



Detection and Classification of Alzheimer’s disease from cognitive impairment with resting-state fMRI

PR. Buvaneswari¹ · R. Gayathri²

Received: 25 March 2021 / Accepted: 17 August 2021 / Published online: 3 September 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Detection and classification of Alzheimer’s disease (AD) are a demanding field of research in medicine throws light on innovative approach in detecting and classifying AD from cognitive impairment with resting-state functional magnetic resonance imaging (rsfMRI). The goal of this research is chiefly aimed to diagnose mild cognitive impairment (MCI) patients who essentially need support for medical intervention. A new concept is presented in classifying AD and MCI from rsfMRI using Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset. The images are preprocessed using some advanced technique to eliminate noise and parameter variations, and the preprocessed images are used for extracting the raw features. The rsfMRI is applied for feature selection processes in order to reduce feature dimensions using principal component analysis (PCA). The proposed kernel-based PCA-support vector regression (SVR) includes t-distributed stochastic neighbor embedding (tSNE) and polynomial kernel-based tSNE that are separately handled by significantly merging correlated local and class features. The kernel PCA method analysis the new features explicitly based on nonlinear mapping function in the data points of high-dimensional search. The kernel PCA method is suitable to analysis the new feature and feature importance in AD classification. The proposed kernel SVR method has the advantage of effectively analyzing the high-dimensional data to provide linear relationship and suitable to apply in MCI and AD data. The PCA method is applied for feature reduction process due to its capacity to select the relevant features and effectively analyzing the individual features. The proposed kernel-SVR method has the advantage of selecting the relevant features and avoid overfitting problem in the classifier. The SVR uses reduced features that are obtained from different reduction methods for classification of AD and MCI, a polynomial kernel based. The results showed that the proposed kernel-based PCA-SVR showed better average accuracy values 98.53% for kernel PCA when compared the existing models hippocampal visual features of 79.15% and deep neural network of 80.21%

Keywords Alzheimer’s Disease Neuroimaging Initiative · Kernel-based principal component analysis · Polynomial kernel-based PCA · Resting-state functional magnetic resonance imaging · Support vector regression · t-distributed stochastic neighbor embedding

List of Symbols

A	Matrix
D	Dimension of the space
d	Degree matrix
f_w	Factors

K	Kernel matrix
k_g	Gaussian kernel
L_g	Gyri surface areas
L^s	Sulci surface areas
P	Eigen decomposition
p_{ij}	Resemblance
Q	Eigen decomposition
u_α	An eigenvector of C
W	Weighted coefficient
w^G, w^W, w^g, w^s	Weighted vectors
X^G	Gray volumes
X^W	White volumes
x	Input image
x_N	Normalized image

✉ PR. Buvaneswari
bhuvaneswari@saveetha.ac.in

R. Gayathri
rgayathri@svce.ac.in

¹ Department of ECE, Saveetha Engineering College, Chennai, TamilNadu, India

² Department of ECE, Sri Venkateswara College of Engineering, Chennai, TamilNadu, India

x_{\max}	Maximum intensity in the image
x_{\min}	Minimum intensity in the image
y_1, y_2, \dots, y_N	Class of data

Greek symbols

α_j	Least square solution for mapping
β_i	Blending coefficients
ϕ	Mapping function
σ_j	Variance bandwidth
$\psi(A)$	Matrix polynomial
$\emptyset_G, \emptyset_W, \emptyset_g,$ and \emptyset_s	Kernel-based projection
ξ_n, ξ'_n	Slack variables

1 Introduction

Nearly 15% of 65 years' age-old persons are affected by MCI and 50% of them develop Alzheimer's disease within a half-decade. As it is very critical to diagnose AD at an initial stage, many types of researches have been intensely conducted on exploring the possibilities of prediction. Degradation of various brain regions is observed in AD, and thus, it is also termed a disconnection syndrome which is a type of neurodegenerative disorder [1]. The neuron network of the healthy brain is modified in the cases of MCI that progresses to AD. Cognitive function of the mentally healthy elderly people should be observed before initiating the treatment by diagnosing the conversion from MCI to AD in mentally healthy elderly people is major precedence for AD investigation [2, 3]. The AD patients with initial symptoms of MCI are cognitively impaired but appear like a healthy person with simpler memory functions. The structural sMRI is used to describe accurately the brain features like cortical thickness, areas, volumes, and the curvatures of gyri and sulci. These brain parameters have been extensively utilized to examine any modification of brain anatomy during the transformation from healthy to Alzheimer's disease. In general, AD patients have unstable memories and abnormal cognitive abilities and it is challenging to identify AD patients at initial stage, the diagnosis is on demand [4, 5]. Nevertheless, the diagnosis of the signs of MCI development continues to be an uninvestigated endeavor.

Many types of researches aimed to investigate AD driven modifications of the brain anatomy, neuroimaging types, namely sMRI, rsfMRI, diffusion tensor imaging (DTI), positron emission tomography (PET), resting-state functional MRI (rsfMRI) and so on, have been observed to be more instructional in establishing the biomarkers for MCI to AD development. Several earlier types of researchers used a single type of neuroimaging in the

identification of AD [6–9], and a similar approach is carried out to increase the prediction accuracy of AD compared to certain previous single image technique approaches [10–14]. The rsfMRI images have proven to be more effective means in determining the morbid physiology of functional connectome among AD patients as well as among neuropsychiatric or neurological patients [15–18]. However, in the existing models, diagnosing the patients with serious signs becomes complex because of the difference in demography and clinical trials; further, it becomes more complex due to largely varying and uncommon symptomatic trends that prevail among MCI patients [19]. Further, the natural connectome at rest establishes the transport links for information about the tasks. Therefore, in the proposed kernel-based PCA-SVR which is an ensemble rsfMRI has proven to effectively detecting AD. By using polynomial kernel-based PCA for feature selection process that overcomes the problem occurred in the existing researches. So, the researchers have attempted to show the capability of rsfMRI in the detection of MCI and AD patients [20–23]. Existing methods in AD classification have the limitation of irrelevant feature selection that creates overfitting problem in the classification. The selection of relevant features selection in rsfMRI data helps to avoid overfitting problem in the classifier. The proposed method applies kernel-PCA method for the feature selection to select the relevant features and applied to kernel-SVR for effective classification. The proposed kernel-SVR method overcomes the limitation of overfitting problem in existing method by relevant features selection for classification. The proposed kernel-SVR method has improved the accuracy of classification of AD up to 98.7% and existing deep learning method 84.97% accuracy in classification.

The organization of the present research work is as follows: Sect. 2 is the literature review section that surveys various existing techniques involved in AD detection using f-MRI. The Sect. 3 describes about the proposed ensemble-based PCA for dimensionality reduction for AD and MCI detection. Section 4 describes about the results and discussions for the present research work by comparing the proposed and existing models. The Sect. 5 describes about the conclusion and further improvement required in future for the research.

2 Literature survey

There are various researches undergone for AD detection using MRI techniques that are as follows:

Grieder et al. [9] developed default mode network (DMN) complexity and cognitive decline in mild AD. The Alzheimer's disease has yielded lower global DMN-MSE

than that of control group. The regional effects were localized at the left and it was true for most scales at the right hippocampi. However, the developed model used small sample size which was not enough in finding the differences among the mild AD and HC increased the error rate.

Eskildsen et al. [15] developed ADNI cohort using patterns of cortical thinning for prediction of AD in subjects with mild cognitive impairment. The predictions of accuracies were obtained from different stages from MCI to AD that improved by learning atrophy patterns for designing different stages of disease progressions. However, the main limitation of AD was that it involved vivo data for uncertain diagnosis which was the reason for probable AD diagnosis will confirm the diagnosis. Also, the patterns for neurodegeneration were found in many studies which has various uncertainties faced difficulties during correct diagnoses.

Peng et al. [17] developed a structured sparsity regularized multiple kernel learning for AD diagnosis. The developed regularizer was enforced on the kernel weights that are concisely selected from the feature set present in each of the homogeneous group. There was the presence of heterogeneous features that took an advantages from the dense norms. However, a general framework was assumed without any prior knowledge about the group of features and their importance was still required in the model.

Ahmed et al. [18] performed an automatic classification of AD subjects from MRI using hippocampal visual features. The main contribution of the model was that it considered visual features from the most involved region in AD (hippocampal area) and used a late fusion to increase precision results. The experiments showed that combining hippocampus features and cerebrospinal fluid (CSF) amount classification gave better accuracy especially when discriminating between AD and MCI when compared with only hippocampal visual features.

Dyrba et al. [19] performed multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. The developed model used fiber tract integrity as that measured diffusion tensor imaging (DTI), GM volume derived from structural MRI, and the graph-theoretical measures 'local clustering coefficient' and 'shortest path length' derived from resting-state functional MRI (rsfMRI) for the evaluation of the utility of the three imaging methods in automated multimodal image diagnostics, to assess their individual performance, and the level of concordance between them. However, the data used in the study were from AD patients in mild to moderate stages of the disease, but patients were not further stratified by disease severity.

Nguyen Thanh Du et al. [21] developed a 3D-deep learning based automatic diagnosis for AD using resting-state fMRI. To improve MMSE regression performance, feature optimization methods including least absolute shrinkage and selection operator and support vector machine-based recursive feature elimination (SVM-RFE) were utilized for the research. However, in the developed model 16 ICAs were selected as input data points for deep learning, but it was difficult to determine exactly what kinds of components impacted the broader neural network.

Problem Statement Existing methods in AD classification have the limitation of overfitting problem, lower efficiency in the large dataset, and requires the balance dataset to train the model. The existing methods have higher misclassification in the imbalance dataset due to the models bias to single-class. Some of the existing methods were higher performance in detection and poor performance in the classification of AD. Visual features were not sufficient to classify the AD due to similarity of the AD to the MCI. Deep learning models in AD classification suffer from the overfitting problem due to inconsistent feature selection.

The aforementioned surveys discussed about the existing methodologies involved for AD detection automatically using f-MRI and also stated the limitations of their studies. Therefore, the present research work performs an automatic classification method that uses the data from rsfMRI is aimed to classify three groups of patients between AD, MCI, and healthy individuals. The research aims to apply a polynomial kernel-based PCA technique to detect more, particularly the MCI patients among stable patients.

3 Materials and techniques

3.1 Proposed approach

A comprehensive method suggested is presented in Fig. 1. ADNI rsfMRI images of 210 patients (68 AD, 69 MCI, and 73 HC) are subjected to this research. Once the preprocessing of rsfMRI images, the ROI atlas is applied to detect the region of interests (ROIs) of the brain, particularly the gray matter, white matter, gyri, and sulci curvatures, and the mean matrix was computed using the volume and surface area of selected brain regions. Global and local features are computed. Nilearn is utilized for preprocessing and volumetric segmenting of brain regions from rsfMRI images. The areas and volumes of the gray and white matters, gyri and sulci in surface area, based on ROI atlas, are computed the features in proposed methods. The rsfMRI local and global features are merged to produce a feature vector of every subject. Four feature selection methods are used to determine an optimal feature subset for

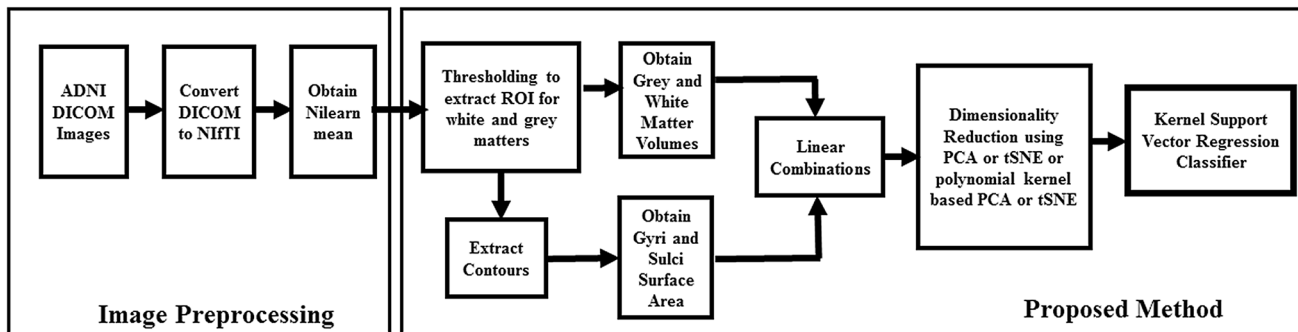


Fig. 1 Flow diagram for the proposed kernel-based PCA-SVR for AD detection

support vector regression (SVR) [20]. The SVR is trained to classify AD, MCI, and healthy applying reduced rsfMRI features derived from PCA, kernel PCA, tSNE, and kernel tSNE. The aim of the research is to classify the AD and MCI based on the rsfMRI data and to reduce the overfitting problem. The input data are applied with normalization for better representation, and ROI method is applied for segmentation. The kernel-PCA method is applied for the feature selection, and PCA method is applied for feature reduction. The selected feature is applied in kernel-SVR to classify the AD and MCI on rsfMRI data.

3.2 Description of subjects

In the present research, most of the subjects listed in ADNI are used from rsfMRIs Alzheimer’s Disease Neuroimaging Initiative (ADNI) image dataset: 68 AD patients with mean age 74.6 years and of 38 females, 69 MCI patients with a mean age of 72.2 years and 39 females, and 73 healthy individuals with a mean age of 73.3 years and 37 females. The AD patients, who are used in this research, get the scores of 20–30 on Mini-Mental State Examination (MMSE) and 0.5–1.0 on clinical dementia rating (CDR). The MCI patients get the scores of 25–32 on MMSE; loss of objective memory deprivation is quantified through education-adjusted scores of 23–29 on Wechsler Memory Scale Logical Memory II and 0.5 on CDR, non-appearance of notable impairment levels in other cognition spheres, necessarily is maintained in day to day activities, and lack

of dementia. The healthy cases are non-demented, non-MCI, and non-depressed and get the scores of 23–32 on MMSE and almost 0 on CDR. Demographic data of the subjects are depicted in Table 1. Analysis of variance (ANOVA) is applied to observe the data variance based on the group, and Fisher method is applied to analysis the significant set in the data. ANOVA measures the mean of the group, and Fisher is measured as the ratio of variation between the sample means and variation within the samples.

3.3 Image collection and preprocessing

The functional MRI images are compiled to the ADNI scanning standards. The fMRI images with DICOM format are available in ADNI. In order to preprocess them in Nilearn, the dicom format is converted in NifTI format. Then, typical preprocessing methods were employed on rsfMRI images utilizing Nilearn python library. The preprocessing involved to mitigate the motional effects [21] and to remove skull, and other non-brain tissue using watershed algorithm [24]. For segmenting the white matter, deeper gray matter, gyri and sulci structures [24], and detecting the boundaries of white and gray matters by normalizing the intensity, and discriminating the borders between gray and white matters, and between gray and CSF, the gradients by intensity were applied in order to precisely locate the tissues classes. The intensity normalization in the input images is given in Eq. (1).

Table 1 Demographic and clinical details

	Number	Male/Female	Age	MMSE Score	CDR Score
AD	68	30/38	74.6 ± 8.18	20.68 ± 2.85	0.036 ± 0.18
MCI	69	30/39	72.2 ± 12.24	27.11 ± 2.03	0.5 ± 0.02
Healthy	73	36/37	74.3 ± 11.76	29.24 ± 1.82	0.94 ± 0.34
<i>P</i> value	–	0.64*	0.86*	0.00028**	0.0001**

*Fisher test, **ANOVA Test

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x represents the input image, x_N represents the normalized image, x_{\max} represents the maximum intensity in the image, and x_{\min} represents the minimum intensity in the image. The AD-based features are needed to be extracted from the preprocessed images, and therefore, next the feature extraction process is undergone.

3.4 Extracting features from rsfMRI data

The features are extracted from the preprocessed rsfMRI using a vector that uses local and global measures are computed based on mean vectors of compiled rsfMRI. The local measure is clustering coefficients of the ROI structures, whereas the global measures are clustering coefficients of the transition between ROI structures. The surface area, contour curvatures, and the volumes of gray and white matters of ROI are used as features of the proposed algorithm. To perform the measurements between subjects comparatively, a structural normalization is carried out. The measurements of volume and area are divided by the sum of all volume and area of ROI, respectively. The cortical thickness and the contours are structurally normalized. Once the extraction of features is done, the features are normalized for every subject separately. The extracted features might create complexity due to higher dimensional data, and thus, the feature selection process is carried out further.

3.5 Feature selection

The extracted features are used for recognizing the pattern and solving the classification problems, but at the same time, handling higher dimensional data becomes a complex case, specifically in the applications of neuroimaging as the sample availability is less with a huge number of features. The selection of features is generally carried out to detect appropriate features, thus decreasing their dimensionality and improving the model abstraction. An effective selection method of features becomes a key element for a machine learning technique wherein the features with higher dimensions. Here, four kinds of feature selection techniques, namely principal component analysis (PCA), kernel-based PCA, t-distributed stochastic neighborhood embedding (t-SNE), and kernel-based t-SNE, were applied. Eventually, the kernel-based PCA method was so effective in recognizing the AD and specifically the AD prodromal condition [20].

The kernel PCA applies a mapping function to transform the design space to the feature. A mapping function is denoted as $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^L$, where L can be arbitrarily large

without through explicitly defining ϕ . For input dataset X , the new covariance matrix $P_{L \times L}$ is $P = \frac{1}{M} \Phi \Phi^T$, where columns vectors $\vec{\phi}(\vec{x}^i) \in \mathbb{R}^L$ form $\Phi \in \mathbb{R}^{L \times M}$. The Eigen decomposition P becomes $P = V \Lambda V^T$, where eigenvectors of $\Lambda \in \mathbb{R}^{L \times L}$ and P forms $V \in \mathbb{R}^{L \times L}$. The Eigen problem long and expensive is depend on value of L . The kernel trick helps to solve this problem. The kernel matrix $K \in \mathbb{R}^{M \times M}$ is given in Eq. (2).

$$K = \Phi^T \Phi \quad (2)$$

An RBF kernel function is considered without generality loss, as given in Eq. (3).

$$k(\vec{x}^i, \vec{x}^j) = \exp\left(-\frac{\|\vec{x}^i - \vec{x}^j\|_2^2}{2\sigma^2}\right) \quad (3)$$

3.6 Feature reduction methods

Feature reduction technique further applied in the selected feature to remove the highly correlated features and this avoid overfitting problem in the classifier. The kernel-PCA method selects the features from rsfMRI and selected features were applied in feature reduction methods. In the feature selection process, to detect appropriate features, decreasing their dimensionality and improving the model abstraction PCA, t-SNE have been applied to recognize patterns by combining the diverse features into vectors obtained from particular modalities [25]. PCA computation involves lesser effort, and thus, it is used for any real-time uses. PCA is the best-known technique for decreasing the size of a higher dimensional feature array, particularly in nonsupervisory learning.

3.6.1 Theory of PCA

The selected features are applied in the PCA method as x_i to reduces the features and avoid high variance in the model. PCA is a factorial analysis technique statistically based to identify the significant variables from distributed data using absolute variance. PCA originally finds the optimal variations in the data that are characterized by a value set are arranged in one-dimensional form. Such data numerical values can be of integer or floating-point representations and can be either discrete or continuous [26, 27]. To derive lower-dimensional data, the data are mapped on to an earlier maximal variance eigenvector direction [19]. For data with large dimensions, few dimensions persist to be zero data length. In such a case, it is onerous to show their eigenvectors that projected will inhabit in the subspace ranged by the data.

$$\lambda_\alpha u_\alpha = Cu_\alpha = \frac{1}{N} \sum x_i x_i^T u_\alpha = \frac{1}{N} \sum (x_i^T u_\alpha) x_i; \tag{4}$$

In which $u_\alpha = \sum \frac{(x_i^T u_\alpha)}{N \lambda_\alpha} x_i = \sum \alpha_i^\alpha x_i$

u_α represents an eigenvector of C .

Eventually, a new set of data t is determined, and thus, projection onto newly reduced space can be derived by the means

$$u_\alpha^T t = \sum_i \alpha_i^\alpha x_i^T t = \sum_i \alpha_i^\alpha K(x_i t) \tag{5}$$

The Eq. (2) is quite a usual one and more basic for numerous kernel methods. A method is needed to compute the K matrix. This indicates that a kernel is required to substitute for $x_i \rightarrow \Phi(x_i)$ and to describe $K_{ij} = \Phi(x_i)\Phi(x_j)^T$, where $\Phi(x_i) = \Phi_{ix}$.

3.6.2 Theory of kernel-based PCA

It is complex particularly to center the features in their space. Nevertheless, the resulting algorithm will depend on a kernel, and thus, feature centering becomes possible [28]. The kernel that does centering can be computed accordingly with non-centered kernel and nonessential features as stated

$$K_c(t_i, x_j) = \left[\Phi(t_i) - \sum_k \Phi(x_k) \right] \left[\Phi(x_j) - \sum_l \Phi(x_l) \right]^T \tag{6}$$

The above expression can be modified by carrying out the identical calculation using $K_c(t_i, x_j)$ and $K_c(x_i, x_j)$

$$K_c(t_i, x_j) = K(t_i, x_j) - k(t_i)1_j^T - 1_i k(x_j)^T + k_i 1_i 1_j^T \tag{7}$$

3.6.3 Theory of t-distributed Stochastic Neighborhood Embedding (tSNE)

A tSNE is computed with two main steps. At first, tSNE draws a probability distribution across pair of high dimension objects in such a way similar objects which have high probability will get selected most likely while dissimilar objects which have a substantially low probability will get unselected most likely. Next, tSNE reflects the probability distribution of identical data in lower-dimensional space, and thus, it decreases relatively the Kullback–Leibler divergence among two neighboring data clusters located in the space [29].

The tSNE intends to learn the feature data present in $y_1, y_2, \dots, y_N, y_i \in R^{D \times N}$ where D is the dimension of the space and identifies the resemblance between data at best

through p_{ij} . Here, the target disease type labels y are associated with features that combined linear, namely gray volumes (X^G), white volumes (X^W), gyri surface areas (L^g) and sulci surface areas (L^s) by following expressions

$$y = X^{GT} w^G + e^G \tag{8}$$

$$y = X^{WT} w^W + e^W \tag{9}$$

$$y = A^{gT} w^g + e^g \tag{10}$$

$$y = A^{sT} w^s + e^s \tag{11}$$

In which $w^G \in R^D, w^W \in R^D, w^g \in R^D$ and $w^s \in R^D$ are the vectors of weighted coefficients of feature vectors, whereas $e^G \in R^N, e^W \in R^N, e^g \in R^N$ and $e^s \in R^N$ are the vectors comprise noises obtained individually from regular normal distributions.

In itself, it estimates the resemblances q_{ij} between data x_i^G and x_i^W, x_j^G and x_j^W, a_i^g and a_i^s, a_j^g and a_j^s , using similar approach. q_{ij} is thus represented as

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq i} \left(1 + \|y_i - y_k\|^2\right)^{-1}} \tag{12}$$

The heavy-tailed student t-distribution has 1 degree of freedom (df) and hence like Cauchy distribution, this is used to evaluate common features among lower-dimensional data to allow dissimilar features to be described distantly in data space, i.e., it sets q_{ii} to zero.

The data values y_i that situated in data space are detected by optimizing the divergence of the Kullback–Leibler between the distributions of P and Q to minimum

$$K_L(P||Q) = \sum_{i \neq j} p_{ij}^G \log \frac{p_{ij}^G}{q_{ij}} + \sum_{i \neq j} p_{ij}^W \log \frac{p_{ij}^W}{q_{ij}} + \sum_{i \neq j} p_{ij}^g \log \frac{p_{ij}^g}{q_{ij}} + \sum_{i \neq j} p_{ij}^s \log \frac{p_{ij}^s}{q_{ij}} \tag{13}$$

In order to obtain the optimal minimum divergence of Kullback–Leibler over data, gradient descent is performed. This optimization results better to identify common features between high and low dimensional data.

The optimization problem is solved using Eqs. (12) and (13) that apply the gradient technique [30, 31]. The features of gray and white matters are selected according to the presence of nonzero elements that correspond to weighted coefficients in $W = [w^G w^W w^g w^s]$.

3.6.4 Theory of kernel-based tSNE

On expanding non-parametric-based dimensional reduction technique, t-SNE is a type of categorical projection

described by setting as $x \rightarrow f_w(x) = y$ and it will be optimizing the factors of f_w instead of mapping on coordinates. This performance expansion from a non-parametric type to a parametric type was used in numerous descriptions [32]. f_w is supplied to a deep auto-encoder to train initially in a normal way, and subsequently, the parameters are fine-adjusted enabling cost function optimization of tSNE by projecting the data to the mapped space. Due to the high flexibility of the deep learning model, kernel tSNE allows the model to achieve good accuracies using a substantial amount of data through training that is accomplished perfectly. As a result, several parameters are constituted in deep auto-encoders and consequential mapped space usually appears complex since the vast amount of data and time needed for training. The notion of parametric type applying in the optimization of the cost function through non-parametric-based dimensionality reduction is evident in tSNE and confirmed its performance in respect of piecewise linear functions. Using simpler functions, a fairly simple concept is derived to train with less amount of data in lesser time. But, the flexibility of the resulting map is confined as opposed to the complete tSNE method since the local nonlinear property may not be captured during local linear mapping. As the first step in the operation of kernel tSNE, projection of f_w is carried out applying linear Gaussians function, where its coefficients are expected to be trained based on its cost function by directly using a false inverse of train data while projecting in tSNE.

The map $f_w = y$ based on kernel tSNE is represented as follows

$$x \rightarrow y(x) = \sum_j \alpha_j \frac{k_g(x, x_j)}{\sum_l k_g(x, x_l)} \tag{14}$$

$\alpha_j \in Y$ are the parameters of data in the projected space, x_j is a vector contains sample data, j takes account of subset X' sampled from data interval $[x_1, \dots, x_m]$. k_g is defined as a Gaussian kernel and σ_j as variance bandwidth.

$$k_g(x, x_j) = e^{-\left(\frac{0.5\|x-x_j\|^2}{\sigma_j^2}\right)} \tag{15}$$

For a confined lesser bandwidth, t-SNE appears the same for the inputs derived from X' . Eventually, α_j decides the projection of y_j on x_j . For remaining \mathbf{x} , an interpolation is performed proportionally according to the length of \mathbf{x} of x_i that sampled from X' . The mapping of this type constitutes an extensive linear map in the manner that the training is carried out particularly in a simpler method for sampled x_i . Then, α_j is computed using the least square solution for mapping. As \mathbf{A} constitutes the coefficients α_j , the entries of after normalization of Gram matrix K are represented as follows

$$[K]_{i,j} = k_g(x_i, x_j) / \sum_l k_g(x_i, x_l) \tag{16}$$

\mathbf{Y} denotes the mapping matrix with y_i elements. Then, least square error is

$$\sum_i \|y_i - y(x_i)\|^2 \tag{17}$$

α_j is represented as follows

$$A = K^{-1} \cdot Y \tag{18}$$

With K^{-1} as K inverse.

A traditional tSNE is employed on subsets of X' to obtain vectors for training. This categorical systemic method is utilized to obtain projecting parameters. Using obtained projection, the whole \mathbf{X} is then projected linearly by adopting y projections. It can be considered as another form of extended dimension reduction. σ_i is projection variance bandwidth that occurs to be an important parameter since it determines the flexibility and coarseness of consequent kernel-based projection. The better technique explores these parameters through the selection of σ_i as a distance measure of x_i with its closest neighbor present in X' , the scaling parameter usually takes the value positive close to zero. This parameter is acquired innately as a minimal measure in a way the elements comprise in K will be in the designated range.

3.6.5 Theory of polynomial kernel

Polynomial filtering [33, 34] method enables to influence of PCA or tSNE models with a closer approximation to a low rank. $\psi(A)$ specifies a matrix polynomial of A matrix of degree d , and it is represented as follows

$$\psi(A) = \xi_d A^d + \xi_{d-1} A^{d-1} + \dots \dots \dots \xi_1 A + \xi_0 \tag{19}$$

If assuming that A is normal which described by the property $A^T A = A A^T$, it results from $A = Q \Lambda Q^T$ as its eigen decomposition

$$\psi(A) = \psi(Q \Lambda Q^T) = Q \psi(\Lambda) Q^T \tag{20}$$

Therefore, the polynomial applies to a results in a polynomial of eigenvalues. This can be comfortably represented in applying polynomial filtering to derive directly the similarity vector by avoiding eigenvalue computations completely.

The \mathbf{x} data are changed using $\tilde{x} = x - \mu$. To determine the similarity magnitude \mathbf{s} , ψ polynomial of $\tilde{A}^T \tilde{A}$ is described as follows

$$\begin{aligned}
 s &= \psi(\widetilde{A}^T \widetilde{A}) \widetilde{A}^T \widetilde{x} \\
 &= \psi(V \Sigma^T \Sigma V^T) V \Sigma^T U^T \widetilde{x} \\
 &= V \psi(\Sigma^T \Sigma) V^T V \Sigma^T U^T \widetilde{x} \\
 &= V \psi(\Sigma^T \Sigma) \Sigma^T U^T \widetilde{x}
 \end{aligned}
 \tag{21}$$

Choosing $\psi(t)$, the polynomial appropriately will enable us to understand this approach as a compromise between the correlation and PCA. If polynomial ψ is unconstrained, one can use any kind of function. While $\psi(t) = 1 \forall x$, $\psi(\Sigma^T \Sigma)$ changes to a resemblance operator. Hence, the discussed approach is similar to a correlation method.

The data can be estimated using polynomial filtering. Therefore, applying polynomial filtering to PCA or tSNE can result in an almost identical alike eigenvalue decomposition method by devolving the expensiveness of eigenvalue decomposition. Further, the condition to hold back the matrices of PCA or tSNE may be completely removed being the demand to restore these matrices when the subspace used proactively learn the differences. The cutoff option is fairly similar to the issue of choosing the parameter k in the PCA or tSNE method. However, the salient difference between PCA and tSNE is choosing a higher k value in PCA or tSNE may result from these techniques expensive. But when selecting a high cutoff in polynomial filtering will drastically reduce the expense [35].

3.6.6 Theory of kernel-based support vector regression classifier

The opted features are supplied to a kernel support vector regression (kSVR) [36, 37] classifier, thereby the complementary features, the GM and WM volume and, gyri and sulci surface areas are integrated. After opting the features, the lowered dimensional N sampled training data $\{\widehat{X}_n^G, \widehat{X}_n^W, \widehat{A}_n^g, \widehat{A}_n^s\}_{n=1}^N$ including the corresponding label produce the vector $\{y\}_{n=1}^N$. The kSVR applies simpler equation that uses ε a robust loss function. Assuming $\widehat{A}_n^g = \widehat{X}_n^g$ and $\widehat{A}_n^s = \widehat{X}_n^s$, then

$$\min_{w^G, w^W, w^g, w^s, b, \xi, \xi'} \frac{1}{2} \sum_{i \in \{G, W, g, s\}} \beta_i \|W^i\|^2 + C \sum_{n=1}^N \xi_n + C \sum_{n=1}^N \xi'_n
 \tag{22}$$

where

$$\begin{cases} \sum_{i \in \{G, W, g, s\}} \beta_i \left((W^i)^T \vartheta_i \widehat{X}_n^i + b \right) - y_n \leq \varepsilon + \xi_n \\ y_n - \sum_{i \in \{G, W, g, s\}} \beta_i \left((W^i)^T \vartheta_i \widehat{X}_n^i + b \right) \leq \varepsilon + \xi'_n \\ \xi_n, \xi'_n > 0 \quad n = 1, 2, \dots, N \end{cases}
 \tag{23}$$

w^G, w^W, w^g, w^s are the weighted vectors, $\vartheta_G, \vartheta_W, \vartheta_g$, and ϑ_s are the kernel-based projection functions of the features, β_i are blending coefficients with the bounds $\beta_i \geq 0$ and $\sum_{i \in \{G, W, g, s\}} \beta_i = 1$. Slack variables are denoted by ξ_n, ξ'_n , and bias is denoted by b . The dual function is obtained as follows

$$\begin{aligned}
 \max_{\alpha, \alpha'} & -\frac{1}{2} \sum_{n,m=1}^N (\alpha'_n \\
 & - \alpha_n) (\alpha'_m - \alpha_m) \sum_{i \in \{G, W, g, s\}} \beta_i k_i \left(\widehat{X}_n^i, \widehat{X}_m^i \left(\widehat{A}_n^i, \widehat{A}_m^i \right) \right) \\
 & - \varepsilon \sum_{n=1}^N (\alpha'_n + \alpha_n) + \sum_{n=1}^N (\alpha'_n - \alpha_n) y_n
 \end{aligned}
 \tag{24}$$

where $\sum_{n=1}^N (\alpha'_n - \alpha_n) = 0$ and $0 \leq \alpha_n, \alpha'_n \leq C$, $n = 1, 2, \dots, N$. $(\widehat{X}_n^i, \widehat{X}_m^i)$ and $(\widehat{A}_n^i, \widehat{A}_m^i)$ are newly developed vectors of low dimension in testing the rsfMRI. A polynomial described by $\vartheta_G, \vartheta_W, \vartheta_g$, and ϑ_s is applied. The classifier performs the classification using the equation [30] after the train of kSVR.

$$\begin{aligned}
 f(\widehat{X}^G, \widehat{X}^W, \widehat{A}^g, \widehat{A}^s) \\
 = \sum_{n=1}^N (\alpha'_n - \alpha_n) k_n \left[\left((\widehat{X}_n^G, \widehat{X}_n^W, \widehat{A}_n^g, \widehat{A}_n^s), (\widehat{X}^G, \widehat{X}^W, \widehat{A}^g, \widehat{A}^s) \right) \right] + b
 \end{aligned}
 \tag{25}$$

Above equation provides the classification of AD based on selected features and the metrics are evaluated from the classified result. Using various PCA-based feature dimensionality reduction techniques, the complexity of the features is reduced.

4 Experimental setup

The present section explained about the result and discussion of the proposed framework. In the study, MATLAB (2018A) was utilized for experimental evaluation. In the scenario, the proposed kernel-based PCA-SVR performance was compared with dissimilar classifiers and some prior research works on the database ADNI. So, to determine the advantage of the proposed method during classification for AD, MCI, and HC from rsfMRI images, a considerable number of experiments are carried out.

5 Results and discussion

The proposed technique will be selecting both global and local resembling features. The experiments are carried out by combining the volume and area features. The training is then carried out with rsfMRI images supply to the kSVR classifier. For training the classifier, 210×7 variable data are applied by combining. The combined performance of the classifier and the feature reduction methods largely influence the accuracy of the classifier. The features of gray matter, white matter, gyri, and sulci are considered here with equal weights in classifying the mental disease. The proposed regression classifier applies the polynomial kernel.

5.1 Preprocessing for sliced rsfMRI

The preprocessing of the rsfMRI used Nilearn technique results in the volume data of gray and white matters and area data of gyri and sulci. The preprocessing involved to

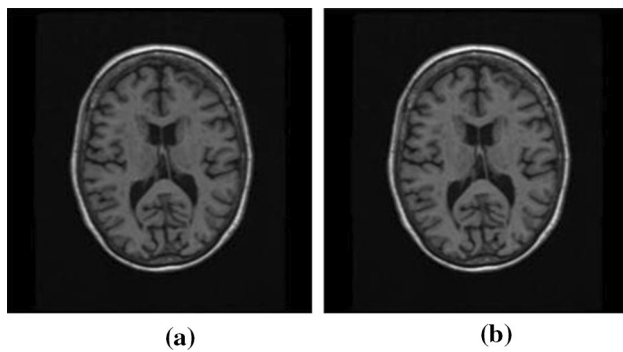


Fig. 2 a Input image, and b normalized image

remove cerebrospinal fluid (CSF) and to mitigate the emotional effects. The input and normalized images are shown in Fig. 2.

Then, the ROI atlas was applied to detect the region of interests (ROIs) of the brain particularly the gray matter, white matter, gyri, and sulci curvatures, and the mean matrix was computed using the volume and surface area of the selected brain regions. These data were collected in excel and those depicted in Table 2. These collected data from 210 subjects are obtained from ADNI and their particulars are described in Table 1. The information of subjects, namely subject ID (SUB_ID), Age, Gender, Intelligent score, and Intelligence Test, can be acquired from the LONI website. The AD patients with the scores of 20–30 on MMSE and 0.5–1.0 on CDR, MIC patients with the scores of 25–32 on MMSE, the scores of 23–29 on Wechsler Memory Scale Logical Memory II and 0.5 on CDR, and the healthy cases with the scores of 23–32 on MMSE and almost 0 on CDR are used.

Figure 3 shows the correlation table using the data collected in excel. The correlation between the white matter volume and gyri surface area + 0.7, with gray matter volume is + 0.6 and with total surface area is + 0.8. The correlation between gray matter volume and sulci surface area is + 0.8 and with total surface area is + 0.9. The correlation between white matter volume and gyri to sulci surface area is negatively correlated by 0.5.

Figure 4 shows the scatter matrix using the data collected in excel. White matter volume shows the linear relationships of an almost same slope with gray matter volume and gyri surface area for the cases of MCI and AD. White matter volume shows the linear relationships of the larger slope with sulci surface area and total surface area of AD compared with MCI and gyri surface area for the cases

Table 2 The data that extracted after image preprocessing (12 Samples)

Label	White matter volume (WV)	Gray matter volume (GV)	Gyri surface area (GA)	Sulci surface area (SA)	Total surface area (TSA)	GV/SV	SA/GA
AD	241,007	54,859	14,287	7011	5359	0.228	0.491
	243,231	54,872	14,298	7397	5378	0.226	0.517
	261,029	59,472	15,507	7608	5837	0.228	0.491
	278,827	64,069	16,720	7819	6300	0.23	0.468
MCI	80,686	26,478	4536	1011	565	0.328	0.223
	95,756	31,377	5398	1206	618	0.328	0.223
	112,149	36,651	6250	1442	690	0.327	0.231
	128,531	41,918	7095	1680	763	0.326	0.237
CN	2113	67,966	13,332	6823	3965	32.166	0.512
	3103	68,021	13,367	7158	3969	21.921	0.535
	1089	68,024	13,370	7296	3973	62.465	0.546
	1131	68,034	13,377	7431	4009	60.154	0.556

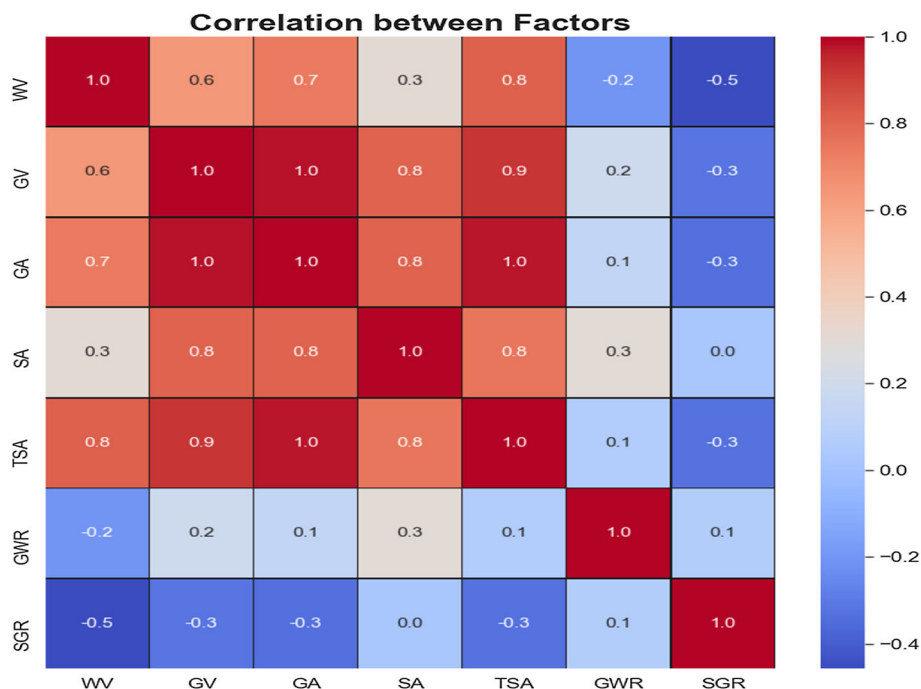


Fig. 3 The correlation table is obtained using the data provided in Table 2. GA, SA, and TSA that observed in the table are gyri area, sulci area, and total surface area

of MCI and AD. White matter volume shows complete independence with gray matter volume, gyri surface area, sulci surface area, and total surface area in the case of HC. Gray matter volume shows the linear relationships of the almost same slope with gyri surface area in all the cases. Gray matter volume shows the linear relationships of the larger slope with sulci surface area and total surface area in AD case in comparison with MCI and HC cases. Gyri to sulci surface area ratio decreases exponentially with gray matter volume in HC.

5.2 Segmenting sliced rsfMRI

In a classification of brain images, the high-dimensional features normally affect available data. Therefore, decreasing the dimensionality and thereby picking up the feature is of enormous significance and gain interest. Here, features like gray and white matter volumes, and gyri and sulci surface area are utilized and supplied to kernel free and kernel-dependent PCA and tSNE to select features and classified using kSVR. The area shaded pixel counts of gray and white matters, and contour pixel counts of gyri and sulci are obtained from rsfMRI. The chosen features have shown correlation better, and supply them by incorporating complementary data for reduction method and regression classifier. Therefore, two kinds of the feature selection procedure are used for dimensionality reduction by including the polynomial kernel, which this attempt will

have an advantage to solve the problems of kernel-less dimensionality reduction, for example, less likely to consider globally correlated features.

Figures 5, 6 and 7 show the preprocessing outputs of rsfMRI (z -axis cut the coordinate value of -200) of AD, MCI, and HC cases, respectively. Figures 5, 6 and 7a depict actual brain scan images with skull and other brain tissue. Figures 5, 6 and 7b show an image after stripping the skull. Figures 5, 6 and 7c show Nilearn ROI atlas. Figs. 5, 6 and 7d show the physical mask. Figures 5, 6 and 7e show the segmentation of gray matter and gyri surface from the atlas image. Figures 5, 6 and 7f show the segmentation of white matter and sulci surface Threshold segmentation is applied to extract contour so to evaluate the surface area and to extract area to evaluate volume. The segmented images are applied for the kernel-PCA for feature selection process and selected features are applied to kernel-SVR for classification.

5.3 Performance of feature dimensionality reduction

For reducing the highly correlated feature dimension, this research applied PCA [30], third-degree polynomial-based kernel PCA, tSNE [31], and third-degree polynomial-based kernel tSNE. The proposed kernel-PCA method provides the linear relation for the features and linear relation is easy for SVR to classify the data. The concentrated features here

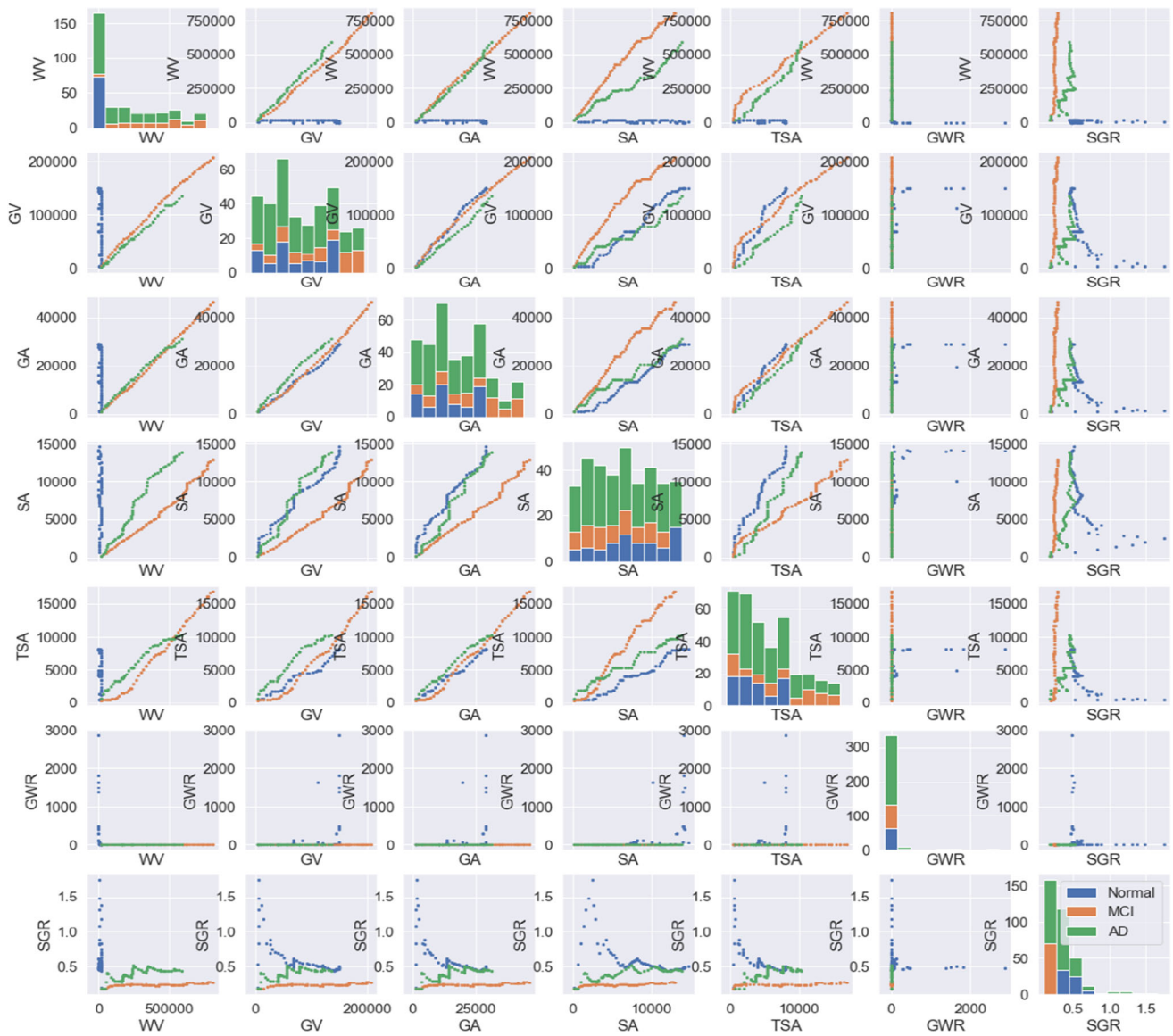


Fig. 4 The scatter matrix is obtained using the data provided in Table 2

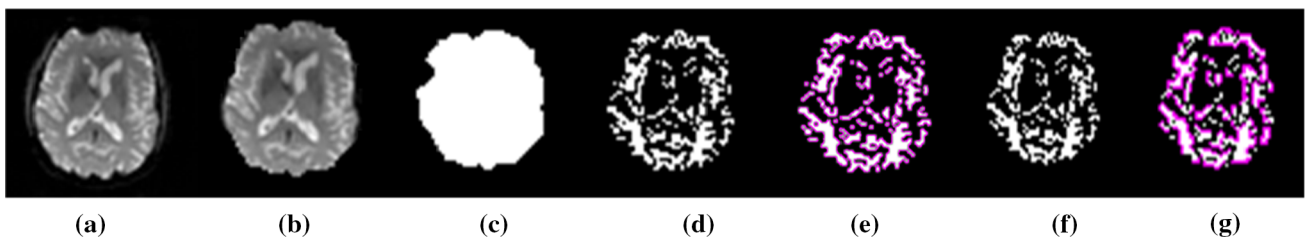


Fig. 5 The outputs of image preprocessing of rsfMRI of AD case **a** actual brain image **b** skull removed image **c** ROI atlas **d** physical mask **e** segmented gray matter and gyri contour **f** segmented white matter and sulci contour



Fig. 6 The outputs of image preprocessing of rsfMRI of MCI case **a** actual brain image **b** skull removed image **c** ROI atlas **d** physical mask **e** segmented gray matter and gyri contour **f** segmented white matter and sulci contour

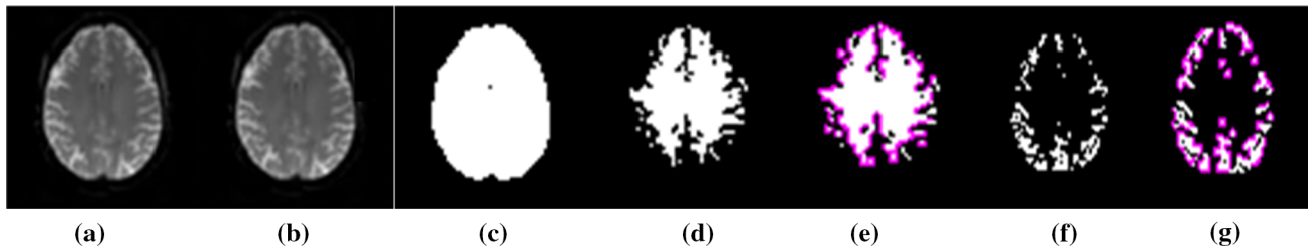


Fig. 7 The outputs of image preprocessing of rsfMRI of HC case **a** actual brain image **b** skull removed image **c** ROI atlas **d** physical mask **e** segmented gray matter and gyri contour **f** segmented white matter and sulci contour

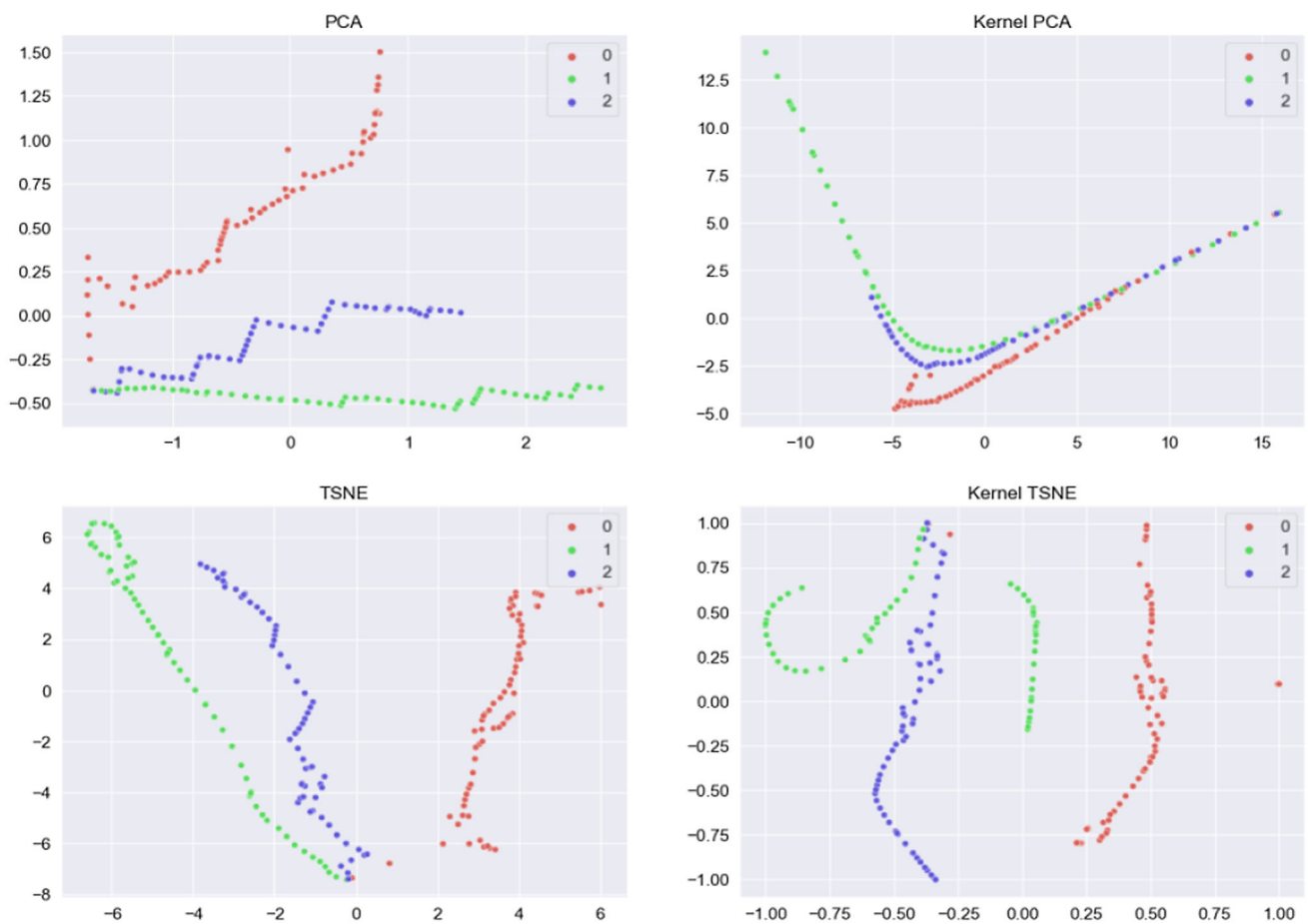


Fig. 8 Multi-dimensional plots of different feature reduction techniques applied on data using PCA, third-degree polynomial-based kernel PCA, tSNE, and third-degree polynomial-based kernel tSNE

Table 3 Comparison of proposed different dimensionality reduction methods used in classification of Alzheimer’s disease using kernel SVR

	PCA			Kernel PCA			tSNE			Kernel tSNE		
	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD
Sensitivity	91	100	94	99.2	98.6	98.6	93	68	98	91	84	82
Specificity	100	92	100	98.8	98.9	98.3	100	96	83	100	87	92
Precision	100	87	100	99.2	99.4	98.9	100	90	74	100	76	83
Accuracy	97	95	98	98.5	98.4	98.7	97	87	88	96	86	89

Table 4 Comparison of proposed different dimensionality reduction methods used in classification of Alzheimer’s disease using kernel-PCA

	Kernel SVR			SVR			KNN			DNN		
	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD
Sensitivity	99.2	98.6	98.6	99.02	97	97	94	65	96	89	83	84
Specificity	98.8	98.9	98.3	97	98	98	98	94	81	99	85	90
Precision	99.2	99.4	98.9	98	98	97	98	89	72	99	74	82
Accuracy	98.5	98.4	98.7	97	94	97	96	86	86	97	83	88

Table 5 Comparison of the proposed model with different feature optimization methods

	Kernel PCA			Relief			ICA			LDA		
	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD
Sensitivity	99.2	98.6	98.6	98.01	96	97.21	91	70	96	89	82	81
Specificity	98.8	98.9	98.3	97	97.2	97.3	99	94	81	99	86	90
Precision	99.2	99.4	98.9	98	98.21	96.9	98	89	73	98	82	82
Accuracy	98.5	98.4	98.7	97	98.14	96.7	96	85	86	97	84	84

are gray matter and white matter volumes, gyri and sulci surface area, and their ratios. Figure 8 depicts multi-dimensional scatter plots of different reduction techniques. In Fig. 8, X-axis denotes the Principal Components 1 and Y-axis denotes the Principal Components 2.

5.4 The performance of kernel-based SVR classifier

The extracted seven features of 210 are supplied to four types of reduction methods. The output from reduction methods is used for training and testing the performance of kernel-based SVR. The proposed kernel-SVR method has the higher performance compared to existing methods in AD classification. The proposed kernel-SVR method has the advantage of selecting the relevant features based on kernel-PCA and avoid overfitting problem in the classification. The proposed kernel-SVR method has the accuracy of 98.7%, and existing deep learning [18] method has 84.97% accuracy. The existing methods have the limitation of overfitting problem that affects the performance of the model. The performances of kernel SVR supplied with output of reduction methods shown in Table 3. Further, the Table 4 shows the proposed method using kernel PCA has

given better result in comparison with the previous works of Jialin Peng [17], developed a structured sparsity regularized multiple kernel learning for AD diagnosis that achieved accuracy of 96.1%, but the general framework was assumed without any prior knowledge about the group of features and their importance was still required in the model. Beheshti et al. [16] applied genetic algorithm for the AD classification and Dyrba et al. [19] applied multi-kernel SVM for the AD classification. Olfa Ben Ahmed [18], performed an automatic classification of AD subjects from MRI using hippocampal visual features achieved accuracy of 87%, but the CSF amount classification failed to give better accuracy especially when discriminating between AD and MCI due to higher dimensionality of data. Nguyen Thanh Duc [21] used 3D-deep learning based automatic diagnosis for AD that achieved accuracy of 84.97%, but it was difficult to determine exactly what kinds of components impacted the broader neural network. The used structured sparsity regularized multiple kernel learning, hippocampal visual features, and 3D-deep learning based automatic diagnosis for reducing the dimensionality of large datasets with less computational complexity. The proposed kernel SVR is a most efficient algorithm for managing large amount of datasets with gives better

Table 6 Comparative analysis of the proposed model with existing techniques

Methods	Dataset	Sensitivity	Specificity	Precision	Accuracy
Kernel PCA	CN	99.2	98.8	99.2	98.5
	MCI	98.6	98.9	99.4	98.4
	AD	98.6	98.3	98.9	98.7
Peng et al. [17]	CN	85.6	69.8	NA	80.3
	MCI	65.9	82.7	NA	76.9
	AD	97.3	94.9	NA	96.1
Ahmed et al. [18]	CN	70.73	81	NA	78.22
	MCI	75	83.34	NA	72.23
	AD	75.5	70	NA	87
Duc et al. [21]	CN	67.73	95.9	NA	86.2
	MCI	67.2	96.86	NA	85.79
	AD	65.63	90.03	NA	84.97
Beheshti et al. [16]	CN	94.3	95.4	97.1	97.4
	MCI	93.2	93.5	97.4	96.2
	AD	93.6	94.5	96.3	96.5
Dyrba et al. [19]	CN	97.5	96.2	96.3	97.6
	MCI	96.1	96.3	97.2	97.5
	AD	96.5	96.4	95.6	95.7

Structured sparsity regularized multiple kernel learning [17], hippocampal visual features [18], 3D-deep learning based automatic diagnosis [21], genetic algorithm [16], multi-kernel SVM [19]

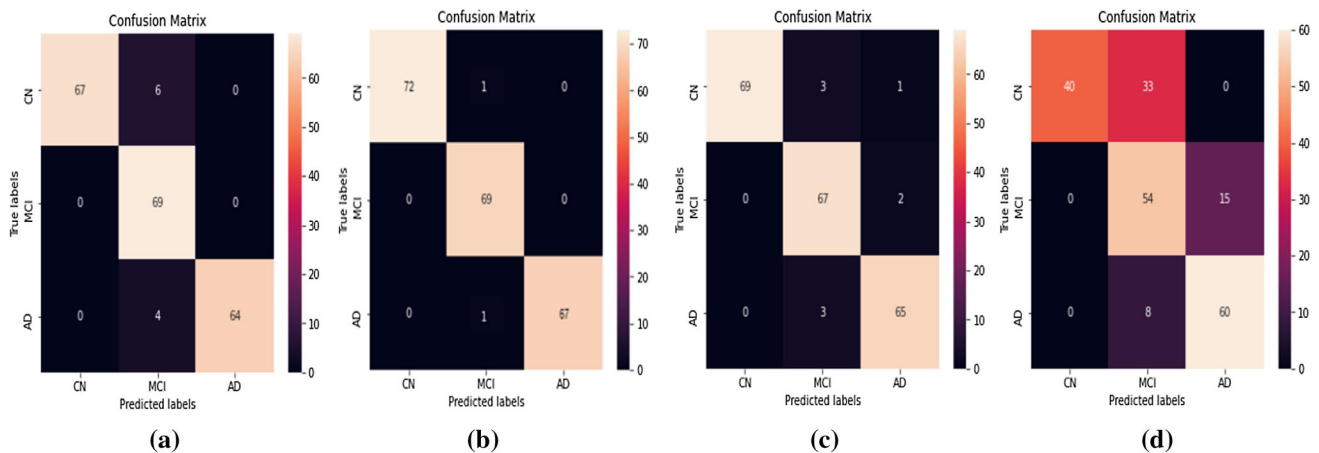


Fig. 9 Confusion matrices of the kSVR classifier while classifying the AD, MCI, and CN **a** uses the output from PCA **b** uses the output from kernel PCA **c** uses the output from tSNE **d** uses the output from kernel tSNE

prediction and high classification accuracy results (Tables 5 and 6 and Fig. 9).

6 Conclusion

The suggested method is to classify Alzheimer's disease utilizing the reduced features derived from the ROI atlas of rsfMRI images. This classification is done using high correlation between different features of gray matter and

white matter volumes along with the surface area of gyri and sulci. These features are obtained by preprocessing the rsfMRI. Then, the features are supplied to feature selection techniques to reduce their dimensions, i.e., transforming initial data space to more correlated lower dimension data space by efficiently overlaying parameters over four fundamental features. Nilearn-based ROI detection reduces the influence of the presence of noises and effectively improves the detection of edges and contours which are important to estimate volume and surface area. A model is

proposed to derive initially the highly correlated lower-dimensional features from actual basic features that are extracted from rsfMRI and then is supplied to kernel-based SVR classifier to diagnose AD and MCI. The proposed kernel-SVR method selects the relevant features from the rsfMRI data based on kernel-PCA and reduces the overfitting problem based on relevant features. The proposed kernel-SVR method has higher performance in the classification of AD and MCI data. In comparison with the existing techniques, the currently suggested technique has no limitations in using rsfMRI obtained from various modes to handle scan parameter differences, and thus, the built classifier model can predict AD and MCI. The proposed model uses more correlated identical data through feature reduction for precise diagnosis of AD and MCI. The present research contributes better option to detect AD and MCI from rsfMRI that can be used for the crucial needs of medical interventions. The proposed kernel-SVR method has lower performance in the imbalance dataset, and deep learning method is applied in future work to overcome the problem.

Funding This study was not funded by any organization.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT (2011) Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med* 1:a006189
- Huang L, Jin Y, Gao Y, Thung K-H, Shen D (2016) Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiol Aging* 46:180–191
- Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Alzheimer's Disease Neuroimaging (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104:398–412
- Hojjati SH, Ebrahimzadeh A, Khazaei A, Babajani-Feremi A, Alzheimer's Disease Neuroimaging I (2017) Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. *J Neurosci Methods* 282:69–80
- Khazaei A, Ebrahimzadeh A, Babajani-Feremi A (2015) Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clin Neurophysiol* 126:2132–2141
- Khazaei A, Ebrahimzadeh A, Babajani-Feremi A (2016) Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging Behav* 10:799–817
- Lin Q, Rosenberg MD, Yoo K, Hsu TW, O'Connell TP, Chun MM (2018) Resting-state functional connectivity predicts cognitive impairment Related to Alzheimer's disease. *Front Aging Neurosci* 10:94
- Ito T, Kulkarni KR, Schultz DH, Mill RD, Chen RH, Solomyak LI et al (2017) Cognitive task information is transferred between brain regions via resting-state network topology. *Nat Commun* 8:1027
- Griender M, Wang DJJ, Dierks T, Wahlund LO, Jann K (2018) Default mode network complexity and cognitive decline in mild Alzheimer's disease. *Front Neurosci* 12:770
- Binnewijzend MA, Schoonheim MM, Sanz-Arigita E, Wink AM, van der Flier WM, Tolboom N et al (2012) Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging* 33:2018–2028
- Lindemer ER, Salat DH, Smith EE, Nguyen K, Fischl B, Greve DN et al (2015) White matter signal abnormality quality differentiates mild cognitive impairment that converts to Alzheimer's disease from nonconverters. *Neurobiol Aging* 36:2447–2457
- Pagani M, Giuliani A, Oberg J, Chincarini A, Morbelli S, Brugnolo A et al (2016) Predicting the transition from normal aging to Alzheimer's disease: a statistical mechanistic evaluation of FDG-PET data. *Neuroimage* 141:282–290
- Mateos-Perez JM, Dadar M, Lacalle-Aurioles M, Iturria-Medina Y, Zeighami Y, Evans AC (2018) Structural neuroimaging as clinical predictor: a review of machine learning applications. *Neuroimage Clin* 20:506–522
- Tong T, Gray K, Gao QQ, Chen L, Rueckert D, Initia ADN (2017) Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recogn* 63:171–181
- Eskildsen SF, Coupé P, García-Lorenzo D, Fonov V, Pruessner JC, Collins DL et al (2013) Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* 65:511–521
- Beheshti I, Demirel H, Matsuda H, A. s. D. N. Initiative (2017) Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput Biol Med* 83:109–119
- Peng JL, Zhu XF, Wang Y, An L, Shen DG (2019) Structured sparsity regularized multiple kernel learning for Alzheimer's disease diagnosis. *Pattern Recogn* 88:370–382
- Ahmed OB et al (2015) Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. *Multimed Tools Appl* 74(4):1249–1266
- Dyrba M, Grothe M, Kirste T, Teipel SJ (2015) Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. *Hum Brain Mapp* 36:2118–2131
- Ryali S, Supekar K, Abrams DA, Menon V (2010) Sparse logistic regression for whole-brain classification of fMRI data. *Neuroimage* 51:752–764
- Duc NT, Ryu S, Qureshi MNI, Choi M, Lee KH, Lee B (2020) 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI. *Neuroinformatics* 18(1):71–86
- Zareapoor M, Shamsolmoali P, Jain DK, Wang H, Yang J (2018) Kernelized support vector machine with deep learning: an efficient approach for extreme multiclass dataset. *Pattern Recogn Lett* 115:4–13
- Jain DK, Zhang Z, Huang K (2020) Multi angle optimal pattern-based deep learning for automatic facial expression recognition. *Pattern Recogn Lett* 139:157–165
- Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK et al (2004) A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22:1060–1075
- Haghighat M, Abdel-Mottaleb M, Alhalabi W (2016) Discriminant correlation analysis: real-time feature level fusion for

- multimodal biometric recognition. *IEEE Trans Inf Forensics Secur* 11:1984–1996
26. Sehgal S, Singh H, Agarwal M, Bhasker V, Shantanu (2014) Data analysis using principal component analysis. In: International conference on medical imaging, m-Health and emerging communication systems (MedCom). IEEE
 27. Song F, Guo Z, Mei D (2010) Feature selection using principal component analysis. In: 2010 International conference on system science, engineering design and manufacturing informatization. IEEE
 28. Schölkopf B, Smola A, Müller KR (1997) Kernel principal component analysis. In: Gerstner W, Germond A, Hasler M, Nicoud JD (eds) *Artificial neural networks—ICANN'97*. ICANN 1997. Lecture notes in computer science, vol 1327. Springer, Berlin
 29. Van der Maaten L, Hinton G (2008) Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* 9:2579–2605
 30. Sidhu G, Asgarian N, Greiner R, Brown M (2012) Kernel principal component analysis for dimensionality reduction in fMRI-based diagnosis of ADHD. *Front Syst Neurosci* 12(74):1–16
 31. Oliveira FHM, Machado ARP, Andrade AO (2018) On the use of t-distributed stochastic neighbor embedding for data visualization and classification of individuals with Parkinson's disease. *Comput Math Methods Med* 8019232:17
 32. Gisbrecht A, Schulz A, Hammer B (2015) Parametric nonlinear dimensionality reduction using kernel t-SNE. *Neurocomputing* 147:71–82
 33. Gisbrecht A, Lueks W, Mokbel B, Hammer B (2012) Out-of-sample kernel extensions for nonparametric dimensionality reduction. In: European symposium on artificial neural networks, computational intelligence and machine learning. Bruges (Belgium), pp 25–27
 34. Lin S, Zeng J (2019) Fast learning with polynomial kernels. *IEEE Trans Cybern* 9(10)
 35. Samosir RS, Gaol FL, Abbas BS, Sabarguna BS, Heryadi Y (2019) Comparison between linear and polynomial kernel function for ovarian cancer classification. In: The 3rd international conference on computing and applied informatics 2018, *Journal of Physics: Conf. Series*, vol 1235
 36. Qu H, Zhang Y (2016) A new kernel of support vector regression for forecasting high-frequency stock returns. *Math Probl Eng* 9
 37. Jain DK, Dubey SB, Choubey RK, Sinhal A, Arjaria SK, Jain A, Wang H (2018) An approach for hyperspectral image classification by optimizing SVM using self organizing map. *J Comput Sci* 25:252–259

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.