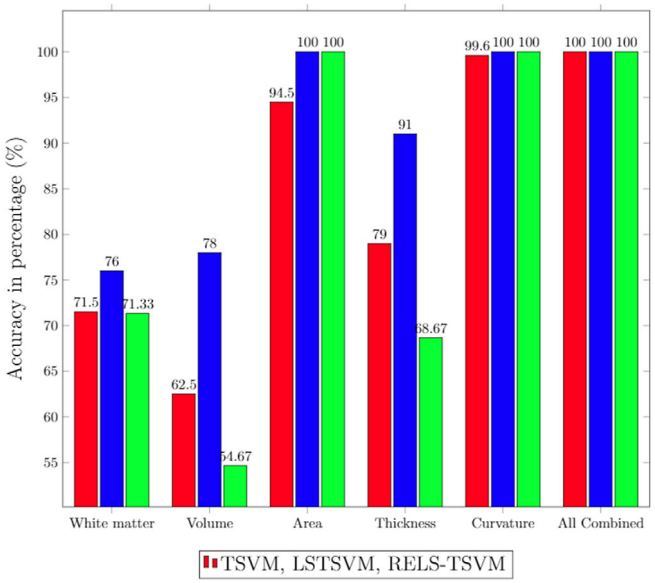

FIGURE 14 Classification accuracy for AD versus MCI

TABLE 7 Comparison of the proposed approach with existing approaches

Approach	Features	Groups	Classifiers	Accuracy (%)
Hosseini-Asl, Gimel'farb, and El-Baz (2016)	Volumetric, intensity and cortical thickness	NC versus AD	Random Forest	83.00
		MCI versus AD		68.00
		HC versus MCI		67.00
		HC versus MCI versus AD		54.00
Zhang, Yu, Jiang, Liu, and Tong (2012)	Textural	AD versus NC	1-NN	89.0
			ANN	98.50
Zheng, Yao, Xie, Fan, and Hu (2018)	Volume, surface area cortical thickness sulcal depth	NC versus AD	SVM	98.70
		NC versus MCI		97.93
		MCI versus AD		73.82
		sMCI versus pMCI		67.92
Proposed approach	Cortical features		TSVM	100
			LSTSVM	100
		AD versus NC	RELS-TSVM	100
			TSVM	100
			LSTSVM	100
		AD versus MCI	RELS-TSVM	100
		NC versus MCI	TSVM	99.5
			LSTSVM	100
		NC versus MCI	RELS-TSVM	100

6.3.2 | Classification results for NC and MCI

For the classification of NC from MCI, we used 50 subjects from each group. Table 5 shows the performance measures such as acc, sen, and spe for feature sets obtained using TSVM, LSTSVM, and RELS-TSVM classifiers for NC and MCI. For every classifier input, we divided each group into a set of 40 and 10 subjects. For each group that is, NC and MCI) a set of 40 subjects was used to train the classifiers, and a set of 10 was used to evaluation classification performance. Figure 13 shows the classification acc for NC and MCI. It shows that the volume features had the lowest acc of 50.5% when TSVM was used for classification. Cortical curvature and all features combined had the highest acc of 100% when LSTSVM and RELS-TSVM were used. Therefore, combining all features increased not only the classification accuracy but also improved other performance

parameters such as sen and spe. In the case of volume, the highest acc of 76% was obtained using RELS-TSVM. The lowest acc of 90% and the highest acc of 98.5% were obtained by TSVM and LSTSVM, respectively, when cortical areas were used as features. When all features were combined, sen and spe of 100% were obtained. Table 5 shows the performance parameters for the various feature sets.

6.3.3 | Classification results for AD and MCI

Table 6 shows the classification results for AD and MCI subjects. For this analysis, we used 50 subjects from each group. For training classifiers and testing, each group was divided into sets of 40 and 10 subjects, respectively. All three classifiers (i.e., TSVM, LSTSVM, and RELS-TSVM) were trained, and the performance measures were evaluated using the predicted values of class levels. Figure 14 shows the classification accuracies obtained by TSVM, LSTSVM, and RELS-TSVM classifiers using the various feature sets of AD and MCI. An acc of 100% was obtained by all the three classifiers when cortical curvature was used as a feature. The same acc was obtained when all features were combined. The lowest acc of 54.67% was obtained for volumetric analysis using RELS-TSVM. For WM, the lowest acc of 71.33% was obtained by RELS-TSVM, while the highest acc of 74% was obtained by LSTSVM. Comparatively higher acc values were obtained for cortical area features. For all features combined, sen of 100% and spe of 100% were obtained. Table 6 summarized all of the performance parameters.

Overall, our results show that LSTSVM and RELS-TSVM perform better than TSVM in most cases. Among the five cortical feature sets, WM, volume, area, thickness, and curvature performed better. In addition, the volume of different regions cannot produce satisfactory results. It can be concluded that the normalization technique we used, combined with classifiers such as LSTSVM and RELS-TSVM, is a better option for overall classification between AD-NC, AD-MCI, and NC-MCI due to the high acc values. Hence, our proposed approach is highly effective for the classification of AD. Table 7 compares our proposed approach with existing approaches.

7 | CONCLUSIONS

The accurate diagnosis of AD and MCI is essential for both research and patient care. To delay or alleviate the progression of AD, preventive measures play an important role. The classification task is a major challenge because of the small number of training samples compared to the large number of features. The proposed approach performs better than some existing methods for AD classification in terms of acc, sen, and spe. In our study, we investigated TSVM, LSTSVM, and RELS-TSVM for the classification of AD, MCI, and NC. For the classification of AD, LSTSVM performs better than TSVM, and RELS-TSVM performs better than both TSVM and LSTSVM. Further, the classification acc of AD improves by applying SFFS for feature selection.

In the future, we plan to analyze the combined effects of more modalities, such as PET and functional MRI (fMRI) with sMRI for the classification of AD. In addition, other modalities such as the combination of EEG and MRI can be analysed. We also plan to investigate the more recent neuro-imaging modality magnetoencephalography (MEG) in our future work.

ACKNOWLEDGEMENTS

This work is supported by the Council of Scientific and Industrial Research (CSIR), New Delhi, India under the Extra Mural Research (EMR) Scheme, Grant No. 22 (0751)/17/EMR-II and the Science and Engineering Research Board (SERB) under the Ramanujan Fellowship Scheme, Grant No. SB/S2/RJN-001/2016. We gratefully acknowledge the Indian Institute of Technology Indore for providing facilities and support. We are thankful to CSIR, New Delhi, India for providing a research fellowship to Mr. R.U. Khan.

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 Defence Award Number W81XWH-12-2-0012). 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 and Development LLC.; Johnson & Johnson Pharmaceutical Research & 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 provides 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 was 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.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Riyaj Uddin Khan https://orcid.org/0000-0002-9124-3985
Mohammad Tanveer https://orcid.org/0000-0002-5727-3697

REFERENCES

Alzheimer's, A. (2015). Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*, 332.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*, *3*, 186–191.

Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., & Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of Aging*, *32*, 2322–e19.

Davatzikos, C., Fan, Y., Wu, X., Shen, D., & Resnick, S. M. (2008). Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging*, *29*, 514–523.

Davatzikos, C., Resnick, S. M., Wu, X., Parmpi, P., & Clark, C. M. (2008). Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. *Neuroimage*, *41*, 1220–1227.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Albert, M. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980.

Fan, Y., Batmanghelich, N., Clark, C. M., Davatzikos, C., & Alzheimer's Disease Neuroimaging Initiative. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage*, *39*, 1731–1743.

Fan, Y., Resnick, S. M., Wu, X., & Davatzikos, C. (2008). Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. *Neuroimage*, *41*, 277–285.

Fan, Y., Shen, D. and Davatzikos, C. (2005) *Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM*. Paper presented at: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 1–8). Springer.

Farzan, A., Mashohor, S., Ramli, A. R., & Mahmud, R. (2015). Boosting diagnosis accuracy of Alzheimer's disease using high dimensional recognition of longitudinal brain atrophy patterns. *Behavioural Brain Research*, *290*, 124–130.

Filipovych, R., Davatzikos, C., & Alzheimer's Disease Neuroimaging Initiative. (2011). Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). *Neuroimage*, *55*, 1109–1119.

Gaur, P., Kaushik, G., Pachori, R. B., Wang, H., & Prasad, G. (2019). Comparison analysis: single and multichannel EMD-based filtering with application to BCI. In *Machine intelligence and signal analysis* (pp. 107–118). Singapore: Springer.

Gaur, P., Pachori, R. B., Wang, H. and Prasad, G. (2015) *An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface*. Paper presented at Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN) (pp. 1–7). IEEE.

Gaur, P., Pachori, R. B., Wang, H. and Prasad, G. (2016a) *Enhanced motor imagery classification in EEG-BCI using multivariate EMD based filtering and CSP features*. Paper presented at Proceedings of the International Brain-Computer Interface (BCI) Meeting 2016.

Gaur, P., Pachori, R. B., Wang, H. and Prasad, G. (2016b) *A multivariate empirical mode decomposition based filtering for subject independent BCI*. Paper presented at Proceedings of the 2016 27th Irish Signals and Systems Conference (ISSC) (pp. 1–7). IEEE.

Gaur, P., Pachori, R. B., Wang, H., & Prasad, G. (2018). A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. *Expert Systems with Applications*, *95*, 201–211.

Gaur, P., Pachori, R. B., Wang, H., & Prasad, G. (2019). An automatic subject specific intrinsic mode function selection for enhancing two-class EEG based motor imagery-brain computer interface. *IEEE Sensors Journal*, *19*, 6938–6947.

Gautam, C., Mishra, P. K., Tiwari, A., Richhariya, B., Pandey, H. M., Wang, S., et al. (2020). Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data. *Neural Networks*, *123*, 191–216.

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.-S., … Eustache, F. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage*, *47*, 1476–1486.

Golland, P. and Fischl, B. (2003) Permutation tests for classification: towards statistical significance in image-based studies. Paper presented at *Biennial International Conference on Information Processing in Medical Imaging*, 330–341. Springer.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., Johnson, S. C., & Alzheimer's Disease Neuroimaging Initiative. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage*, *48*, 138–149.

Horn, J.-F., Habert, M.-O., Kas, A., Malek, Z., Maksud, P., Lacomblez, L., … Fertil, B. (2009). Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images. *Artificial Intelligence in Medicine*, *47*, 147–158.

Hosseini-Asl, E., Gimel'farb, G. and El-Baz, A. (2016) Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. arXiv preprint arXiv:1607.00556.

Hu, K., Wang, Y., Chen, K., Hou, L., & Zhang, X. (2016). Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. *Neurocomputing*, *175*, 132–145.

Jayadeva, Khemchandani, R., & Chandra, S. (2007). Twin support vector machines for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*, 905–910.

Kamathe, R. S., & Joshi, K. R. (2018). A novel method based on independent component analysis for brain MR image tissue classification into CSF, WM and GM for atrophy detection in Alzheimer's disease. *Biomedical Signal Processing and Control*, *40*, 41–48.

Kar, S., & Majumder, D. D. (2019). A novel approach of diffusion tensor visualization based Neuro fuzzy classification system for early detection of Alzheimer's disease. *Journal of Alzheimer's Disease Reports*, **40**,1–18.

Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A., Segovia, F., & Alzheimer's Disease Neuroimaging Initiative. (2015). Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, *151*, 139–150.

Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., … Frackowiak, R. S. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain*, *131*, 681–689.

Kulkarni, N., & Bairagi, V. (2017). Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier. *IETE Journal of Research*, *63*, 11–22.

Kumar, M. A., & Gopal, M. (2009). Least squares twin support vector machines for pattern classification. *Expert Systems with Applications*, *36*, 7535–7543.

López, M., Ramírez, J., Górriz, J., Salas-Gonzalez, D., Alvarez, I., Segovia, F., & Puntonet, C. (2009). Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules. *Electronics Letters*, *45*, 389–391.

Mazaheri, A., Segaert, K., Olichney, J., Yang, J.-C., Niu, Y.-Q., Shapiro, K., & Bowman, H. (2018). EEG oscillations during word processing predict MCI conversion to Alzheimer's disease. *NeuroImage: Clinical*, *17*, 188–197.

Misra, C., Fan, Y., & Davatzikos, C. (2009). Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *Neuroimage*, *44*, 1415–1422.

Nasiri, J. A., Charkari, N. M., & Mozafari, K. (2014). Energy-based model of least squares twin support vector machines for human action recognition. *Signal Processing*, *104*, 248–257.

Ortiz, A., Górriz, J. M., Ramírez, J., Martínez-Murcia, F. J., & Alzheimer's Disease Neuroimaging Initiative. (2013). LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease. *Pattern Recognition Letters*, *34*, 1725–1733.

Richhariya, B., & Tanveer, M. (2020). A reduced universum twin support vector machine for class imbalance learning. *Pattern Recognition*, **102**, 107150.

Richhariya, B., Tanveer, M., Rashid, A., & Alzheimer's Disease Neuroimaging Initiative. (2020). Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomedical Signal Processing and Control*, *59*, 101903.

Shao, Y.-H., Zhang, C.-H., Wang, X.-B., & Deng, N.-Y. (2011). Improvements on twin support vector machines. *IEEE Transactions on Neural Networks*, *22*, 962–968.

Tanveer, M. (2015a). Application of smoothing techniques for linear programming twin support vector machines. *Knowledge and Information Systems*, *45*, 191–214.

Tanveer, M. (2015b). Newton method for implicit Lagrangian twin support vector machines. *International Journal of Machine Learning and Cybernetics*, *6*, 1029–1040.

Tanveer, M. (2015c). Robust and sparse linear programming twin support vector machines. *Cognitive Computation*, *7*, 137–149.

Tanveer, M., Gautam, C., & Suganthan, P. N. (2019). Comprehensive evaluation of twin SVM based classifiers on UCI datasets. *Applied Soft Computing*, *83*, 105617.

Tanveer, M., Khan, M. A., & Ho, S.-S. (2016). Robust energy-based least squares twin support vector machines. *Applied Intelligence*, *45*, 174–186.

Tanveer, M., & Pachori, R. B. (2019). *Machine intelligence and signal analysis* (Vol. 748). New York, NY: Springer.

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C.-T. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *28*, 1–35.

Tanveer, M., Sharma, A., & Suganthan, P. N. (2019). General twin support vector machine with pinball loss function. *Information Sciences*, *494*, 311–327.

Tanveer, M., & Shubham, K. (2017). Smooth twin support vector machines via unconstrained convex minimization. *Filomat*, *31*, 2195–2210. JSTOR.

Tanveer, M., Tiwari, A., Choudhary, R., & Jalan, S. (2019). Sparse pinball twin support vector machines. *Applied Soft Computing*, *78*, 164–175.

Tian, Y., Qi, Z., Ju, X., Shi, Y., & Liu, X. (2014). Nonparallel support vector machines for pattern classification. *IEEE Transactions on Cybernetics*, *44*, 1067–1079.

Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., & Jiang, T. (2007). Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. *Human Brain Mapping*, *28*, 967–978.

Ye, D. H., Pohl, K. M., Davatzikos, C. (2011) *Semi-supervised pattern classification: application to structural MRI of Alzheimer's disease*. Paper presented at Proceedings of the 2011 International Workshop on Pattern Recognition in NeuroImaging (PRNI), (pp. 1–4). IEEE.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, *55*, 856–867.

Zhang, J., Yu, C., Jiang, G., Liu, W., & Tong, L. (2012). 3D texture analysis on MRI images of Alzheimer's disease. *Brain Imaging and Behavior*, *6*, 61–69.

Zhang, Y., Wang, S., & Dong, Z. (2014). Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Progress in Electromagnetics Research*, *144*, 171–184.

Zheng, W., Yao, Z., Xie, Y., Fan, J., & Hu, B. (2018). Identification of Alzheimer's disease and mild cognitive impairment using networks constructed based on multiple morphological brain features. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*, 887–897.

AUTHOR BIOGRAPHIES

Riyaj Uddin Khan is working as Senior Research Fellow (SRF) in Discipline of Mathematics, Indian Institute of Technology, Indore, India. He received his B.Tech. and M.Tech. degree from Aligarh Muslim University, Aligarh, India, in 2015 and 2017 respectively. His research interest includes bio medical signal processing, machine learning, brain-computer interface, neurodegenerative diseases, computer-aided medical diagnosis, and Alzheimer's disease (AD).

Mohammad Tanveer (SM'18) is Associate Professor and Ramanujan Fellow at the Discipline of Mathematics of the Indian Institute of Technology, Indore. Prior to that, he spent one year as a Postdoctoral Research Fellow at the Rolls-Royce@NTU Corporate Lab of the Nanyang Technological University, Singapore. His research interests include machine learning, deep learning, optimization, biomedical engineering and applications to Alzheimer's disease and dementias. He has published over 40 referred journal papers of international repute. He is the recipient of the 2016 DST-Ramanujan Fellowship in Mathematical Sciences and 2017 SERB-Early Career Research Award in Engineering Sciences which are the prestigious awards of INDIA at early career level. He is a Senior Member of IEEE, professional member of ACM, editorial review board member of Applied Intelligence, Springer, Guest Editor of ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Applied Soft Computing, Elsevier, Multimedia Tools and Applications, Springer and Associate Editor for IEEE SMC 2019. He has also co-edited one book inSpringer on machine intelligence and signal analysis. He has also been lead organizer/ general chair and invited/plenary/keynote speaker in many international conferences and Symposiums. He was the Co-Chair of Special Session Proposal in IEEE SSCI 2018. Tanveer is currently the Principal Investigator of 07 major research projects funded by Government of India including Department of Science and Technology (DST), Science & Engineering Research Board (SERB), Council of Scientific & Industrial Research (CSIR).

Ram Bilas Pachori (SM'16) received the B.E.(Hons.) degree in electronics and communication engineering from Rajiv Gandhi Technological University, Bhopal, India, in 2001, and theM.Tech. and Ph.D. degrees in electrical engineering from the Indian Institute of Technology (IIT) Kanpur, Kanpur, India, in 2003 and 2008, respectively. He was a Postdoctoral Fellow with the Charles Delaunay Institute, University of Technology of Troyes, Troyes, France, from 2007 to 2008. He served as an Assistant Professor at Communication Research Center, International Institute of Information Technology, Hyderabad, India, from 2008 to 2009. He served as an Assistant Professor at the Discipline of Electrical Engineering, IIT Indore, Indore, India, from 2009 to 2013. He was a Visiting Scholar with the Intelligent Systems Research Center, Ulster University, Northern Ireland,U.K., in December 2014. He was an Associate Professor with the Discipline of Electrical Engineering, IIT Indore, from 2013 to 2017, where he has been working as a Professor since 2017. He is also an Associated Faculty with the Discipline of Biosciences and Biomedical Engineering, IIT Indore. He has also been a Visiting Professor with the School of Medicine, Faculty of Health and Medical Sciences, Taylor's University, Subang Jaya, Malaysia, since December 2018. He has more than 170 publications which include journal articles, conference articles, books, and book chapters. His publications have around 5800 citations,H index of 41, and i10 index of 98 (Google Scholar, April 2020).He has served on review boards for more than 95 scientific journals and served for scientific committees of various national and international conferences. His research interests are in the areas of biomedical signal processing, non-stationary signal processing, speech signal processing, signal processing for communications, computer-aided medical diagnosis, and signal processing for mechanical systems. He is a Fellow of IETE. He is an Associate Editor of Electronics Letters and Biomedical Signal Processing and Control journal; and an Editor of IETE Technical Review Journal.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Khan RU, Tanveer M, Pachori RB, Alzheimer's Disease Neuroimaging Initiative (ADNI). A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging. *Expert Systems*. 2020;1–22. https://doi.org/10.1111/exsy.12566