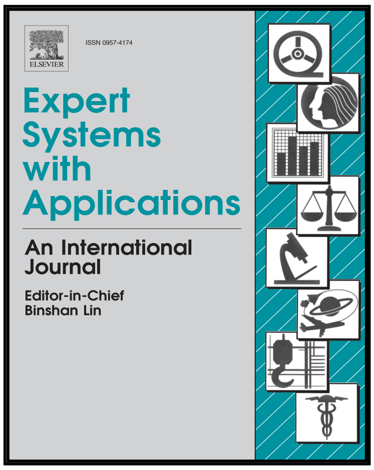
Accepted Manuscript



EEG signal classification using universum support vector machine

B. Richhariya , M. Tanveer

PII: S0957-4174(18)30210-0
DOI: 10.1016/j.eswa.2018.03.053
Reference: ESWA 11900

To appear in: *Expert Systems With Applications*

Received date: 3 November 2017
Revised date: 25 February 2018
Accepted date: 26 March 2018

Highlights

- Electroencephalogram signal classification is performed using universum learning.
- Support vector machine classifier uses prior information from interictal signals.
- Many feature extraction techniques are used for comparing the algorithms.
- Universum support vector machine is used first time for seizure classification.

EEG signal classification using universum support vector machine

B. Richhariya[1], M. Tanveer[2]

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

Abstract

Support vector machine (SVM) has been used widely for classification of electroencephalogram (EEG) signals for the diagnosis of neurological disorders such as epilepsy and sleep disorders. SVM shows good generalization performance for high dimensional data due to its convex optimization problem. The incorporation of prior knowledge about the data leads to a better optimized classifier. Different types of EEG signals provide information about the distribution of EEG data. To include prior information in the classification of EEG signals, we propose a novel machine learning approach based on universum support vector machine (USVM) for classification. In our approach, the universum data points are generated by selecting universum from the EEG dataset itself which are the interictal EEG signals. This removes the effect of outliers on the generation of universum data. Further, to reduce the computation time, we use our approach of universum selection with universum twin support vector machine (UTSVM) which has less computational cost in comparison to traditional SVM. For checking the validity of our proposed methods, we use various feature extraction techniques for different datasets consisting of healthy and seizure signals. Several numerical experiments are performed on the generated datasets and the results of our proposed approach are compared with other baseline methods. Our proposed USVM and proposed UTSVM show better generalization performance compared to SVM, USVM, Twin SVM (TWSVM) and UTSVM. The proposed UTSVM has achieved highest classification accuracy of 99 % for the healthy and seizure EEG signals.

Keywords: Universum, interictal, support vector machine, twin support vector machine.

[1] phd1701241001@iiti.ac.in

[2] Corresponding author
E-mail addresses: tanveergouri@gmail.com, mtanveer@iiti.ac.in

1. Introduction

Electroencephalogram (EEG) signal classification is a major challenge in the field of machine learning and signal processing. EEG is widely used non-invasive technique for the detection of various types of brain disorders such as epileptic seizures and sleep disorders. In epilepsy, the extent of disease ranges from partial to generalized seizures which are reflected in their respective EEG. The different types of EEG signals are shown in fig. 2. For the better feature extraction and classification of EEG signals, several signal processing techniques have been used by researchers. Among the various feature extraction techniques, wavelet transform is one of the frequently used methods. In wavelet transform, the frequency domain features are extracted from the signal with good localization in time which is in contrast to the Fourier transform where the signal analysis is done mainly in the frequency domain. In wavelet analysis, the approximation and decomposition coefficients are used to form the feature vector as shown in fig. 3. The different families of wavelet are used for specific type of signals to get better characteristics of that signal. Adeli et al. (2003) proposed a computer aided diagnosis (CAD) method for epilepsy using discrete wavelet transform (DWT). They used Daubechies wavelet with db-4 as the mother wavelet for the feature extraction. Rosso et al. (2005) used orthogonal decimated discrete wavelet transform (ODWT) for detecting maturational changes associated with childhood absence epilepsy. Ocak (2008) performed the classification of EEG signals using wavelet packet analysis and genetic algorithm. Daubechies wavelet-2 is used for the classification of five different EEG signals (Guler & Ubeyli, 2005). Subasi and Gursoy (2010) used principal component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA) for the feature extraction, and support vector machine (SVM) for classification.

The proper selection of classification techniques is very crucial for the automated diagnosis of patients having neurological diseases. Among the various classification algorithms, support vector machines (SVMs) (Cortes and Vapnik, 1995) have emerged as a powerful classification technique. SVM solves a convex optimization problem which leads to a globally optimal solution. This is in contrast to artificial neural network (ANN) that suffers from the problem of local minima. SVM also has a lower VC (Vapnik-Chervonenkis) dimension that enables it to classify high dimensional data with less optimizing parameters. Many researchers have used SVM in the classification of EEG signals (Ma et al., 2016) and for the diagnosis of neurological diseases like epilepsy (Liu et al., 2012, Nicolaou, & Georgiou, 2012, Zavar, & Rahati, 2011). Guo et al. (2011) performed the classification of mental tasks from the analysis of EEG signals using SVM. Least squares support vector machine (LSSVM) (Suykens & Vandewalle, 1999) is used in (Li & Wen, 2009, Bajaj & Pachori, 2012, Sharma & Pachori, 2015, Joshi & Pachori, 2014) for the detection of epilepsy. LSSVM is used for classification of EEG signal with a clustering based approach (Li & Wen, 2011). For multiclass classification of EEG signals, Guler and Ubeyli (2007) proposed a support vector machine based model and showed that SVM gives better classification accuracy for EEG signals as compared to probabilistic neural network (PNN) and multilayer perceptron neural network (MLPNN).

Weston et al. (2006) proposed a universum support vector machine (USVM) to give prior information to the classifier about the distribution of data. The universum data points do not belong to any of the classes and lie within an $\varepsilon-$insensitive tube between the two classes. This approach is also called as 'learning through contradiction'. In USVM, along with the hinge loss it involves an $\varepsilon-$insensitive loss function. This universum based approach has been applied to various real world applications. Long and Tang (2016) performed the classification of investor sentiments using universum support vector machine. Gao et al. (2008) used universum SVM for prediction of translation initiation in proteins. They used two approaches for selecting the universum: one is based on uniform distribution of noise and other using random averaging of the data points. Hao and Zhang (2013) proposed an ensemble universum support vector machine for the detection of Alzheimer's disease from brain imaging data by using the patients with mild cognitive impairment (MCI) as the universum. Text classification is also performed using universum data (Liu et al., 2016).

The major challenge with universum based approach is the proper selection of universum data points. In Weston et al. (2006), the universum data is selected based on similarity of digits in digit classification. For example, digit '3' is chosen as universum for classifying '5' and '8' since its shape is similar to both '5' and '8'. Chapelle et al. (2008) presented an analysis for the selection of proper universum data. In (Bai & Cherkassky, 2008), universum samples are generated for classification of faces using the random averaging approach where the average of the pixels of two faces is used as the universum. In (Chen & Zhang, 2009), an in-between-universum (IBU) approach is proposed for the proper selection of universum. The practical conditions for choosing the universum data are given in (Cherkassky, Dai, 2009, Cherkassky et al., 2011). In the recent decade some nonparallel SVMs such as generalized eigenvalue proximal support vector machine (GEPSVM) (Mangasarian & Wild, 2006) and twin support vector machine (TWSVM) (Jayadeva et al, 2007) are proposed to reduce the computational complexity of standard SVM. Inspired by the work of TWSVM, some scholars proposed variants of TWSVM (Kumar & Gopal, 2009, Shao et al., 2011, Qi et al., 2013, Tanveer, 2015a,b, Wang et al., 2015, Khemchandani et al., 2016, Tanveer et al., 2016, Xu et al., 2017) to improve the performance and reduce the computational complexity of TWSVM. TWSVM is used for the first time in this work for the classification of seizure EEG signals,. Qi et al. (2012) proposed a universum twin support vector machine (UTSVM) to reduce the computational complexity of USVM and used the random averaging approach for universum selection. Xu et al. (2016) also used the random averaging scheme for selecting the universum data. Since the random averaging approach suffers from the effect of outliers, the method of generation of universum data depends solely on the type of application and is currently an area of research.

Motivated by the work on universum support vector machine in (Long & Tang, 2016, Gao et al., 2008, Hao & Zhang, 2013), we propose a novel approach of selecting the universum in the classification of EEG signals for seizure detection. Since universum based support vector machines have not been used for the classification of EEG signals, we also present an application of USVM and UTSVM for EEG signals. For the classification of EEG signals in the healthy and seizure (ictal) classes, the interictal EEG signals are chosen as the universum which corresponds to the EEG recording for the time period in between the seizures in a patient with epilepsy. Our approach of EEG classification is tested for different datasets that are generated using various feature extraction techniques, and the results are compared with other existing methods.

In this work, all vectors are taken as column vectors. The inner product of two vectors is represented by: $a^t b$ where a and b are the vectors of $n-$dimensional real space R^n, and a^t is the transpose of a. $||a||$ and $||G||$ represent the 2-norm of a vector a and a matrix G respectively. e denotes the vector of ones of dimension m. I represents the identity matrix of appropriate size.

The rest of this paper is organized as follows: Section 2 discusses the formulations of USVM and UTSVM. Section 3 elaborates our proposed approach of USVM and UTSVM. Several numerical experiments are performed on the datasets generated from EEG signals using different feature extraction techniques for the discussed and proposed approach in section 4. Finally, section 5 gives the conclusions and possible future directions.

2. Related Work

In this section, we briefly review USVM and UTSVM. For detailed description, the interested readers are referred to (Weston et al., 2006, Qi et al., 2012).

2.1 Universum Support Vector Machine

In case of USVM (Weston et al., 2006), the universum data points are used to provide prior knowledge about the distribution of data. The universum data is used as a constraint to lie within an $\varepsilon-$insensitive tube between the margins of the SVM hyperplane. The formulation of USVM is written as follows:

$$\min_{w,b,\xi,\psi} \frac{1}{2}||w||^2 + C\left(\sum_{i=1}^{m}\xi_i\right) + C_u \sum_{j=1}^{2|u|}\psi_j$$

$$\text{s. t.} \quad y_i(\varphi(x_i)^t w + b) \geq 1 - \xi_i,$$

$$y_j(\varphi(x_j)^t w + b) \geq -\varepsilon - \psi_j,$$

$$\xi_i \geq 0,\ \psi_j \geq 0,\ \forall i = 1,\ldots,m,\ \forall j = 1,\ldots,2|u|, \tag{1}$$

where C and C_u are the penalty parameters, ξ, ψ are slack variables, $\varphi : R^n \rightarrow R^p$ is the mapping to higher dimension where $p > n$, ε is the tolerance value for the universum, m is the total number of samples and u is the total number of universum points.

In the optimization problem of USVM we take the set of universum points twice with target values as $+1$ and -1.

The Lagrangian of the objective function (1) is given as

$$L = \frac{1}{2}||w||^2 + C\left(\sum_{i=1}^{m}\xi_i\right) + C_u\left(\sum_{j=1}^{2|u|}\psi_j\right) - \sum_{i=1}^{m}\lambda_i(y_i(\varphi(x_i)^t w + b) - 1 + \xi_i) - \sum_{i=1}^{m}\eta_i\xi_i$$
$$-\sum_{j=1}^{2|u|}\alpha_j(y_j(\varphi(x_j)^t w + b) + \varepsilon + \psi_j) - \sum_{j=1}^{2|u|}\beta_j\psi_j, \tag{2}$$

where λ, η, α and β are the Lagrange multipliers.

The dual formulation of USVM after applying the KKT conditions is written as

$$\max_{\alpha} \ -\sum_{i=1}^{m+2|u|}\mu_i\alpha_i + \frac{1}{2}\sum_{i=1}^{m+2|u|}\sum_{j=1}^{m+2|u|}\alpha_i\alpha_j K(x_i,x_j)\, y_i^t y_j$$

$$\text{s. t.} \quad 0 \leq \alpha_i \leq C,\ \forall i = 1,\ldots,m$$

$$\mu_i = 1,\ \forall i = 1,\ldots,m$$

$$0 \leq \alpha_i \leq C_u,\ \forall i = m+1,\ldots,m+2|u|$$

$$\mu_i = -\varepsilon,\ \forall i = m+1,\ldots,m+2|u|$$

$$\text{and} \quad \sum_{i=1}^{m+2|u|}\alpha_i y_i = 0,$$

where $K(x_i,x_j) = \varphi(x_i)^t\varphi(x_j)$ is the kernel function and α is the vector containing the Lagrange multipliers.

The decision function is given as

$$f(x) = sign\left(\sum_{i=1}^{m+2|u|}\alpha_i y_i\ K(x_i,x) + b\right). \tag{3}$$

2.2 Universum Twin Support Vector Machine

In UTSVM (Qi et al., 2012), two smaller quadratic programming problems (QPPs) are solved instead of one large QPP as in the case of USVM. This makes UTSVM computationally efficient in comparison to USVM. The universum data points are added in the constraints of each QPP.

Let us consider the input matrices X_1 and X_2 having size $p \times n$ and $q \times n$ respectively where p is the number of data points of 'class 1' and q is the number of data points belonging to 'class 2'. U is the matrix representing the universum data points of size $r \times n$. The total number of data samples is $m = p + q$ with r universum points and n is the dimension of each data point.

The nonlinear UTSVM comprises of the following pair of minimization problems:

$$\min_{w_1, b_1, \xi, \psi} \frac{1}{2} \| K(X_1, D^t) w_1 + e_1 b_1 \|^2 + C_1 e_2^t \xi + C_u \psi$$

$$\text{s. t.} \quad -(K(X_2, D^t) w_1 + e_2 b_1) + \xi \geq e_2, \; \xi \geq 0 \tag{4}$$

$$(K(U, D^t) w_1 + e_2 b_1) + \psi \geq (-1 + \varepsilon) e_u, \; \psi \geq 0$$

$$\min_{w_2, b_2, \eta, \psi^*} \frac{1}{2} \| K(X_2, D^t) w_2 + e_2 b_2 \|^2 + C_2 e_1^t \eta + C_u \psi^*$$

$$\text{s. t.} \quad (K(X_1, D^t) w_2 + e_1 b_2) + \eta \geq e_1, \; \eta \geq 0 \tag{5}$$

$$-(K(U, D^t) w_2 + e_u b_1) + \psi^* \geq (-1 + \varepsilon) e_u, \; \psi^* \geq 0,$$

where ξ, ψ, η, ψ^* are the slack variables; C_1, C_2 and C_u are the penalty parameters; $D = [X_1; X_2]$; ε is the tolerance value for the universum; e_1, e_2 are vectors of suitable dimensions having all values as 1's and $K(x^t, D^t) = (k(x, x_1), \ldots, k(x, x_m))$ is a row vector in R^m space.

The Lagrangians of problems (4) & (5) can be expressed as

$$L_1 = \frac{1}{2} \| K(X_1, D^t) w_1 + e_1 b_1 \|^2 + C_1 e_2^t \xi + C_u \psi + \alpha_1^t ((K(X_2, D^t) w_1 + e_2 b_1) - \xi + e_2) - \beta_1^t \xi$$
$$- \mu_1^t ((K(U, D^t) w_1 + e_2 b_1) + \psi + (1 - \varepsilon) e_u) - \gamma_1^t \psi \tag{6}$$

$$L_2 = \frac{1}{2} \| K(X_2, D^t) w_2 + e_2 b_2 \|^2 + C_2 e_1^t \eta + C_u \psi^* + \alpha_2^t ((-K(X_1, D^t) w_2 - e_1 b_2) - \eta + e_1) - \beta_2^t \eta$$
$$+ \mu_2^t ((K(U, D^t) w_2 + e_1 b_2) - \psi^* - (1 - \varepsilon) e_u) - \gamma_2^t \psi^* \tag{7}$$

where $\alpha_1 = (\alpha_{11},...,\alpha_{1q})^t$, $\beta_1 = (\beta_{11},...,\beta_{1q})^t$, $\mu_1 = (\mu_{11},...,\mu_{1r})^t$, $\gamma_1 = (\gamma_{11},...,\gamma_{1r})^t$, $\alpha_2 = (\alpha_{21},...,\alpha_{2p})^t$, $\beta_2 = (\beta_{21},...,\beta_{2p})^t$, $\mu_2 = (\mu_{21},...,\mu_{2r})^t$ and $\gamma_2 = (\gamma_{21},...,\gamma_{2r})^t$ are the Lagrange multipliers.

The Wolfe duals of Eq. (6) and (7) are obtained by applying the KKT necessary and sufficient conditions as

$$\max_{\alpha_1,\mu_1} \; e_2^t\alpha_1 - \frac{1}{2}(\alpha_1^t T - \mu_1^t O)(S^t S)^{-1}(T\alpha_1^t - O\mu_1^t) + (\varepsilon - 1)e_u{}^t\mu_1 \tag{8}$$

$$\text{s. t.} \qquad 0 \le \alpha_1 \le C_1,\; 0 \le \mu_1 \le C_u,$$

$$\max_{\alpha_2,\mu_2} \; e_1^t\alpha_2 - \frac{1}{2}(\alpha_2^t S - \mu_2^t O)(T^t T)^{-1}(S^t\alpha_2 - O^t\mu_2) + (\varepsilon - 1)e_u{}^t\mu_2 \tag{9}$$

$$\text{s. t.} \qquad 0 \le \alpha_2 \le C_2,\; 0 \le \mu_2 \le C_u,$$

where $S = [K(\mathrm{X}_1, D^t) \;\; e_1]$, $T = [K(\mathrm{X}_2, D^t) \;\; e_2]$ and $O = [K(\mathrm{U}, D^t) \;\; e_u]$.

The nonlinear hyperplanes $K(x^t, D^t)w_1 + b_1 = 0$ and $K(x^t, D^t)w_2 + b_2 = 0$ are obtained by using parameters w and b from the following equations (10) and (11),

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(S^t S)^{-1}(T^t\alpha_1 - O^t\mu_1), \tag{10}$$

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (T^t T)^{-1}(S^t\alpha_2 - O^t\mu_2). \tag{11}$$

To avoid the ill-conditioning in the calculation of inverse $(S^t S)^{-1}$ and $(T^t T)^{-1}$, we add a regularization term δI to the matrices in (10) and (11) as $(S^t S + \delta I)^{-1}$ and $(T^t T + \delta I)^{-1}$ to make them positive definite where δ is a small positive value.

For a data point $x \in R^n$, it is assigned to a class $'i'$ on the basis of the following decision function

$$class\ i = \min |K(x^t, D^t)w_i + b_i| \text{ for } i = 1,2. \tag{12}$$

Among the various algorithms used for the classification of EEG signals, SVM is the most widely used technique. This is due to its better generalization performance for various kinds of data as compared to algorithms like ANN which suffers from the problem of locally minimal solution. Many researchers have used SVM as a classification technique for the classification of EEG data. However, there is no prior information about the distribution of data in the formulation of SVM. As a result of this, the classifier is trained in the same manner for all types of datasets without having the knowledge about the data distribution. On the other hand, universum learning provides this prior information to the SVM classifier so that the hyperplane separating the classes aligns itself according to the distribution of data. Moreover, the twin SVM based approach has the advantage of being insensitive to class imbalanced data (Jayadeva et al., 2017) in comparison to traditional SVM and LSSVM. So, UTSVM gives the advantage of prior information with less training time as compared to SVM with less sensitivity to class imbalanced data. Also, our proposed approach of selecting the universum from the interictal EEG signals gives proper prior information without getting affected by outliers as compared to other techniques like random averaging of data points.

LSSVM is also used by many researchers for the classification of EEG signals since it solves a system of linear equations and therefore is more efficient in terms of computational time as compared to SVM. Our universum based approach takes more training time than the traditional methods due to the addition of more data points i.e, universum in the constraints of the optimization problem.

3. Proposed approach

In many of the classification approaches for EEG signals, the prior information about the distribution of EEG data is not used. Due to this, the classification techniques are not able to give better generalization performance even if the most efficient feature extraction technique is used. The universum based approach actually gives some prior information in the construction of the classifier. So we used a universum based approach with support vector machine to classify the EEG signals. Further, in the datasets generated from the EEG signals, many data points behave as outliers, especially in case of seizure signal as shown in fig. 4 and 5. Consequently, the traditional approach of universum based support vector machine based on random averaging is not so efficient in giving the prior information. The outlier data points affect the generation of the universum points in the random averaging approach which leads to incorrect classification.

Our approach of universum support vector machine (USVM) takes the universum points from the EEG dataset itself. We take the interictal or seizure free signals from the EEG dataset (Andrzejak et al., 2001) as the universum. Since the variation of the signal in the seizure free state comes in between the variation of healthy and seizure EEG signals, this gives the required prior information to the support vector machine classifier in a more efficient manner. Moreover, there are no outliers in the universum data since our universum data is not generated from the training data and thus there is no effect of noise from the training data. A comparison of our proposed approach with the traditional random averaging scheme is illustrated in fig. 4 and 5 where the universum data points of our proposed approach lie in between the two classes.

Further, we use our approach with universum twin support vector machine (UTSVM) which is a more efficient technique in terms of computational complexity. A brief illustration of our methodology is given in fig. 1.

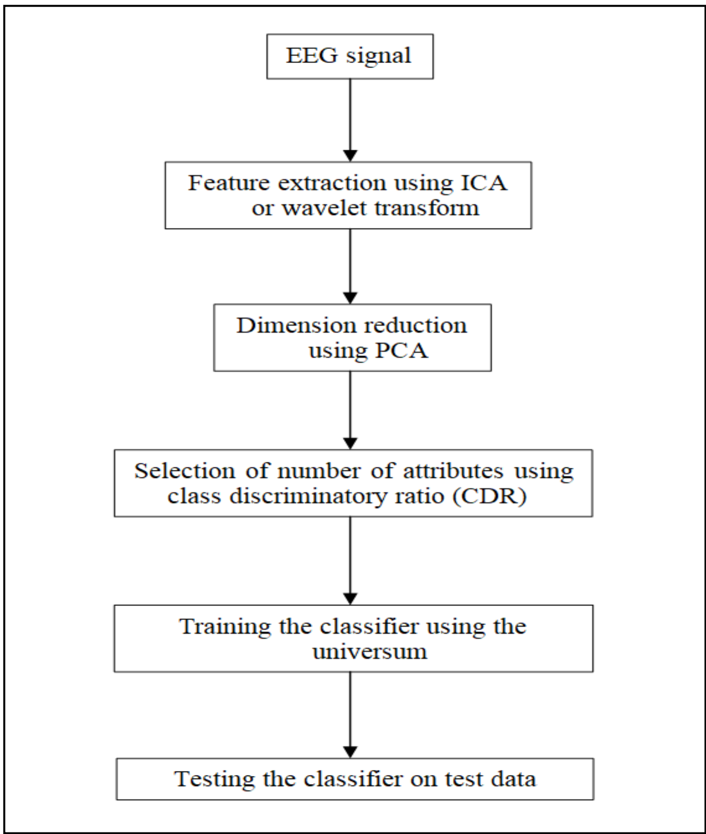


Figure 1: Proposed approach

The steps involved in our proposed approach for classification of EEG signals are as follows:

(i) Choose a feature extraction technique and extract the features from the training data consisting of healthy and seizure data points.

(ii) Extract the features from the universum points which are taken from seizure free dataset.

(iii) Reduce the dimension of the feature vector using PCA (Wagner, 2012) and class discriminatory ratio (CDR) (Bartlett et al., 2002).

(iv) Train the model using training data with the universum.

(v) Test the model by using steps (ii) and (iii) and the classifier.

In this work, different feature extraction techniques are used to extract the appropriate features from the datasets such as principal component analysis (PCA), independent component analysis (ICA) and wavelet transform with different families of wavelet such as db1, db2, db4, db6 and Haar wavelet.

4. Numerical Experiments

In this section, numerical experiments are performed for the classification of EEG signals of healthy state and seizure. The EEG dataset is taken from (Andrzejak et al., 2001) which is available online. The dataset consists of five sets viz. Z, O, N, F and S. Each set contains 100 single-channel EEG signals sampled at a sampling rate of 173.61 Hz and of 23.6 seconds duration. The sets Z and O are surface EEG recordings of five healthy volunteers with eyes open and closed respectively. The sets N and F are recordings of five patients in the interictal state and the region of recording is the hippocampal formation of the opposite hemisphere of the brain in N and the epileptogenic zone in F. The set S is for the ictal state consisting of seizure recordings from all the recording sites exhibiting ictal activity. The mode of EEG recording is intra-cranial for N, F and S. For all the EEG signals, same 128-channel amplifier system is used with an average common reference.

In the numerical experiments, the training and testing set consists of 50 samples each, chosen from the sets Z, O and S each containing 100 samples. In our proposed approach, the universum is chosen from the set N which contains the interictal EEG signals. For the cross-validation, we use interleaving of samples in the training data from the two classes for all the datasets. For feature extraction, various techniques are applied including principal component analysis (PCA), independent component analysis (ICA) and wavelet transform. In case of wavelet transform, several families of wavelets are applied with different levels of decomposition as used in the available literature. Discrete wavelet transform (DWT) is implemented using different families of wavelet on specific levels of decomposition. The set of the approximation and decomposition coefficients is taken as the feature vector. The level of decomposition is set at level-3 for Daubechies wavelet- db2, db4 and Haar wavelet. For db1 and db6 wavelets level-2 decomposition is used. In case of ICA and wavelet transform, PCA is applied for the dimension reduction. The implementation of ICA is same as in (Bartlett et al., 2002) (ICA Architecture1). The class discriminatory ratio (CDR) is used to sort the PCA components and to choose the most relevant PCA components. To check the effectiveness of our proposed method, the results of our proposed method for universum are compared with the SVM, LSSVM and USVM with random averaging scheme. In case of UTSVM, we made a comparison with TWSVM and UTSVM with random averaging.

All computations were carried-out on a PC running on Windows 10 OS with 64 bit, 3.60 GHz Intel® core™ i7-7700 processor having 16 GB of RAM under MATLAB R2008b environment. MOSEK optimization toolbox (http://www.mosek.com) is used to solve the formulations of SVM, USVM, TWSVM and UTSVM. For nonlinear case, we used Gaussian kernel $k(a,b) = \exp\left(-\frac{1}{2\sigma^2}\|a-b\|^2\right)$ where vector $a,b \in R^m$ and σ is the kernel parameter.

The value of the parameters $C = C_1 = C_2 = C_u$ is taken from the set $\{10^{-5},...,10^{5}\}$ for all the cases. For USVM, proposed USVM, UTSVM and proposed UTSVM, the number of universum samples i.e. u is taken from the set $\{10, 20, 30, 40\}$ and ε is chosen by varying values from the set $\{0.1, 0.2, 0.3, 0.5, 0.6, 0.7\}$. For the selection of the optimal parameters, 5-fold cross-validation is used. In our proposed approaches, universum is selected from the set N of the EEG database and for the existing universum methods random averaging is used for generating the universum data. The value of σ is calculated as per the following formula (Tsang et al., 2006) in all the methods,

$$\sigma = \frac{1}{N^2} \sum_{i,j=1}^{N} \| x_i - x_j \|^2$$

where x_i represents each data point and N is the total number of data points.

For all the datasets, the number of attributes are decided on the basis of two factors, (a) variance accounted for (Wagner, 2012) and (b) class discriminatory ratio (CDR). The approach of calculating CDR of components is taken from (Bartlett et al., 2002) as

$$r = \frac{\sigma_{between}}{\sigma_{within}},$$

where $\sigma_{between} = \sum_{i}^{c} \left(\bar{x}_{\mathrm{i}} - \bar{x} \right)^2$ is the variance of the $'c'$ class means and $\sigma_{between} = \sum_{i}^{c} \sum_{j}^{c} \left(x_{\mathrm{ij}} - \bar{x}_{\mathrm{j}} \right)^2$ is the sum of the within class variance of all the $'c'$ classes.

The plots for variance and CDR are shown in fig. 6 and fig. 7 for Z & S dataset using PCA and ICA respectively. In fig. 8, the generalization performance of the proposed approach for UTSVM is compared with the random averaging approach for Z&S using PCA, O&S using PCA, O&S using ICA and O&S using wavelet feature extraction technique.

The results for all the proposed and baseline methods are shown in terms of prediction accuracy and training time in Table 1 & Table 3. One can observe from Table 1 that our proposed approach outperforms USVM with random averaging, LSSVM and SVM in terms of accuracy. It can be observed in Table 2 that LSSVM performs better than SVM and USVM.

From Table 3, it is evident that our proposed approach is showing better generalization performance for almost all the datasets as compared to TWSVM and UTSVM. In terms of training time, our proposed approach is comparable with respect to the existing universum based methods. It is also noticeable from Table 1 and 3 that the universum based approaches take more computation time as compared to traditional algorithms such as SVM, LSSVM and TWSVM. This additional time is due to the incorporation of universum data points which can be traded for the generalization performance. LSSVM takes very less computation time since it solves a system of linear equations. It is noticeable in Table 1 and 3 that the existing universum based approaches viz. USVM and UTSVM which use random averaging for universum have not performed better than the other algorithms. This is because the seizure data contains noisy data points and thus the generated universum data do not reflect the distribution of data. On the other hand, our proposed approach of selecting the universum from the interictal EEG signals gives better accuracy in most of the datasets since there is no effect of noise which justifies its applicability for classification of seizure and healthy EEG signals.

One can notice from Table 1 that our proposed approach has not performed better for all the datasets. So, we analyze the comparative performance of our proposed approach with the existing approaches. The average ranks of SVM, LSSVM, USVM and proposed USVM on the basis of accuracy is shown in Table 2. One can notice from Table 2 that the average rank of our proposed USVM is lowest among all the methods. We perform the Friedman test with the corresponding post-hoc test (Demsar, 2006) for the statistical comparison of the performance of the 4 algorithms using 14 datasets. We assume all the methods are equivalent under null hypothesis. The Friedman statistic is computed as

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right],$$

where k is the number of methods and N is the number of datasets.

$$\chi_F^2 = \frac{12\times 14}{4\times(4+1)}\left[(3^2+2.3214^2+3.5^2+1.1786^2) - \frac{4\times(4+1)^2}{4}\right] \cong 25.4352,$$

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)\chi_F^2} = \frac{(14-1)\times 25.4352}{14\times(4-1)-25.4352} \cong 19.9615,$$

where F_F is the distribution according to the F-distribution with $(3, 3\times 13) = (3, 39)$ degrees of freedom with 4 methods and 14 datasets. The critical value of $F(3,39)$ is 2.8451 for the level of significance at $\alpha = 0.05$. Since the value of $F_F = 16.4633 > 2.8451$ so we reject the null hypothesis. For pair-wise comparison of methods, we perform the Nemenyi post-hoc test. The significant difference between the methods is checked by computing the critical difference (CD) at $p = 0.10$ which should differ by at least $2.291\sqrt{\frac{4\times(4+1)}{6\times 14}} \approx 1.1179$.

The differences between the average ranks of SVM, LSSVM and USVM with proposed USVM are $(3-1.1786=1.8214)$, $(2.3214-1.1786=1.1428)$ and $(3.5-1.1786=2.3214)$ respectively. Since, for all the methods, the difference of ranks is greater than 1.1179 so we conclude that our proposed USVM is significantly better than SVM, LSSVM and USVM.

The accuracy values are shown with the training time for the proposed UTSVM with TWSVM and UTSVM in Table 3. One can observe that our proposed UTSVM has shown better generalization performance in most of the cases. Table 4 shows the average ranks of TWSVM, UTSVM and proposed UTSVM based on accuracy values. Our proposed UTSVM has the lowest rank among all the methods. We further performed the Friedman statistics with the corresponding post-hoc test to find the significant difference between TWSVM, UTSVM and proposed UTSVM. The Friedman statistic is computed using Table 4 under null hypothesis as:

$$\chi_F^2 = \frac{12\times 14}{3\times(3+1)}\left[(2.3571^2 + 2.4286^2 + 1.2143^2) - \frac{3\times(3+1)^2}{4}\right] \cong 12.9996$$

$$F_F = \frac{(14-1)\times 12.9996}{14\times(3-1)-12.9996} \cong 11.266$$

Since the value of F_F is more than the critical value of $F(2,26)$ i.e. 3.3690 for the level of significance at $\alpha = 0.05$. Thus we reject the null hypothesis. Further, the pair-wise comparisons are performed by using the Nemenyi post-hoc test. The difference between the methods should be more than the critical difference (CD) at $p = 0.10$, calculated using the critical value as $2.052\sqrt{\frac{3\times(3+1)}{6\times 14}} \approx 0.7756.$

The difference between the average ranks of our proposed UTSVM with TWSVM and UTSVM are $(2.3571-1.2143=1.1428)$ and $(2.4286-1.2143=1.2143)$ which are greater than 0.7756. Hence, the performance of our proposed UTSVM is significantly better than TWSVM and UTSVM. It is noticeable from Table 2 and 4 that our proposed UTSVM is showing highest generalization performance as compared to the existing methods. The highest accuracy for Z & S is obtained as 99 % in the case of ICA feature extraction with our proposed UTSVM. For O & S, the highest accuracy is found with ICA feature extraction technique using our proposed UTSVM.

The accuracy value for different selections of number of universum points is shown in fig. 8 for (a) Z&S using PCA, (b) O&S using PCA, (c) O&S using ICA and (d) O&S using wavelet (db4) feature extraction technique. It can be seen that in all the cases our proposed approach is giving higher accuracy in comparison to the traditional approaches. Also the effect of outliers is clearly visible in fig. 8 (c) and (d) for the random averaging approach where the accuracy decreases for some sets of the universum. This justifies our selection of the universum.

Fig. 9 illustrates the accuracy comparison of different algorithms for the classification of seizure and non-seizure data using different feature extraction techniques. In fig. 10, the insensitivity performance of our proposed approach of USVM is shown for the parameters C and ε. It can be observed that the proposed USVM gives high accuracy for higher values of C and ε. The insensitivity performance of our proposed approach with UTSVM is shown in fig. 11. It is evident from fig. 11 that our proposed UTSVM gives better generalization performance for lesser values of C and ε .

5. Conclusions

On the basis of the experimental results, it can be stated that our universum based approach gives better generalization performance for the classification of EEG signals as compared to the existing approaches. Our method of selection of universum points has proved to be a promising approach for the classification of healthy and seizure EEG signals. Also, the effect of outliers on the universum is reduced by using the universum from the EEG dataset itself i.e., the seizure free EEG signal. The distribution of interictal (seizure free) signals provides prior information about the distribution of healthy and seizure signals and also lies in between the two classes. Based on the experimental results, it is evident that universum twin support vector machine (UTSVM) is better in comparison to other support vector machine algorithms for EEG signal classification. Among the different feature extraction techniques, ICA shows the best results using our proposed approach with 99 % accuracy.

The proposed work also gives a comparison of the different SVM based algorithms for the classification of EEG signals. It is evident from the experimental results that other variants of SVM such as TWSVM and UTSVM give good generalization and computational performance and are applicable for the classification of EEG signals. The universum based SVM approach needs to be applied to other diseases which are diagnosed using EEG signals with the proper selection of universum. In future, our universum based approach of EEG classification can be improved in terms of computational time. Our proposed universum based approach can be extended to multiclass classification of EEG signals using EEG datasets generated with different feature extraction techniques.

Acknowledgements

This work was supported by Science and Engineering Research Board (SERB) as Early Career Research Award grant no. ECR/2017/000053 and Department of Science and Technology as Ramanujan fellowship grant no. SB/S2/RJN-001/2016. We gratefully acknowledge the Indian Institute of Technology Indore for providing facilities and support. We are thankful to the Ministry of Human Resource Development (MHRD), Govt. of India for providing Teaching Assistantship (TA) fellowship to Mr. Bharat Richhariya.

References

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*, *123*(1), 69-87.

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, *64*(6), 061907.

Bai, X., & Cherkassky, V. (2008, June). Gender classification of human faces using inference through contradictions. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on* (pp. 746-750). IEEE.

Bajaj, V., & Pachori, R. B. (2012). Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine*, *16*(6), 1135-1142.

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, *13*(6), 1450-1464.

Chapelle, O., Agarwal, A., Sinz, F. H., & Schölkopf, B. (2008). An analysis of inference with the universum. In *Advances in Neural Information Processing Systems* (pp. 1369-1376).

Chen, S., & Zhang, C. (2009, July). Selecting Informative Universum Sample for Semi-Supervised Learning. In *IJCAI* (Vol. 6, pp. 1016-1021).

Cherkassky, V., & Dai, W. (2009, September). Empirical study of the Universum SVM learning for high-dimensional data. In *International Conference on Artificial Neural Networks* (pp. 932-941). Springer, Berlin, Heidelberg.

Cherkassky, V., Dhar, S., & Dai, W. (2011). Practical conditions for effectiveness of the universum learning. *IEEE Transactions on Neural Networks*, *22*(8), 1241-1255.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20 (3), 273-297.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(Jan), 1-30.

Gao, T., Tian, Y., Shao, X., & Deng, N. (2008). Accurate prediction of translation initiation sites by Universum SVM. In *The 2nd International Symposium on Optimization and Systems biology* (pp. 279-286).

Guo, L., Wu, Y., Zhao, L., Cao, T., Yan, W., & Shen, X. (2011). Classification of mental task from EEG signals using immune feature weighted support vector machines. *IEEE Transactions on Magnetics*, *47*(5), 866-869.

Guler, I., & Ubeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of Neuroscience Methods*, *148*(2), 113-121.

Guler, I., & Ubeyli, E. D. (2007). Multiclass support vector machines for EEG-signals classification. *IEEE Transactions on Information Technology in Biomedicine*, *11*(2), 117-126.

Hao, X., & Zhang, D. (2013, September). Ensemble universum SVM learning for multimodal classification of Alzheimer's disease. In *International Workshop on Machine Learning in Medical Imaging* (pp. 227-234). Springer, Cham.

Jayadeva, Khemchandani, R., & Chandra, S. (2007). Twin support vector machines for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(5), 905-910.

Jayadeva, Khemchandani, R., & Chandra, S. (2017). Applications Based on TWSVM. In *Twin Support Vector Machines* (pp. 193-206). Springer, Cham.

Joshi, V., Pachori, R. B., & Vijesh, A. (2014). Classification of ictal and seizure-free EEG signals using fractional linear prediction. *Biomedical Signal Processing and Control*, *9*, 1-5.

Khemchandani, R., Goyal, K., & Chandra, S. (2016). TWSVR: regression via twin support vector machine. *Neural Networks*, 74, 14-21.

Kumar, M.A., & Gopal, M. (2009). Least squares twin support vector machines for pattern classification. *Expert Systems with Applications*, 36(4), 7535-7543.

Li, Y., & Wen, P. (2009, July). Classification of EEG signals using sampling techniques and least square support vector machines. In *International Conference on Rough Sets and Knowledge Technology* (pp. 375-382). Springer, Berlin, Heidelberg.

Li, Y., & Wen, P. P. (2011). Clustering technique-based least square support vector machine for EEG signal classification. *Computer Methods and Programs in Biomedicine*, *104*(3), 358-372.

Liu, C. L., Hsaio, W. H., Lee, C. H., Chang, T. H., & Kuo, T. H. (2016). Semi-supervised text classification with universum learning. *IEEE Transactions on Cybernetics*, *46*(2), 462-473.

Liu, Y., Zhou, W., Yuan, Q., & Chen, S. (2012). Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*(6), 749-755.

Long, W., Tang, Y. R., & Tian, Y. J. (2016). Investor sentiment identification based on the universum SVM. *Neural Computing and Applications*, 1-10.

Ma, Y., Ding, X., She, Q., Luo, Z., Potter, T., & Zhang, Y. (2016). Classification of motor imagery EEG signals with support vector machines and particle swarm optimization. *Computational and Mathematical Methods in Medicine*, *2016*.

Mangasarian, O.L., & Wild, E.W. (2006). Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE transactions on pattern analysis and machine intelligence*, 28(1), 69-74.

Nicolaou, N., & Georgiou, J. (2012). Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Systems with Applications*, *39*(1), 202-209.

Ocak, H. (2008). Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal Processing*, *88*(7), 1858-1867.

Qi, Z., Tian, Y., & Shi, Y. (2013). Robust twin support vector machine for pattern classification. *Pattern Recognition*, 46(1), 305-316.

Qi, Z., Tian, Y., & Shi, Y. (2012). Twin support vector machine with universum data. *Neural Networks*, *36*, 112-119.

Rosso, O. A., Hyslop, W., Gerlach, R., Smith, R. L. L., Rostas, J. A. P., & Hunter, M. (2005). Quantitative EEG analysis of the maturational changes associated with childhood absence epilepsy. Physica A: Statistical Mechanics and its Applications, 356(1), 184-189.

Shao, Y.H., Zhang, C.H., Wang, X.B., & Deng, N.Y. (2011). Improvements on twin support vector machines. *IEEE transactions on neural networks*, 22(6), 962-968.

Sharma, R., & Pachori, R. B. (2015). Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Systems with Applications*, *42*(3), 1106-1117.

Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, *37*(12), 8659-8666.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, *9*(3), 293-300.

Tanveer, M. (2015a). Robust and sparse linear programming twin support vector machines. *Cognitive Computation*, 7(1), 137-149.

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

Tanveer, M., Khan, M.A., & Ho, S.S. (2016). Robust energy-based least squares twin support vector machines. *Applied Intelligence*, 45(1), 174-186.

Tsang, I. W., Kocsor, A., & Kwok, J. T. (2006, August). Efficient kernel feature extraction for massive data sets. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data mining* (pp. 724-729). ACM.

Wagner, P. (2012). Face Recognition with GNU Octave/MATLAB.

Wang, Z., Shao, Y.H., Bai, L., & Deng, N.Y. (2015). Twin support vector machine for clustering. *IEEE transactions on neural networks and learning systems*, 26(10), 2583-2588.

Weston, J., Collobert, R., Sinz, F., Bottou, L., & Vapnik, V. (2006, June). Inference with the universum. In *Proceedings of the 23rd International Conference on Machine learning* (pp. 1009-1016). ACM.

Xu, Y., Chen, M., & Li, G. (2016). Least squares twin support vector machine with Universum data for classification. *International Journal of Systems Science*, *47*(15), 3637-3645.

Xu, Y., Yang, Z., & Pan, X. (2017). A novel twin support-vector machine with pinball loss. *IEEE transactions on neural networks and learning systems*, 28(2), 359-370.

Zavar, M., Rahati, S., Akbarzadeh-T, M. R., & Ghasemifard, H. (2011). Evolutionary model selection in a wavelet-based support vector machine for automated seizure detection. *Expert Systems with Applications*, *38*(9), 10751-10758.

ACCEPTED MANUSCRIPT

Dataset (Train size, Test size)	Feature Extraction Method	SVM Accuracy (%) (C, σ) Time (s)	LSSVM Accuracy (%) (C, σ) Time (s)	USVM Accuracy (%) $(C, \sigma, \varepsilon, u)$ Time (s)	Proposed USVM Accuracy (%) $(C, \sigma, \varepsilon, u)$ Time (s)
Z & S (100× 50, 100× 50)	PCA	69 (10^2, 21180.7) 0.09954	73 (10^5, 21180.7) 0.03017	69 (10^2, 21180.7, 0.2, 15) 0.17169	**77** (10^3, 21180.7, 0.3, 15) 0.16816
Z & S (100× 15, 100× 15)	ICA	80 (10^3, 69.919) 0.09963	79 (10^4, 69.919) 0.02733	**81** (10^3, 69.919, 0.2, 10) 0.14142	79 (10^2, 69.919, 0.1, 10) 0.13958
Z & S (100× 50, 100× 50)	Wavelet (db4)	69 (10^2, 21415.9) 0.09859	73 (10^5, 21415.9) 0.02984	69 (10^2, 21415.9, 0.1, 10) 0.14816	**76** (10^4, 21415.9, 0.3, 10) 0.14314
Z & S (100× 50, 100× 50)	Wavelet (Haar)	69 (10^2, 21196.3) 0.10009	72 (10^4, 21196.3) 0.0295	69 (10^3, 21196.3, 0.2, 30) 0.25236	**84** (10^4, 21196.3, 0.7, 30) 0.25148
Z & S (100× 50, 100× 50)	Wavelet (db2)	69 (10^2, 21315.9) 0.10182	73 (10^5, 21315.9) 0.02925	69 (10^2, 21315.9, 0.1, 10) 0.14347	**76** (10^4, 21315.9, 0.5, 10) 0.14183
Z & S (100× 50, 100× 50)	Wavelet (db6)	69 (10^2, 21503.3) 0.09742	71 (10^2, 21503.3) 0.03005	69 (10^2, 21503.3, 0.1, 10) 0.14553	**77** (10^4, 21503.3, 0.5, 10) 0.14177
Z & S (100× 50, 100× 50)	Wavelet (db1)	69 (10^2, 20956.4) 0.09981	74 (10^5, 20956.4) 0.031	69 (10^2, 20956.4, 0.1, 10) 0.14285	**78** (10^4, 20956.4, 0.1, 10) 0.14328
O & S (100× 50, 100× 50)	PCA	72 (10^1, 20400) 0.10182	69 (10^2, 20400) 0.02975	67 (10^3, 20400, 0.1, 40) 0.32457	**75** (10^1, 20400, 0.3, 40) 0.30981
O & S (100× 50, 100× 50)	ICA	72 (10^2, 105.268) 0.10174	74 (10^5, 105.268) 0.03955	72 (10^3, 105.268, 0.3, 20) 0.2129	**76** (10^2, 105.268, 0.6, 20) 0.18707

O & S (100× 30, 100× 30)	Wavelet (db4)	71 (10^1, 20139.2) 0.09831	71 (10^1, 20139.2) 0.0284	70 (10^2, 20139.2, 0.1, 40) 0.32568	**75** (10^1, 20139.2, 0.3, 40) 0.31293
O & S (100× 50, 100× 50)	Wavelet (Haar)	70 (10^2, 19800.4) 0.10273	70 (10^2, 19800.4) 0.04194	69 (10^3, 19800.4, 0.2, 40) 0.32389	**75** (10^1, 19800.4, 0.3, 40) 0.31326
O & S (100× 50, 100× 50)	Wavelet (db2)	68 (10^2, 20074.4) 0.09935	69 (10^2, 20074.4) 0.03094	67 (10^2, 20074.4, 0.1, 40) 0.31621	**75** (10^1, 20074.4, 0.3, 40) 0.31399
O & S (100× 50, 100× 50)	Wavelet (db6)	69 (10^1, 19984.8) 0.09922	70 (10^2, 19984.8) 0.02976	69 (10^2, 19984.8, 0.1, 40) 0.31894	**77** (10^0, 19984.8, 0.1, 40) 0.31528
O & S (100× 50, 100× 50)	Wavelet (db1)	71 (10^1, 20412.5) 0.10013	69 (10^2, 20412.5) 0.03019	68 (10^2, 20412.5, 0.3, 40) 0.32286	**76** (10^0, 20412.5, 0.1, 40) 0.31789

Table 1: Performance comparison of proposed USVM with SVM, LSSVM and USVM for classification of seizure and healthy EEG signals using Gaussian kernel.

Dataset	Feature Extraction Method	SVM	LSSVM	USVM	Proposed USVM
Z & S	PCA	3.5	2	3.5	1
Z & S	ICA	2	3.5	1	3.5
Z & S	Wavelet (db4)	3.5	2	3.5	1
Z & S	Wavelet (Haar)	3.5	2	3.5	1
Z & S	Wavelet (db2)	3.5	2	3.5	1
Z & S	Wavelet (db6)	3.5	2	3.5	1
Z & S	Wavelet (db1)	3.5	2	3.5	1
O & S	PCA	2	3	4	1
O & S	ICA	3.5	2	3.5	1
O & S	Wavelet (db4)	2.5	2.5	4	1
O & S	Wavelet (Haar)	2.5	2.5	4	1
O & S	Wavelet (db2)	3	2	4	1
O & S	Wavelet (db6)	3.5	2	3.5	1
O & S	Wavelet (db1)	2	3	4	1
Average Rank		3	2.3214	3.5	1.1786

Table 2: Average ranks of SVM, LSSVM, USVM and proposed USVM on classification accuracy for seizure and healthy EEG signals using Gaussian kernel.

Dataset (Train size, Test size)	Feature Extraction Method	TWSVM Accuracy (%) (C, σ) Time (s)	UTSVM Accuracy (%) $(C, \sigma, \varepsilon, u)$ Time (s)	Proposed UTSVM Accuracy (%) $(C, \sigma, \varepsilon, u)$ Time (s)
Z & S (100×50, 100×50)	PCA	82 (10^-5, 21180.7) 0.0191	89 (10^0, 21180.7, 0.7, 30) 0.02529	**90** (10^1, 21180.7, 0.1, 30) 0.02599
Z & S (100×15, 100×15)	ICA	94 (10^0, 69.919) 0.01607	95 (10^-2, 69.919, 0.6, 10) 0.01804	**99** (10^-5, 69.919, 0.1, 10) 0.01756
Z & S (100×50, 100×50)	Wavelet (db4)	82 (10^-5, 21415.9) 0.01805	78 (10^1, 21415.9, 0.3, 30) 0.02533	**91** (10^1, 21415.9, 0.1, 30) 0.02509
Z & S (100×50, 100×50)	Wavelet (Haar)	79 (10^-5, 21196.3) 0.01819	80 (10^0, 21196.3, 0.6, 20) 0.02217	**88** (10^1, 21196.3, 0.1, 20) 0.02255
Z & S (100×50, 100×50)	Wavelet (db2)	82 (10^-5, 21315.9) 0.01854	89 (10^0, 21315.9, 0.7, 30) 0.02468	**90** (10^1, 21315.9, 0.1, 30) 0.0251
Z & S (100×50, 100×50)	Wavelet (db6)	80 (10^-5, 21503.3) 0.01813	81 (10^0, 21503.3, 0.7, 20) 0.022	**87** (10^1, 21503.3, 0.1, 20) 0.02306
Z & S (100×50, 100×50)	Wavelet (db1)	80 (10^-5, 20956.4) 0.01832	**89** (10^0, 20956.4, 0.7, 30) 0.02431	88 (10^1, 20956.4, 0.1, 30) 0.0256
O & S (100×50, 100×50)	PCA	79 (10^-4, 20400) 0.01826	80 (10^-4, 20400, 0.6, 40) 0.02571	**84** (10^-2, 20400, 0.6, 40) 0.02601
O & S (100×50, 100×50)	ICA	94 (10^0, 105.268) 0.01699	90 (10^-1, 105.268, 0.6, 10) 0.01823	**95** (10^-1, 105.268, 0.1, 10) 0.01942
O & S (100×30, 100×30)	Wavelet (db4)	**84** (10^-3, 20139.2) 0.01822	78 (10^0, 20139.2, 0.2, 20) 0.02271	**84** (10^-3, 20139.2, 0.1, 20) 0.02236
O & S (100×50, 100×50)	Wavelet (Haar)	**82** (10^-3, 19800.4) 0.01874	79 (10^0, 19800.4, 0.3, 10) 0.02041	**82** (10^-3, 19800.4, 0.1, 10) 0.01991
O & S (100×50, 100×50)	Wavelet (db2)	**83** (10^-3, 20074.4) 0.01829	78 (10^-1, 20074.4, 0.5, 10) 0.02005	**83** (10^-3, 20074.4, 0.1, 10) 0.02022
O & S (100×50, 100×50)	Wavelet (db6)	80 (10^-4, 19984.8) 0.02598	77 (10^0, 19984.8, 0.3, 40) 0.02778	**85** (10^-2, 19984.8, 0.7, 40) 0.02615
O & S (100×50, 100×50)	Wavelet (db1)	**84** (10^-3, 20412.5) 0.01862	79 (10^-1, 20412.5, 0.5, 10) 0.02023	**84** (10^-3, 20412.5, 0.1, 10) 0.02013

Table 3: Performance comparison of proposed UTSVM with TWSVM and UTSVM for classification of seizure and healthy EEG signals using Gaussian kernel.

Dataset	Feature Extraction Method	TWSVM	UTSVM	Proposed UTSVM
Z & S	PCA	3	2	1
Z & S	ICA	3	2	1
Z & S	Wavelet (db4)	2	3	1
Z & S	Wavelet (Haar)	3	2	1
Z & S	Wavelet (db2)	3	2	1
Z & S	Wavelet (db6)	3	2	1
Z & S	Wavelet (db1)	3	1	2
O & S	PCA	3	2	1
O & S	ICA	2	3	1
O & S	Wavelet (db4)	1.5	3	1.5
O & S	Wavelet (Haar)	1.5	3	1.5
O & S	Wavelet (db2)	1.5	3	1.5
O & S	Wavelet (db6)	2	3	1
O & S	Wavelet (db1)	1.5	3	1.5
Average Rank		2.3571	2.4286	1.2143

Table 4: Average ranks of TWSVM, UTSVM and proposed UTSVM on classification accuracy for seizure and healthy EEG signals using Gaussian kernel.

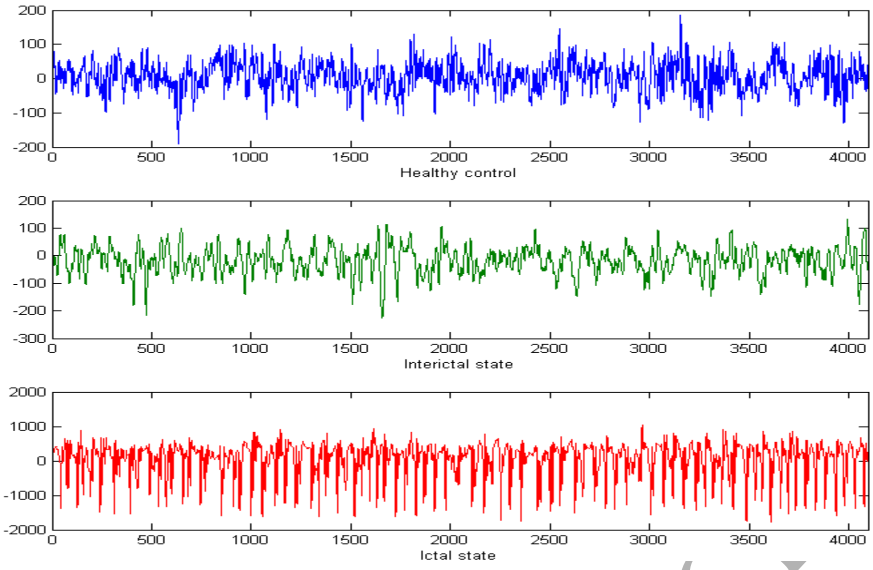


Figure 2: EEG signals of healthy control, interictal (seizure free) and ictal (seizure) state.

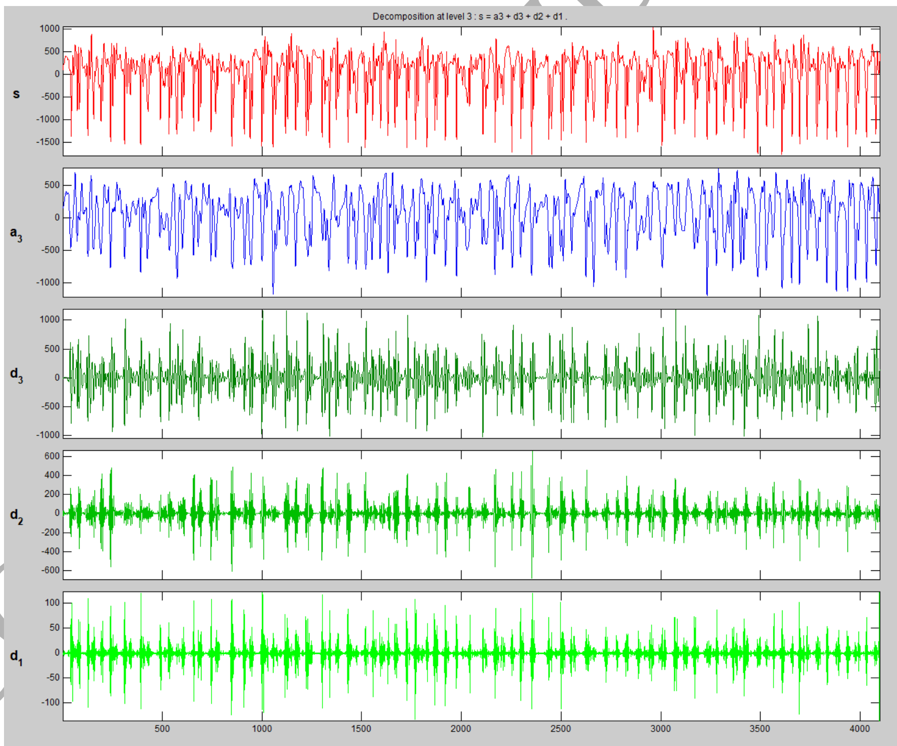
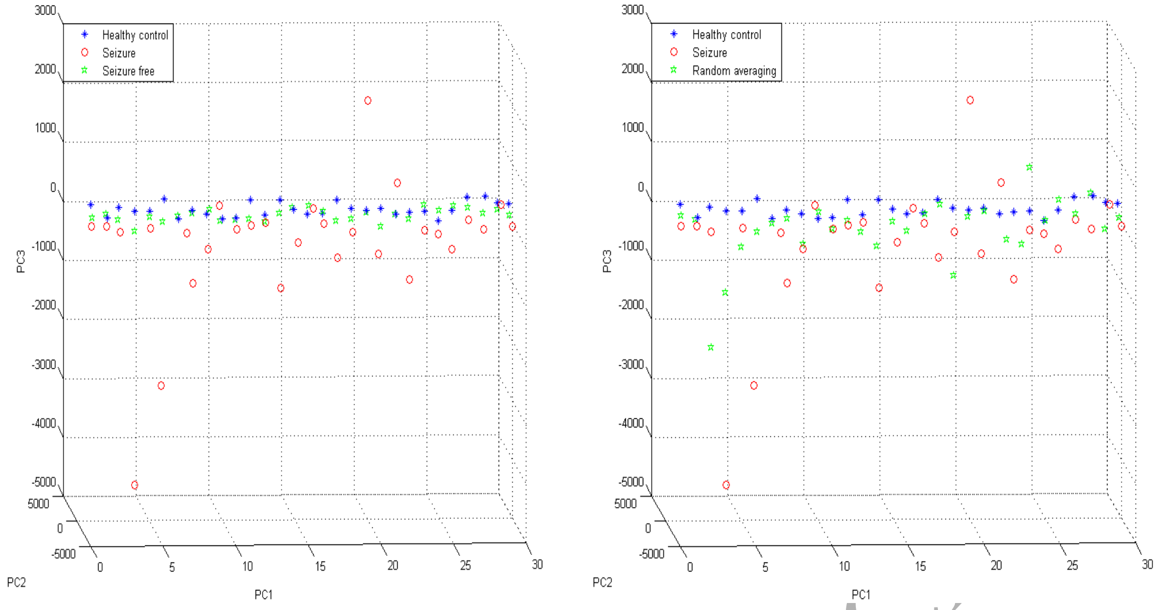
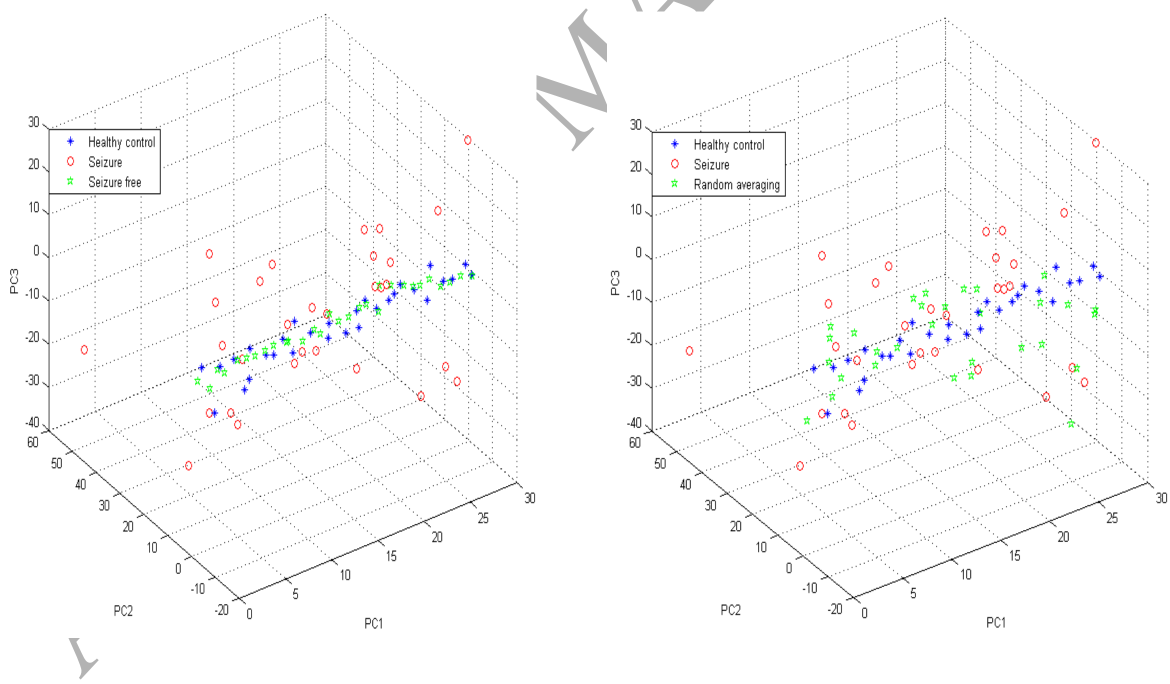


Figure 3: Discrete wavelet decomposition of EEG signal at 3rd level of decomposition using Daubechies-4 wavelet.



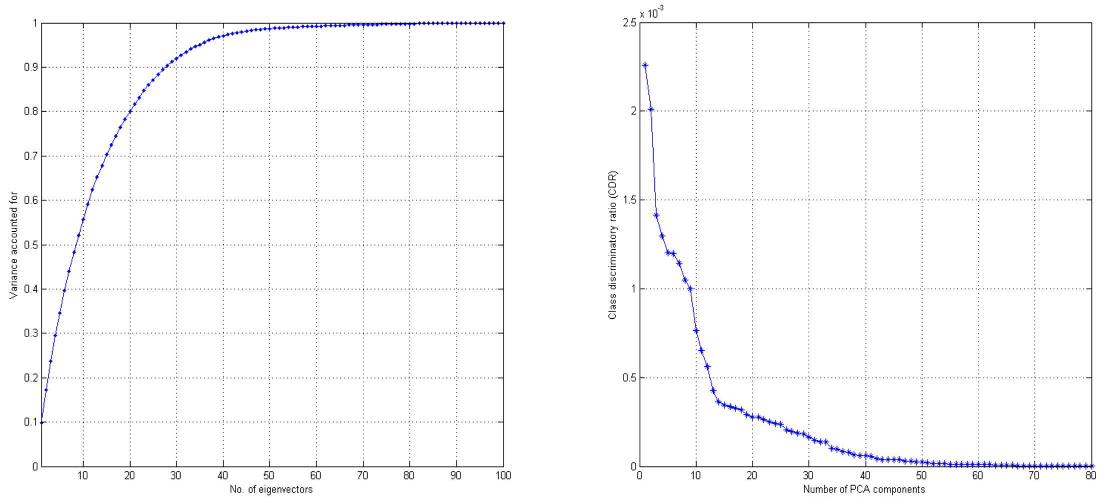
(a) (b)

Figure 4: Distribution of data points of set Z set as healthy control, S as seizure using PCA up to 3 principal components in (a) for proposed method i.e., using N (seizure free) data points as universum and in (b) random averaging is used for generating the universum.



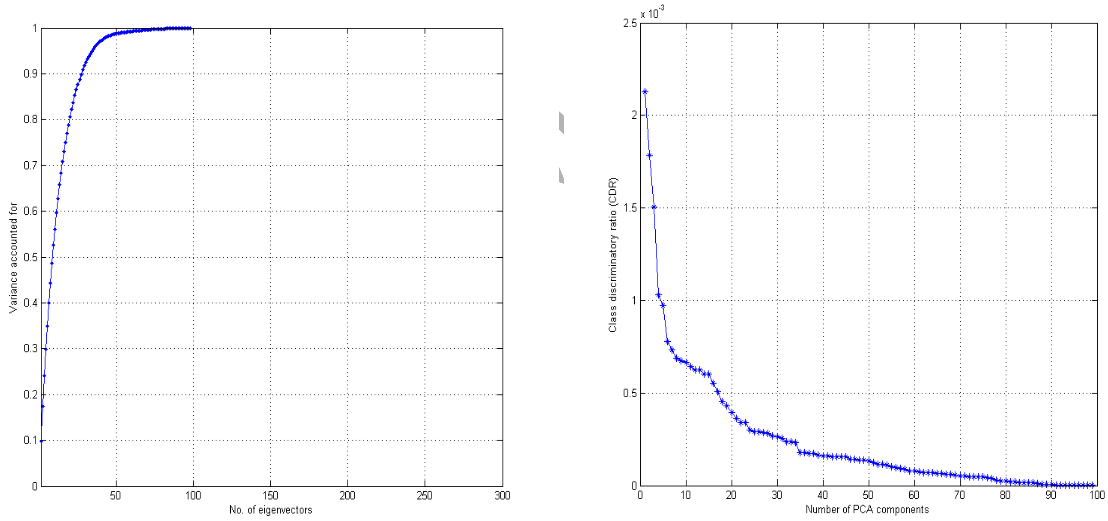
(a) (b)

Figure 5: Distribution of data points of set Z set as healthy control, S as seizure using ICA up to 3 principal components (PCs) in (a) the proposed approach using seizure free data points as universum and (b) universum data points generated using random averaging.



(a) (b)

Figure 6: (a) Variance of data points (b) class discriminatory ratio vs. number of PCA components for Z & S dataset using PCA feature extraction technique.



(a) (b)

Figure 7: (a) Variance of data points (b) class discriminatory ratio vs number of PCA components for Z & S dataset using ICA feature extraction technique.

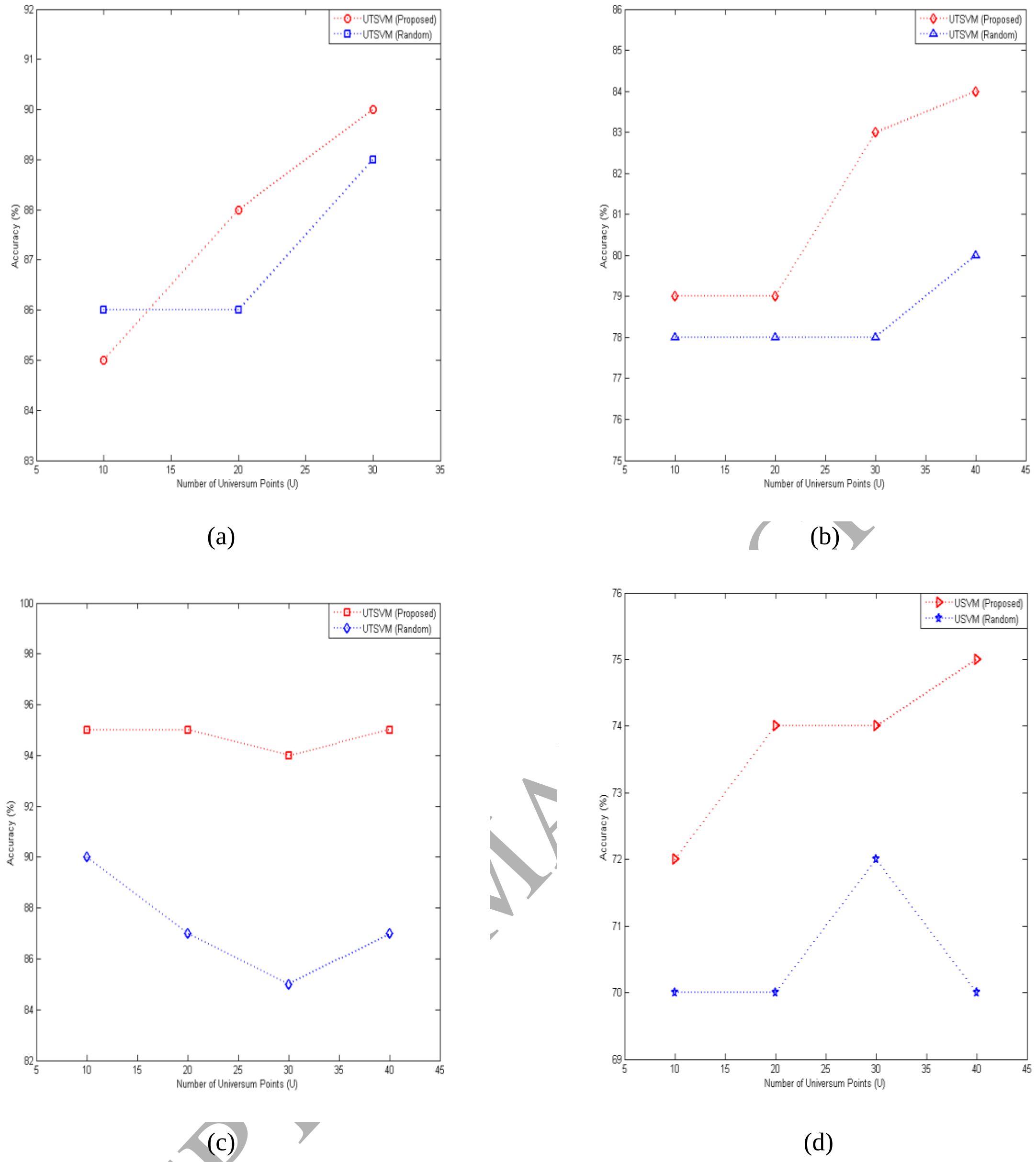
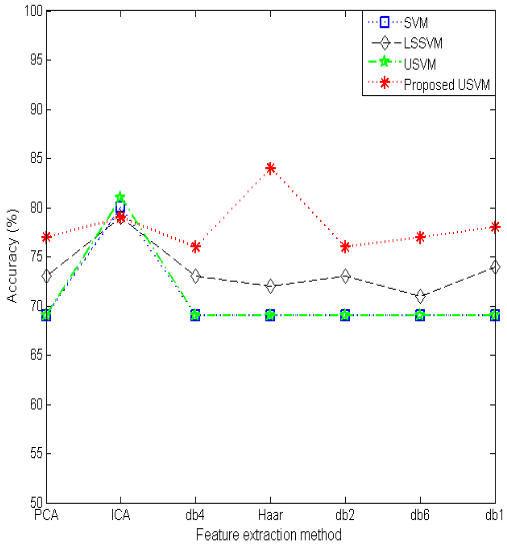
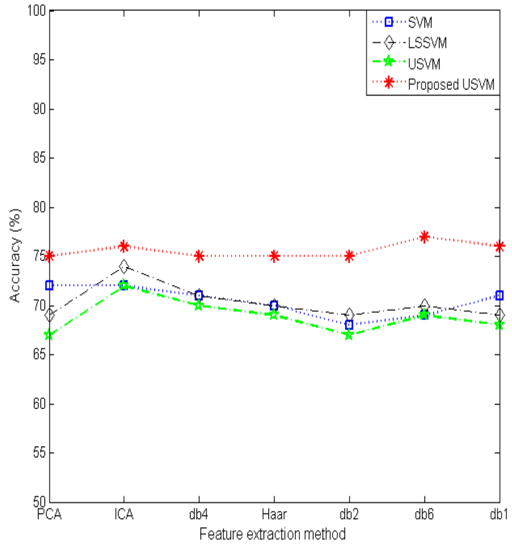


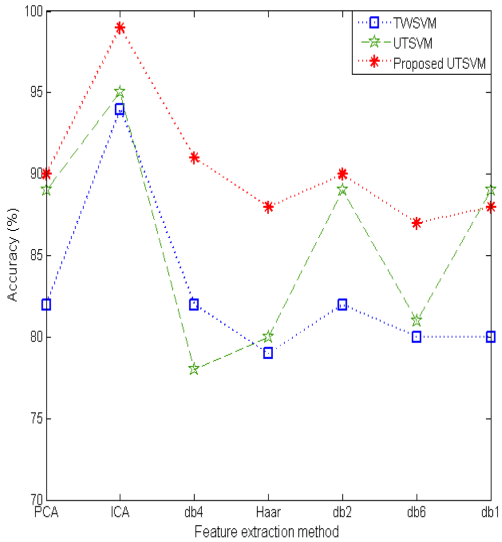
Figure 8: Performance comparison of proposed approach for UTSVM with the random averaging method on (a) Z&S using PCA, (b) O&S using PCA, (c) O&S using ICA and (d) O&S using wavelet (db4) feature extraction technique.



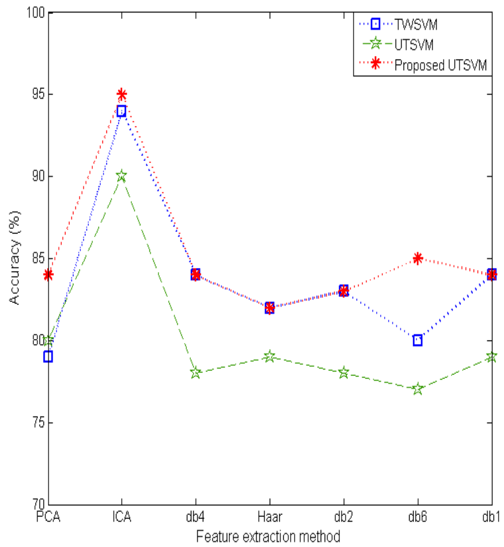
(a)



(b)

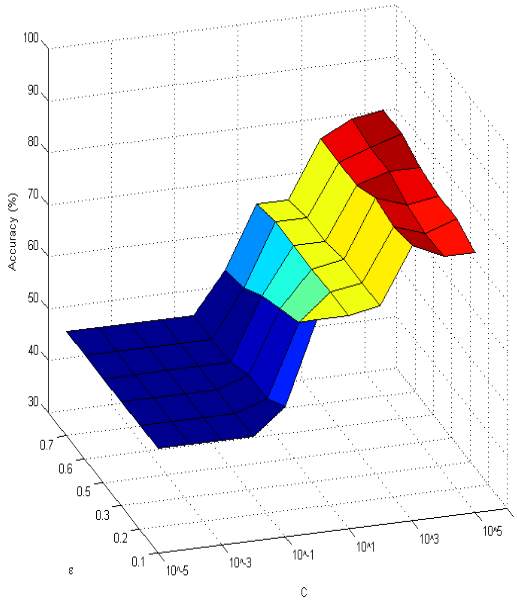


(c)

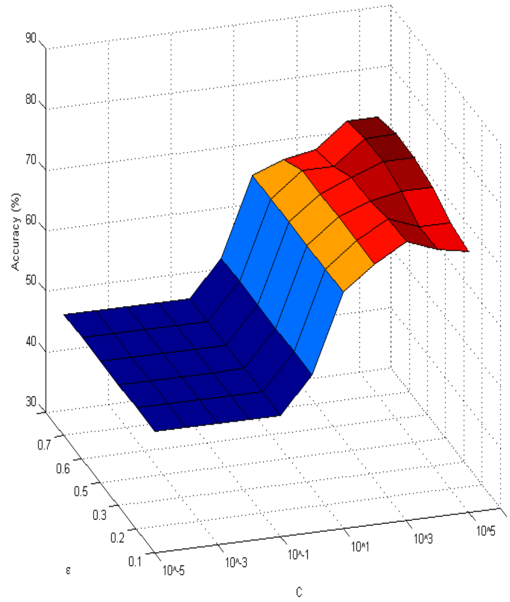


(d)

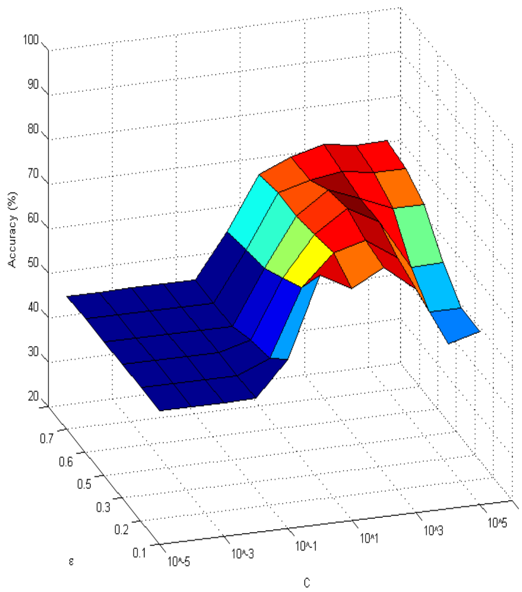
Figure 9: Accuracy comparison for classification of EEG signals using different algorithms with Gaussian kernel. SVM based algorithms for classification on (a) Z&S and (b) O&S datasets, and TWSVM based algorithms on (c) Z&S and (d) O&S datasets using different feature extraction techniques.



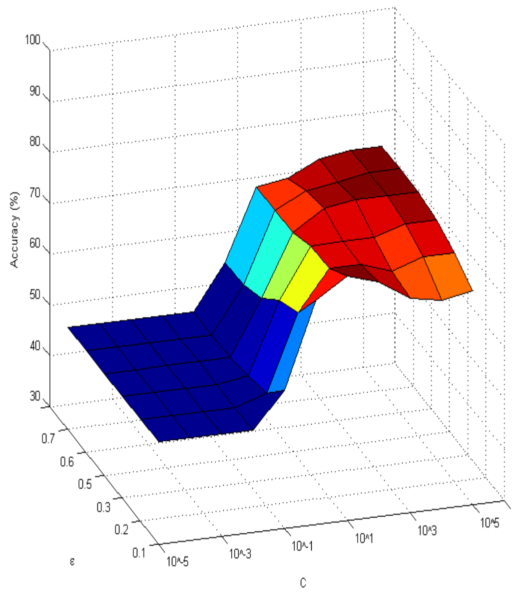
(a) Z&S with Haar



(b) Z&S with db4

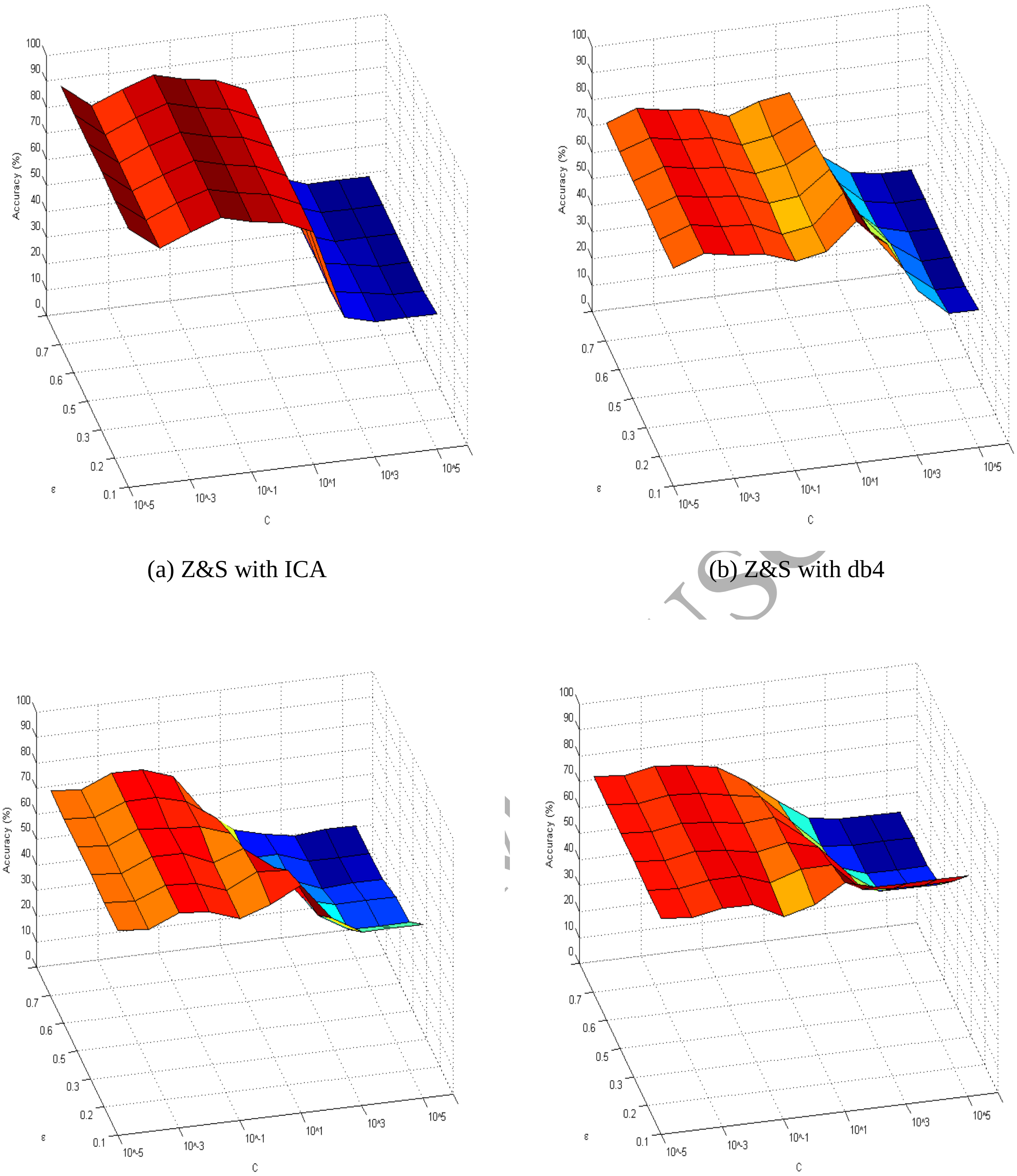


(c) O&S with db1



(d) O&S with db4

Figure 10: Insensitivity performance of proposed USVM for classification of seizure and healthy EEG signals to the user specified parameters (C, ε) using Gaussian kernel.



(a) Z&S with ICA

(b) Z&S with db4

(c) O&S with PCA

(d) O&S with db4

Figure 11: Insensitivity performance of proposed UTSVM for classification of seizure and healthy EEG signals to the user specified parameters (C, ε) using Gaussian kernel.