Biomedical Signal Processing and Control 59 (2020) 101903

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE)

B. Richhariya, M. Tanveer*,
A.H. Rashid, for the Alzheimer's Disease Neuroimaging Initiative[1]

Discipline of Mathematics, Indian Institute of Technology Indore, Simrol, Indore 453552, India

ARTICLE INFO

Article history:
Received 22 June 2019
Received in revised form 27 January 2020
Accepted 15 February 2020
Available online 25 February 2020

Keywords:
Universum
Alzheimer's disease
MRI
Feature selection
Prior information
Support vector machine

ABSTRACT

Alzheimer's disease is one of the most common causes of death in today's world. Magnetic resonance imaging (MRI) provides an efficient and non-invasive approach for diagnosis of Alzheimer's disease. Efficient feature extraction techniques are needed for accurate classification of MRI images. Motivated by the work on support vector machine based recursive feature elimination (SVM-RFE) [16], we propose a novel feature selection technique to incorporate prior information about data distribution in the recursive feature elimination process. Our method is termed as universum support vector machine based recursive feature elimination (USVM-RFE). The proposed method provides global information about data in the RFE process as compared to the local approach of feature selection in SVM-RFE. We also present the application of feature selection and classification algorithms on both voxel based as well as volume based morphometry analysis of structural MRI images (ADNI database). Feature selection is performed using MRI data of brain tissues such as gray matter, white matter, and cerebrospinal fluid. USVM-RFE provides improvement over SVM-RFE in classification of control normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) subjects. Moreover, better accuracy is obtained by USVM-RFE with lesser number of features in comparison to SVM-RFE. This leads to identification of prominent brain regions for feature selection and classification of MRI images. The highest accuracies obtained by our method for classification of CN vs AD, CN vs MCI, and MCI vs AD are 100%, 90%, and 73.68%, respectively.

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease, primarily affecting the elderly population. It is also a leading cause of dementia. Currently, 50 million people are affected worldwide by dementia, and is expected to be 152 million by the year 2050 [1]. Diagnosis of AD is a formidable task that requires a lot of expertise, and a thorough examination of patient data. Examination of such a vast amount of data is prone to human error. The application of machine learning in detection of Alzheimer's disease has shown very promising results, and has been a very active topic of research in the last decade [2]. Different feature extraction techniques have been utilized by researchers to classify Alzheimers' data using machine learning methods [3]. This research has also been aided by the availability of open source neurological and genetic data from sources like ADNI (adni.loni.usc.edu), AIBL (aibl.csiro.au) and OASIS (www.oasis-brains.org).

Feature selection plays an important role in classification of data. In case of AD classification, many researchers have used voxel based morphometry (VBM) analysis of MRI images [3,4]. In Alzheimer's disease, there is atrophy in gray matter (GM) and white matter (WM) tissues which leads to increase in cerebrospinal fluid (CSF) volume. However, most of the researchers considered GM and WM features [5,6] using VBM analysis. The CSF voxel features are used for classification in few works [7,8]. Volumetric analysis is also used in many researches on AD classification [9,10].

Support vector machine (SVM) [11] has been a widely used machine learning technique for classification of AD [3,12,13] and other diseases [14,15]. SVM classifies data based on the maximum

* Corresponding author.
E-mail addresses: phd1701241001@iiti.ac.in (B. Richhariya), tanveergouri@gmail.com, mtanveer@iiti.ac.in (M. Tanveer), ashrafrashid@iiti.ac.in (A.H. Rashid).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.bspc.2020.101903
1746-8094/

margin principle. It utilizes the structural risk minimization (SRM) principle for regularization, while reducing the empirical risk. This leads to good generalization performance with less overfitting of data. Also, the optimization problem of SVM gives a globally optimal solution in comparison to techniques like artificial neural network (ANN) which suffer from the problem of local minimum. In 2002, Guyon et al. [16] proposed a feature selection method for cancer classification, termed as support vector machine based recursive feature elimination (SVM-RFE). In SVM-RFE, the feature weights are assigned on the basis of SVM classification. The features are selected from the dataset in a recursive manner using weights from SVM classifier.

SVM-RFE is a popular feature selection technique, and has been used in many applications [4,17]. To reduce the computation cost of SVM-RFE, twin support vector machine based RFE (TWSVM-RFE) is proposed in [18]. SVM-RFE for multiclass classification is also proposed in [19]. An improvement on SVM-RFE based on T-test analysis is presented in [20]. However, SVM is agnostic towards the distribution of data [21]. As a consequence, some important features get eliminated in the SVM-RFE process. Moreover, SVM-RFE uses a greedy approach, and selects features based on information available at each iteration. To encode meaningful information about the distribution of data, we used universum samples in the RFE process. The universum samples are in the same domain as the problem at hand, and hence provide prior information about the distribution of data. Universum samples do not belong to the dataset, but are related to the classification problem. For example, in EEG classification, seizure free data samples have been used as universum for classification of healthy and seizure samples [22]. Universum support vector machine (USVM) [21,23,24] has shown good generalization performance in comparison to SVM, and is applied in various applications [22,25]. The idea of universum is to choose the complexity of the model as per the data distribution [21]. In order to improve SVM-RFE algorithm, we propose a novel universum based recursive feature elimination algorithm (USVM-RFE) for classification problems. The main contributions of our work are listed below:

- A novel feature selection technique is proposed, i.e., USVM-RFE to incorporate prior knowledge about the data distribution in the feature selection process.
- USVM-RFE provides global information about the data distribution from universum in each iteration of feature selection. This is in contrast to SVM-RFE which selects features on basis of local information from each feature set.
- We provide a comprehensive analysis of feature selection for diagnosis of Alzheimer's disease based on both VBM as well as volumetric analysis of structural MRI images.
- VBM analysis is presented using DARTEL approach [26,27] on training and testing data separately. This approach is helpful in real world scenarios where testing data is not available beforehand. This procedure is applied for all cases, i.e., control normal (CN) vs Alzheimer's disease (AD), CN vs mild cognitive impairment (MCI), and MCI vs AD using all the three tissue features viz. GM, WM, and CSF voxels.
- Moreover, in the available literature, universum learning is applied on Alzheimer's disease for classification of CN vs AD only [28]. So, we present the application of proposed USVM-RFE as well as SVM-RFE on classification of CN vs AD, CN vs MCI, and MCI vs AD.
- We present a discussion on the brain regions selected by proposed USVM-RFE for detection of Alzheimer's disease.

The rest of the paper is organized as follows: Section 2 describes the materials and methods used in this work, while Section 3 gives a brief explanation of related work. The proposed algorithm is presented in Section 4. The experimental results are shown in Section 5, and discussion on results is presented in Section 6. Finally, the conclusions with future work are given in Section 7.

Table 1
Subject demographics.

Diagnosis	Age	Gender	MMSE
CN	76.65 ± 4.30	39M/11F	29.02 ± 1.15
MCI	75.23 ± 7.02	26M/24F	26.9 ± 1.96
AD	75.60 ± 6.58	28M/22F	23.62 ± 2.24

The mathematical notations used in this work are as follows: All vectors are assumed as column vectors. X is a matrix containing the data points belonging to class '1' and '−1' of size $p \times n$ and $q \times n$ respectively. U represents universum data points having dimension $r \times n$. Total number of data points are represented by $l = p + q$, where n is the dimension of each data point. $\|x\|$ and $\|X\|$ represents the 2-norm of a vector x and matrix X respectively.

2. Materials and methods

2.1. Data

All data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in the year 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The main objective of ADNI is to analyze the effectiveness of neuroimaging techniques like magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical neuropsychological tests to estimate the onset of Alzheimer's disease from the state of mild cognitive impairment. For more information, visit www.adni-info.org.

2.2. Image acquisition

A total of 150 T1-weighted structural MRI (sMRI) images are downloaded from ADNI database. Each of the categories, i.e., CN, MCI, and AD comprises of 50 subjects. The subjects are within the age range of 60–90 with mean age of 75.83, and standard deviation of 6.07. The Mini-Mental State Examination (MMSE) score of subjects is in the range of 17–30 with mean and standard deviation of 26.51 ± 2.88. The detailed demographics of cohort are given in Table 1.

Images of the following specifications are acquired from the ADNI archive: field strength = 1.5 T; description = MP-RAGE; acquisition = 3D; pulse sequence = RM; slice thickness = 1.2 mm; flip angle = 8°; acquisition plane = sagittal, manufacturer = GE medical systems.

We also downloaded 817 sMRI images from the ADNI baseline dataset [29,30] to verify the applicability of the proposed model.

2.3. Voxel based morphometry (VBM)

A frequently used neuroimaging toolbox, i.e., Statistical Parametric Mapping (SPM) version 12 (Wellcome Trust Centre for Neuroimaging, University College London, UK) is used to perform the VBM analysis. The preprocessed data is used for three different binary classification tasks, i.e., CN vs AD, CN vs MCI, and MCI vs AD. The scans from each subject category were randomly divided into training and testing sets of 40 and 10 images respectively.

All raw images are aligned in the same coordinate space by setting the origin of the raw scans manually to the anterior commissure (AC), and registering with SPM's single subject T1 template.

The registered images are processed by SPM's unified segmentation routine to segment the images into GM, WM and CSF, and create the template using DARTEL [26] approach. The template is used to normalize the images into the Montreal Neurological Institute (MNI) space with modulation. A Gaussian kernel with full width at half maximum (FWHM) of 8 mm is used for smoothing. All the images are transformed to a dimension of $121 \times 145 \times 121$ with a voxel size of 1.5 mm^3.

A two sample T-test is used for finding the statistically significant voxels by keeping subject age and gender as covariates in the general linear model (GLM) [31]. The differences of individual head sizes are controlled by introducing total intracranial volume (TIV) as a covariate of no interest [32] using the analysis of covariance (ANCOVA)-by-subject approach. The T-test analysis is done by using a p-value of 0.05 with family-wise error (FWE) correction, and an extent threshold of 0 adjacent voxels. The complete approach is shown in Fig. S.1.

The specified GLM is given in supplementary Fig. S.2. The voxels of interest (VOI) retrieved after statistical analysis of training images are used as masks for specifying voxel coordinates that are significantly different between subject groups [33]. Figs. S.3 and S.4 show significant voxels of CN vs AD, and CN vs MCI analysis respectively.

In some previous works [34,35], all the acquired MRI images were used for creating DARTEL template. Here, in this work we use DARTEL pipeline for training and testing phase separately. This procedure is followed in all the cases, i.e., CN vs AD, CN vs MCI, and MCI vs AD with different features, i.e., GM, WM, and CSF. As a whole, the DARTEL approach is used on 18 sets of data (9 training, 9 testing). This leads to the generation of distinct subject specific templates for training and testing to take care of real world scenarios where testing images may not be available beforehand. The mask from the training set is applied on the testing images for feature extraction [33].

2.4. Volume based morphometry (VolBM)

For VolBM analysis, Freesurfer's recon-all pipeline (version 6.0.1) [9,36] is used on structural MRI images. Out of 150 MRI images, 1 MCI image failed to process in Freesurfer. So, feature selection was performed on 149 images. We extracted 23 subcortical tissue volumes (SCV), 34 WM tissue volumes (WMV), and 34 cortical thickness (CT) measures of every subject. To check the performance of our model on an independent dataset, we downloaded 817 sMRI images from ADNI baseline dataset [29,30], out of which 4 images failed to process through the Freesurfer pipeline. Thus, our baseline dataset includes 228 CN, 398 MCI, and 187 AD images.

Thickness measures from both the brain hemispheres are added together to form the cortical thickness features. A similar approach is used for volumetric features. The volumetric features are normalized by dividing by TIV of the subjects [9,10]. The list of all the neurological features obtained from Freesurfer is given in supplementary Table S.1, and their variations are illustrated in Fig. S.5.

After extracting significant voxel features from the masked images using SPM, PCA [33] and F-score [37] algorithms are implemented for dimensionality reduction and feature selection respectively.

3. Related work

In this section we briefly discuss the formulation of SVM, and describe the SVM-RFE algorithm.

3.1. Support vector machine (SVM)

The formulation of SVM [11] is written as follows:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}\xi_i \\ s.t. \quad & y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \\ & \xi_i \geq 0, \quad \forall i = 1,2,\ldots,l, \end{aligned} \tag{1}$$

where l is the total number of data points, $c > 0$ is penalty parameter, ξ_i is slack variable, y_i is the class label, and $\phi : R^n \rightarrow R^p$ is the function mapping from n to p dimension where $p > n$.

The parameters w and b are obtained by solving the dual formulation [11,23] of quadratic programming problem (QPP) (1).

The classifier is given as

$$f(x) = sign(w^T x + b). \tag{2}$$

3.2. Support vector machine based recursive feature elimination (SVM-RFE)

SVM-RFE is proposed by Guyon et al. [16] to perform gene selection for cancer classification. In SVM-RFE, the weight vector w contains the weightage of each feature of the dataset. Therefore, SVM-RFE algorithm uses weights of the SVM classifier as a ranking criteria. Using this approach, the features are eliminated in order to find the most important features for classification.

The SVM-RFE algorithm is an iterative algorithm which eliminates the features and generates datasets with different number of features. These datasets are then classified by SVM using the standard cross-validation technique. The feature elimination process of SVM-RFE is described in [16].

3.3. Universum support vector machine (USVM)

The formulation of USVM [21] is given as follows:

$$\begin{aligned} \min_{w,b,\xi,\eta} \quad & \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}\xi_i + c_u\sum_{j=1}^{2r}\eta_j \\ s.t. \quad & y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \\ & y_j(w^T\phi(x_j)+b) \geq -\epsilon-\eta_j, \\ & \xi_i \geq 0, \quad \eta_j \geq 0, \quad \forall i = 1,2,\ldots,l, \\ & \forall j = 1,2,\ldots,2r, \end{aligned} \tag{3}$$

where l is the total number of data points, $c > 0$, $c_u > 0$ are penalty parameters, ξ_i and η_j are slack variables, $y_{i,j}$ is class label, $\phi : R^n \rightarrow R^p$ is the function mapping from n to p dimension where $p > n$, and r is the number of universum samples.

The dual formulation of QPP (3) is written by applying the K.K.T. conditions as

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{l+2r}\mu_i\alpha_i - \frac{1}{2}\sum_{i=1}^{l+2r}\sum_{j=1}^{l+2r}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j) \\ s.t. \quad & 0 \leq \alpha_i \leq c, \quad \mu_i = 1, \quad \forall i = 1,2,\ldots,l, \\ & 0 \leq \alpha_i \leq c_u, \quad \mu_i = -\epsilon, \\ & \forall i = l+1, l+2, \ldots, l+2r, \\ & \sum_{i=1}^{l+2r}\alpha_i y_i = 0, \end{aligned} \tag{4}$$

where $\alpha_i \geq 0$ is the Lagrange multiplier [11,21].

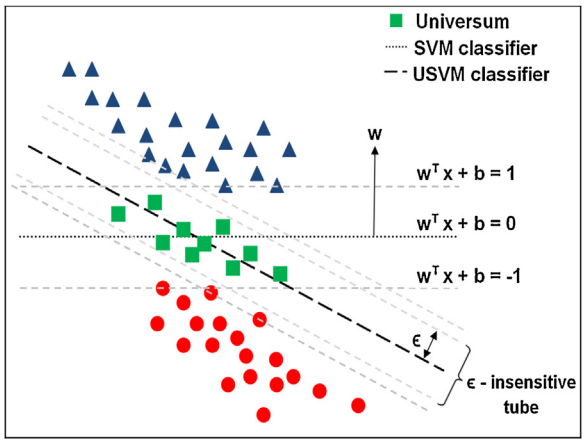


Fig. 1. Universum data.

The classifier is given by Eq. (2), and the weight vector is obtained as $w = \sum_{i=1}^{l+2r} \alpha_i y_i x_i$.

4. Proposed universum support vector machine based recursive feature elimination (USVM-RFE)

The SVM-RFE algorithm lacks the knowledge about the distribution of data. Moreover, in the recursive process, the SVM classifier eliminates features on basis of weights for the maximal margin. In order to incorporate knowledge about the data distribution, we use universum samples as shown in Fig. 1. The resulting universum based algorithm uses this prior information about distribution of data in the recursive elimination of features leading to better feature selection.

4.1. Universum data

The universum data is used to align the classifier with the data distribution. As shown in Fig. 1, the classifier generated by USVM is better aligned to classify the data points. This helps in the classification of testing data. Without this knowledge of the data distribution, the SVM classifier only tries to maximize the margin. This results in reduced generalization performance of the model. In this work, we generated universum samples using random averaging of data points [38,22].

4.2. Iterative procedure

In our USVM-RFE, we use the universum data points in each iteration. This results in selection of important features due to addition of universum constraints. The universum data points are constrained to lie within an ϵ-insensitive tube as shown in QPP (3). Since USVM-RFE is a wrapper [17] method, the process of USVM-RFE is divided into three phases: parameter selection, feature elimination, and classification. Our universum based USVM-RFE algorithm is described in Algorithm 1.

We used k-fold cross validation for selecting optimal parameters for feature elimination. Then, the features are eliminated in an iterative manner.

Algorithm 1. Proposed USVM-RFE

Inputs:
Training data
$X = [x_1, x_2, \ldots, x_l]^T$
Class labels
$Y = [y_1, y_2, \ldots, y_l]^T$

Universum data
$U = [u_1, u_2, \ldots, u_r]^T$

Process:
Find optimal parameters for recursive process using k-fold cross validation.
Feature set
$S = [1, 2, \ldots, n]$
Feature ranked list
$R = [\ ]$
Repeat until $S = [\ ]$
Feature selection
$X = X(:, S)$
$U = U(:, S)$
Train SVM classifier using parameters obtained in step 11
$\alpha = USVM_{train}(X, Y, U)$
Compute the weight vector with dimension of length(S)
$w = \sum_{i=1}^{l+2r} \alpha_i y_i x_i$
Compute ranking criteria
$C_i = (w_i)^2, \quad i = 1, 2, \ldots, length(S)$
Find feature with smallest rank
$f = argmin(C)$
Update feature ranked list
$R = [S(f), R]$
Eliminate the feature with smallest rank
$S = [1 : f - 1, f + 1 : length(S)]$

Output:
Feature ranked list R.

Lastly, the classification is performed on each of the feature subsets. For computational efficiency, more than one feature can be eliminated in one iteration in our proposed USVM-RFE.

Our proposed USVM-RFE provides global information about data distribution as compared to the greedy approach in SVM-RFE. However, proper selection of universum is a topic of research [21,22].

Our universum based feature selection is useful for applications such as MRI image classification. In MRI images, only few voxels corresponding to specific regions are helpful in the classification. So, we present the application of USVM-RFE on classification of Alzheimer's disease. Feature selection is performed on VBM voxels as well as volumetric features for identifying brain regions with neurodegeneration. USVM is trained on the feature sets obtained after each iteration using k-fold cross validation, and tested on testing data. The feature set with highest classification accuracy is selected as the optimal feature set.

5. Numerical experiments

The experiments are carried out on feature sets obtained from VBM as well as VolBM analysis. For VBM, SPM version 12 is used, while for VolBM, Freesurfer version 6.0.1 is used to process the images. The softwares used in generation of 2D and 3D brain overlays are: ITK Snap (v3.8.0-beta) [39], Paraview (v5.6.0) [40], Mricron (www.nitrc.org/projects/mricron) and Mricrogl (v1.0.20180623) (www.nitrc.org/projects/mricrogl). We used WFU_PickAtlas (v3.0.5b) for selecting ROIs from AAL atlas [41] in supplementary Fig. S.6. The experiments for classification are performed in MATLAB 2008b environment using MOSEK optimization toolbox (http://www.mosek.com/).

The image processing is carried out on a workstation with Windows 10 OS, 64-bit, running on 2.30 GHz Intel® Xeon processor, and 128 GB RAM. The numerical experiments for classification were carried out on a PC with Windows 10 OS, 64 bit, with 3.60 GHz Intel® core™ i7-7700 processor, and 16 GB of RAM. The optimal parameters for the recursive process are obtained using 5-fold cross validation in all datasets. Linear kernel is used in both SVM and USVM for feature extraction as well as classification. The number of PCA components is chosen so as to account for 99% of variance in the data. The number of F-score features is selected as 500. In all cases, the features in training and testing data are normalized

Table 2
Performance comparison of proposed USVM-RFE with SVM-RFE based on classification accuracy (%) for CN vs AD on reduced VBM feature sets. Bold values indicate highest accuracy for the dataset, and underlined values show highest accuracy of the algorithm.

Features (%)	Number of features	Proposed USVM-RFE			SVM-RFE		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Gray matter							
1	264	**80**	80	80	70	70	70
5	1323	**75**	90	60	70	70	70
10	2628	**85**	90	80	70	70	70
15	3952	**85**	90	80	75	60	90
20	5304	**90**	90	90	75	70	80
30	7953	**90**	90	90	80	80	80
50	13,045	**90**	90	90	75	70	80
White matter							
1	26	**70**	60	80	60	40	80
5	133	70	60	80	**75**	80	70
10	265	**75**	80	70	**75**	80	70
15	398	**70**	70	70	**70**	70	70
20	534	**70**	70	70	**70**	70	70
30	794	70	80	60	**75**	80	70
50	1333	**70**	70	70	**70**	80	60
CSF							
1	18	**70**	50	90	65	40	90
5	90	**65**	40	90	**65**	40	90
10	180	**65**	40	90	**65**	40	90
15	269	**65**	40	90	**65**	40	90
20	359	**65**	40	90	**65**	40	90
30	535	**60**	40	80	**60**	40	80
50	895	65	50	80	**75**	80	70

to a mean of 0 and standard deviation of 1 using Z-score [10]. The parameter selection is performed based on the following settings.

5.1. Parameter settings

In some previous works, fixed value of c is chosen for feature selection [7,16]. However, in this work we used different types of features with varying dimensions. So, in all the VBM features, i.e., GM, WM, and CSF grid search is used for obtaining optimal parameters. The values of penalty parameters $c = c_u$ are selected from the range $\{10^{-6}, 10^{-5}, \ldots, 10^{0}\}$ for SVM and USVM. This range is selected to avoid overfitting of classifier [42] for high dimensional datasets. For USVM, number of universum samples, i.e., u is selected from the set $\{0.1, 0.2\}$, and ϵ is selected from $\{0.3, 0.5, 0.7\}$.

In the RFE phase of VBM, the features are reduced by percentage (per) of feature size for computational efficiency [16] using the following criteria:

$$Repeat \quad until \quad (No\_of\_features \geq ceil \quad (0.01 * Total\_feature))$$

$$\{$$

$$No\_of\_features = per * No\_of\_features$$

$$if \quad (No\_of\_features < ceil \quad (0.1 * Total\_feature))$$

$$per = 0.995;$$

$$else \quad if(No\_of\_features < ceil \quad (0.5 * Total\_feature))$$

$$per = 0.99;$$

$$else \quad (No\_of\_features < ceil \quad (0.7 * Total\_feature))$$

$$per = 0.98;$$

$$\}$$

where *ceil* is the ceiling function.

In all VolBM feature sets, i.e., CT, SCV, and WMV, $c = c_u$ is chosen from $\{10^{-6}, 10^{-5}, \ldots, 10^{4}\}$ in both feature selection and classification phase. For USVM, for parameter selection, u is selected from the set $\{0.1, 0.3, 0.45\}$, and ϵ is selected from $\{0.6, 0.7, 0.8\}$. In the classification phase, to take care of features of different dimensions u is selected from the set $\{0.1, 0.15, 0.35, 0.45\}$, and ϵ is selected from $\{0.3, 0.5, 0.6, 0.8\}$. In the RFE process, the features are reduced one at a time, due to less size of VolBM feature set.

To check the performance of proposed USVM-RFE on other applications, we performed experiments on two UCI [43] biomedical datasets, i.e., Wpbc (Breast Cancer Wisconsin Prognostic), and Wdbc (Breast Cancer Wisconsin Diagnostic). The parameters c, c_u are selected by varying values from the set $\{10^{-5}, 10^{-4}, \ldots, 10^{5}\}$ for SVM and USVM. For USVM, the values for u and ϵ are set as 0.3 and 0.5 respectively.

5.2. Results

Experimental results on classification of different subject groups, i.e., CN, MCI, and AD are shown for different features sets in the following subsections.

5.2.1. VBM features

We performed experiments on feature selection from VBM features obtained from SPM toolbox for both SVM-RFE and proposed USVM-RFE.

CN vs AD: The comparison of classification accuracy for CN vs AD is presented in Table 2 for GM, WM, and CSF features sets. It is observable that for GM features, proposed USVM-RFE outperforms SVM-RFE w.r.t. accuracy and sensitivity in all the 7 reduced feature sets. This is a result of prior knowledge in USVM-RFE about the data distribution. However, the accuracy of USVM-RFE reached a maximum of 90% for 20% features, and then declined in lower dimensional features due to loss of informative voxels. For WM, both methods have similar performance. In case of CSF, USVM-RFE performed better for lower dimensional feature set (1% features), while SVM-RFE is having high accuracy for high dimensional feature set (50% features).

The classification accuracy of SVM and USVM for full feature sets of VBM is shown in Table 3. The optimal parameters for RFE process are also shown. These parameters are utilized to perform the feature elimination process. Moreover, Table 3 also shows the accuracies obtained after feature reduction by PCA, and F-score.

Table 3
Performance comparison of USVM with SVM is shown based on classification accuracy (%) for CN vs AD on VBM features. The optimal parameters are shown in parentheses.

		USVM			SVM		
Features	Number of features	Accuracy ($c = c_u, \epsilon, u$)	Sensitivity	Specificity	Accuracy (c)	Sensitivity	Specificity
Gray matter							
All features	26,524	**85** (10^{-3}, 0.3, 0.2)	90	80	70 (10^{-6})	80	60
PCA	59	75 (10^{0}, 0.3, 0.2)	80	70	**80** (10^{-1})	90	70
F-score	500	**70** (10^{-4}, 0.5, 0.2)	70	70	**70** (10^{-4})	70	70
White matter							
All features	2675	**75** (10^{-5}, 0.3, 0.2)	90	60	**75** (10^{-5})	90	60
PCA	26	60 (10^{-2}, 0.3, 0.2)	40	80	**65** (10^{-1})	50	80
F-score	500	**70** (10^{-4}, 0.3, 0.2)	70	70	**70** (10^{-4})	70	70
CSF							
All features	1802	70 (10^{-4}, 0.3, 0.1)	60	80	**75** (10^{-4})	70	80
PCA	23	**60** (10^{0}, 0.5, 0.1)	30	90	**60** (10^{2})	30	90
F-score	500	**60** (10^{-4}, 0.5, 0.1)	40	80	**60** (10^{-4})	40	80

Table 4
Performance comparison of proposed USVM-RFE with SVM-RFE is shown based on classification accuracy (%) for CN vs MCI on reduced feature sets of VBM. Bold values indicate highest accuracy for the dataset, and underlined values show highest accuracy of the algorithm.

		Proposed USVM-RFE			SVM-RFE		
Features (%)	Number of features	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Gray matter							
1	64	**70**	80	60	65	70	60
5	322	**75**	90	60	70	80	60
10	647	**70**	80	60	**70**	80	60
15	969	70	80	60	**75**	90	60
20	1285	**70**	80	60	**70**	80	60
30	1925	**70**	80	60	**70**	80	60
50	3179	**70**	80	60	**70**	80	60
White matter							
1	26	**50**	20	80	45	10	80
5	133	**60**	60	60	**60**	60	60
10	265	**65**	60	70	**65**	60	70
15	398	55	30	80	**65**	60	70
20	534	40	10	70	**55**	40	70
30	794	**50**	20	80	45	20	70
50	1333	**55**	30	80	45	10	80
CSF							
1	14	**90**	100	80	80	100	60
5	70	**85**	100	70	**85**	100	70
10	140	**85**	100	70	80	100	60
15	212	80	100	60	**85**	100	70
20	281	80	100	60	**85**	100	70
30	424	**85**	90	80	**85**	100	70
50	700	**85**	100	70	80	100	60

From Table 3, one can observe that proposed USVM-RFE performs better than PCA and F-score in the feature selection.

CN vs MCI: For CN vs MCI, the results for feature selection are shown in Table 4. It is visible that in case of GM, USVM-RFE gives accuracy of 75% with sensitivity of 90% for lesser features, i.e., 5%, as compared to SVM-RFE (15% features). However, for CSF voxels, highest accuracy of 90% is obtained by proposed USVM-RFE for 1% features, whereas SVM-RFE obtains lower accuracy (85%) with a higher feature size (5%). In case of WM also, USVM-RFE has outperformed SVM-RFE in most cases. Table 5 shows the classification results on full feature sets. One may notice that USVM has performed better than SVM for full features on CSF features. SVM and USVM have also shown lesser accuracies than PCA and F-score in some cases. This may be attributed to overfitting of SVM and USVM for high dimensional feature sets [42]. However, both algorithms have shown improvement on accuracy in the RFE process.

MCI vs AD: Table 6 shows the classification accuracies for MCI vs AD on reduced VBM feature sets. One can notice that USVM-RFE performs better than SVM-RFE for lower dimensional features of GM. In case of WM, SVM-RFE performed better than USVM-RFE. This may be attributed to improper universum data generated using random averaging scheme. Since, the MCI vs AD dataset is non-linear in nature, random averaging may generate improper universum data. In case of CSF, both SVM-RFE and USVM-RFE have shown low classification accuracy. However, USVM-RFE has per-

Table 5
Performance comparison of USVM with SVM is shown based on classification accuracy (%) for CN vs MCI on VBM features. The optimal parameters are shown in parentheses.

		USVM			SVM		
Features	Number of features	Accuracy ($c = c_u, \epsilon, u$)	Sensitivity	Specificity	Accuracy (c)	Sensitivity	Specificity
Gray matter							
All features	6478	**70** (10^{-5}, 0.3, 0.1)	80	60	**70** (10^{-4})	80	60
PCA	37	65 (10^{1}, 0.3, 0.2)	50	80	**70** (10^{1})	50	90
F-score	500	**75** (10^{-4}, 0.7, 0.1)	90	60	70 (10^{-3})	80	60
White matter							
All features	2675	**45** (10^{-3}, 0.7, 0.1)	10	80	**45** (10^{-3})	10	80
PCA	28	**55** (10^{1}, 0.3, 0.1)	10	100	**55** (10^{1})	10	100
F-score	500	**60** (10^{-3}, 0.3, 0.1)	60	60	**60** (10^{-3})	60	60
CSF							
All features	1416	**85** (10^{-4}, 0.3, 0.2)	100	70	80 (10^{-4})	90	70
PCA	22	60 (10^{1}, 0.7, 0.1)	70	50	**65** (10^{1})	70	60
F-score	500	75 (10^{-4}, 0.3, 0.1)	80	70	**85** (10^{-3})	100	70

Table 6
Performance comparison of proposed USVM-RFE with SVM-RFE is shown based on classification accuracy (%) for MCI vs AD on reduced feature sets of VBM. Bold values indicate highest accuracy for the dataset, and underlined values show highest accuracy of the algorithm.

Tissue features (%)	Number of features	Proposed USVM-RFE			SVM-RFE		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Gray matter							
1	264	**60**	50	70	55	40	70
5	1323	**65**	60	70	55	40	70
10	2628	**65**	60	70	60	50	70
15	3952	**60**	50	70	**60**	50	70
20	5304	**60**	50	70	**60**	50	70
30	7953	55	50	60	**60**	50	70
50	13,045	**60**	50	70	**60**	50	70
White matter							
1	26	45	50	40	**50**	60	40
5	133	**55**	70	40	50	60	40
10	265	**50**	100	0	**50**	60	40
15	398	50	100	0	**65**	100	30
20	534	50	100	0	**65**	90	40
30	794	50	90	10	**60**	100	20
50	1333	55	90	20	**65**	90	40
CSF							
1	18	**45**	0	90	35	40	30
5	90	**40**	0	80	**40**	40	40
10	180	**45**	0	90	40	50	30
15	269	**45**	0	90	**45**	60	30
20	359	40	0	80	**45**	70	20
30	535	**50**	40	60	45	70	20
50	895	**55**	50	60	45	50	40

formed better in comparison to SVM-RFE. Moreover, Table 7 shows the performance on full features sets. It can be seen that USVM performs better than SVM in most cases. Moreover, in Table 6 the classification accuracy of MCI vs AD is low in both SVM-RFE and USVM-RFE. This is attributed to the relatively difficult classification problem of MCI vs AD in comparison to CN vs AD, and CN vs MCI [33]. The discriminating regions in CN vs AD are more as compared to MCI vs AD. This is due to similar distribution of data points in MCI and AD subjects [44].

Remark. No significant voxels were found in case of WM features of CN vs MCI, and all cases of MCI vs AD. Thus, the masks obtained from CN vs AD experiments were used to select region of interests (ROI) in these cases [33].

5.2.2. VolBM features

For comprehensive analysis of MRI data, VolBM analysis is also used in this work for extracting features from the MRI images. Experiments are performed on volumetric and thickness measures obtained from Freesurfer toolbox. Table 8 shows the accuracy and optimal parameters of USVM and SVM on classification of CN vs AD, CN vs MCI, and MCI vs AD on VolBM features. The set of all VolBM features, i.e., CT, SCV, and WMV are used for feature selection. It is clearly visible from Table 8 that for all the VolBM features, USVM outperforms SVM in terms of accuracy.

Table 7
Performance comparison of USVM with SVM is shown based on classification accuracy (%) for MCI vs AD on VBM features. The optimal parameters are shown in parentheses.

		USVM			SVM		
Features	Number of features	Accuracy ($c = c_u, \epsilon, u$)	Sensitivity	Specificity	Accuracy (c)	Sensitivity	Specificity
Gray matter							
All features	26,524	**70** (10^{-6}, 0.7, 0.2)	60	80	60 (10^{-5})	60	60
PCA	60	**50** (10^{-6}, 0.5, 0.1)	30	70	**50** (10^{-2})	30	70
F-score	500	**65** (10^{-4}, 0.3, 0.1)	60	70	60 (10^{-3})	50	70
White matter							
All features	2675	55 (10^{-5}, 0.3, 0.2)	70	40	**65** (10^{-3})	90	40
PCA	26	**40** (10^{0}, 0.5, 0.2)	50	30	40 (10^{-1})	50	30
F-score	500	40 (10^{2}, 0.3, 0.1)	80	0	**50** (10^{0})	100	0
CSF							
All features	1802	**45** (10^{-1}, 0.3, 0.2)	0	90	35 (10^{-3})	40	30
PCA	24	**35** (10^{-2}, 0.3, 0.2)	10	60	**35** (10^{0})	10	60
F-score	500	**55** (10^{1}, 0.3, 0.2)	10	100	40 (10^{-2})	50	30

Table 8
Performance comparison of USVM with SVM is shown based on classification accuracy (%) for VolBM features. Bold values indicate highest accuracy for the dataset. The optimal parameters for RFE process using SVM and USVM are shown in parentheses.

		USVM			SVM		
Features	Number of features	Accuracy ($c = c_u, \epsilon, u$)	Sensitivity	Specificity	Accuracy (c)	Sensitivity	Specificity
(a) CN vs AD							
CT	34	**75** (10^{-2}, 0.7, 0.45)	50	100	**75** (10^{-2})	50	100
SCV	23	**90** (10^{-2}, 0.6, 0.1)	90	90	**90** (10^{-2})	90	90
WMV	34	**50** (10^{1}, 0.6, 0.45)	50	50	40 (10^{2})	30	50
CT + SCV + WMV	91	**85** (10^{-2}, 0.6, 0.3)	70	100	80 (10^{-2})	60	100
(b) CN vs MCI							
CT	34	**68.42** (10^{-2}, 0.6, 0.1)	77.78	60	**68.42** (10^{-2})	77.78	60
SCV	23	**78.95** (10^{0}, 0.8, 0.45)	88.89	70	**78.95** (10^{0})	88.89	70
WMV	34	**84.21** (10^{0}, 0.8, 0.3)	100	70	78.95 (10^{0})	100	60
CT + SCV + WMV	91	**84.21** (10^{0}, 0.6, 0.1)	88.89	80	**84.21** (10^{0})	88.89	80
(c) MCI vs AD							
CT	34	**57.89** (10^{-2}, 0.6, 0.3)	40	77.78	**57.89** (10^{-2})	30	88.89
SCV	23	**73.68** (10^{-2}, 0.6, 0.45)	80	66.67	68.42 (10^{-2})	80	55.56
WMV	34	**63.16** (10^{1}, 0.6, 0.45)	60	66.67	57.89 (10^{-1})	40	77.78
CT + SCV + WMV	91	**63.16** (10^{-2}, 0.6, 0.3)	50	77.78	**63.16** (10^{-2})	50	77.78

The accuracy of USVM for CN vs AD classification for different feature sets of RFE process is shown in Fig. 2. One can see that accuracy of proposed USVM-RFE is better than SVM-RFE. Moreover, the variation in accuracy of proposed USVM-RFE is less than SVM-RFE. This is the result of universum data, which helps the classifier to follow the data distribution. The classification accuracy and F1 scores on reduced VolBM features using optimal parameters are shown in Figs. 3–5. One may observe in Fig. 3(a) that USVM-RFE achieves 100% accuracy in 5 feature sets for CN vs AD, with lowest feature set having 5% of total features.

Also, the variation of accuracy looks similar in both algorithms with higher accuracy in proposed USVM-RFE. The F1 scores follow similar trend as accuracy in Fig. 3(b). For CN vs MCI in Fig. 4(a) proposed USVM-RFE achieves highest accuracy of 84.21% for 5% features, while SVM-RFE requires 100% features for achieving the same accuracy. Fig. 4(b) shows the corresponding F1 scores.

In case of MCI vs AD in Fig. 5, the accuracy of proposed USVM-RFE is not better than SVM-RFE in all the cases. This is the result of non-linear nature of MCI vs AD data shown in supplementary Fig. S.7. This leads to generation of improper universum data not

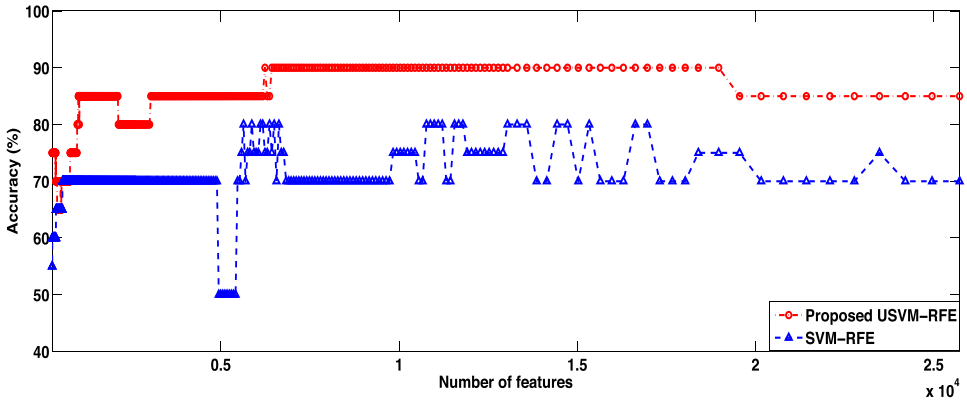


Fig. 2. Comparison of classification accuracy of SVM-RFE and proposed USVM-RFE on different feature sets obtained in RFE process on VolBM features for CN vs AD.

Table 9
Performance comparison of USVM with SVM is shown based on classification accuracy (%) for UCI datasets. Bold values indicate highest accuracy for the dataset.

Dataset (Train size, Test size)	USVM Accuracy (c, c_u, ϵ, u) Time (s)	SVM Accuracy (c) Time (s)
Wpbc	**79.69**	78.13
(130 × 33, 64 × 33)	(10, 1, 0.5, 0.3)	(1)
	0.36443	0.13955
Wdbc	**98.43**	97.81
(250 × 30, 319 × 30)	(0.01, 0.001, 0.5, 0.3)	(1)
	1.35046	0.51699

lying between the two classes. However, proposed USVM-RFE has shown highest classification of 73.68% for MCI vs AD with 13% and 15% features. This is better than SVM-RFE which achieved highest accuracy of 68.42% on 30% features.

We present a formula for calculating score for feature ranking in RFE based methods. The feature score of the ith feature is calculated as per the following:

$$Score\ \ (i) = \sum_{j=1}^{f} (r_{lowest} - r_{ij}), \tag{5}$$

where r_{ij} is the rank of ith feature in jth feature set, r_{lowest} is the lowest rank value in largest feature set, and f is the total number of features.

By using above mentioned formula, more weightage is given to features which survive up to the smallest feature sets. Tables 14–16 show scores of different features for classification of CN vs AD, CN vs MCI, and MCI vs AD respectively. The discussion on these scores is presented in Section 6.2.

5.2.3. ADNI baseline dataset

In order to verify the comparative performance of proposed USVM-RFE with existing algorithms, we conducted experiments on ADNI baseline dataset [29,30]. The VolBM features are used to compare the classification performance of proposed USVM-RFE with SVM-RFE and TWSVM-RFE.

The comparative performance of proposed USVM-RFE for classification CN vs AD is shown in Table 10. It is clearly observable that proposed USVM-RFE is performing better than existing algorithms in most of the feature sets. The highest accuracy obtained by USVM-RFE is 89.2% with a sensitivity of 84.87% for 15% features.

In case of CN vs MCI in Table 11, again the proposed USVM-RFE outperfomed SVM-RFE and TWSVM-RFE in most of the feature sets. Moreover, the highest accuracy of proposed USVM-RFE, i.e., 72.61% is obtained for just 3% of total features. Also, for MCI vs AD in Table 12, proposed USVM-RFE performed better than other algorithms in many of the feature sets. However, the highest accuracy obtained by proposed USVM-RFE is same as TWSVM-RFE, i.e., 71.88%, but with higher sensitivity. However, USVM-RFE is

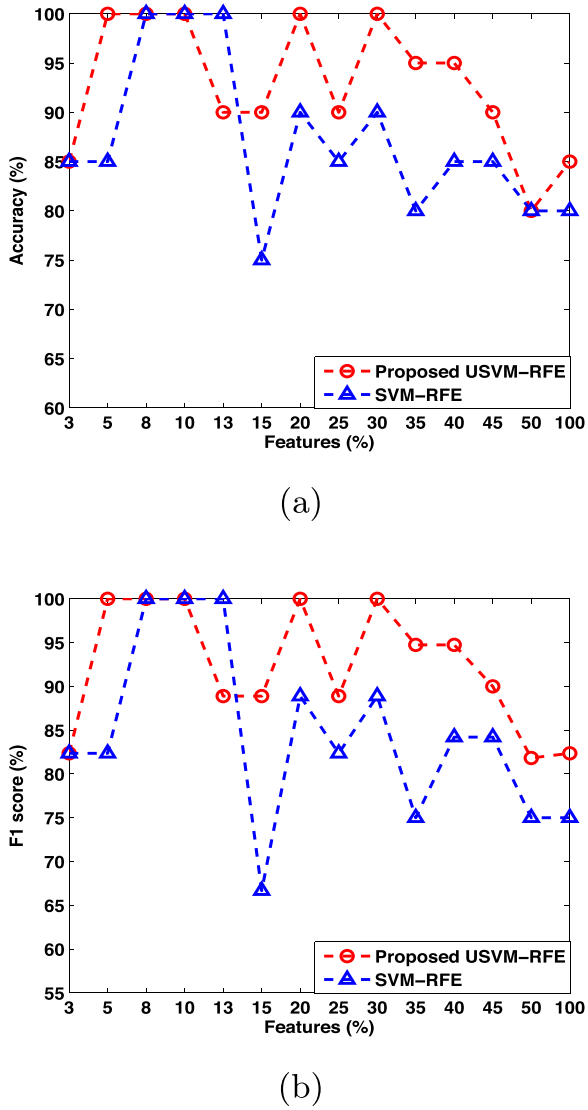


Fig. 3. Plot showing comparison of (a) accuracy, and (b) F1 score of SVM-RFE, and proposed USVM-RFE for classification of CN vs AD using VolBM features.

Table 10
Comparison of performance of proposed USVM-RFE in terms of accuracy (%) with existing algorithms on ADNI baseline dataset for CN vs AD using VolBM features. Bold values indicate highest accuracy for the dataset, and underlined values show highest accuracy of the algorithm.

Tissue features (%)	Proposed USVM-RFE			SVM-RFE			TWSVM-RFE		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
CN vs AD									
3	83.6	71.43	94.66	83.2	72.27	93.13	**84**	72.27	94.66
5	**86.4**	78.99	93.13	**86.4**	78.99	93.13	85.2	82.35	87.79
8	**84.4**	74.79	93.13	82.8	71.43	93.13	**84.4**	78.15	90.08
10	**86.4**	78.99	93.13	84.8	77.31	91.6	85.2	77.31	92.37
13	86.8	80.67	92.37	**87.2**	79.83	93.89	85.2	80.67	89.31
15	**89.2**	84.87	93.13	86.8	78.15	94.66	83.6	78.99	87.79
20	**84.4**	76.47	91.6	87.2	78.99	94.66	83.2	73.95	91.6
25	**86.8**	77.31	95.42	85.2	77.31	92.37	81.6	82.35	80.92
30	**86.4**	78.15	93.89	**86.4**	74.79	96.95	82	73.95	89.31
35	**88.8**	83.19	93.89	85.2	73.95	95.42	81.2	73.95	87.79
40	**86.4**	74.79	96.95	84.8	73.11	95.42	78.8	79.83	77.86
45	**86.8**	74.79	97.71	84.4	71.43	96.18	79.6	77.31	81.68
50	**86**	74.79	96.18	85.6	74.79	95.42	79.2	78.15	80.15

Table 11
Comparison of performance of proposed USVM-RFE in terms of accuracy (%) with existing algorithms on ADNI baseline dataset for CN vs MCI using VolBM features. Bold values indicate highest accuracy for the dataset, and underlined values show highest accuracy of the algorithm.

Tissue features (%)	Proposed USVM-RFE			SVM-RFE			TWSVM-RFE		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
CN vs MCI									
3	**72.61**	86.67	47.79	69.41	75.83	58.09	71.28	67.08	78.68
5	**72.34**	84.58	50.74	71.28	86.25	44.85	69.68	86.25	40.44
8	69.41	82.08	47.06	**71.01**	86.25	44.12	69.95	82.5	47.79
10	**70.74**	76.25	61.03	70.21	82.92	47.79	70.21	82.5	48.53
13	**70.74**	81.25	52.21	69.95	80.83	50.74	**70.74**	79.17	55.88
15	**72.07**	82.92	52.94	69.95	80	52.21	68.35	75.83	55.15
20	**72.34**	81.25	56.62	70.74	82.5	50	68.62	73.75	59.56
25	70.74	85.42	44.85	**71.54**	84.17	49.26	68.09	73.33	58.82
30	69.41	82.92	45.59	**70.74**	87.08	41.91	67.82	86.67	34.56
35	68.88	73.33	61.03	**70.74**	86.25	43.38	66.49	78.75	44.85
40	**69.41**	76.25	57.35	**69.41**	86.67	38.97	67.02	62.08	75.74
45	**70.21**	78.75	55.15	69.41	86.25	39.71	68.09	76.67	52.94
50	**71.28**	80.42	55.15	68.62	85.42	38.97	68.88	81.25	47.06

Table 12
Comparison of performance of proposed USVM-RFE in terms of accuracy (%) with existing algorithms on ADNI baseline dataset for MCI vs AD using VolBM features. Bold values indicate highest accuracy for the dataset, and underlined values show highest accuracy of the algorithm.

Tissue features (%)	Proposed USVM-RFE			SVM-RFE			TWSVM-RFE		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
MCI vs AD									
3	**70.45**	36.7	85.6	68.47	24.77	88.07	69.03	0	100
5	**70.17**	38.53	84.36	69.32	55.05	75.72	63.64	64.22	63.37
8	70.17	40.37	83.54	69.6	39.45	83.13	**70.74**	17.43	94.65
10	**71.88**	41.28	85.6	71.59	40.37	85.6	**71.88**	25.69	92.59
13	**71.59**	42.2	84.77	68.18	42.2	79.84	68.75	34.86	83.95
15	**70.17**	46.79	80.66	68.47	44.95	79.01	67.05	33.95	81.89
20	**70.74**	54.13	78.19	69.89	40.37	83.13	63.92	53.21	68.72
25	**71.02**	55.05	78.19	66.76	40.37	78.6	62.5	51.38	67.49
30	**68.75**	54.13	75.31	67.33	38.53	80.25	60.23	52.29	63.79
35	**69.89**	52.29	77.78	67.05	39.45	79.42	62.5	54.13	66.26
40	**70.45**	52.29	78.6	67.61	39.45	80.25	65.34	49.54	72.43
45	**71.02**	54.13	78.6	68.18	44.95	78.6	65.06	52.29	70.78
50	**70.74**	49.54	80.25	68.18	48.62	76.95	61.65	49.54	67.08

having higher accuracy than TWSVM-RFE in lesser sized feature sets.

The results for the classification performance of USVM, SVM and TWSVM for all the features are shown in Table 13. One can observe that USVM is performing better than the other algorithms for all the cases. This is due to the introduction of universum samples, which provide prior information about AD data. The optimal parameters used in the RFE process are selected from Table 13.

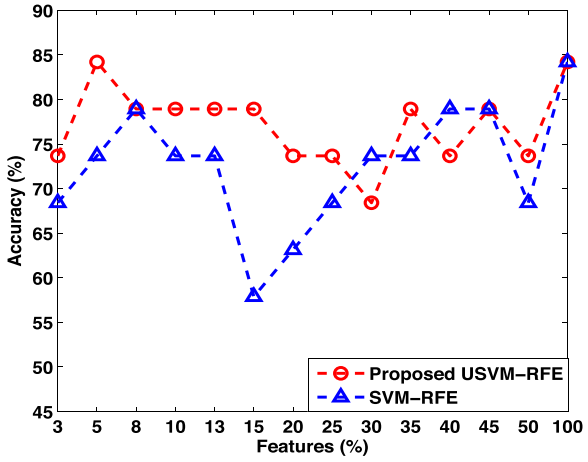
5.2.4. UCI datasets

Experimental results on classification of cancer patients in Wpbc and Wdbc are shown in Fig. 6. The optimal parameters for RFE process are shown in Table 9.

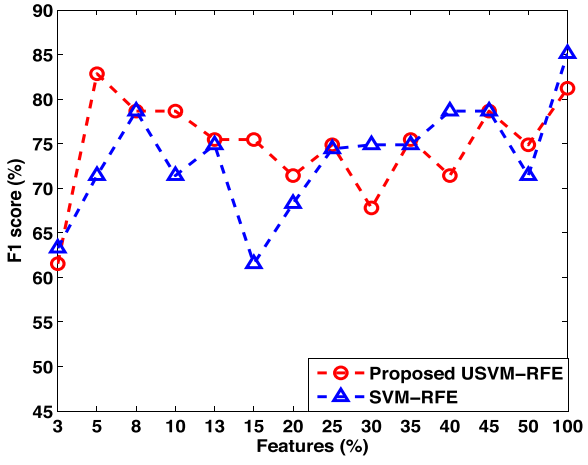
One can notice in Fig. 6(a) that proposed USVM-RFE performs better than SVM-RFE for lower dimensional feature sets. Also, it can be observed that USVM-RFE recovers to 82.81% accuracy for 13% features immediately after a decline to accuracy

Table 13
Comparison of performance of USVM in terms of accuracy (%) with other existing algorithms on ADNI baseline dataset using all VolBM features. Bold values indicate highest accuracy for the dataset.

Dataset (Train size, Test size)	USVM			SVM			TWSVM		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
CN vs AD	**87.6**	81.51	93.13	87.2	78.15	95.42	68	77.31	59.54
($165 \times 91, 250 \times 91$)	(10^{-2}, 0.6, 0.3)			(10^{-2})			(10^{-1})		
CN vs MCI	**70.21**	81.25	50.74	68.35	83.75	41.18	63.83	68.75	55.15
($250 \times 91, 376 \times 91$)	(10^{-2}, 0.8, 0.3)			(10^{-2})			(10^{1})		
MCI vs AD	**72.73**	39.45	87.65	64.77	45.87	73.25	58.24	60.55	57.2
($233 \times 91, 352 \times 91$)	(10^{-2}, 0.6, 0.3)			(10^{0})			(10^{-6})		



(a)



(b)

Fig. 4. Plot showing comparison of (a) accuracy, and (b) F1 score of SVM-RFE, and proposed USVM-RFE for classification of CN vs MCI using VolBM features.

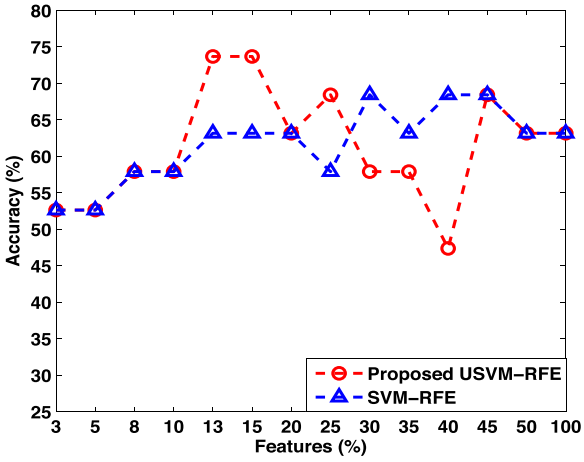
of 65.63%. This is better than SVM-RFE for the same feature sets. Similarly, better performance is shown by USVM-RFE for Wdbc dataset in Fig. 6(b). The comparison of training time is shown in Table 9. It can be seen that USVM takes more time than SVM. This is due to the inclusion of universum data. The addition of universum data leads to better generalization performance. So, the training time is a tradeoff for classification accuracy.

6. Discussion

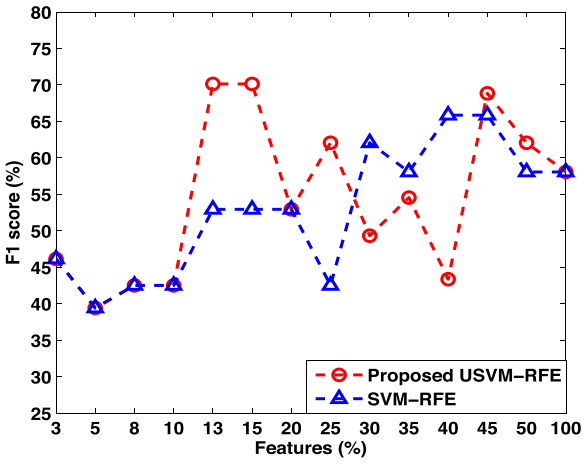
In this section, we present the discussion on reduced features obtained from VBM and VolBM analysis by SVM-RFE, and proposed USVM-RFE for CN vs AD, CN vs MCI, and MCI vs AD cases. The different regions of brain affected in Alzheimer's disease are illustrated in supplementary Fig. S.6.

6.1. VBM feature selection

The features selected from VBM analysis are shown in Figs. 7–9, and the region names of the clusters are given in supplementary Figs. S.8–S.13.

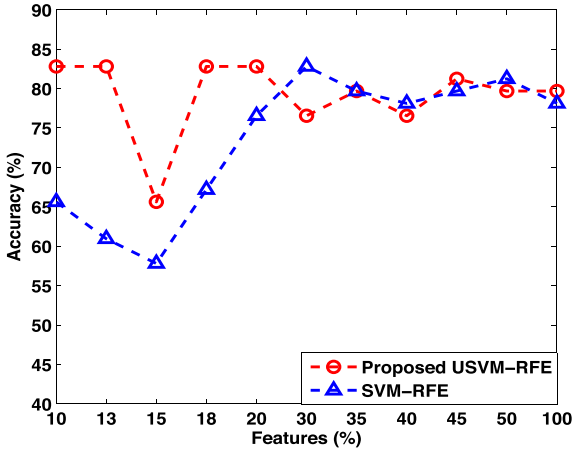


(a)

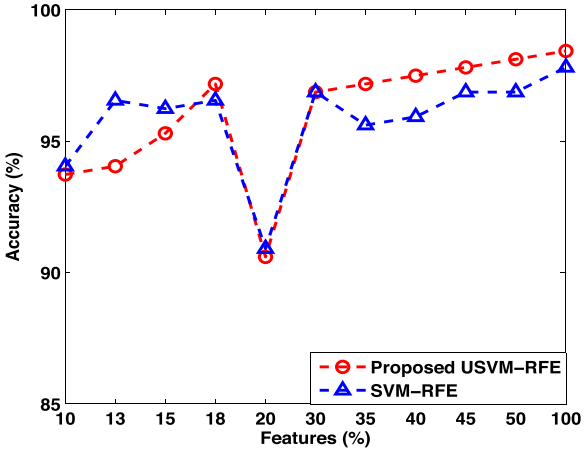


(b)

Fig. 5. Plot showing comparison of (a) accuracy, and (b) F1 score of SVM-RFE, and proposed USVM-RFE for classification of MCI vs AD using VolBM features.



(a)



(b)

Fig. 6. Plot showing comparison of classification accuracy of SVM-RFE, and proposed USVM-RFE for (a) Wpbc, and (b) Wdbc datasets.

In case of CN vs AD, our proposed USVM-RFE has shown better classification accuracy for less number of voxels as compared to SVM-RFE. The feature set with least size and highest accuracy is illustrated in Fig. 7 for both the algorithms. For SVM-RFE, the most overlapping GM regions with AAL atlas from supplementary Fig. S.8 are: amygdala (left), amygdala (right), hippocampus (left), parahippocampal (left), parahippocampal (right), and hippocampus (right).

For proposed USVM-RFE, the regions from supplementary Fig. S.9 are: amygdala (left), hippocampus (left), parahippocampus (left), parahippocampus (right), and temporal pole mid (left). This is in accordance with previous studies [45].

For CN vs MCI, Fig. 8 shows the regions of increased CSF. The brain regions proximal to area of increased CSF are given in supplementary Figs. S.10 and S.11 for SVM-RFE and proposed USVM-RFE respectively. The regions selected by SVM-RFE are: cingulate anterior (left), temporal pole mid (left), hippocampus (left), temporal pole superior (right), and parahippocampal (right). For proposed USVM-RFE, the regions are: cingulate anterior (left), temporal pole superior (right), temporal pole mid (left), temporal pole mid (right), and parahippocampal (right). This is a result of atrophy in the temporal regions [46]. The CSF volumes of our cohort also support this finding. The mean and standard deviation of CSF volume (mm^3) of CN, MCI, and AD are 1306.43 ± 410.93, 1493.98 ± 388.67 and 1652.78 ± 634.80 respectively.

This indicates that CSF voxels are useful for classification of CN vs MCI. Moreover, the number of CSF voxels selected by proposed USVM-RFE for highest accuracy are lesser than SVM-RFE. This helps in localizing the regions with change in CSF volume.

In MCI vs AD, there is significant atrophy in the left temporal lobe as compared to right in both SVM-RFE and USVM-RFE features as shown in Fig. 9. The regions for voxels selected by SVM-RFE are: amygdala (left), parahippocampal (left), fusiform (left), hippocampus (left), temporal pole superior (left), and temporal inferior (left) (For details, refer supplementary Fig. S.12). Proposed USVM-RFE selected fewer voxels with higher classification accuracy as compared to SVM-RFE. The corresponding regions are: amygdala (left), parahippocampal (left), fusiform (left), temporal pole superior (left), and hippocampus (left) (For details refer Fig. S.13). Moreover, voxels selected by proposed USVM-RFE are mostly in the left side of brain. This is due to asymmetric atrophy of left and right side of brain in AD [47,48]. The variation of amygdala and hippocampus volume in left and right side of brain in our cohort is shown in supplementary Fig. S.14.

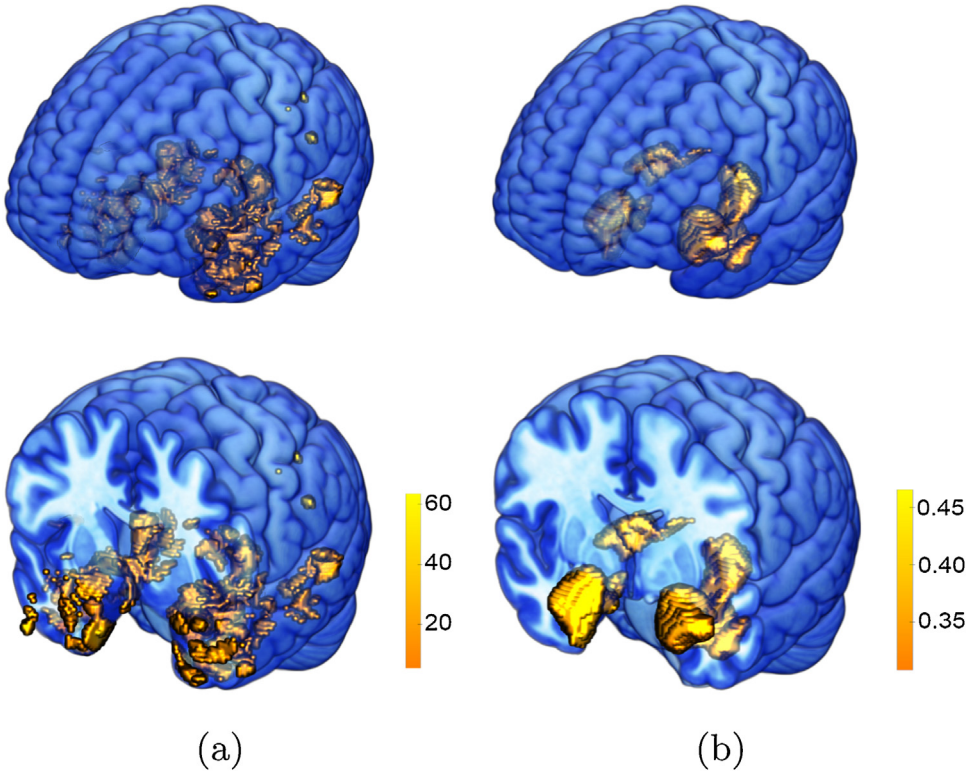


Fig. 7. Illustration of reduced GM voxels (VBM) obtained after feature elimination process of (a) proposed USVM-RFE and (b) SVM-RFE for CN vs AD. The variation of weights is shown using heat map.

In most of the cases, GM features provided better accuracy due to significant atrophy of GM regions in the brain [49]. Fig. 10 illustrates the variation of accuracy with feature set for SVM-RFE and proposed USVM-RFE. It can be seen that proposed USVM-RFE selects features from all the regions of brain as compared to SVM-RFE for CN vs AD. This shows that SVM-RFE is having local information about the dataset in every iteration, while proposed USVM-RFE has global information, i.e., distribution of data. This results in the selection of discriminative voxels from different regions of brain. This illustrates the fact that for classification of Alzheimer's disease, optimal features are needed from different regions of brain.

6.2. VolBM feature selection

The volumetric features selected by SVM-RFE and proposed USVM-RFE for CN vs AD are shown in Table 14. It can be seen that amygdala volume is having highest rank in both SVM-RFE and proposed USVM-RFE. This is due to atrophy of amygdala in AD subjects [50].

Other highest ranking features for SVM-RFE and USVM-RFE are parahippocampal thickness, entorhinal thickness, hippocampus volume, and inferior parietal thickness which are in accordance with other studies [51,52]. This also correlates with our VBM based feature analysis. The decrease in measures of these regions is also illustrated as box plots in supplementary Fig. S.14.

However, one can see that proposed USVM-RFE assigned higher rank to hippocampus volume than inferior parietal thickness as compared to SVM-RFE. This justifies better feature selection by proposed USVM-RFE, as hippocampus is one of the prominent features for CN vs AD classification [53]. For CN vs MCI, Table 15 shows that inferior parietal thickness is the most discriminative feature for SVM-RFE, while USVM-RFE ranked parahippocampal thickness as the most optimal. Other studies also suggested thinning of parahippocampal gyrus [45] in MCI patients.

For MCI vs AD, Table 16 shows the most significant features as superior temporal and entorhinal thickness by proposed USVM-RFE and SVM-RFE respectively. This was also shown in previous studies [52,54]. One can observe that parahippocampal thickness is selected as discriminative feature in all cases, i.e., CN vs AD, CN vs MCI, and MCI vs AD. The reason for this is the fact that parahippocampal region encloses the brain structures affected in Alzheimer's disease [55].

From the results, it is evident that proposed USVM-RFE gives higher classification accuracy than SVM-RFE due to its prior knowledge about the distribution of data. Moreover, the features selected by proposed USVM-RFE are also in accordance with the literature which justifies its applicability.

7. Conclusions and future work

We proposed a universum based technique for feature selection. This provides an improvement over SVM-RFE algorithm by giving prior information about data. On the basis of our analysis on Alzheimer's disease, proposed USVM-RFE has performed better than SVM-RFE in most of the cases for classification of CN, MCI, and AD subjects. Moreover, we presented an approach of using VBM on training and testing phase separately. This is useful in real world scenarios. We provided an analysis of the feature extraction methods for MRI images, i.e., voxel based and volume based features. On the basis of our work on VBM and VolBM features, USVM-RFE achieved relatively better classification accuracy of CN vs AD and MCI vs AD in VolBM as 100% and 73.68% respectively. For CN vs MCI, VBM features provided highest accuracy of 90% with CSF measures. According to our feature analysis, amygdala volume is a prominent discriminative feature for detection of Alzheimers.

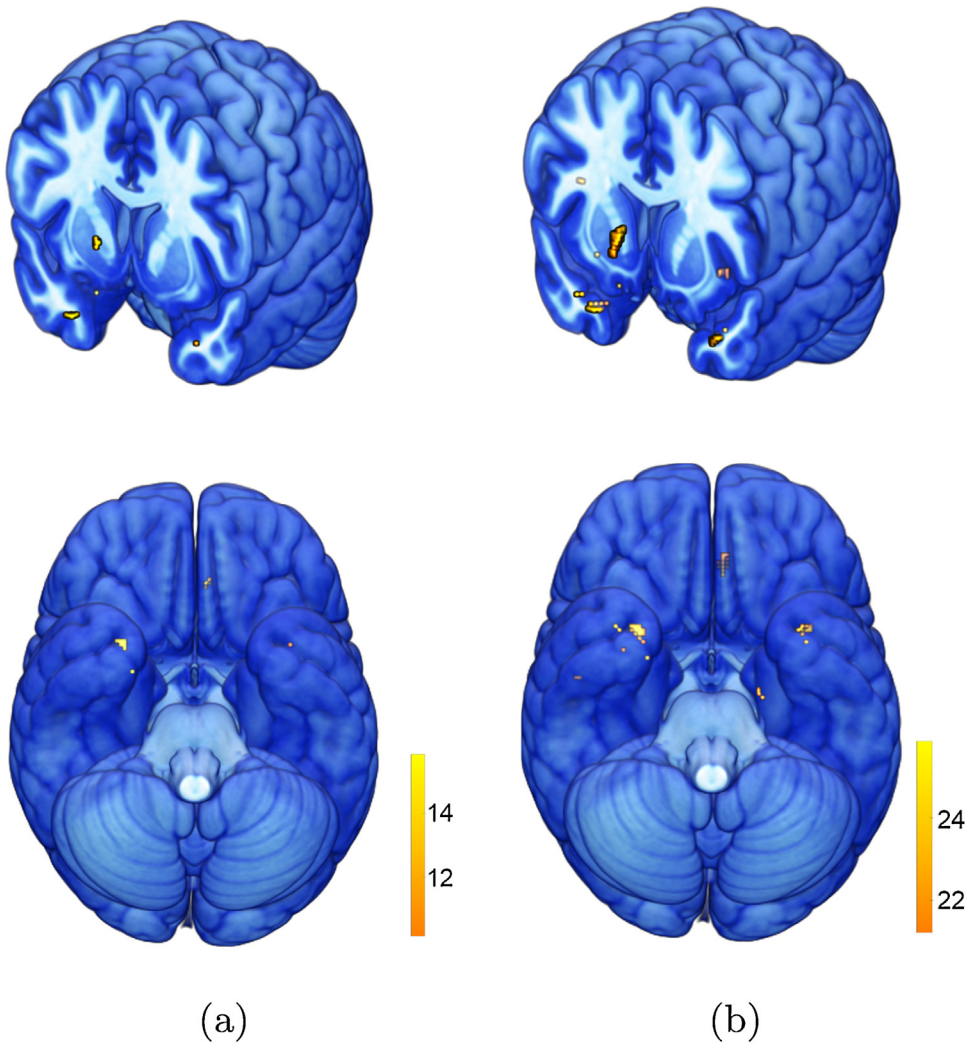


Fig. 8. Illustration of reduced CSF voxels (VBM) obtained after feature elimination process of (a) proposed USVM-RFE and (b) SVM-RFE for CN vs MCI. The variation of weights is shown using heat map.

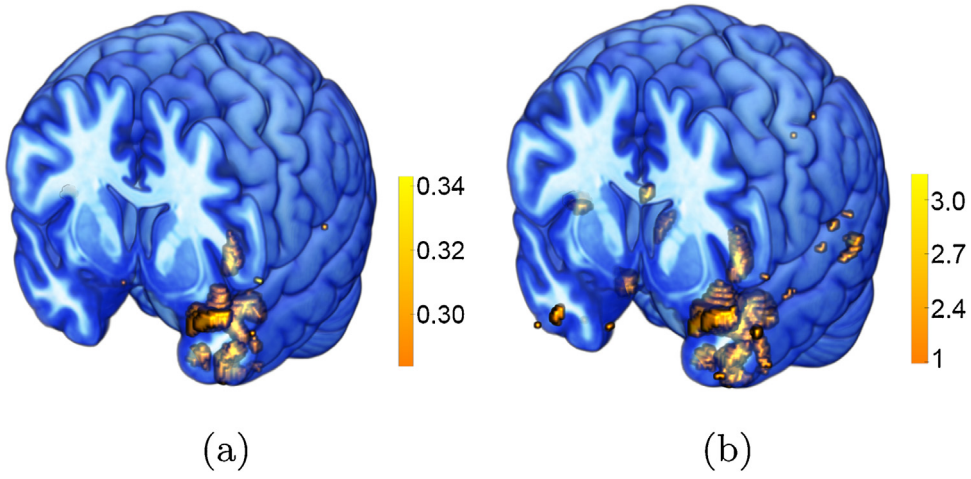


Fig. 9. Illustration of reduced GM voxels (VBM) obtained after feature elimination process of (a) proposed USVM-RFE and (b) SVM-RFE for MCI vs AD. The variation of weights is shown using heat map.

Table 14
Comparison of feature sets selected by proposed USVM-RFE with SVM-RFE from VolBM features on classification of CN vs AD. The features are represented by their feature IDs from supplementary Table S.1.

Proposed USVM-RFE									SVM-RFE								
Features (%)	13	10	8	5	3	Overall	Score	Feature name	Features (%)	13	10	8	5	3	Overall	Score	Feature name
Accuracy (%)	90	100	100	100	85				Accuracy (%)	100	100	100	85	85			
Feature rank									Feature rank								
1	48	48	48	48	48	**48**	50	Amygdala volume	**1**	48	48	48	48	48	**48**	50	Amygdala volume
2	53	20	20	20	11	**20**	35	Parahippocampal thickness	**2**	20	20	13	20	11	**20**	35	Parahippocampal thickness
3	20	53	53	11		**11**	35	Entorhinal thickness	**3**	53	53	20	11		**11**	35	Entorhinal thickness
4	71	71	11	13		**53**	25	Hippocampus volume	**4**	11	13	53	13		**13**	29	Inferior parietal thickness
5	11	13	71			**13**	23	Inferior parietal thickness	**5**	13	12	11			**53**	23	Hippocampus volume
6	13	11	13			**71**	20	WM entorhinal volume	**6**	71	11	12			**12**	14	Fusiform thickness
7	12	12	12			**12**	12	Fusiform thickness	**7**	5	71	5			**5**	11	Temporal pole thickness
8	10	10				**10**	6	Precuneus thickness	**8**	12	5				**71**	9	WM entorhinal volume
9	5	5				**5**	4	Temporal pole thickness	**9**	57	57				**57**	4	Cortex volume
10	15					**15**	1	Isthmus cingulate thickness	**10**	8					**8**	1	Caudal middle frontal thickness
11	57					**57**	0	Cortex volume	**11**	36					**36**	0	Fourth ventricle volume

Table 15
Comparison of feature sets selected by proposed USVM-RFE with SVM-RFE from VolBM features on CN vs MCI classification. The features are represented by their feature IDs from Table S.1.

Proposed USVM-RFE									SVM-RFE								
Features (%)	13	10	8	5	3	Overall	Score	Feature name	Features (%)	13	10	8	5	3	Overall	Score	Feature name
Accuracy (%)	78.95	78.95	78.95	84.21	73.68				Accuracy (%)	73.68	73.68	78.95	73.68	68.42			
Feature rank									Feature rank								
1	50	20	50	20	20	**20**	47	Parahippocampal thickness	**1**	13	13	13	13	48	**13**	49	Inferior parietal thickness
2	20	50	3	3	50	**50**	46	Cerebellum WM volume	**2**	63	50	48	48	13	**48**	42	Amygdala volume
3	34	3	20	50		**3**	31	Middle temporal thickness	**3**	40	40	40	50		**40**	31	Ventral DC volume
4	17	47	17	47		**47**	21	Accumbens area volume	**4**	48	48	50	40		**50**	30	Cerebellum WM volume
5	77	17	47			**17**	20	Lateral orbitofrontal thickness	**5**	50	63	63			**63**	21	WM middle temporal volume
6	3	77	82			**34**	15	Insula thickness	**6**	9	20	9			**9**	14	Cuneus thickness
7	13	6	34			**77**	11	WM lingual volume	**7**	20	9	20			**20**	13	Parahippocampal thickness
8	82	34				**82**	10	WM pars opercularis volume	**8**	16	67				**16**	5	Lateral occipital thickness
9	84	82				**6**	4	Transverse temporal thickness	**9**	67	16				**67**	5	WM temporal pole volume
10	47					**13**	4	Inferior parietal thickness	**10**	88					**88**	1	WM rostral anterior cingulate volume
11	6					**84**	2	WM pericalcarine volume	**11**	68					**68**	0	WM transverse temporal volume

Table 16
Comparison of feature sets selected by proposed USVM-RFE with SVM-RFE from VolBM features on MCI vs AD classification. The features are represented by their feature IDs from Table S.1.

Proposed USVM-RFE											SVM-RFE									
Features (%)	15	13	10	8	5	3	Overall	Score	Feature name		Features (%)	15	13	10	8	5	3	Overall	Score	Feature name
Accuracy (%)	73.68	73.68	57.89	57.89	52.63	52.63					Accuracy (%)	63.16	63.16	57.89	57.89	52.63	52.63			
Feature rank											Feature rank									
1	74	74	74	4	4	4	**4**	67	Superior temporal thickness		**1**	74	11	11	74	4	4	**11**	68	Entorhinal thickness
2	11	4	11	11	11	11	**11**	65	Entorhinal thickness		**2**	11	74	4	11	11	11	**4**	65	Superior temporal thickness
3	4	11	4	74	74		**74**	56	WM lateral occipital volume		**3**	4	4	74	4	12		**74**	45	WM lateral occipital volume
4	20	68	20	68	20		**20**	43	Parahippocampal thickness		**4**	68	20	20	20	20		**20**	44	Parahippocampal thickness
5	68	20	68	20			**68**	34	WM transverse temporal volume		**5**	20	68	68	12			**12**	38	Fusiform thickness
6	12	12	48	12			**12**	27	Fusiform thickness		**6**	12	1	12	68			**68**	32	WM transverse temporal volume
7	19	19	12	2			**2**	18	Inferior temporal thickness		**7**	1	12	48	3			**3**	19	Middle temporal thickness
8	48	1	1				**48**	15	Amygdala volume		**8**	48	48	3				**1**	17	Bankssts thickness
9	2	2	2				**19**	12	Lingual thickness		**9**	3	3	1				**48**	16	Amygdala volume
10	50	48					**1**	11	Bankssts thickness		**10**	2	18					**18**	3	Medial orbitofrontal thickness
11	85	85					**85**	4	WM pars orbitalis volume		**11**	85	50					**50**	3	Cerebellum WM volume
12	1						**50**	3	Cerebellum WM volume		**12**	50						**2**	3	Inferior temporal thickness
13	82						**82**	0	WM pars opercularis volume		**13**	18						**85**	2	WM pars orbitalis volume

One of the important advantages of our universum based algorithm is the global or holistic approach in feature selection as compared to SVM-RFE. This provides robustness to our USVM-RFE in each iteration of feature elimination. Moreover, we found that there is significant atrophy in left side of brain in Alzheimer's disease. This result can be used to develop neuropsychological tests which require functions performed by left hemisphere of brain. This may result in early detection of Alzheimers through a non-invasive approach.

The proposed USVM-RFE has the limitation of higher training time due to the universum. Also, more research is needed in the proper selection of universum, as it is dependent on the classification problem. Moreover, our USVM-RFE can be modified for multiclass classification problems, and used on various other real world applications.

CRediT authorship contribution statement

B. Richhariya: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - review & editing, Visualization. **M. Tanveer:** Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization. **A.H. Rashid:** Conceptualization, Methodology, Validation, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

We are thankful to the anonymous reviewers for their constructive comments for the improvement of the paper.

This work is supported by Department of Science and Technology, India under Ramanujan fellowship scheme grant no. SB/S2/RJN-001/2016, Council of Scientific & Industrial Research (CSIR), New Delhi, India under Extra Mural Research (EMR) scheme grant no. 22(0751)/17/EMR-II, and Science and Engineering Research Board (SERB), India under Early Career Research Award scheme grant no. ECR/2017/000053. We gratefully acknowledge the Indian Institute of Technology Indore for providing facilities and support. We are thankful to the Indian Institute of Technology Indore for providing Institute fellowship to Mr. Bharat Richhariya.

The collection of data and sharing of this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904), and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The funding for ADNI is provided by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. The dissemination of ADNI data is carried out by the Laboratory for Neuro Imaging at the University of Southern California.

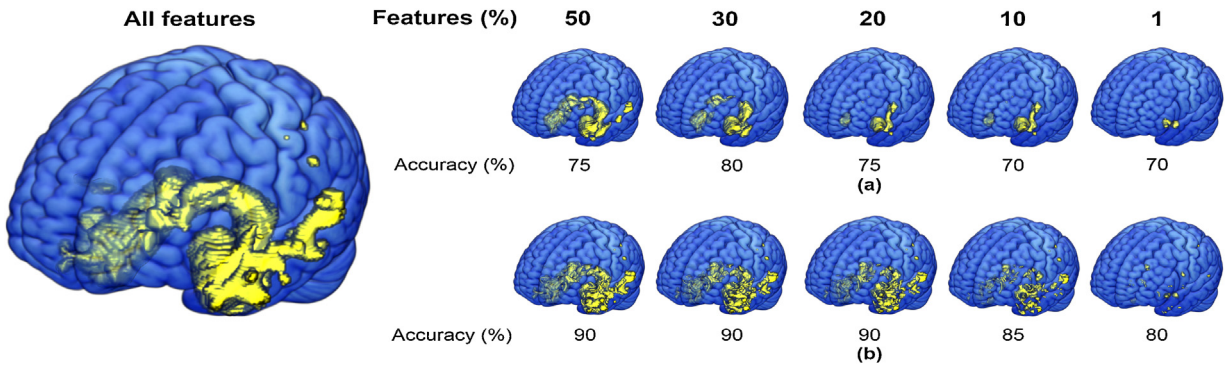


Fig. 10. Comparison of feature elimination process based on classification accuracy of (a) SVM-RFE and (b) proposed USVM-RFE on CN vs AD for GM (VBM) features.

Declaration of Competing Interest

None declared.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.bspc.2020.101903.

References

[1] C. Patterson, The State of the Art of Dementia Research: New Frontiers, 2018, World Alzheimer's Report 2018.
[2] E. Pellegrini, et al., Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review, Alzheimer's Dementia: Diagn. Assess. Dis. Monit. 10 (2018) 519–535.
[3] M. Tanveer, B. Richhariya, R. Khan, A. Rashid, P. Khanna, M. Prasad, C. Lin, Machine learning techniques for the diagnosis of Alzheimer's disease: a review, ACM Trans. Multimedia Comput. Commun. Appl. (2019) (in press).
[4] P. Vemuri, et al., Differential diagnosis of neurodegenerative dementias using structural MRI, Alzheimer's Dementia: J. Alzheimer's Assoc. 5 (2009) P16.
[5] L. Mesrob, B. Magnin, O. Colliot, M. Sarazin, V. Hahn-Barma, B. Dubois, P. Gallinari, S. Lehéricy, S. Kinkingnéhun, H. Benali, Identification of atrophy patterns in Alzheimer's disease based on SVM feature selection and anatomical parcellation, International Workshop on Medical Imaging and Virtual Reality (2008) 124–132.
[6] L. Khedher, J. Ramírez, J.M. Górriz, A. Brahim, F. Segovia, Alzheimer's Disease Neuroimaging Initiative, Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images, Neurocomputing 151 (2015) 139–150.
[7] P. Vemuri, J.L. Gunter, M.L. Senjem, J.L. Whitwell, K. Kantarci, D.S. Knopman, B.F. Boeve, R.C. Petersen, C.R. Jack Jr., Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies, NeuroImage 39 (2008) 1186–1197.
[8] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, Alzheimer's Disease Neuroimaging Initiative, Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database, NeuroImage 56 (2011) 766–781.
[9] E. Westman, J.-S. Muehlboeck, A. Simmons, Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion, NeuroImage 62 (2012) 229–238.
[10] L. Sørensen, C. Igel, A. Pai, I. Balas, C. Anker, M. Lillholm, M. Nielsen, Alzheimer's Disease Neuroimaging Initiative, Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry, NeuroImage: Clin. 13 (2017) 470–482.
[11] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[12] S. Lahmiri, A. Shmuel, Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease, Biomed. Signal Process. Control 52 (2019) 414–419.
[13] R.S. Kamathe, K.R. Joshi, A novel method based on independent component analysis for brain MR image tissue classification into CSF, WM and GM for atrophy detection in Alzheimer's disease, Biomed. Signal Process. Control 40 (2018) 41–48.
[14] L. Sharma, R. Tripathy, S. Dandapat, Multiscale energy and eigenspace approach to detection and localization of myocardial infarction, IEEE Trans. Biomed. Eng. 62 (2015) 1827–1837.
[15] R.K. Tripathy, A. Bhattacharyya, R.B. Pachori, A novel approach for detection of myocardial infarction from ECG signals of multiple electrodes, IEEE Sens. J. 19 (2019) 4509–4517.
[16] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422.
[17] M.-L. Huang, Y.-H. Hung, W. Lee, R. Li, B.-R. Jiang, SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier, Sci. World J. 2014 (2014).
[18] Z.-M. Yang, J.-Y. He, Y.-H. Shao, Feature selection based on linear twin support vector machines, Proc. Comput. Sci. 17 (2013) 1039–1046.
[19] X.-W. Chen, X. Zeng, D. van Alphen, Multi-class feature selection for texture classification, Pattern Recogn. Lett. 27 (2006) 1685–1691.
[20] S. Mishra, D. Mishra, SVM-BT-RFE: An improved gene selection framework using Bayesian T-test embedded in support vector machine (recursive feature elimination) algorithm, Karbala Int. J. Mod. Sci. 1 (2015) 86–96.
[21] J. Weston, R. Collobert, F. Sinz, L. Bottou, V. Vapnik, Inference with the universum, Proceedings of the 23rd International Conference on Machine learning (2006) 1009–1016.
[22] B. Richhariya, M. Tanveer, EEG signal classification using universum support vector machine, Expert Syst. Appl. 106 (2018) 169–182.
[23] B. Richhariya, A. Sharma, M. Tanveer, Improved universum twin support vector machine, 2018 IEEE Symposium Series on Computational Intelligence (SSCI) (2018) 2045–2052.
[24] B. Richhariya, M. Tanveer, A reduced universum twin support vector machine for class imbalance learning, Pattern Recogn. 102 (2020), http://dx.doi.org/10.1016/j.patcog.2019.107150.
[25] B. Richhariya, D. Gupta, Facial expression recognition using iterative universum twin support vector machine, Appl. Soft Comput. 76 (2019) 53–67.
[26] J. Ashburner, A fast diffeomorphic image registration algorithm, NeuroImage 38 (2007) 95–113.
[27] H. Matsuda, Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease, Aging Dis. 4 (2013) 29.
[28] X. Hao, D. Zhang, Ensemble universum SVM learning for multimodal classification of Alzheimer's disease, International Workshop on Machine Learning in Medical Imaging (2013) 227–234.
[29] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach. Intell. (2018).
[30] R.Y. Lo, W.J. Jagust, Predicting missing biomarker data in a longitudinal study of Alzheimer disease, Neurology 78 (2012) 1376–1382.
[31] S. Keller, U. Wieshmann, C. Mackay, C. Denby, J. Webb, N. Roberts, Voxel based morphometry of grey matter abnormalities in patients with medically intractable temporal lobe epilepsy: effects of side of seizure onset and epilepsy duration, J. Neurol. Neurosurg. Psychiatry 73 (2002) 648–655.
[32] L.M. O'Brien, D.A. Ziegler, C.K. Deutsch, J.A. Frazier, M.R. Herbert, J.J. Locascio, Statistical adjustments for brain size in volumetric neuroimaging studies: some practical implications in methods, Psychiatry Res.: Neuroimaging 193 (2011) 113–122.
[33] I. Beheshti, H. Demirel, H. Matsuda, Alzheimer's Disease Neuroimaging Initiative, 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 (2017) 109–119.
[34] Z. Xiao, Y. Ding, T. Lan, C. Zhang, C. Luo, Z. Qin, Brain MR image classification for Alzheimer's disease diagnosis based on multifeature fusion, Comput. Math. Methods Med. 2017 (2017).
[35] V. Youssofzadeh, B. McGuinness, L.P. Maguire, K. Wong-Lin, Multi-kernel learning with dartel improves combined MRI-PET classification of Alzheimer's disease in AIBL data: group and individual analyses, Front. Human Neurosci. 11 (2017) 380.
[36] M. Reuter, N.J. Schmansky, H.D. Rosas, B. Fischl, Within-subject template estimation for unbiased longitudinal image analysis, NeuroImage 61 (2012) 1402–1418.
[37] Y.-W. Chen, C.-J. Lin, Combining SVMs with various feature selection strategies, in: Feature Extraction, Springer, 2006, pp. 315–324.
[38] Z. Qi, Y. Tian, Y. Shi, Twin support vector machine with universum data, Neural Netw. 36 (2012) 112–119.

[39] P.A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, NeuroImage 31 (2006) 1116–1128.
[40] J. Ahrens, B. Geveci, C. Law, Paraview: an end-user tool for large data visualization, Visual. Handb. 717 (2005).
[41] J.A. Maldjian, P.J. Laurienti, R.A. Kraft, J.H. Burdette, An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets, NeuroImage 19 (2003) 1233–1239.
[42] C. Chu, A.-L. Hsu, K.-H. Chou, P. Bandettini, C. Lin, A.D.N. Initiative, Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images, NeuroImage 60 (2012) 59–70.
[43] D. Dua, C. Graff, UCI Machine Learning Repository, 2017.
[44] D. Schmitter, et al., An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease, NeuroImage: Clin. 7 (2015) 7–17.
[45] S. Krumm, S.L. Kivisaari, A. Probst, A.U. Monsch, J. Reinhardt, S. Ulmer, C. Stippich, R.W. Kressig, K.I. Taylor, Cortical thinning of parahippocampal subregions in very early Alzheimer's disease, Neurobiol. Aging 38 (2016) 188–196.
[46] Y. Fan, N. Batmanghelich, C.M. Clark, C. Davatzikos, Alzheimer's Disease Neuroimaging Initiative, Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline, NeuroImage 39 (2008) 1731–1743.
[47] S. Derflinger, C. Sorg, C. Gaser, N. Myers, M. Arsic, A. Kurz, C. Zimmer, A. Wohlschl“ager, M. M”uhlau, Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized, J. Alzheimer's Dis. 25 (2011) 347–357.
[48] O. Bugiani, J. Constantinidis, B. Ghetti, C. Bouras, F. Tagliavini, Asymmetrical cerebral atrophy in Alzheimer's disease, Clin. Neuropathol. 10 (1991) 55–60.
[49] P.S. Weston, I.J. Simpson, N.S. Ryan, S. Ourselin, N.C. Fox, Diffusion imaging changes in grey matter in Alzheimer's disease: a potential marker of early neurodegeneration, Alzheimer's Res. Ther. 7 (2015) 47.
[50] S.P. Poulin, R. Dautoff, J.C. Morris, L.F. Barrett, B.C. Dickerson, Alzheimer's Disease Neuroimaging Initiative, Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity, Psychiatry Res.: Neuroimaging 194 (2011) 7–13.
[51] A. Du, et al., Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease, J. Neurol. Neurosurg. Psychiatry 71 (2001) 441–447.
[52] L. Velayudhan, et al., Entorhinal cortex thickness predicts cognitive decline in Alzheimer's disease, J. Alzheimer's Dis. 33 (2013) 755–766.
[53] D.H. Adler, et al., Characterizing the human hippocampus in aging and Alzheimer's disease using a computational atlas derived from ex vivo MRI and histology, Proc. Natl. Acad. Sci. U.S.A. 115 (2018) 4252–4257.
[54] B.C. Dickerson, R.A. Sperling, Functional abnormalities of the medial temporal lobe memory system in mild cognitive impairment and Alzheimer's disease: insights from functional MRI studies, Neuropsychologia 46 (2008) 1624–1635.
[55] G.W. Van Hoesen, J.C. Augustinack, J. Dierking, S.J. Redman, R. Thangavel, The parahippocampal gyrus in Alzheimer's disease: clinical and preclinical neuroanatomical correlates, Ann. N. Y. Acad. Sci. 911 (2000) 254–274.