Universum least squares twin parametric-margin support vector machine

B. Richhariya
Discipline of Mathematics
Indian Institute of Technology Indore
Simrol, Indore, 453552, India
phd1701241001@iiti.ac.in

M. Tanveer
Discipline of Mathematics
Indian Institute of Technology Indore
Simrol, Indore, 453552, India
mtanveer@iiti.ac.in

for the Alzheimer's Disease Neuroimaging Initiative*
adni.loni.usc.edu

***Abstract*—Universum based algorithms involve universum samples in the classification problem to improve the generalization performance. In order to provide prior information about data, we utilized universum data to propose a novel classification algorithm. In this paper, a novel parametric model for universum based twin support vector machine is presented for classification problems. The proposed model is termed as universum least squares twin parametric-margin support vector machine (ULSTPMSVM). The solution of ULSTPMSVM involves a system of linear equations. This makes the ULSTPMSVM efficient w.r.t. training time. In order to verify the performance of the proposed model, various experiments are carried out on real world benchmark datasets. Statistical tests are performed to verify the significance of the proposed method. The proposed ULSTPMSVM performed better than existing algorithms in terms of classification accuracy and training time for most of the datasets. Moreover, an application of proposed ULSTPMSVM is presented for classification of Alzheimer's disease data.**

***Index Terms*—Universum, twin parametric model, prior knowledge, magnetic resonance imaging, Alzheimer's disease.**

I. Introduction

Support vector machine (SVM) [1] is a prominent machine learning algorithm. One of the most efficient classification technique is the twin support vector machine (TWSVM) [2] algorithm. TWSVM constructs twin hyperplanes instead of one, thereby reducing the computation cost of the SVM model. A similar approach is followed to propose the twin support vector regression algorithms, using regularization [3] and weighting based approaches [4]. To improve the generalization performance of TWSVM, various models have been proposed in the past. Linear programming twin support vector machines are proposed by formulating unconstrained optimization problems. Regularization is incorporated with a smoothing technique to propose a smooth linear programming twin support vector machine (SLPTSVM) [5]. Moreover, an iterative algorithm is proposed as Newton method for linear programming twin support vector machine (NLPTSVM) [6]. An implicit Lagrangian twin support vector machines is proposed [7] using Newton method with a generalized Hessian approach.

A noise insensitive loss function is used to propose a general twin support vector machine using pinball loss (Pin-GTSVM) [8]. Further, a sparse model of Pin-GTSVM is proposed as sparse pinball twin support vector machine (SPTWSVM) [9]. To improve the robustness, an improved sparse twin support vector machine (ISPTSVM) is proposed [10] by including the structural risk minimization (SRM) principle. An alternative model termed as twin parametric-margin support vector machine (TPMSVM) [11] is also proposed to improve the generalization performance of TWSVM. Some other twin parametric based SVM models are proposed, such as fuzzy based models [12], [13] to improve the generalization performance.

To reduce the training time of TWSVM, a least squares twin support vector machine (LSTSVM) [14] is proposed in the past. The solution of least squares based algorithms involve a system of linear equations. This makes the LSTSVM algorithm efficient in comparison to TWSVM, which involve a pair of quadratic programming problems (QPPs). Various improvements on LSTSVM are proposed in the literature. For class imbalance learning, a robust fuzzy least squares twin support vector machine (RFLSTSVM-CIL) [15] is proposed to deal with imbalanced datasets. For noisy data, a least squares K-nearest neighbor-based weighted multi-class twin SVM [16] is formulated for multi-class classification. In a comprehensive evaluation [17] of 179 classifiers, the best classification accuracy is obtained by the robust energy based LSTSVM (RELS-TSVM) [18] algorithm.

Weston et al. [19] proposed a universum based support vector machine algorithm. The idea of universum based SVM (USVM) is to introduce prior knowledge of data distribution in the optimization problem of SVM. Due to due to higher generalization performance, universum based algorithms are applied in various applications such as EEG signal classification [20] to facial expression detection [21]. However, the inclusion of additional data points leads to increase in the computation cost of the model [22]–[24]. To remove these drawbacks, a twin support vector machine with universum data (UTSVM) [25] is proposed. Further, a least squares twin

*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.

support vector machine with universum data (ULSTSVM) [26] is also proposed to reduce the computation time of UTSVM. Recently, an efficient angle based universum least squares twin support vector machine (AULSTSVM) [27] is proposed for classification problems. Some sample screening techniques [28], [29] are proposed to reduce the training time in universum based learning.

Shao et al. [30] proposed a least squares twin parametric-margin support vector machine (LSTPMSVM), where the classifier is determined by parametric-margin based hyperplanes. In the spirit of LSTPMSVM, we propose a novel universum based least squares twin parametric-margin support vector machine (ULSTPMSVM). The key contributions of this paper are as follows:

- The proposed ULSTPMSVM involves an alternative formulation as compared to the existing ULSTSVM algorithm.
- The solution of ULSTPMSVM involves the solution of system of linear equations. This makes it an efficient and novel universum based classification algorithm.
- The applicability of proposed ULSTPMSVM is shown by experiments on biomedical data.
- Statistical tests are performed to verify the difference between the proposed and existing algorithms.
- Experiments are performed on Alzheimer's disease (AD) [31] data i.e. structural magnetic resonance imaging (sMRI) images to classify the different types of subjects.

In this work, all vectors are treated as column vectors, and the 2-norm of a vector x is denoted by $\|x\|$. The rest of the paper is organized as follows: Section II presents the related work, while section III discusses the formulation of the proposed algorithm. The results of numerical experiments are presented in section IV. Lastly, section V concludes the paper with future work.

II. Related work

In this section, we discuss the formulations of LSTSVM and ULSTSVM algorithm in brief.

A. Least squares twin support vector machine (LSTSVM)

The QPPs of LSTSVM [14] for non-linear case are described as

$$\begin{aligned}\min_{w_1,b_1,\eta_1} \quad & \frac{1}{2}\|K(A,D^T)w_1+e_1b_1\|^2+\frac{c_1}{2}\eta_1^T\eta_1 \\ s.t. \quad & -(K(B,D^T)w_1+e_2b_1)+\eta_1=e_2, \end{aligned} \tag{1}$$

$$\begin{aligned}\min_{w_2,b_2,\eta_2} \quad & \frac{1}{2}\|K(B,D^T)w_2+e_2b_2\|^2+\frac{c_2}{2}\eta_2^T\eta_2 \\ s.t. \quad & K(A,D^T)w_2+e_1b_2+\eta_2=e_1, \end{aligned} \tag{2}$$

where $c_i, i=1,2$ are positive parameters, $\eta_i, i=1,2$ denote the slack variables, $K(.,D^T)$ is the kernel matrix where $D=[A;B]$, and e_1,e_2 represent vectors consisting of ones of suitable dimensions.

By using the constraints in their respective objective functions, we get

$$\begin{aligned}\min_{w_1,b_1} \quad & \frac{1}{2}\|K(A,D^T)w_1+e_1b_1\|^2, \\ & +\frac{c_1}{2}\|K(B,D^T)w_1+e_2b_1+e_2\|^2, \end{aligned} \tag{3}$$

$$\begin{aligned}\min_{w_2,b_2} \quad & \frac{1}{2}\|K(B,D^T)w_2+e_2b_2\|^2 \\ & +\frac{c_2}{2}\|-(K(A,D^T)w_2+e_1b_2)+e_1\|^2. \end{aligned} \tag{4}$$

Taking the gradient of QPP (3) w.r.t. w_1 and b_1, we get

$$\begin{aligned}& K\big(A,D^T\big)^T\big(K(A,D^T)w_1+e_1b_1\big) \\ & +c_1K\big(B,D^T\big)^T\big(K(B,D^T)w_1+e_2b_1+e_2\big)=0, \end{aligned} \tag{5}$$

$$\begin{aligned}& e_1^T\big(K(A,D^T)w_1+e_1b_1\big) \\ & +c_1e_2^T\big(K(B,D^T)w_1+e_2b_1+e_2\big)=0 \end{aligned} \tag{6}$$

Combining Eqs. (5) and (6) and solving, we get

$$[w_1 \;\; b_1]^T=-\left(G^TG+\frac{1}{c_1}H^TH\right)^{-1}G^Te_2, \tag{7}$$

where $H=[K(A,D^T) \;\; e_1]$, and $G=[K(B,D^T) \;\; e_2]$. Similarly, using Eq. (4), we get

$$[w_2 \;\; b_2]^T=\left(H^TH+\frac{1}{c_2}G^TG\right)^{-1}H^Te_1. \tag{8}$$

For a testing data point x, the class is assigned using the following decision function

$$class\ (x)=arg\min_{i=1,2}\frac{|K(x,D^T)w_i+e_ib_i|}{\|w_i\|}. \tag{9}$$

B. Least squares twin support vector machine with universum data (ULSTSVM)

The QPPs of non-linear ULSTSVM [25] are written as follows:

$$\begin{aligned}\min_{w_1,b_1,\eta_1,\psi_1} \quad & \frac{1}{2}\|K(A,D^T)w_1+e_1b_1\|^2+\frac{c_1}{2}\eta_1^T\eta_1 \\ & +\frac{c_3}{2}(\|w_1\|^2+b_1^2)+\frac{c_u}{2}\psi_1^T\psi_1 \\ s.t. \quad & -(K(B,D^T)w_1+e_2b_1)+\eta_1=e_2, \\ & K(U,D^T)w_1+e_ub_1+\psi_1=(-1+\epsilon)e_u, \end{aligned} \tag{10}$$

$$\begin{aligned}\min_{w_2,b_2,\eta_2,\psi_2} \quad & \frac{1}{2}\|K(B,D^T)w_2+e_2b_2\|^2+\frac{c_2}{2}\eta_2^T\eta_2 \\ & +\frac{c_4}{2}(\|w_2\|^2+b_2^2)+\frac{c_u}{2}\psi_2^T\psi_2 \\ s.t. \quad & K(A,D^T)w_2+e_1b_2+\eta_2=e_1, \\ & -(K(U,D^T)w_2+e_ub_2)+\psi_2=(-1+\epsilon)e_u, \end{aligned} \tag{11}$$

where $c_i, i = 1, 2$ are positive parameters, $\eta_i, \psi_i, i = 1, 2$ represent the slack variables, and e_1, e_2, e_u denote the vectors of ones of suitable dimensions.

Substituting the constraints in their objective functions, we get

$$\begin{aligned}\min_{w_1,b_1} \quad & \frac{1}{2}\|K(A,D^T)w_1 + e_1b_1\|^2 + \frac{c_3}{2}(\|w_1\|^2 + b_1^2) \\ & + \frac{c_1}{2}\|K(B,D^T)w_1 + e_2b_1 + e_2\|^2 \\ & + \frac{c_u}{2}\| - (K(U,D^T)w_1 + e_ub_1) + (-1+\epsilon)e_u\|^2, \end{aligned} \tag{12}$$

$$\begin{aligned}\min_{w_2,b_2} \quad & \frac{1}{2}\|K(B,D^T)w_2 + e_2b_2\|^2 + \frac{c_4}{2}(\|w_2\|^2 + b_2^2) \\ & + \frac{c_2}{2}\| - (K(A,D^T)w_2 + e_1b_2) + e_1\|^2 \\ & + \frac{c_u}{2}\|(K(U,D^T)w_2 + e_ub_2) + (-1+\epsilon)e_u\|^2. \end{aligned} \tag{13}$$

Taking the gradient of QPP (12) w.r.t. w_1 and b_1 and solving [26], we get

$$\begin{aligned}[w_1 \;\; b_1]^T = & -\big(P^TP + c_1Q^TQ + c_3I + c_uR^TR\big)^{-1}\big(c_1Q^Te_2 \\ & + c_u(1-\epsilon)R^Te_u\big), \end{aligned} \tag{14}$$

where $P = [K(A,D^T) \quad e_1]$, $Q = [K(B,D^T) \quad e_2]$, and $R = [K(U,D^T) \;\; e_u]$. Similarly, using Eq. (13), we get

$$\begin{aligned}[w_2 \;\; b_2]^T = & \big(Q^TQ + c_2P^TP + c_4I + c_uR^TR\big)^{-1}\big(c_2P^Te_1 \\ & + c_u(1-\epsilon)R^Te_u\big). \end{aligned} \tag{15}$$

For a testing data point x, the class is assigned using Eq. (9).

III. Proposed Universum Least Squares Twin Parametric-Margin Support Vector Machine (ULSTPMSVM)

In this section, we present the formulation of proposed ULSTPMSVM for the linear and non-linear cases.

A. Linear ULSTPMSVM

The optimization problem of linear ULSTPMSVM is written as

$$\begin{aligned}\min_{w_1,b_1,\eta_1,\psi_1} \quad & \frac{1}{2}(\|w_1\|^2 + b_1^2) + \nu_1 e_2^T(Bw_1 + e_2b_1) + \frac{c_1}{2}\eta_1^T\eta_1 \\ & + \frac{c_u}{2}\psi_1^T\psi_1 \\ s.t. \quad & Aw_1 + e_1b_1 = \eta_1, \\ & Uw_1 + e_ub_1 + (1-\epsilon)e_u = \psi_1, \end{aligned} \tag{16}$$

$$\begin{aligned}\min_{w_2,b_2,\eta_2,\psi_2} \quad & \frac{1}{2}(\|w_2\|^2 + b_2^2) - \nu_2 e_1^T(Aw_2 + e_1b_2) + \frac{c_2}{2}\eta_2^T\eta_2 \\ & + \frac{c_u}{2}\psi_2^T\psi_2 \\ s.t. \quad & Bw_2 + e_2b_2 = \eta_2, \\ & Uw_2 + e_ub_2 - (1-\epsilon)e_u = \psi_2, \end{aligned} \tag{17}$$

where $c_i, i = 1, 2, c_u$ are positive parameters, and $\eta_i, \psi_i, i = 1, 2$ represent the slack variables.

Using the constraints of Eqs. (16) and (17) in their respective objective functions, we get

$$\begin{aligned}\min_{w_1,b_1} \quad & \frac{c_1}{2}(\|Aw_1 + e_1b_1\|^2) + \nu_1 e_2^T(Bw_1 + e_2b_1) \\ & + \frac{1}{2}(\|w_1\|^2 + b_1^2) + \frac{c_u}{2}(\|Uw_1 + e_ub_1 + (1-\epsilon)e_u\|^2), \end{aligned} \tag{18}$$

$$\begin{aligned}\min_{w_2,b_2} \quad & \frac{c_2}{2}(\|Bw_2 + e_2b_2\|^2) - \nu_2 e_1^T(Aw_2 + e_1b_2) \\ & + \frac{1}{2}(\|w_2\|^2 + b_2^2) + \frac{c_u}{2}(\|Uw_2 + e_ub_2 - (1-\epsilon)e_u\|^2). \end{aligned} \tag{19}$$

Now, taking the gradient of QPP (18) w.r.t. w_1 and b_1 and equating to 0, we get

$$\begin{aligned} & c_1A^T(Aw_1 + e_1b_1) + \nu_1 B^Te_2 + w_1 \\ & + c_uU^T(Uw_1 + e_ub_1 + (1-\epsilon)e_u) = 0, \end{aligned} \tag{20}$$

$$\begin{aligned} & c_1e_1^T(Aw_1 + e_1b_1) + \nu_1 e_2^Te_2 + b_1 \\ & + c_ue_u^T(Uw_1 + e_ub_1 + (1-\epsilon)e_u) = 0. \end{aligned} \tag{21}$$

Combining Eqs. (20) and (21) and solving, we get

$$\begin{aligned}[w_1 \;\; b_1]^T = & -(c_1H^TH + c_uO^TO + I)^{-1}(\nu_1G^Te_2 \\ & + (1-\epsilon)c_uO^Te_u), \end{aligned} \tag{22}$$

where $H = [A; \;\; e_1]$, $G = [B; \;\; e_2]$, and $O = [U; \;\; e_u]$. Similarly, using Eq. (19), we get

$$\begin{aligned}[w_2 \;\; b_2]^T = & (c_2G^TG + c_uO^TO + I)^{-1}(\nu_2H^Te_1 \\ & + (1-\epsilon)c_uO^Te_u). \end{aligned} \tag{23}$$

B. Non-linear ULSTPMSVM

The formulation of non-linear ULSTPMSVM involves kernel generated surfaces. The optimization problem comprises of the following two QPPs,

$$\begin{aligned}\min_{w_1,b_1,\eta_1,\psi_1} \quad & \frac{1}{2}(\|w_1\|^2 + b_1^2) + \nu_1 e_2^T(K(B,D^T)w_1 + e_2b_1) \\ & + \frac{c_1}{2}\eta_1^T\eta_1 + \frac{c_u}{2}\psi_1^T\psi_1 \\ s.t. \quad & K(A,D^T)w_1 + e_1b_1 = \eta_1, \\ & K(U,D^T)w_1 + e_ub_1 + (1-\epsilon)e_u = \psi_1, \end{aligned} \tag{24}$$

$$\begin{aligned}\min_{w_2,b_2,\eta_2,\psi_2} \quad & \frac{1}{2}(\|w_2\|^2 + b_2^2) - \nu_2 e_1^T(K(A,D^T)w_2 + e_1b_2) \\ & + \frac{c_2}{2}\eta_2^T\eta_2 + \frac{c_u}{2}\psi_2^T\psi_2 \\ s.t. \quad & K(B,D^T)w_2 + e_2b_2 = \eta_2, \\ & K(U,D^T)w_2 + e_ub_2 - (1-\epsilon)e_u = \psi_2, \end{aligned} \tag{25}$$

where $c_i, i = 1, 2, c_u$ are positive parameters, $K(., D^T)$ is the kernel matrix where $D = [A; B]$, and $\eta_i, \psi_i, i = 1, 2$ represent the slack variables.

Substituting the constraints of Eqs. (24) and (25) in their respective objective functions, we get

$$\begin{aligned}\min_{w_1,b_1} \quad & \frac{1}{2}(\|w_1\|^2+b_1^2)+\frac{c_1}{2}(\|K(A,D^T)w_1+e_1b_1\|^2) \\ & +\nu_1 e_2^T(K(B,D^T)w_1+e_2b_1) \\ & +\frac{c_u}{2}(\|K(U,D^T)w_1+e_ub_1+(1-\epsilon)e_u\|^2), \quad (26)\end{aligned}$$

$$\begin{aligned}\min_{w_2,b_2} \quad & \frac{1}{2}(\|w_2\|^2+b_2^2)+\frac{c_2}{2}(\|K(B,D^T)w_2+e_2b_2\|^2) \\ & -\nu_2 e_1^T(K(A,D^T)w_2+e_1b_2) \\ & +\frac{c_u}{2}(\|K(U,D^T)w_2+e_ub_2-(1-\epsilon)e_u\|^2). \quad (27)\end{aligned}$$

Solving similar to the linear case, we get

$$\begin{aligned}[w_1 \;\; b_1]^T =& -(c_1P^TP+c_uR^TR+I)^{-1}(\nu_1Q^Te_2 \\ & +(1-\epsilon)c_uR^Te_u), \quad (28)\end{aligned}$$

where $P=[A; \;\; e_1]$, $Q=[B; \;\; e_2]$, and $R=[U; \;\; e_u]$.
Similarly, using Eq. (27), we get

$$\begin{aligned}[w_2 \;\; b_2]^T =& (c_2Q^TQ+c_uR^TR+I)^{-1}(\nu_2P^Te_1 \\ & +(1-\epsilon)c_uR^Te_u). \quad (29)\end{aligned}$$

The decision function of proposed ULSTPMSVM is same as in Eq. (9).

C. Time complexity

The time complexity of proposed ULSTPMSVM is lesser than existing algorithms such as TWSVM and ULSTSVM. In comparison to TWSVM where QPPs are solved, proposed ULSTPMSVM solves a system of linear equations, leading to lesser computation cost [30]. Moreover, in comparison to ULSTSVM, time complexity of proposed ULSTPMSVM is lesser because ULSTSVM involves an additional matrix multiplication term in its solution (Eqs. 14 and 15), as compared to proposed ULSTPMSVM in Eqs. 28 and 29. However, the computation cost of proposed ULSTPMSVM is higher than LSTSVM. This is due to the incorporation of universum data points in the proposed ULSTPMSVM algorithm.

IV. Experimental results

In this section, we show the results of experiments carried out on real world datasets, along with an application on Alzheimer's disease. The real world datasets are taken from UCI [32], and KEEL repository [33].

All MRI images used in this work were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was started in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The main objective of ADNI is to find out the effectiveness of neuroimaging techniques like 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.

A. Experimental setup

All the experiments are carried out on a PC using Windows 10 OS 64 bit, 3.60 GHz Intel® core™ i7-7700 processor, 16 GB of RAM and MATLAB R2008b environment. To solve the QPPs in case of TWSVM, we utilized MOSEK optimization toolbox (http://www.mosek.com). 5-fold cross-validation is performed for all the methods to obtain the optimal hyperparameters. For non-linear case, Gaussian kernel is employed in all the cases.

For experiments on real world datasets, 50% data is used for training. The parameters c and ν are chosen from the set $\{10^{-5},10^{-4},...,10^{5}\}$, while μ is chosen from $\{2^{-5},2^{-4},...,2^{5}\}$. The value for ϵ is selected from $\{0.1,0.4,0.7\}$. The universum data is generated by performing random averaging of data points in all the cases [20], [21], [25].

In case of Alzheimer's disease dataset, 40% of the samples are used for training. Freesurfer's recon-all pipeline (version 6.0.1) [34] is applied for processing the structural MRI (sMRI) images. The total intracranial volume (TIV) is used for normalizing the volumetric features of brain in all the subjects [35]–[37].

B. Real world datasets

In Table I, performance comparison of proposed ULSTPMSVM is shown with existing algorithms on 18 real world benchmark datasets. The existing algorithms used for comparison in this work are TWSVM [2], LSTSVM [14], LSTPMSVM [30], and ULSTSVM [26]. One can observe that the proposed ULSTPMSVM is showing lowest average rank on the basis of accuracy. In terms of training time, the time taken by proposed ULSTPMSVM is comparable or lesser than the existing algorithms. It is noticeable that the training time of TWSVM is the highest. This is because the solution of TWSVM involves a pair of QPPs, as compared to systems of linear equations in least squares based algorithms.

C. Statistical analysis

To check the statistical difference, the Friedman test [38] is performed with the corresponding posthoc test. First, we assume that there is no difference between the methods. Now, the χ_F^2 value is calculated for Friedman test using average ranks r_i from Table I as,

$$\chi_F^2=\frac{12\times 18}{5(5+1)}\left[\sum_{i=1}^{5}r_i^2-\frac{5(5+1)^2}{4}\right],$$

$$\begin{aligned}\chi_F^2=\frac{12\times 18}{5(5+1)}\Bigg[&(3.8889^2+3.4167^2+3.0278^2+3.0556^2 \\ &+1.6111^2)-\frac{5(5+1)^2}{4}\Bigg]\approx 20.8605.\end{aligned}$$

The F_F value is calculated as

$$F_F=\frac{(18-1)(20.8605)}{18\times(5-1)-20.8605}\approx 6.9345.$$

TABLE I: Comparison of proposed ULSTPMSVM with existing algorithms on classification of real world datasets using Gaussian kernel function.

Dataset (Sample size)	TWSVM Accuracy (%) (c, μ) Time (s)	LSTSVM Accuracy (%) (c, μ) Time (s)	LSTPMSVM Accuracy (%) (c, ν, μ) Time (s)	ULSTSVM Accuracy (%) (c_1, c_2, ϵ, μ) Time (s)	Proposed ULSTPMSVM Accuracy (%) (c, ν, ϵ, μ) Time (s)
Ecoli-0-1_vs_5 (242 × 6)	94.2149 (10^{-5}, 2^5) 0.0442	90.9091 (10^{-3}, 2^5) 0.0033	95.0413 (10^{-2}, 10^1, 2^4) 0.0035	95.0413 (10^{-5}, 10^{-1}, 0.1, 2^4) 0.0035	**97.5207** (10^1, 10^2, 0.1, 2^5) 0.0035
Ecoli-0-1-4-7_vs_5-6 (334 × 6)	97.006 (10^{-3}, 2^5) 0.0135	94.6108 (10^{-1}, 2^5) 0.0016	97.6048 (10^0, 10^1, 2^4) 0.0018	**98.8024** (10^0, 10^{-3}, 0.7, 2^5) 0.0017	**98.8024** (10^1, 10^{-2}, 0.1, 2^5) 0.0016
Ecoli-0-2-6-7_vs_3-5 (226 × 7)	93.8053 (10^{-3}, 2^5) 0.0079	**97.3451** (10^2, 2^5) 0.0007	96.4602 (10^2, 10^2, 2^5) 0.0007	92.9204 (10^{-1}, 10^{-5}, 0.4, 2^5) 0.0009	96.4602 (10^{-1}, 10^{-2}, 0.1, 2^4) 0.0007
Ecoli-0-3-4-6_vs_5 (206 × 7)	97.0874 (10^{-5}, 2^5) 0.0071	**98.0583** (10^4, 2^5) 0.0006	95.1456 (10^5, 10^2, 2^5) 0.0009	**98.0583** (10^{-1}, 10^{-5}, 0.7, 2^5) 0.0007	**98.0583** (10^1, 10^{-1}, 0.7, 2^5) 0.0007
Ecoli-0-6-7_vs_3-5 (224 × 7)	91.0714 (10^{-3}, 2^4) 0.0106	94.6429 (10^1, 2^5) 0.0007	91.9643 (10^0, 10^1, 2^4) 0.0008	91.0714 (10^{-2}, 10^{-3}, 0.7, 2^5) 0.0008	**95.5357** (10^1, 10^{-1}, 0.1, 2^5) 0.0009
Ecoli4 (336 × 7)	97.6331 (10^{-2}, 2^2) 0.0133	97.6331 (10^{-1}, 2^4) 0.0013	97.6331 (10^3, 10^5, 2^1) 0.0015	**98.2249** (10^{-2}, 10^{-5}, 0.1, 2^3) 0.0015	**98.2249** (10^5, 10^3, 0.4, 2^1) 0.0019
Glass-0-1-6_vs_2 (194 × 9)	89.6907 (10^{-4}, 2^1) 0.0064	91.7526 (10^{-3}, 2^3) 0.0008	**92.7835** (10^5, 10^2, 2^{-1}) 0.0006	87.6289 (10^{-3}, 10^{-5}, 0.4, 2^2) 0.0006	92.7835 (10^{-5}, 10^{-3}, 0.4, 2^{-2}) 0.0006
Glass-0-4_vs_5 (92 × 9)	91.4894 (10^{-1}, 2^4) 0.0058	**100** (10^{-1}, 2^3) 0.0003	95.7447 (10^4, 10^5, 2^4) 0.0005	**100** (10^{-1}, 10^{-5}, 0.1, 2^2) 0.0002	97.8723 (10^3, 10^0, 0.4, 2^0) 0.0002
Heart-stat (270 × 13)	66.9118 (10^0, 2^5) 0.0077	62.5 (10^{-5}, 2^5) 0.0013	65.4412 (10^0, 10^1, 2^5) 0.001	63.2353 (10^1, 10^2, 0.4, 2^5) 0.0009	**67.6471** (10^{-1}, 10^{-2}, 0.7, 2^5) 0.001
Led7digit-0-2-4-5-6-7-8-9_vs_1 (444 × 7)	91.8919 (10^{-4}, 2^1) 0.0236	93.6937 (10^0, 2^2) 0.0024	**96.8468** (10^{-2}, 10^{-1}, 2^0) 0.0026	92.3423 (10^{-1}, 10^{-5}, 0.1, 2^2) 0.0026	92.7928 (10^2, 10^2, 0.1, 2^1) 0.0026
Ecoli-0-1-4-6_vs_5 (282 × 6)	98.5816 (10^{-3}, 2^5) 0.0103	98.5816 (10^3, 2^5) 0.001	98.5816 (10^1, 10^1, 2^4) 0.001	98.5816 (10^{-2}, 10^{-3}, 0.1, 2^5) 0.001	**99.2908** (10^0, 10^{-2}, 0.1, 2^5) 0.001
Ecoli2 (336 × 7)	91.716 (10^{-1}, 2^0) 0.0116	86.9822 (10^{-2}, 2^0) 0.0014	90.5325 (10^{-2}, 10^{-1}, 2^{-3}) 0.0016	**94.0828** (10^0, 10^1, 0.1, 2^{-2}) 0.0013	92.3077 (10^1, 10^{-2}, 0.4, 2^{-1}) 0.0018
Glass4 (282 × 6)	94.4444 (10^{-5}, 2^0) 0.0084	96.2963 (10^{-2}, 2^1) 0.001	94.4444 (10^0, 10^2, 2^{-1}) 0.0007	96.2963 (10^{-2}, 10^{-5}, 0.7, 2^1) 0.0007	**97.2222** (10^2, 10^{-1}, 0.7, 2^1) 0.0008
Brwisconsin (683 × 9)	98.538 (10^{-4}, 2^3) 0.032	98.2456 (10^1, 2^4) 0.0064	98.538 (10^{-1}, 10^0, 2^4) 0.0066	98.2456 (10^{-4}, 10^{-4}, 0.4, 2^3) 0.0069	**98.8304** (10^3, 10^3, 0.1, 2^5) 0.0068
Ecoli3 (336 × 7)	91.716 (10^{-1}, 2^0) 0.012	86.9822 (10^{-2}, 2^0) 0.0016	90.5325 (10^{-2}, 10^{-1}, 2^{-3}) 0.0021	**94.0828** (10^0, 10^1, 0.1, 2^{-2}) 0.0016	92.3077 (10^1, 10^{-2}, 0.4, 2^{-1}) 0.0015
Yeast1vs7 (460 × 8)	91.7391 (10^{-1}, 2^{-2}) 0.0219	93.0435 (10^0, 2^{-1}) 0.003	93.4783 (10^{-4}, 10^{-2}, 2^{-1}) 0.0032	93.4783 (10^{-2}, 10^{-3}, 0.4, 2^{-2}) 0.0031	**93.913** (10^1, 10^{-1}, 0.1, 2^{-1}) 0.0042
Ecoli0137vs26 (312 × 7)	95.5128 (10^{-2}, 2^{-2}) 0.0108	96.1538 (10^{-1}, 2^{-1}) 0.0015	**97.4359** (10^{-5}, 10^{-4}, 2^{-2}) 0.0013	96.1538 (10^1, 10^3, 0.4, 2^{-1}) 0.0013	96.7949 (10^{-1}, 10^{-2}, 0.4, 2^{-1}) 0.0012
Votes (436 × 16)	94.4954 (10^{-1}, 2^5) 0.0141	94.4954 (10^0, 2^4) 0.0023	94.4954 (10^3, 10^2, 2^4) 0.0025	94.0367 (10^1, 10^{-4}, 0.7, 2^4) 0.0025	**94.9541** (10^1, 10^0, 0.4, 2^2) 0.0025
Average rank	3.8889	3.4167	3.0278	3.0556	**1.6111**

This F-distribution involves $\big(5-1,(5-1)(18-1)\big)=(4,68)$ degrees of freedom. Thus, for the significance level at $\alpha = 0.05$, the critical value for $F(4,68)$ is 2.5066. Since, $F_F = 6.9345 > 2.5066$, the null hypothesis is rejected.

To check the pairwise difference between the proposed ULSTPMSVM and existing algorithms, we perform the Nemenyi posthoc test [38]. For significant pairwise difference between the methods at significance level of $\alpha = 0.10$, the average ranks of the algorithms shown in Table I should differ by atleast $2.459\sqrt{\frac{5(5+1)}{6\times18}} \approx 1.296$. The pairwise difference between the methods is shown in Table II.

TABLE II: Pairwise significant difference between proposed ULSTPMSVM and existing algorithms.

Statistical difference	TWSVM	LSTSVM	LSTPMSVM	ULSTSVM
Proposed ULSTPMSVM	**Yes**	**Yes**	**Yes**	**Yes**

It can be stated from Table II that the proposed ULSTPMSVM is significantly better than TWSVM, LSTSVM, LSTPMSVM, and ULSTSVM algorithms.

D. Insensitivity performance

The insensitivity analysis of proposed ULSTPMSVM is presented in fig. 1. The variation of accuracy w.r.t. the penalty parameter c and ν is shown for 4 datasets viz. Ecoli-0-1-4-7_vs_5-6, Ecoli-0-6-7_vs_3-5, Glass4, and Votes.

One can observe that the accuracy of proposed ULSTPMSVM increases with higher value of c, while ν does not have much effect on the accuracy. However, the accuracy is slightly higher for larger values of ν.

E. Discussion

Experiments show the advantages of proposed ULSTPMSVM over existing algorithms. In comparison to the existing ULSTSVM algorithm, the formulation of proposed ULSTPMSVM is novel and efficient. Moreover, in comparison to other algorithms involving QPPs such as TWSVM or UTSVM [25], proposed ULSTPMSVM requires very less computation time. Since, there are not many least squares based twin SVM models using universum, the proposed approach can be beneficial for many applications.

F. Alzheimer disease classification

In Alzheimer disease data, we have considered three classes namely control normal (CN), Alzheimer's disease (AD), and mild cognitive impairment (MCI) [39], [40]. We include 50 sMRI images of CN and AD each, and 49 sMRI images of MCI, since one MCI image failed to process. The performance of proposed ULSTPMSVM and existing algorithms on classification of Alzheimer data is shown in Table III. One can see that proposed ULSTPMSVM performed better than other algorithms in 2 out of 3 datasets i.e. CN vs MCI, and MCI vs AD.

The highest accuracy of proposed ULSTPMSVM in MCI vs AD indicates that it may be used for the early diagnosis of Alzheimer's disease. Moreover, the proposed ULSTPMSVM can be used for other diseases such as epilepsy, where the universum data is selected from the dataset itself [20], [23]. This may lead to higher classification accuracy for such problems.

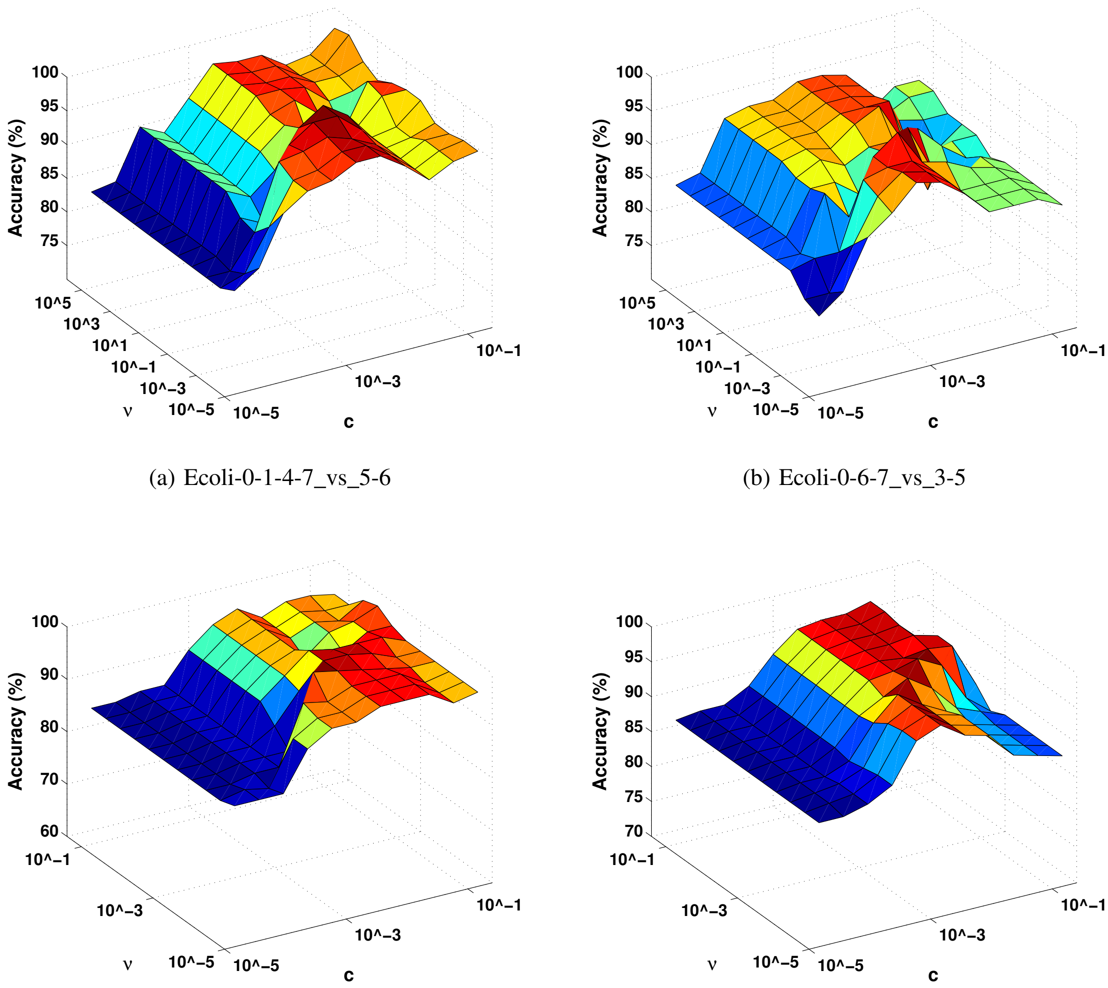
V. Conclusion

A novel universum based algorithm is proposed in this work termed as universum least squares twin parametric-margin support vector machine (ULSTPMSVM). The proposed algorithms show high generalization performance with lesser training time in comparison to existing algorithms. The formulation of ULSTPMSVM is an alternative approach towards universum based learning. The optimization problem of ULSTPMSVM involves a parametric model, solved by a system of linear equations. In terms of statistical difference in the generalization performance, proposed ULSTPMSVM turns out to be significantly better than the existing algorithms. In future, ULSTPMSVM can be applied by selecting universum samples from different classes of the classification problem. Moreover, the proposed ULSTPMSVM performed well for classification of Alzheimer's disease, showing its applicability on real world biomedical applications. For implementation, the codes of proposed algorithm will be publicly available on the author's Github page i.e., https://github.com/mtanveer1/.

Acknowledgement

This work was supported by Science and Engineering Research Board (SERB), INDIA under Ramanujan fellowship grant no. SB/S2/RJN-001/2016 and Early Career Research Award grant no. ECR/2017/000053. It is also supported by Council of Scientific & Industrial Research (CSIR), New Delhi, INDIA under Extra Mural Research (EMR) Scheme grant no. 22(0751)/17/EMR-II. We gratefully acknowledge the Indian Institute of Technology Indore for providing facilities and support. A sincere thanks 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



(a) Ecoli-0-1-4-7_vs_5-6

(b) Ecoli-0-6-7_vs_3-5

(c) Glass4

(d) Votes

Fig. 1: Plots showing insensitivity performance for the penalty parameter $c = c_1 = c_2$ with $\nu = \nu_1 = \nu_2$ for proposed ULSTPMSVM using Gaussian kernel function.

TABLE III: Performance comparison of proposed ULSTPMSVM on classification of Alzheimers data.

Dataset	TWSVM Accuracy (%)	LSTSVM Accuracy (%)	LSTPMSVM Accuracy (%)	ULSTSVM Accuracy (%)	Proposed ULSTPMSVM Accuracy (%)
CN vs AD	85	80	80	85	76.6667
CN vs MCI	74.5763	59.322	**76.2712**	74.5763	**76.2712**
MCI vs AD	61.0169	44.0678	61.0169	42.3729	**64.4068**

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.

References

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[2] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 905–910, 2007.

[3] M. Tanveer, K. Shubham, M. Aldhaifallah, and K. Nisar, "An efficient implicit regularized Lagrangian twin support vector regression," *Applied Intelligence*, vol. 44, no. 4, pp. 831–848, 2016.

[4] M. Tanveer, K. Shubham, M. Aldhaifallah, and S. S. Ho, "An efficient regularized K-nearest neighbor based weighted twin support vector regression," *Knowledge-Based Systems*, vol. 94, pp. 70–87, 2016.

[5] M. Tanveer, "Application of smoothing techniques for linear programming twin support vector machines," *Knowledge and Information Systems*, vol. 45, no. 1, pp. 191–214, 2015.

[6] M. Tanveer, "Robust and sparse linear programming twin support vector machines," *Cognitive Computation*, vol. 7, no. 1, pp. 137–149, 2015.

[7] M. Tanveer, "Newton method for implicit Lagrangian twin support vector machines," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 6, pp. 1029–1040, 2015.

[8] M. Tanveer, A. Sharma, and P. N. Suganthan, "General twin support vector machine with pinball loss function," *Information Sciences*, vol. 494, pp. 311–327, 2019.

[9] M. Tanveer, A. Tiwari, R. Choudhary, and S. Jalan, "Sparse pinball twin support vector machines," *Applied Soft Computing*, vol. 78, pp. 164–175, 2019.

[10] M. Tanveer, T. Rajani, and M. Ganaie, "Improved sparse pinball twin SVM," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 3287–3291.

[11] X. Peng, "TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2678–2692, 2011.

[12] P. Borah and D. Gupta, "A two-norm squared fuzzy-based least squares twin parametric-margin support vector machine," in *Machine Intelligence and Signal Analysis*. Springer, 2019, pp. 119–134.

[13] D. Gupta, P. Borah, and M. Prasad, "A fuzzy based Lagrangian twin parametric-margin support vector machine (FLTPMSVM)," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2017, pp. 1–7.

[14] M. A. Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7535–7543, 2009.

[15] B. Richhariya and M. Tanveer, "A robust fuzzy least squares twin support vector machine for class imbalance learning," *Applied Soft Computing*, vol. 71, pp. 418–432, 2018.

[16] M. Tanveer, A. Sharma, and P. N. Suganthan, "Least squares K-nearest neighbor-based weighted multi-class twin SVM," *Neurocomputing*, vol. (In press), 2020.

[17] M. Tanveer, C. Gautam, and P. N. Suganthan, "Comprehensive evaluation of twin svm based classifiers on UCI datasets," *Applied Soft Computing*, vol. 83, p. 105617, 2019.

[18] M. Tanveer, M. A. Khan, and S.-S. Ho, "Robust energy-based least squares twin support vector machines," *Applied Intelligence*, vol. 45, no. 1, pp. 174–186, 2016.

[19] J. Weston, R. Collobert, F. Sinz, L. Bottou, and V. Vapnik, "Inference with the universum," in *Proceedings of the 23rd International Conference on Machine learning*. ACM, 2006, pp. 1009–1016.

[20] B. Richhariya and M. Tanveer, "EEG signal classification using universum support vector machine," *Expert Systems with Applications*, vol. 106, pp. 169–182, 2018.

[21] B. Richhariya and D. Gupta, "Facial expression recognition using iterative universum twin support vector machine," *Applied Soft Computing*, vol. 76, pp. 53–67, 2019.

[22] W. Long, Y.-r. Tang, and Y.-j. Tian, "Investor sentiment identification based on the universum SVM," *Neural Computing and Applications*, vol. 30, no. 2, pp. 661–670, 2018.

[23] B. Richhariya and M. Tanveer, "A reduced universum twin support vector machine for class imbalance learning," *Pattern Recognition*, vol. 102, p. 107150, 2020.

[24] B. Richhariya, A. Sharma, and M. Tanveer, "Improved universum twin support vector machine," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, pp. 2045–2052.

[25] Z. Qi, Y. Tian, and Y. Shi, "Twin support vector machine with universum data," *Neural Networks*, vol. 36, pp. 112–119, 2012.

[26] Y. Xu, M. Chen, and G. Li, "Least squares twin support vector machine with universum data for classification," *International Journal of Systems Science*, vol. 47, no. 15, pp. 3637–3645, 2016.

[27] B. Richhariya, M. Tanveer, and Alzheimer's Disease Neuroimaging Initiative, "An efficient angle based universum least squares twin support vector machine for pattern classification," *ACM Transactions on Internet Technology (TOIT)*, vol. https://doi.org/10.1145/3387131, 2020.

[28] J. Zhao and Y. Xu, "A safe sample screening rule for universum support vector machines," *Knowledge-Based Systems*, vol. 138, pp. 46–57, 2017.

[29] J. Zhao, Y. Xu, and H. Fujita, "An improved non-parallel universum support vector machine and its safe sample screening rule," *Knowledge-Based Systems*, vol. 170, pp. 79–88, 2019.

[30] Y.-H. Shao, Z. Wang, W.-J. Chen, and N.-Y. Deng, "Least squares twin parametric-margin support vector machine for classification," *Applied Intelligence*, vol. 39, no. 3, pp. 451–464, 2013.

[31] C. Patterson, "The state of the art of dementia research: New frontiers," *World Alzheimer's Report 2018*, 2018.

[32] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[33] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework." *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, 2011.

[34] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *NeuroImage*, vol. 61, no. 4, pp. 1402–1418, 2012.

[35] E. Westman, J.-S. Muehlboeck, and A. Simmons, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion," *NeuroImage*, vol. 62, no. 1, pp. 229–238, 2012.

[36] L. Sørensen, C. Igel, A. Pai, I. Balas, C. Anker, M. Lillholm, M. Nielsen, and 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: Clinical*, vol. 13, pp. 470–482, 2017.

[37] C. Jongkreangkrai, Y. Vichianin, C. Tocharoenchai, H. Arimura, and Alzheimer's Disease Neuroimaging Initiative, "Computer-aided classification of Alzheimer's disease based on support vector machine with combination of cerebral image features in MRI," in *Journal of Physics: Conference Series*, vol. 694, no. 1. IOP Publishing, 2016, 2016, p. 012036.

[38] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine learning Research*, vol. 7, no. Jan, pp. 1–30, 2006.

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

[40] M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, M. Prasad, and C. T. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 1s, pp. 1–35, 2020.