ELSEVIER



Contents lists available at ScienceDirect

# **Knowledge-Based Systems**

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

# KNN weighted reduced universum twin SVM for class imbalance learning

## M.A. Ganaie, M. Tanveer<sup>\*</sup>, for the Alzheimer's Disease Neuroimaging Initiative<sup>1</sup>

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

#### ARTICLE INFO

Article history: Received 19 October 2021 Received in revised form 27 February 2022 Accepted 9 March 2022 Available online 23 March 2022

Keywords: Universum Rectangular kernel Class imbalance Imbalance ratio Twin support vector machine KNN weighted

#### ABSTRACT

In real world problems, imbalance of data samples poses major challenge for the classification problems as the data samples of a particular class are dominating. Problems like fault and disease detection involve imbalance data and hence need attention to avoid the bias towards a particular class. The classification models like support vector machines (SVM) get biased to majority class samples and hence results in misclassification of the minority class samples. SVM suffers as no prior information related to the data is involved in the generation of hyperplanes. Also, local information of the neighbourhood is ignored in SVM samples and thus treats each sample equally for generating the hyperplanes. However, the data points may be contaminated and may mislead the generation of hyperplanes. Inspired by the idea of prior data information and local neighbourhood information, we propose K-nearest neighbour based weighted reduced universum twin SVM for class imbalance learning (KWRUTSVM-CIL). The proposed KWRUTSVM-CIL embodies the local neighbourhood information and uses universum data to balance the classes in class imbalance problems. Local neighbourhood information is incorporated via weight matrix in the objective function. In proposed KWRUTSVM-CIL model, weight vectors are used in the corresponding constraints of the objective functions to exploit the interclass information. The oversampling and undersampling approaches are followed to balance the data in class imbalance problems. Universum data gives prior information of the data. Twin SVM, universum twin SVM, and reduced universum twin SVM for class imbalance implement empirical risk minimization principle and thus may lead to overfitting. However, the proposed KWRUTSVM-CIL model embodies regularization term to maximize the margin and implement the structural risk minimization principle which is the marrow of statistical learning and overcomes the issues of overfitting. Experimental results and the statistical analysis signify that the generalization ability of proposed KWRUTSVM-CIL model is superior in comparison to other twin SVM based models. As an application, we use the proposed KWRUTSVM-CIL model for the diagnosis of Alzheimer's disease and breast cancer disease. The proposed KWRUTSVM-CIL model showed better generalization performance compared to other twin SVM based models in biomedical datasets.

© 2022 Elsevier B.V. All rights reserved.

#### 1. Introduction

Support vector machine (SVM) [1] is the successful algorithm for the classification problems used across the domains like detection of faults [2], disease detection [3,4], face recognition [5,6] and internet traffic classification [7]. Moreover, SVMs have also

\* Corresponding author.

mtanveer@iiti.ac.in (M. Tanveer).

https://doi.org/10.1016/j.knosys.2022.108578 0950-7051/© 2022 Elsevier B.V. All rights reserved. been employed in fuel consumption estimation [8] and wire arc additive manufacturing [9]. SVM via structural risk minimization principle leads to better generalization performance. However, the computational cost of SVM is higher and hence twin support vector machine (TSVM) [10] was proposed. TSVM generates two non-parallel hyperplanes by solving two smaller size quadratic programming problems (QPPs). TSVM generates each hyperplane proximal to the samples of one class and as far as possible from the samples of other class. Least squares twin SVM (LSTSVM) [11] reduced the computation cost of TSVM by solving a pair of linear system of equations instead of QPPs. Peng [12] introduced  $\nu$ twin SVM to improve the sparsity of the TSVM model while as Wang et al. [13], Yan et al. [14] introduced L<sub>1</sub>-norm based TSVM to improve the robustness of the TSVM model. To overcome the effects of outliers and noise, robust L<sub>2,1</sub>-norm enhanced multiweight vector projection SVM [15] and weighted structural twin

E-mail addresses: phd1901141006@iiti.ac.in (M.A. Ganaie),

<sup>&</sup>lt;sup>1</sup> 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.

SVM by local and global information [16] have been proposed. To further improve the generalization performance of the SVM based models, multi-view learning is exploited [17,18]. For multiclass problems, weighted linear loss multiple birth SVM [19] has been proposed.

Weston et al. [20] introduced the concept of universum in SVM, known as USVM, which incorporates prior data distribution information for generating the classifiers. However, USVM is computationally expensive, hence, Qi et al. [21] formulated universum twin SVM (UTSVM) which is an efficient model as it solves QPPs of smaller size compared to USVM. Both USVM and UTSVM require external toolbox to solve the QPPs, hence, Xu et al. [22], Richhariya and Tanveer [23] proposed least squares twin SVM with universum data (ULSTSVM) which solves a system of linear equations for generating the classifiers. Sinz et al. [24] analysed the effect of universum and concluded that selection of universum data is problem specific. Mostly, random averaging approach is followed for the generation of universum data [21,24]. USVM and UTSVM models incorporate universum data to improve the generalization performance of the models, however, it simultaneously increases the computation cost [3].

In addition to noise and outliers, the classification models also suffer in class imbalance problems. Here, the samples of one class (majority class) outnumber the samples of the other class (minority class). The imbalance ratio is the ratio of number of samples of majority class to the samples of minority class. For classification problems with high imbalance ratio of the samples. SVM classification models are biased to the maiority class samples and hence, the samples of minority class are misclassified. In problems like diagnosis of diseases [3,25] the focus is on the classification of minority class samples. To deal with such problem, multiple approaches have been followed. For creating the balance of classes, fuzzy SVMs for class imbalance learning [26,27] assigned fuzzy membership weights, boosting SVM (BSVM) [28] used boosting algorithm, and entropy based fuzzy SVM (EFSVM) [29] used information entropy of data. Also, synthetic minority oversampling technique (SMOTE) [30], random under-sampling approach [31,32], have been proposed to balance the classes. SVMs use kernel trick to handle the non-linear data. However, the aforementioned models generate the weights or synthesize the samples in the feature space instead of kernel space. To get the weights in kernel space, weighted kernel based SMOTE SVM (WKS-SVM) [33] generated the weights in kernel space which improved the generalization performance. Minimum variance embedded weighted kernel extreme learning machine (MVWKELM) and minimum varianceembedded class-specific kernelized ELM (MVCSKELM) [34] exploits the variance of the data for the class imbalance problems. SVM has also been designed for highly imbalanced classification problems [35]. Robust fuzzy least squares TSVM for class imbalance learning (RFLSTSVM-CIL) [36] used imbalance ratio of the samples in generating the fuzzy weights which resulted in improved generalization performance. Tanveer et al. [37], Ganaie and Tanveer [38] proposed general TSVM (Pin-GTSVM) with pinball loss function to handle the noisy data and large scale pinball loss TSVM [39] for large scale datasets. Margin maximization in twin spheres SVM (MMTSSVM) for class imbalance data [40] generated twin hyperspheres instead of hyperplanes. To minimize the effect of noise, margin maximization and volume minimization in hyper-spheres machine with pinball loss [40] and K-nearest neighbour (KNN)-based margin maximization and volume minimization based hyper-sphere machine [41] for class imbalance problems have been proposed. Twin SVM models have also been successfully used in regression problems [42-44].

Reduced support vector machine [45] reduced the kernel matrix by randomly choosing the subset of training data to limit the number of support vectors. For regression problems, reduced twin support vector regression [46] has been proposed. Reduced kernels retain the maximum information [47] and hence, give competitive performance.

To enhance the generalization ability of a classifier and make it robust, similarity or local neighbourhood information is exploited. Half of the information about the data lies in the local neighbourhood and is exploited using KNN [48]. Weighted TSVM uses local similarity information [49] for better generalization. KNN based weighted rough  $\nu$ -twin support vector machine [50], KNN based weighted TSVM [51,52], KNN based structural TSVM [53] and KNN weighted multiclass least squares TSVM [54] exploited the local neighbourhood information by assigning different weights to the samples of each class via KNN method. Ensemble learning [55] approach has been used in TSVM based models [56–58]. For detailed review of TSVM, readers are referred to [59].

The performance of the universum based learning algorithms is better due to the prior information [20,24]. However, incorporation of the universum data leads to increase in computational time. Recently, reduced universum TSVM for class imbalance learning (RUTSVM-CIL) [60] used universum data efficiently to balance the samples of the different class. Here, universum data is incorporated in a manner that it give prior information about the data without increasing the computational cost of the model. Moreover, reduced kernel is used to further reduce the complexity of the models. TSVM, UTSVM and RUTSVM-CIL ignore the local neighbourhood information which leads to the problem that each sample is effecting the separating hyperplane equally, however, each sample effects the separating hyperplane differently. TSVM is not able to use the prior information, however, UTSVM can incorporate the universum data which improve the generalization performance of the model at the cost of complexity. TSVM and UTSVM models may suffer in class imbalance problems due to over representation of the majority class and under representation of the minority class. Also, TSVM, UTSVM and RUTSVM-CIL implements the empirical risk minimization principle which results in the issues of overfitting. Moreover, matrices appearing in the dual formulation of TSVM, UTSVM and RUTSVM-CIL are positive semidefinite. Hence, to overcome the aforementioned issues of TSVM, UTSVM and RUTSVM-CIL models, we formulate KNNbased weighted reduced universum TSVM for class imbalance learning (KWRUTSVM-CIL).

The major highlights of this work are

- To incorporate the local neighbourhood information, *K* nearest neighbour-based weights are used in the proposed KWRUTSVM-CIL.
- Unlike RUTSVM-CIL, UTSVM, TSVM and RFLSTSVM-CIL models which implement empirical risk minimization principle, the proposed KWRUTSVM-CIL model implements the structural risk minimization principle, hence, avoid the issues of overfitting.
- Similar to RUTSVM-CIL, the proposed KWRUTSVM-CIL model incorporates prior information about the data (universum data) to steer the class imbalance problem.
- Wolfe dual of the proposed KWRUTSVM-CIL involves positive definite matrices, while as the matrices in the Wolfe dual of RUTSVM-CIL, UTSVM, TSVM and RFLSTSVM-CIL are positive semidefinite.
- As an application, we use the proposed KWRUTSVM-CIL model for the diagnosis of Alzheimer's disease and breast cancer disease.

The rest of this paper is organized as follows: Section 2 discuss the existing work, Section 3 gives the formulation of the proposed model, its computational complexity and its advantages over

existing models. Section 4 discusses the experimental results. Finally, the conclusions and potential future research directions are given in Section 6.

#### 2. Related work

Suppose the data samples of minority class +1 are denoted by matrix  $A \in \mathbb{R}^{m_1 \times n}$  and the class -1's data points are the samples of majority class given by matrix  $B \in \mathbb{R}^{m_2 \times n}$ . Also, universum data samples are given by U. Here,  $m_1$  and  $m_2$  are the number of samples in positive and negative class, respectively and nrepresents the dimensions of each sample.

#### 2.1. Class imbalance learning

Class imbalance problems pose a real challenge to the machine learning algorithms [61]. Here, the number of samples of one class outnumber the samples of the other class, hence the classification models get biased towards the samples of majority class. Class imbalance problems are predominant in medical domain [62], software defect detection [63], computer vision [64], oil spill detection [65] and breast cancer classification [66]. The conventional classification algorithms get biased towards the dominant class of samples and the samples of other class, due to less representation, get misclassified. Particularly in medical diagnosis such as Alzheimer's disease diagnosis [67] the samples of minority class get misclassified. Hence, the tailored methods for class imbalance learning are drawing attention from the machine learning [68.69].

Multiple techniques have been developed for the class imbalance learning [61,70,71]. Class imbalance learning models are broadly classified as the data level methods, algorithmic level methods and the hybrid approaches [61].

#### 2.1.1. Data level methods

To minimize the imbalance among the samples of different classes, data level methods such as undersampling and oversampling [31,61] have been utilized. In undersampling methods, fraction of samples from the majority class are dropped to get the balance across the samples of different classes. However, these methods such as EasyEnsemble and BalanceCascade [31] leads to loss of information. Oversampling method replicates the samples of minority class, however, these models pose the risk of overfitting [72]. To overcome this risk, synthetic minority oversampling technique (SMOTE) [30] generates the samples of the minority class.

#### 2.1.2. Algorithm level methods

Algorithm level methods [35] adapt the structure of the algorithm directly to handle the class imbalance problems. Another category of algorithm level methods is the cost sensitive methods [73] which give more penalty for the misclassification of the minority class such as weighted SVM [74], weighted Lagrangian twin SVM [75], fuzzy weighted twin SVM [76] and enhanced twin SVM [77].

#### 2.1.3. Hybrid approaches

Hybrid approaches like EasyEnsemble and BalanceCascade [31] combine the benefits of the data level methods and the algorithm level methods for handling the imbalance problems. In EasyEnsemble, random undersampling approach is followed to generate multiple balanced data subsets from the majority class, which are used to train the learners and then combine their outputs. The BalancedCascade learns the models sequentially wherein the samples of majority class being correctly classified are removed from further consideration.

#### 2.2. K Nearest neighbour (KNN) weight generation

KNN method [48] gives the local information of positive and negative classes. For each data point of +1 class, define

$$NH_s(x_k) = \{x_k^i, \text{ if both } x_k^i \text{ and } x_k \text{ are in same class }, 0 \le i \le m_1\},$$
(1)

 $NH_d(x_k) = \{x_k^i, \text{ if } x_k^i \text{ and } x_k \text{ are in different class, } 0 \le i \le m_2\},\$ (2)

where  $NH_s(x_k)$  and  $NH_d(x_k)$  represent the  $m_1$  and  $m_2$  neighbours of  $x_k$  from +1 and -1 classes, respectively. Define the adjacency matrix of class +1 [52], as follows

$$M_{s,ij} = \begin{cases} 1, \text{ if } x_j \in NH_s(x_i) \text{ or } x_i \in NH_s(x_j) \\ 0, \text{ otherwise} \end{cases}$$
(3)

and

$$M_{d,ij} = \begin{cases} 1, \text{ if } x_j \in NH_d(x_i) \text{ or } x_i \in NH_d(x_j) \\ 0, \text{ otherwise.} \end{cases}$$
(4)

When  $M_{s,ij} = 1$  or  $M_{d,ij} = 1$ , undirected edge between the two samples is added in the corresponding graph. For reducing the sample points,  $M_{d,ij}$  is redefined as

$$f_j = \begin{cases} 1 & \text{if } \exists i, M_{d,ij} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$
(5)

2.3. TSVM

TSVM [10] generates two nonparallel hyperplanes for each class of the binary classification problem. Each plane is proximal to one class of data points and as far as possible from the data points of other class.

The separating hyperplanes of TSVM for non-linear case are:

$$f_1(x) = K(x^t, C^t)u_1 + b_1$$
 and  $f_2(x) = K(x^t, C^t)u_2 + b_2$ , (6)

where C = [A; B] and  $K(\cdot)$  is the kernel function which is usually taken as the Gaussian kernel.

The optimization problems of TSVM are:

$$\min_{u_1,b_1,\xi_1} \frac{1}{2} \| K(A, C^t) u_1 + e_1 b_1 \|^2 + c_1 e_2^t \xi_1 
s.t. - (K(B, C^t) u_1 + e_2 b_1) + \xi_1 \ge e_2, \quad \xi_1 \ge 0$$
(7)

and

$$\min_{\substack{u_2, b_2, \xi_2 \\ s.t. \ K(A, C^t)u_2 + e_2b_2 \|^2 + e_2e_1^t \xi_2} \frac{1}{2} \|K(B, C^t)u_2 + e_2b_2\|^2 + e_2e_1^t \xi_2 \le 0,$$
(8)

where  $c_i$  are penalty parameters,  $e_i$  denote vectors of ones with appropriate dimensions and  $\xi_i$  are slack vectors, for i = 1, 2.

Using the necessary and sufficient Karush-Kuhn-Tucker (K.K.T.) conditions, the Wolfe dual of (7) and (8) are given as follows:

$$\max_{\alpha_1} e_2^t \alpha_1 - \frac{1}{2} \alpha_1^t G(H^t H)^{-1} G^t \alpha_1$$
  
s.t.  $0 \le \alpha_1 \le c_1$  (9)  
and

and

$$\max_{\alpha_2} \quad e_1^t \alpha_2 - \frac{1}{2} \alpha_2^t H(G^t G)^{-1} H^t \alpha_2$$
  
s.t.  $0 \le \alpha_2 \le c_2$ , (10)

where  $\alpha_1$  and  $\alpha_2$  are the Lagrangian multipliers, *H*  $\begin{bmatrix} K(A, C^t) & e_1 \end{bmatrix}$  and  $G = \begin{bmatrix} K(B, C^t) & e_2 \end{bmatrix}$ .

Once we solve (9) and (10), the separating hyperplanes are given by

$$\begin{bmatrix} u_1\\b_1 \end{bmatrix} = -\left(H^t H\right)^{-1} G^t \alpha_1,\tag{11}$$

$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = (G^t G)^{-1} H^t \alpha_2. \tag{12}$$

As the matrices  $G^t G$  or  $H^t H$  may be ill conditioned, hence the matrices  $(G^T G)^{-1}$  and  $(H^T H)^{-1}$  are added small positive number  $\delta$  along the diagonal i.e.  $(G^T G + \delta I)^{-1}$  and  $(H^T H + \delta I)^{-1}$ , here *I* is an identity matrix of appropriate dimensions.

#### 2.4. UTSVM

UTSVM [21] incorporated the universum data into TSVM formulation. Similar to TSVM, UTSVM solves two smaller sized QPPs to generate the hyperplanes. The optimization problems of nonlinear UTSVM are given as follows

$$\min_{\substack{u_1,b_1,\xi_1,\eta_1\\u_1,b_1,\xi_1,\eta_1}} \frac{1}{2} \|K(A, C^t)u_1 + e_1b_1\|^2 + c_1e_2^t\xi_1 + c_ue_u^t\eta_1 \\
s.t. - (K(B, C^t)u_1 + e_2b_1) + \xi_1 \ge e_2, \quad \xi_1 \ge 0 \\
(K(U, C^t)u_1 + e_ub_1) + \eta_1 \ge (-1 + \varepsilon)e_u, \quad \eta_1 \ge 0 \quad (13) \\
and \\
\min_{\substack{u_2,b_2,\xi_2,\eta_2\\u_2,b_2,\xi_2,\eta_2}} \frac{1}{2} \|K(B, C^t)u_2 + e_2b_2\|^2 + c_2e_1^t\xi_2 + c_ue_u^t\eta_2$$

$$\min_{u_2, b_2, \xi_2, \eta_2} \frac{1}{2} \| K(B, C^t) u_2 + e_2 b_2 \|^2 + c_2 e_1^t \xi_2 + c_u e_u^t \eta_2$$
s.t.  $(K(A, C^t) u_2 + e_1 b_2) + \xi_2 \ge e_1, \quad \xi_2 \ge 0$   
 $- (K(U, C^t) u_2 + e_u b_2) + \eta_2 \ge (-1 + \varepsilon) e_u, \quad \eta_2 \ge 0,$ 
(14)

where  $c_i$ ,  $c_u$  are penalty parameters,  $e_i$ ,  $e_u$  are vectors of ones with appropriate dimensions,  $\xi_i$ ,  $\eta_i$  are slack variables, for i = 1, 2 and C = [A; B].

Using the necessary and sufficient K.K.T. conditions, the Wolfe dual of (13) and (14) are given as follows:

$$\max_{\substack{\alpha_{1},\beta_{1} \\ \alpha_{1},\beta_{1}}} e_{2}^{t} \alpha_{1} - \frac{1}{2} (\alpha_{1}^{t} G - \beta_{1}^{t} O) (H^{t} H)^{-1} (G^{t} \alpha_{1} - O^{t} \beta_{1}) + (\epsilon - 1) e_{u}^{t} \beta_{1}$$
s.t.  $0 \le \alpha_{1} \le c_{1},$   
 $0 \le \beta_{1} \le c_{u}$  (15)

and

$$\max_{\substack{\alpha_2,\beta_2}} e_2^t \alpha_2 - \frac{1}{2} (\alpha_2^t H - \beta_2^t O) (G^t G)^{-1} (H^t \alpha_2 - O^t \beta_2) + (\epsilon - 1) e_u^t \beta_2$$
  
s.t.  $0 \le \alpha_2 \le c_2$ ,  
 $0 \le \beta_2 \le c_u$ , (16)

where  $\alpha_i$  and  $\beta_i$  are Lagrangian multipliers, for  $i = 1, 2, H = \begin{bmatrix} K(A, C^t) & e_1 \end{bmatrix}, G = \begin{bmatrix} K(B, C^t) & e_2 \end{bmatrix}$ , and  $O = \begin{bmatrix} K(U, C^t) & e_u \end{bmatrix}$ .

Once we solve (15) and (16), the separating hyperplanes are given by

$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(H^t H)^{-1} (G^t \alpha_1 - O^t \beta_1),$$
(17)  
$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = (G^t G)^{-1} (H^t \alpha_2 - O^t \beta_2).$$
(18)

As the matrices  $G^t G$  or  $H^t H$  may be ill conditioned, hence the matrices  $(G^T G)^{-1}$  and  $(H^T H)^{-1}$  are added small positive number  $\delta$  along the diagonal i.e.  $(G^T G + \delta I)^{-1}$  and  $(H^T H + \delta I)^{-1}$ , here *I* is appropriate dimension identity matrix.

The new data point  $x \in R^n$  is given class label as

$$class(x) = \arg\min_{i=1,2} \frac{|K(x^{t}, C^{t})u_{i} + b_{i}|}{\|u_{i}\|}.$$
(19)

#### 2.5. RUTSVM-CIL

Suppose  $B^* \in R^{r \times n}$  is a randomly chosen decreased data sample matrix of the majority class, here  $r = m_1$ . The universum data matrix is  $U \in R^{d \times n}$ , where  $d = m_2 - m_1$ . Also,  $U^* \in R^{g \times n}$  be random subset of U, where  $g = \lceil \frac{r}{2} \rceil$ .

The optimization problem of nonlinear RUTSVM-CIL [71] are:

$$\min_{\substack{u_1,b_1,\xi_1,\psi_1}} \frac{1}{2} \| K(A, C^t) u_1 + e_1 b_1 \|^2 + c_1 e_1^t \xi_1 + c_u e_g^t \psi_1 
s.t. - (K(B^*, C^t) u_1 + e_1 b_1) + \xi_1 \ge e_1, 
(K(U^*, C^t) u_1 + e_g b_1) + \psi_1 \ge (-1 + \varepsilon) e_g, 
\xi_1, \psi_1 \ge 0,$$
(20)

and

$$\min_{u_{2},b_{2},\xi_{2},\psi_{2}} \frac{1}{2} \|K(B,C^{t})u_{2} + e_{2}b_{2}\|^{2} + c_{2}e_{1}^{t}\xi_{2} + c_{u}e_{r}^{t}\psi_{2}$$
s.t.  $(K(A,C^{t})u_{2} + e_{1}b_{2}) + \xi_{2} \ge e_{1},$  (21)  
 $(K(U,C^{t})u_{2} + e_{r}b_{2}) + \psi_{2} \ge (1-\varepsilon)e_{r},$   
 $\xi_{2},\psi_{2} \ge 0,$ 

here  $e_1$ ,  $e_g$  and  $e_r$  are vector of ones of appropriate dimensions. Using the necessary and sufficient K.K.T. conditions, the Wolfe dual of (20) and (21) are given as follows:

$$\max_{\alpha_{1},\mu_{1}} e_{1}^{t} \alpha_{1} - \frac{1}{2} (\alpha_{1}^{t} P^{*} - \mu_{1}^{t} O^{*}) (S^{t} S)^{-1} (P^{*t} \alpha_{1} - O^{*t} \mu_{1}) + (\epsilon - 1) e_{g}^{t} \mu_{1} s.t. \ 0 \le \alpha_{1} \le c_{1}, 0 \le \mu_{1} \le c_{u}$$
(22)

and

$$\max_{\substack{\alpha_{2}, \mu_{2} \\ s.t. \ 0 \le \alpha_{2} \le c_{2}, \\ 0 \le \mu_{2} \le c_{u}, } e_{1}^{t} \alpha_{2} - \frac{1}{2} (\alpha_{2}^{t} S + \mu_{2}^{t} 0) (P^{t} P)^{-1} (S^{t} \alpha_{2} + O^{t} \mu_{2}) + (1 - \epsilon) e_{r}^{t} \mu_{2}$$
s.t.  $0 \le \alpha_{2} \le c_{2},$ 

$$(23)$$

where  $S = [K(A, C^t), e_1]$ ,  $P^* = [K(B^*, C^t), e_1]$ ,  $P = [K(B, C^t), e_2]$ ,  $O^* = [K(U^*, C^t), e_g]$ ,  $O = [K(U, C^t), e_r]$ ,  $C = [A; B^*]$  is the reduced kernel and  $\alpha_i$ ,  $\mu_i$  are the Lagrange multipliers, with i = 1, 2.

Once we solve (22) and (23), the optimal hyperplanes are given as:

$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(S^t S + \delta I)^{-1} (P^{*t} \alpha_1 - O^{*t} \mu_1),$$
(24)

$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = (P^t P + \delta I)^{-1} (S^t \alpha_2 + O^t \mu_2),$$
(25)

where  $\delta$  is a small positive number.

The new data point  $x \in R^n$  is given class label as

$$class(x) = \arg\min_{i=1,2} \frac{|K(x^{t}, C^{t})u_{i} + b_{i}|}{\|u_{i}\|}.$$
(26)

# **3.** Proposed KNN weighted reduced universum twin SVM for CIL (KWRUTSVM-CIL)

Motivated by [21,52,71], we propose KNN weighted reduced universum twin SVM for CIL (KWRUTSVM-CIL). The proposed KWRUTSVM-CIL considers the local neighbourhood information and uses universum data to balance the classes in class imbalance problems. Local neighbourhood information is incorporated via weight matrix. The weight matrix in the objective function of the proposed KWRUTSVM-CIL exploits the intraclass information and weight vector in the constraints of the corresponding objective function exploits the interclass information. The oversampling and undersampling approaches are followed to balance the samples in class imbalance problems. Universum data gives prior information of the data. Unlike TSVM, universum TSVM, and reduced universum TSVM for class imbalance, the proposed KWRUTSVM-CIL model incorporates extra regularization term for margin maximization and to embody the structural risk minimization principle which is the marrow of statistical learning. Also, the matrices appearing in the Wolfe dual of proposed KWRUTSVM-CIL formulation are positive definite.

#### 3.1. Linear KWRUTSVM-CIL

The optimization problem of the proposed KWRUTSVM-CIL are

$$\min_{u_1,b_1,\xi_1,\psi_1} \frac{c_3}{2} (\|u_1\|^2 + b_1^2) + \frac{1}{2} \|D_1(Au_1 + e_1b_1)\|^2 
+ c_1 e_1^t \xi_1 + c_u e_g^t \psi_1 
s.t. - F_1(B^*u_1 + e_1b_1) + \xi_1 \ge F_{v_1}, \quad \xi_1 \ge 0 
H_1(U^*u_1 + e_gb_1) + \psi_1 \ge (-1 + \varepsilon)H_g, \quad \psi_1 \ge 0$$
(27)

and

$$\min_{u_{2},b_{2},\xi_{2},\psi_{2}} \frac{c_{4}}{2} (\|u_{2}\|^{2} + b_{2}^{2}) + \frac{1}{2} \|D_{2}(Bu_{2} + e_{2}b_{2})\|^{2} 
+ c_{2}e_{1}^{t}\xi_{2} + c_{u}e_{d}^{t}\psi_{2} 
s.t. F_{2}(Au_{2} + e_{1}b_{2}) + \xi_{2} \ge F_{v_{2}}, \quad \xi_{2} \ge 0 
H_{2}(Uu_{2} + e_{d}b_{2}) + \psi_{2} \ge (1 - \varepsilon)H_{v}, \quad \psi_{2} \ge 0,$$
(28)

where  $\xi_j$ ,  $\psi_j$  are the slack variables,  $c_i$  and  $c_u$  are the penalty parameters, for j = 1, 2 and i = 1, 2, 3, 4.  $e_1, e_2, e_g$  and  $e_d$ are the vector of ones with appropriate dimensions. Also,  $D_1 = diag(d_1, d_2, \ldots, d_{m_1})$  and  $d_i = \sum_{i=1}^{m_1} M_{s,ij}$ ,  $F_1 = diag(f_1, f_2, \ldots, f_{m_1})$ ,  $H_1 = diag(h_1, h_2, \ldots, h_{m_g})$ ,  $F_{v_1}$  is a vector of diagonal elements of  $F_1$ , and similarly,  $H_g$  and  $H_v$  are defined. In the same manner,  $D_2, F_2, H_2$  and their corresponding vectors  $F_{v_2}, H_v$  are defined. Here,  $d = min(m_1, m_2)$ .

Taking the Lagrangian of (27), we have

$$L = \frac{c_3}{2} (\|u_1\|^2 + b_1^2) + \frac{1}{2} \|D_1(Au_1 + e_1b_1)\|^2 + c_1 e_1^t \xi_1 + c_u e_g^t \psi_1 - \alpha^t (-F_1(B^*u_1 + e_1b_1) + \xi_1 - F_{v_1}) - \beta^t (H_1(U^*u_1 + e_gb_1) + \psi_1 - (-1 + \varepsilon)H_g) - \lambda^t \xi_1 - \sigma^t \psi_1,$$
(29)

where  $\alpha$ ,  $\beta$ ,  $\lambda$  and  $\sigma$  are the Lagrange multipliers.

Using the necessary and sufficient K.K.T. conditions for (29), we have

$$c_3 u_1 + (D_1 A)^t (D_1 (A u_1 + e_1 b_1)) + B^{*t} F_1^t \alpha - U^{*t} H_1^t \beta = 0,$$
(30)

$$c_{3}b_{1} + (D_{1}e_{1})^{t}(D_{1}(Au_{1} + e_{1}b_{1})) + e_{1}^{t}F_{1}^{t}\alpha - e_{g}^{t}H_{1}^{t}\beta = 0, \qquad (31)$$

$$c_1 e_1 - \alpha - \lambda = 0, \qquad (32)$$

$$c_u e_g - \beta - \sigma = 0, \qquad (33)$$

$$\alpha^{t}(-F_{1}(B^{*}u_{1}+e_{1}b_{1})+\xi_{1}-F_{v_{1}})=0, \qquad (34)$$

$$\beta^{t}(H_{1}(U^{*}u_{1} + e_{g}b_{1}) + \psi_{1} - (-1 + \varepsilon)H_{g}) = 0, \qquad (35)$$

$$\lambda^t \xi_1 = 0, \qquad (36)$$

$$\sigma^t \psi_1 = 0. \tag{37}$$

Rewriting (30) and (31), we have

$$c_{3}\begin{bmatrix}u_{1}\\b_{1}\end{bmatrix} + \begin{bmatrix}(D_{1}A)^{t}\\(D_{1}e_{1})^{t}\end{bmatrix} \left(D_{1}\begin{bmatrix}A & e_{1}\end{bmatrix}\begin{bmatrix}u_{1}\\b_{1}\end{bmatrix}\right) + \begin{bmatrix}(F_{1}B^{*})^{t}\\(F_{1}e_{1})^{t}\end{bmatrix}\alpha$$
$$- \begin{bmatrix}(H_{1}U^{*})^{t}\\(H_{1}e_{g})^{t}\end{bmatrix}\beta = 0.$$
(38)

Let  $S = D_1 \begin{bmatrix} A & e_1 \end{bmatrix}$ ,  $T^* = F_1 \begin{bmatrix} B^* & e_1 \end{bmatrix}$  and  $O^* = H_1 \begin{bmatrix} U^* & e_g \end{bmatrix}$ , then we can write (38) as

$$(c_{3}I + S^{t}S)\begin{bmatrix} u_{1} \\ b_{1} \end{bmatrix} + T^{*t}\alpha - O^{*t}\beta = 0,$$
(39)

or 
$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(c_3 I + S^t S)^{-1} (T^{*t} \alpha - O^{*t} \beta).$$
 (40)

Using (40) and the above K.K.T conditions, we have the Wolfe dual of  $\left( 27\right)$  as

$$\max_{\alpha,\beta} F_{v_1}^t \alpha - \frac{1}{2} (\alpha^t T^* - \beta^t O^*) (c_3 I + S^t S)^{-1} (T^{*t} \alpha - O^{*t} \beta) + (\varepsilon - 1) H_g^t \beta s.t. \quad 0 \le \alpha \le c_1, \quad 0 \le \beta \le c_u.$$
(41)

Following the similar procedure, the Wolfe dual of (28) is given as

$$\max_{\eta,\theta} F_{v_2}^t \eta - \frac{1}{2} (\eta^t S + \theta^t O) (c_4 I + T^t T)^{-1} (S^t \eta + O^t \theta) + (1 - \varepsilon) H_d^t \beta s.t. \quad 0 \le \eta \le c_2, \quad 0 \le \theta \le c_u,$$
(42)

where  $T = F_2[A e_1]$  and  $O = H_2[U e_d]$ .

Once we solve (41) and (42), the separating hyperplanes are obtained as follows

$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(S^t S + c_3 I)^{-1} (T^{*t} \alpha - O^{*t} \beta),$$
(43)

$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = (T^t T + c_4 I)^{-1} (S^t \eta + O^t \theta).$$
(44)

The class label of a new data sample  $x \in \mathbb{R}^n$  is assigned as follows:

$$class(x) = \arg\min_{i=1,2} \frac{|x^{t}u_{i} + b_{i}|}{\|u_{i}\|}.$$
(45)

#### 3.2. Non-linear KWRUTSVM-CIL

The optimization problem of the proposed KWRUTSVM-CIL for non-linear case are

$$\min_{u_{1},b_{1},\xi_{1},\psi_{1}} \frac{c_{3}}{2} (\|u_{1}\|^{2} + b_{1}^{2}) + \frac{1}{2} \|D_{1}(K(A, C^{t})u_{1} + e_{1}b_{1})\|^{2} 
+ c_{1}e_{1}^{t}\xi_{1} + c_{u}e_{g}^{t}\psi_{1} 
s.t. - F_{1}(K(B^{*}, C^{t})u_{1} + e_{1}b_{1}) + \xi_{1} \ge F_{v_{1}}, \quad \xi_{1} \ge 0 
H_{1}(K(U^{*}, C^{t})u_{1} + e_{g}b_{1}) + \psi_{1} \ge (-1 + \varepsilon)H_{g}, \quad \psi_{1} \ge 0$$
(46)

and

$$\min_{u_{2},b_{2},\xi_{2},\psi_{2}} \frac{c_{4}}{2} (\|u_{2}\|^{2} + b_{2}^{2}) + \frac{1}{2} \|D_{2}(K(B,C^{t})u_{2} + e_{2}b_{2})\|^{2} 
+ c_{2}e_{1}^{t}\xi_{2} + c_{u}e_{d}^{t}\psi_{2} 
s.t. F_{2}(K(A,C^{t})u_{2} + e_{1}b_{2}) + \xi_{2} \ge F_{v_{2}}, \quad \xi_{2} \ge 0 
H_{2}(K(U,C^{t})u_{2} + e_{d}b_{2}) + \psi_{2} \ge (1 - \varepsilon)H_{v}, \quad \psi_{2} \ge 0,$$
(47)

where  $C = [A; B^*]$  and  $K(\cdot)$  is the reduced kernel.

Following the similar procedure as in the linear case, the Wolfe duals of (46) and (47) are

$$\max_{\alpha,\beta} F_{v_1}^t \alpha - \frac{1}{2} (\alpha^t T^* - \beta^t O^*) (c_3 I + S^t S)^{-1} (T^{*t} \alpha - O^{*t} \beta)$$

Dataset details.		
Datasets	Samples ×Features	IR
abalone9-18	731 × 8	16.4048
aus	690 × 15	1.2476
brwisconsin	683 × 10	1.8577
bupa_or_liver-disorders	$345 \times 7$	1.3793
checkerboard_Data	690 × 15	1.2476
cleve	297 × 14	1.1679
cmc	$1473 \times 10$	1.3418
ecoli-0-1_vs_2-3-5	$244 \times 8$	9.1667
ecoli-0-1_vs_5	$240 \times 7$	11
ecoli-0-1-4-6_vs_5	$280 \times 7$	13
ecoli-0-1-4-7_vs_2-3-5-6	336 × 8	10.5862
ecoli-0-1-4-7_vs_5-6	332 × 7	12.28
ecoli-0-2-3-4_vs_5	202 × 8	9.1
ecoli-0-3-4-6_vs_5	205 × 8	9.25
ecoli-0-3-4-7_vs_5-6	257 × 8	9.28
ecoli-0-4-6_vs_5	$203 \times 7$	9.15
ecoli-0-6-7_vs_3-5	222 × 8	9.0909
ecoli-0-6-7_vs_5	220 × 7	10
ecoli0137vs26	311 × 8	4.7593
ecoli01vs5	240 × 8	11
ecoli2	336 × 8	8.6
ecoli3	336 × 8	8.6
ecoli4	336 × 8	15.8
glass4	$214 \times 10$	15.4615
heart-stat	$270 \times 14$	1.25
iono	351 × 34	1.7857
new-thyroid1	$215 \times 6$	5.1429
pima	768 × 9	1.8657
vehicle1	846 × 19	2.8986
vehicle2	846 × 19	2.8807
votes	435 × 17	1.5893
vowel	988 × 11	9.9778
wpbc	$194 \times 34$	3.2174
yeast-0-2-5-6_vs_3-7-8-9	$1004 \times 9$	9.1414
yeast-0-2-5-7-9_vs_3-6-8	$1004 \times 9$	9.1414
yeast-0-5-6-7-9_vs_4	528 × 9	9.3529
yeast-2_vs_4	514 × 9	9.0784
yeast2vs8	483 × 9	23.15
yeast3	$1484 \times 9$	8.1043
yeast5	1484 × 9	32.7273

$$+ (\varepsilon - 1)H_{g}^{t}\beta$$
  
s.t.  $0 \le \alpha \le c_{1}, \ 0 \le \beta \le c_{u}$  (48)

and

$$\max_{\eta,\theta} F_{v_2}^t \eta - \frac{1}{2} (\eta^t S + \theta^t O) (c_4 I + T^t T)^{-1} (S^t \eta + O^t \theta) + (1 - \varepsilon) H_d^t \beta$$
  
s.t.  $0 \le \eta \le c_2, \ 0 \le \theta \le c_u,$  (49)

where  $S = D_1 [K(A, C^t) e_1]$ ,  $T^* = F_1 [K(B^*, C^t) e_1]$ ,  $O^* = H_1 [K(U^*, C^t) e_g]$ ,  $T = F_2 [K(A, C^t) e_1]$  and  $O = H_2 [K(U, C^t) e_d]$ .

Once we solve (48) and (49), the separating hyperplanes are obtained as follows:

$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -(S^t S + c_3 I)^{-1} (T^{*t} \alpha - O^{*t} \beta),$$
(50)

$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = (T^t T + c_4 I)^{-1} (S^t \eta + O^t \theta).$$
(51)

The class label of a new data sample  $x \in \mathbb{R}^n$  is assigned as follows:

$$class(x) = \arg\min_{i=1,2} \frac{|K(x^{t}, C^{t})u_{i} + b_{i}|}{\|u_{i}\|}.$$
(52)

#### 3.3. Computational complexity

Consider a binary classification problem with *m* number of samples such that  $m_1$  samples belong to the positive class and  $m_2$  samples belong to the negative class ( $m = m_1 + m_2$ ). Assume that we have the class imbalance problem with imbalance ratio,  $IR = \frac{m_2}{m_1}$ . The time complexity of the TSVM is  $(IR^3 + 1)O(m_1)^3$  and RUTSVM-CIL is  $(IR^3 + 3.375)O(m_1)^3$  [71]. The computational complexity of proposed KWRUTSVM-CIL involves the assignment of weights to the samples via KNN graph and solving the QPPs. Thus, main complexity of the model involves the following two steps:

- 1. Weight initialization via KNN graph: Using  $O(m^2 log(m))$  the weights of the weight matrix D with m points can be computed.
- 2. The complexity of optimizing the QPPs of the proposed formulations is same as that of the RUTSVM-CIL model.

Hence, the complexity of the proposed model is relatively increased compared to RUTSVM-CIL model due to the incorporation of the local neighbourhood information.

3.4. Analysis of proposed KWRUTSVM-CIL with respect to TSVM, UTSVM and RUTSVM-CIL models

The advantages of the proposed KWRUTSVM-CIL with respect to TSVM, UTSVM and RUTSVM-CIL models are given as follows

- TSVM, UTSVM and RUTSVM-CIL minimize empirical risk which may result in the issues of overfitting. However, the proposed KWRUTSVM-CIL model minimizes structural risk which overcomes the issues of overfitting. Also, TSVM, UTSVM and RUTSVM-CIL assume that data matrices appearing in the formulation are positive definite. However, in real world scenarios, such assumption may not be valid which results in the issues of singularity. The matrices in the proposed KWRUTSVM-CIL are positive definite and hence, theoretically more stable.
- TSVM, UTSVM and RUTSVM-CIL assume that all the samples are equally important for generating hyperplanes. However, such an assumption may not hold in real world scenarios. Thus, the proposed KWRUTSVM-CIL model exploit neighbourhood information via KNN to give the appropriate weights to the samples which improves the generalization performance.

#### 4. Experimental results

In classification problems, class imbalance problem poses a challenge to the classification algorithms. Such kind of problems have irregularity in the probability distribution of samples corresponding to different classes. These kind of datasets are termed as imbalance datasets [61]. The classification algorithms suffer in problems dominated by the samples of a particular class. The ratio of number of majority class samples to the number of minority class samples is known as imbalance ratio (IR). Mathematically,

$$IR = \frac{\text{Count of majority class data points}}{\text{Count of minority class data points}}.$$
(53)

We analyse the classification performance of TSVM [10], UTSVM [21], RUTSVM-CIL [71], MVKWELM [34], MVCSKELM [34], KW-SMOTE-SVM [33], RFLSTSVM-CIL [36] and proposed KWRUTSVM-CIL model. All the models are evaluated on datasets from UCI machine learning repository [78] and KEEL repository [79] (see Table 1). We follow 5-fold cross validation and the average results of different classification models is given in

Classification performance of baseline models and proposed KWRUTSVM-CIL model on real world KEEL datasets with Gaussian kernel.

Datasets	TSVM [10]	UTSVM [21]	RUTSVM-CIL [71]	MVKWELM [34]	MVCSKELM [34]	KW-SMOTE- SVM	RFLSTSVM- CIL	KWRUTSVM- CIL
						[33]	[36]	
	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD
	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,
	G-Iviean	G-Mean	G-Iviean	G-Iviean	G-Iviean	G-Iviean	G-Iviean	G-Mean
abalone9–18	$0.79 \pm 0.03$	$0.82 \pm 0.06$	$0.88\pm0.06$	$0.66 \pm 0.07$	$0.76 \pm 0.06$	$0.87 \pm 0.06$	$0.85 \pm 0.03$	$0.89 \pm 0.04$
2110	0.5, 0.52	0.46, 0.5	0.51, 0.56	0.18, 0.27	0.4, 0.43	0.39, 0.48	0.42, 0.49	0.53, 0.58
dus	$0.87 \pm 0.04$	$0.87 \pm 0.02$	$0.88 \pm 0.00$	$0.87 \pm 0.03$	$0.87 \pm 0.02$	$0.88 \pm 0.04$	$0.87 \pm 0.04$	$0.89 \pm 0.05$
brwisconsin	0.80, 0.80 0.98 + 0.02	0.80, 0.80 0.98 + 0.02	0.87, 0.87 0.98 + 0.01	0.80, 0.80 0.97 $\pm$ 0.02	0.80, 0.80 0.98 + 0.01	0.87, 0.87 0.97 + 0.02	0.80, 0.80 0.98 + 0.01	0.88, 0.88 0.98 $\pm$ 0.01
biwisconsin	0.97.0.97	0.97, 0.97	0.97, 0.97	0.96.0.96	0.96, 0.96	0.96.0.96	0.97, 0.97	0.97, 0.97
bupa or liver-	$0.71 \pm 0.05$	$0.71 \pm 0.09$	$0.69 \pm 0.08$	$0.59 \pm 0.04$	$0.71 \pm 0.06$	$0.67 \pm 0.02$	$0.7 \pm 0.05$	$0.71 \pm 0.07$
disorders								
	0.67, 0.67	0.64, 0.65	0.67, 0.68	0.54, 0.54	0.66, 0.66	0.64, 0.64	0.65, 0.65	0.66, 0.66
checkerboard	$0.87\pm0.04$	$0.87\pm0.04$	$0.88 \pm 0.06$	$0.87\pm0.03$	$0.87 \pm 0.02$	$0.88\pm0.04$	$0.87\pm0.04$	$\textbf{0.89} \pm 0.05$
	0.86, 0.86	0.86, 0.86	0.87, 0.87	0.86, 0.86	0.86, 0.86	0.87, 0.87	0.86, 0.86	0.88, 0.88
cleve	$0.83 \pm 0.06$	$0.84 \pm 0.05$	$0.86 \pm 0.04$	$0.82 \pm 0.09$	$0.82 \pm 0.04$	$0.83 \pm 0.06$	$0.83 \pm 0.08$	$0.85 \pm 0.03$
cmc	0.82, 0.82	0.83, 0.83	0.85, 0.85	0.81, 0.81	0.8, 0.8	0.82, 0.82	0.81, 0.81	0.83, 0.84 07 $\pm 0.02$
CIIIC	$0.7 \pm 0.05$	$0.7 \pm 0.02$	$0.09 \pm 0.02$	$0.09 \pm 0.04$	$0.09 \pm 0.05$	$0.00 \pm 0.01$	$0.7 \pm 0.04$	$0.7 \pm 0.02$
ecoli-0-	0.03, 0.03 $0.9 \pm 0.04$	0.03, 0.03	0.03, 0.04 0.9 + 0.07	0.04, 0.04 $0.87 \pm 0.08$	0.00, 0.00 $0.89 \pm 0.05$	0.02, 0.02 $0.85 \pm 0.05$	0.04, 0.04 $0.92 \pm 0.05$	0.04, 0.05 $0.92 \pm 0.05$
1 vs 2-3-5				0107 1 0100			0102 1 0100	
	0.8, 0.81	0.88, 0.88	0.63, 0.66	0.61, 0.64	0.71, 0.72	0.7, 0.71	0.82, 0.83	0.8, 0.82
ecoli-0-1_vs_5	$0.92\pm0.07$	$\textbf{0.96} \pm 0.06$	$0.94\pm0.11$	$0.9\pm0.12$	$0.95\pm0.07$	$0.91\pm0.07$	$\textbf{0.96} \pm 0.06$	$\textbf{0.96} \pm 0.05$
	0.87, 0.88	0.84, 0.85	0.87, 0.88	0.61, 0.64	0.92, 0.92	0.78, 0.79	0.86, 0.86	0.86, 0.87
ecoli-0-1-4-	$0.92\pm0.07$	$0.92\pm0.07$	$0.92\pm0.12$	$0.91\pm0.07$	$0.94\pm0.08$	$0.87\pm0.09$	$0.93\pm0.07$	$\textbf{0.95} \pm 0.06$
6_vs_5		0.00.00	0.00.0.00	0.0.00	0.05.0.05	0.00.004	0.01.0.00	0.70.0.70
anali 0 1 1	0.89, 0.9	0.89, 0.9	0.89, 0.89	0.6, 0.63	0.85, 0.85	0.83, 0.84	0.91, 0.92	0.76, 0.78
2 vc 2 3 5 6	$0.91 \pm 0.08$	$0.89 \pm 0.02$	$0.91 \pm 0.06$	$0.80 \pm 0.12$	$0.91 \pm 0.08$	$0.84 \pm 0.09$	$0.92 \pm 0.07$	$0.91 \pm 0.05$
7_V3_2-5-5-0	0.83 0.83	074075	0.68.07	06.062	0.85 0.85	04305	07072	077078
ecoli-0-1-4-	$0.92 \pm 0.05$	0.94 + 0.05	$0.93 \pm 0.05$	$0.89 \pm 0.05$	$0.92 \pm 0.07$	$0.85 \pm 0.03$	$0.92 \pm 0.05$	$0.93 \pm 0.05$
7_vs_5-6								
	0.86, 0.86	0.75, 0.77	0.72, 0.74	0.64, 0.66	0.65, 0.68	0.4, 0.48	0.77, 0.78	0.73, 0.75
ecoli-0-2-3-	$0.92\pm0.11$	$0.95\pm0.06$	$0.92\pm0.12$	$0.9\pm0.08$	$0.94\pm0.11$	$0.91\pm0.07$	$0.95\pm0.06$	$\textbf{0.96} \pm 0.06$
4_vs_5								
	0.85, 0.85	0.79, 0.8	0.85, 0.85	0.65, 0.68	0.84, 0.85	0.81, 0.82	0.79, 0.8	0.85, 0.86
ecoli-0-3-4-	$0.92 \pm 0.12$	$0.94 \pm 0.12$	$0.92 \pm 0.05$	$0.89 \pm 0.06$	$0.91 \pm 0.11$	$0.89 \pm 0.12$	$0.95 \pm 0.05$	<b>0.96</b> ± 0.06
0_VS_3	0.86.0.87	0.80 0.80	0.66.0.69	0.67.0.69	07908	08.08	0.82 0.83	0.88 0.89
ecoli-0-3-4-	$0.93 \pm 0.05$	0.03, 0.03	0.00, 0.03 $0.92 \pm 0.07$	$0.86 \pm 0.08$	$0.92 \pm 0.04$	$0.89 \pm 0.01$	$0.92 \pm 0.08$	$0.93 \pm 0.08$
7_vs_5-6								
	0.85, 0.85	0.8, 0.81	0.69, 0.72	0.64, 0.66	0.69, 0.72	0.82, 0.82	0.87, 0.88	0.75, 0.77
ecoli-0-4-	$0.94\pm0.11$	$0.95\pm0.07$	$0.93\pm0.08$	$0.89\pm0.1$	$0.93\pm0.05$	$0.91\pm0.11$	$\textbf{0.96} \pm 0.06$	$0.95\pm0.07$
6_vs_5								
1: 0.0	0.85, 0.86	0.92, 0.92	0.79, 0.8	0.68, 0.71	0.7, 0.74	0.82, 0.83	0.9, 0.91	0.92, 0.92
ecoli-0-6-	$0.9 \pm 0.16$	$0.91 \pm 0.11$	$0.91 \pm 0.05$	$0.87 \pm 0.09$	$0.89 \pm 0.09$	$0.81 \pm 0.1$	$0.92 \pm 0.12$	<b>0.93</b> ± 0.1
7_08_3-3	0 81 0 84	0.81 0.82	0 76 0 78	0.67.0.69	0 67 0 7	0.45.0.51	072073	0 78 0 81
ecoli-0-6-	$0.91 \pm 0.11$	$0.9 \pm 0.07$	0.92 + 0.07	$0.86 \pm 0.1$	$0.88 \pm 0.06$	0.45, 0.51 $0.85 \pm 0.11$	$0.89 \pm 0.11$	$0.92 \pm 0.07$
7_vs_5								
	0.79, 0.8	0.75, 0.76	0.71, 0.74	0.63, 0.65	0.58, 0.63	0.77, 0.78	0.82, 0.82	0.75, 0.77
ecoli0137vs26	$0.93\pm0.03$	$0.93\pm0.06$	$\textbf{0.94} \pm 0.03$	$\textbf{0.89} \pm \textbf{0.03}$	$\textbf{0.94} \pm 0.02$	$0.88\pm0.04$	$\textbf{0.94} \pm 0.02$	$\textbf{0.94} \pm 0.03$
	0.9, 0.9	0.89, 0.89	0.88, 0.88	0.74, 0.75	0.91, 0.91	0.68, 0.71	0.86, 0.86	0.89, 0.89
ecoli01vs5	$1 \pm 0.01$	$1\pm 0$	$1 \pm 0.01$	$0.98\pm0.02$	$0.99 \pm 0.01$	$0.99 \pm 0.01$	$1 \pm 0.01$	$1 \pm 0.01$
1.0	0.96, 0.96	1,1	0.98, 0.98	0.82, 0.84	0.92, 0.93	0.94, 0.94	0.98, 0.98	0.98, 0.98
ecoli2	$0.9 \pm 0.06$	$0.89 \pm 0.1$	$0.9 \pm 0.03$	$0.87 \pm 0.05$	$0.91 \pm 0.06$	$0.83 \pm 0.04$	$0.91 \pm 0.05$	$0.92 \pm 0.03$
ecoli3	0.00, 0.08 0.9 $\pm$ 0.06	0.00, 0.09 0.01 $\pm$ 0.04	0.0, 0.03 0.0 + 0.03	0.49, 0.30 0.87 $\pm$ 0.05	0.02, 0.00 0.91 $\pm$ 0.06	0.44, 0.52 0.83 $\pm$ 0.04	0.7, 0.72 0.91 $\pm$ 0.05	0.07, 0.7 0.02 + 0.03
	0.66, 0.68	0.7. 0.72	0.6. 0.65	0.49.0.56	0.62. 0.66	0.44. 0.52	0.7. 0.72	0.67.07
ecoli4	$0.97 \pm 0.02$	$0.98 \pm 0$	<b>0.98</b> ± 0.01	$0.92 \pm 0.01$	$0.95 \pm 0.02$	$0.91 \pm 0.03$	$0.95 \pm 0.06$	<b>0.98</b> ± 0.01
	0.68, 0.72	0.77, 0.79	0.78, 0.8	0.43, 0.52	0.59, 0.65	0.43, 0.52	0.69, 0.72	0.79, 0.81
glass4	$\textbf{0.96} \pm 0.07$	$0.94\pm0.12$	$\textbf{0.96} \pm 0.07$	$0.91\pm0.08$	$0.95\pm0.07$	$0.81\pm0.11$	$\textbf{0.96} \pm 0.08$	$\textbf{0.96} \pm 0.04$
	0.91, 0.92	0.85, 0.85	0.83, 0.84	0.53, 0.59	0.75, 0.77	0.39, 0.45	0.88, 0.89	0.7, 0.74
heart-stat	$0.82\pm0.04$	$0.82\pm0.06$	$0.78\pm0.03$	$0.59\pm0.03$	$0.8\pm0.03$	$\textbf{0.84} \pm 0.04$	$0.8\pm0.04$	$0.82\pm0.04$
	0.8, 0.8	0.79, 0.79	0.76, 0.76	0.55, 0.56	0.77, 0.78	0.82, 0.82	0.77, 0.77	0.79, 0.8
10110	$0.95 \pm 0.04$	$0.95 \pm 0.03$	$0.91 \pm 0.03$	$0.84 \pm 0.03$	$0.93 \pm 0.04$	$0.88 \pm 0.03$	$0.95 \pm 0.02$	$0.94 \pm 0.02$
new-thyroid1	0.94, 0.94 1 + 0.01	0.93, 0.93 0.99 + 0.01	0.09, 0.9 <b>1</b> + 0	0.0, 0.01 0.97 + 0.02	0.32, 0.32 0 99 + 0 01	0.00, 0.07 $0.9 \pm 0.05$	0.94, 0.94 1 + 0.01	1+0
new myroidi	0.99, 0.99	0.95, 0.95	1,1	0.9, 0.9	0.97, 0.97	0.75, 0.77	0.99, 0.99	1,1

(continued on next page)

#### M.A. Ganaie, M. Tanveer and A.D.N. Initiative

Datasets	TSVM [10]	LITSVM [21]	RUTSVM_CII	MVKWEIM	MVCSKELM	KW-SMOTE_	REI STSVM-	KWBI ITSVM-
Dalasels			[71]		[34]	SVM	CII	CII
			[/1]	[34]		[33]	[36]	CIL
	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD
	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,
	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean
pima	$0.72\pm0.07$	$0.74\pm0.04$	$0.74\pm0.05$	$0.66\pm0.04$	$0.72\pm0.07$	$\textbf{0.75}\pm0.02$	$0.73\pm0.04$	$0.74\pm0.04$
	0.62, 0.62	0.66, 0.68	0.67, 0.68	0.57, 0.57	0.63, 0.64	0.67, 0.67	0.66, 0.67	0.67, 0.68
vehicle1	$0.85\pm0.02$	$0.84\pm0.02$	$0.85\pm0.02$	$0.61\pm0.04$	$0.82\pm0.04$	$0.78\pm0.02$	$0.82\pm0.03$	$\textbf{0.86} \pm 0.02$
	0.72, 0.74	0.7, 0.72	0.71, 0.73	0.44, 0.44	0.67, 0.7	0.64, 0.66	0.69, 0.71	0.73, 0.75
vehicle2	$0.99\pm0$	$0.98\pm0.02$	$0.98\pm0.01$	$0.9\pm0.03$	$0.97\pm0.02$	$0.93\pm0.01$	$0.98\pm0.01$	$\textbf{0.99} \pm 0.02$
	0.98, 0.98	0.97, 0.97	0.97, 0.97	0.82, 0.82	0.95, 0.95	0.83, 0.84	0.97, 0.97	0.97, 0.97
votes	$\textbf{0.97} \pm 0.02$	$0.97\pm0.02$	$0.97\pm0.01$	$0.94\pm0.02$	$0.96\pm0.02$	$0.96\pm0.01$	$0.96\pm0.02$	$\textbf{0.97}\pm0.02$
	0.95, 0.95	0.96, 0.96	0.96, 0.96	0.93, 0.93	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.96, 0.96
vowel	$1\pm 0$	$1\pm 0$	$0.99\pm0.01$	$0.94\pm0.01$	$1\pm 0$	$0.89\pm0.04$	$1\pm0$	$1\pm 0$
	1,1	1,1	0.99, 0.99	0.62, 0.67	1,1	0.81, 0.81	1,1	0.98, 0.98
wpbc	$0.68\pm0.09$	$\textbf{0.72}\pm0.06$	$0.69\pm0.05$	$0.58\pm0.12$	$0.7\pm0.1$	$0.72 \pm 0.08$	$0.7\pm0.09$	$0.69\pm0.07$
	0.51, 0.51	0.55, 0.57	0.51, 0.54	0.39, 0.41	0.52, 0.55	0.54, 0.57	0.53, 0.54	0.52, 0.54
yeast-0-2-5-	$0.78\pm0.04$	$0.79\pm0.04$	$0.81\pm0.03$	$0.78\pm0.04$	$0.81\pm0.05$	$0.72\pm0.11$	$0.81\pm0.05$	$0.82 \pm 0.02$
6_vs_3-7-8-9								
	0.64, 0.64	0.58, 0.59	0.6, 0.6	0.43, 0.47	0.59, 0.59	0.38, 0.43	0.59, 0.59	0.63, 0.63
yeast-0-2-5- 7-9_vs_3-6-8	$\textbf{0.92}\pm0.03$	$0.91\pm0.03$	$0.91\pm0.04$	$0.85\pm0.05$	$0.91\pm0.02$	$0.89\pm0.04$	$\textbf{0.92}\pm0.03$	$\textbf{0.92}\pm0.04$
	0.77, 0.78	0.8, 0.8	0.77, 0.78	0.52, 0.56	0.75, 0.76	0.62, 0.65	0.73, 0.74	0.79, 0.79
yeast-0-5-6-	$0.81\pm0.08$	$0.8\pm0.06$	$\textbf{0.83}\pm0.1$	$0.8\pm0.04$	$0.82\pm0.07$	$0.76\pm0.04$	$0.81\pm0.05$	$\textbf{0.83} \pm 0.11$
7-9_vs_4								
	0.55, 0.56	0.43, 0.49	0.53, 0.56	0.42, 0.48	0.46, 0.51	0.41, 0.45	0.52, 0.55	0.56, 0.59
yeast-2_vs_4	$\textbf{0.91} \pm 0.07$	$\textbf{0.91} \pm 0.05$	$0.89\pm0.05$	$0.85\pm0.06$	$0.89\pm0.07$	$0.89\pm0.06$	$\textbf{0.91} \pm 0.06$	$0.89\pm0.04$
	0.69, 0.71	0.76, 0.77	0.71, 0.72	0.53, 0.57	0.76, 0.76	0.77, 0.77	0.74, 0.75	0.78, 0.79
yeast2vs8	$\textbf{0.82} \pm 0.14$	$\textbf{0.82} \pm 0.14$	$0.81\pm0.14$	$0.77\pm0.11$	$0.79\pm0.07$	$0.77\pm0.11$	$\textbf{0.82}\pm0.14$	$0.8\pm0.14$
	0.64, 0.65	0.61, 0.62	0.57, 0.58	0.68, 0.7	0.41, 0.44	0.68, 0.7	0.72, 0.74	0.7, 0.73
yeast3	$0.92\pm0.04$	$0.92\pm0.04$	$\textbf{0.93} \pm 0.03$	$0.85\pm0.03$	$0.91\pm0.03$	$0.89\pm0.04$	$0.92\pm0.02$	$\textbf{0.93}\pm0.02$
	0.72, 0.74	0.75, 0.77	0.72, 0.74	0.5, 0.56	0.68, 0.7	0.75, 0.75	0.7, 0.72	0.74, 0.76
yeast5	$0.97\pm0.01$	$0.97\pm0.03$	$0.97\pm0.01$	$0.92\pm0.01$	$0.97\pm0.01$	$\textbf{0.98} \pm 0.01$	$\textbf{0.98} \pm 0.01$	$\textbf{0.98} \pm 0.01$
	0.49, 0.57	0.61, 0.66	0.54, 0.61	0.29, 0.41	0.55, 0.62	0.6, 0.66	0.67, 0.71	0.61, 0.66
Average AUC	0.8891	0.8952	0.893	0.8364	0.8854	0.8539	0.895	0.9028
Average Rank	4.29	3.6	3.84	7.51	4.79	6.51	3.49	1.98
Overall	[2, 0, 0]	[6, 0, 0]	[1, 0, 0]	[0, 0, 20]	[0, 0, 0]	[3, 0, 17]	[5, 0, 2]	[20, 0, 0]
Win-Tie-Loss								

Here, STD denotes the standard deviation.

#### Table 3

Nemenyi post ł	oc significant d	ifference of classi	fication models based of	on their performance	in real world KEEL d	atasets with Gaussian kerne	el.
Significance	TSVM [10]	UTSVM [21]	RUTSVM-CIL [71]	MVKWELM [34]	MVCSKELM [34]	KW-SMOTE-SVM [33]	RFLSTSVM-CIL [36]
Proposed	Yes	No	Yes	Yes	Yes	Yes	No
Table 4		: <b>6</b> - 1 : - <b>1</b> : - :	G		:	e with Coursien laseral	
Nemenyi post r	loc significant d	inerence of classi	fication models based of	on their performance	in biomedical dataset	s with Gaussian kernel.	
Significance	TSVM [10]	UTSVM [21]	RUTSVM-CIL [71]	MVKWELM [34]	MVCSKELM [34]	KW-SMOTE-SVM [33]	RFLSTSVM-CIL [36]
Proposed	Yes	No	Yes	Yes	No	Yes	Yes

#### Table 5

Pair-Wise Win-1	tie-loss sign tes	t of classification	models based on their	performance in real	world KEEL datasets	with Gaussian kernel.	
Significance	TSVM [10]	UTSVM [21]	RUTSVM-CIL [71]	MVKWELM [34]	MVCSKELM [34]	KW-SMOTE-SVM [33]	RFLSTSVM-CIL [36]
Proposed	Yes	Yes	Yes	Yes	Yes	Yes	Yes

#### Table 6

Pair-Wise Win-t	ie-loss sign test	of classification	models based on their	performance in biom	edical datasets with (	Gaussian kernel.	
Significance	TSVM [10]	UTSVM [21]	RUTSVM-CIL [71]	MVKWELM [34]	MVCSKELM [34]	KW-SMOTE-SVM [33]	RFLSTSVM-CIL [36]
Proposed	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2. Here, the data is partitioned randomly into 5 subsets out of which one partition is reserved for testing and the rest partitions are involved in training the model. This process is repeated 5 times and average accuracy of the folds is taken as the performance measure. All the classification models have been implemented in MATLAB R2017b on a system with configuration Intel (R) Core (TM) i7-6700 CPU @ 3.40 GHZ with 8 GB of RAM.

We use Gaussian kernel  $K(x, y) = exp^{-\frac{1}{2\mu^2} ||x-y||^2}$ , where  $\mu$  is the kernel parameter.

The classification models involve proper tuning of parameters for better performance. We used standard grid search approach to tune the optimal from the following range of parameters:  $c_0 = [0.5, 1, 1.5, 2, 2.5], \mu = [10^{-5}, 10^{-4}, \dots, 10^4, 10^5], c_u =$ 



Fig. 1. The sensitivity evaluation of the proposed KWRUTSVM-CIL classifier based on user specified parameters with Gaussian kernel.

 $c_i = [10^{-5}, 10^{-4}, \dots, 10^4, 10^5]$ , for i = 1, 2, 3, 4. To reduce the computations, we use  $c_1 = c_2 = c_u, c_3 = c_4$ .

For generating the universum data, we used random averaging method [71]. In random averaging method, same number of samples are randomly chosen from the given classes and the selected samples are averaged to get the universum data. For generating the KNN weights [54], we followed procedure given above in Section 2.2.

The evaluation of the baseline classifiers and the proposed KWRUTSVM-CIL is performed using area under receiver operating characteristics (ROC) curve i.e. (AUC) or accuracy and other measures which are defined as:

Accuracy, AUC = 
$$\frac{TP + TN}{TP + FP + TN + FN},$$
(54)

Sensitivity or Recall = 
$$\frac{TP}{TP + FN}$$
, (55)

$$Precision = \frac{Tr}{TP + FP},$$
(56)

$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall},$$
(57)

$$G-mean = \sqrt{Precision \times Recall},$$
(58)

where *TP* denotes true positive, *TN* : true negative, *FP* : false positive and *FN* : false negative. The different hyperparameters of a given classifier generates multiple classifiers which are represented by a point in ROC curve. The optimal hyperparameters are the parameters corresponding to the point (a classifier in ROC curve) in northwest of the AUC curve. The optimal point

Gaussiali Kerner via vviicoko	JII Sigileu Talik test.			
	R+	R-	<i>p</i> -value	Hypothesis (0.05)
TSVM	720	100	0.00001	Rejected
UTSVM	623	197	0.00424	Rejected
RUTSVM-CIL	674.5	66.5	0.00001	Rejected
MVKWELM	820	0	0.00001	Rejected
MVCSKELM	806	14	0.00001	Rejected
KW-SMOTE-SVM	792.5	27.5	0.00001	Rejected
RFLSTSVM-CIL	588	153	0.00164	Rejected

Statistical comparison of classification models with respect to the proposed KWRUTSVM-CIL on real world KEEL datasets with Gaussian kernel via Wilcoxon signed rank test.

The *p*-value is calculated from paired Wilcoxon test.

R+ (R-) present the sum of positive (negative) ranks, respectively.

(classifier) is the classifier corresponding to the optimal hyperparameters and the same information is reported in above given measures.

Alzheimer's data is obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database http://adni.loni.usc.edu/. In 2003, ADNI with Michael W. Weiner, MD as the Principal investigator was established. The motive of ADNI is to analyse the onset of Alzheimer's disease via Neuroimaging approaches involving positron emission tomography (PET), magnetic resonance imaging (MRI) and other biological markers. For details, we refer to www.adni-info.org. We used ADNI baseline dataset [80,81] and downloaded 817 structural MRI (sMRI) images. For Volume based morphometry (VolBM) of the images, we used Freesurfer's reconall pipeline (version6.0.1) [82,83]. Out of 817 images, 4 images failed to process via Freesurfer pipeline. Thus, we use control normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) subjects with 228,398 and 187 images respectively.

The BreakHis histopathological breast cancer image dataset [84] contains 1240 images with 400X magnification. The dataset involves 2 classes: malignant class and benign class. Benign category is subdivided into phyllodes tumour (PT), adenosis (ADN), tubular adenoma (TA) and fibroadenoma (FA) having 115, 106, 130 and 237 images, respectively. Malignant category is subdivided into mucinous carcinoma (MC), ductal carcinoma (DC), papillary carcinoma (PC) and lobular carcinoma (LC) which contains 169, 208, 138 and 137 images, respectively. The images are converted into grey level images and up to 3 levels of decomposition of Daubechies-4 (db4) wavelet is applied on each image [85]. The concatenation of approximation coefficients and detail coefficients is used as the feature vector. The information of biomedical data is presented in Table 9.

#### 4.1. Real world datasets

The experimental results of TSVM [10], UTSVM [21], RUTSVM-CIL [71], MVKWELM [34], MVCSKELM [34], KW-SMOTE-SVM [33], RFLSTSVM-CIL [36] and proposed KWRUTSVM-CIL with Gaussian kernel are presented in Table 2.

From Table 2, the average accuracy of TSVM [10], UTSVM [21], RUTSVM-CIL [71], MVKWELM [34], MVCSKELM [34], KW-SMOTE-SVM [33], RFLSTSVM-CIL [36] and the proposed KWRUTSVM-CIL classifiers are 0.8891, 0.8952, 0.893, 0.8364, 0.8854, 0.8539, 0.895 and 0.9028, respectively. The proposed KWRUTSVM-CIL classifier demonstrated superior performance in comparison with the existing models. According to overall win-tie-loss analysis, the TSVM model wins in 2 datasets, UTSVM wins in 6 datasets, RUTSVM-CIL models win in 1 dataset, KW-SMOTE-SVM and RFLSTSVM-CIL win in 3 and 5 datasets, respectively. The proposed KWRUTSVM-CIL model wins in 20 datasets. Thus, the proposed KWRUTSVM-CIL model is the successful classifier with majority of the wins. Since average accuracy may be biased as higher performance in one dataset may compensate the others, hence, we rank the classification models on each dataset to evaluate them. From Table 2, one can see that the average rank of TSVM [10], UTSVM [21], RUTSVM-CIL [71], MVKWELM [34], MVCSKELM [34], KW-SMOTE-SVM [33], RFLSTSVM-CIL [36] and proposed KWRUTSVM-CIL are 4.29, 3.6, 3.84, 7.51, 4.79, 6.51, 3.49, and 1.98, respectively. Thus, the proposed KWRUTSVM-CIL model emerged as the best classifier with lowest average rank.

For evaluation of the models via statistical tests, we perform statistical analysis. We used Friedman test with corresponding Nemenyi post hoc test for the comparison of the models. Every classifier in Friedman test is given a rank on a dataset with the worse performing classifier assigned higher rank and vice versa. The average of the rank across different datasets corresponding to a given model is calculated and is known as the average rank of the classifier. The classifier with smallest average rank is the best performing model while as the model with highest rank is the model with lowest performance. Under the null hypothesis, the average rank of the classification models is equal and hence show similar performance. Let *N* number of datasets are used to evaluate the performance of *k* number of classification models. The Friedman statistic follows  $\chi^2$  distribution with (k-1) degrees of freedom and is given as follows:

$$\chi_F^2 = \frac{12N}{k(k+1)} \Big[ \sum_j R_j^2 - \frac{k(k+1)^2}{4} \Big],$$
(59)

where  $R_j$  is the average rank of the *j*th classifier. As  $\chi_F^2$  is undesirably conservative and thus a better statistic  $F_F$  with *F*-distribution (k-1, (k-1)(N-1)) degrees of freedom is given. Mathematically,

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2}.$$
(60)

After simple calculation with N = 40 and k = 8, we get  $\chi_F^2 = 146.23$  and  $F_F = 42.6327$  follows F-distribution with (7, 273) degrees of freedom. For significance level  $\alpha = 0.05$ , the critical values of  $F_F(7, 273)$  is 2.045. Since,  $F_F = 42.6327 > 2.045$  hence, null hypothesis gets rejected. Thus, difference is significant among the given classifiers. To check the significant difference between the classifiers, Nemenyi post-hoc test is used. For Nemenyi test, the critical difference (CD) is given by

$$CD = q_{\alpha} \sqrt{\frac{k \times (k+1)}{6N}}.$$
(61)

After simple calculation, with  $q_{0.05} = 3.031$  the critical difference is 1.6601. If the average rank of two classifiers differ at least by the CD then the difference between two classifiers is significant with the lower rank classifier performing better than the other classifiers. Tables 3 and 4 gives the significant difference of the models on KEEL and biomedical datasets, respectively. It is evident that the proposed KWRUTSVM-CIL is performing significantly better than the existing models (except UTSVM and RFLSTSVM-CIL) in KEEL datasets. Fig. 2 gives the pictorial representation of the significant difference of the models.

Performance comparison of proposed KWRUTSVM-CIL model with different values of parameter K on real world KEEL datasets with Gaussian kernel.

Datasets	2	3	4	5	6	7	8	9	10
	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD
	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,
	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean
abalone9-18	$0.82\pm0.05$	$0.86\pm0.07$	$0.87\pm0.08$	$0.89\pm0.04$	$0.89\pm0.04$	$0.88\pm0.04$	$0.88\pm0.04$	$0.87\pm0.05$	$0.87\pm0.05$
	0.58, 0.61	0.58, 0.61	0.56, 0.59	0.52, 0.57	0.48, 0.54	0.47, 0.53	0.48, 0.54	0.46, 0.52	0.45, 0.51
aus	$0.8\pm0.05$	$0.82\pm0.05$	$0.82\pm0.05$	$\textbf{0.83} \pm \textbf{0.04}$	$\textbf{0.83} \pm \textbf{0.04}$	$0.83\pm0.05$	$0.85\pm0.04$	$0.86\pm0.05$	$0.86\pm0.05$
	0.76, 0.77	0.78, 0.79	0.79, 0.8	0.8, 0.81	0.8, 0.81	0.81, 0.82	0.83, 0.83	0.84, 0.84	0.85, 0.85
brwisconsin	$0.93 \pm 0.02$	$0.94 \pm 0.02$	$0.95 \pm 0.01$	$0.96 \pm 0.01$	$0.96 \pm 0$	$0.97 \pm 0.01$	$0.97 \pm 0.01$	$0.98 \pm 0.01$	$0.98 \pm 0.01$
huna or liver	0.92, 0.92	0.93, 0.93	0.94, 0.94	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.96, 0.96	0.97, 0.97	0.97, 0.97 0.71 $\pm$ 0.06
disorders	0.05 ± 0.07	$0.00 \pm 0.00$	0.00 ± 0.00	$0.03 \pm 0.00$	0.05 ± 0.07	$0.00 \pm 0.00$	$0.05 \pm 0.00$	0.7 ± 0.00	0.71 ± 0.00
uisoraers	0.51, 0.54	0.59, 0.6	0.59, 0.6	0.62, 0.62	0.61, 0.62	0.61, 0.61	0.63, 0.64	0.64, 0.64	0.66, 0.66
checkerboard	$0.8 \pm 0.05$	$0.82 \pm 0.05$	$0.82\pm0.05$	$0.83 \pm 0.04$	$0.83 \pm 0.04$	$0.83 \pm 0.05$	$0.85 \pm 0.04$	$0.86 \pm 0.05$	$0.86 \pm 0.05$
	0.76, 0.77	0.78, 0.79	0.79, 0.8	0.8, 0.81	0.8, 0.81	0.81, 0.82	0.83, 0.83	0.84, 0.84	0.85, 0.85
cleve	$0.79\pm0.05$	$0.82\pm0.03$	$0.84\pm0.03$	$0.83\pm0.04$	$0.85\pm0.03$	$0.85\pm0.03$	$0.85\pm0.04$	$0.85\pm0.04$	$0.85\pm0.03$
	0.74, 0.76	0.79, 0.8	0.81, 0.82	0.81, 0.81	0.83, 0.83	0.84, 0.84	0.83, 0.83	0.83, 0.83	0.83, 0.84
стс	$0.63 \pm 0.02$	$0.67 \pm 0$ 0.55 0.58	$0.67 \pm 0.01$	$0.68 \pm 0.01$	$0.68 \pm 0.01$	$0.68 \pm 0.02$	$0.69 \pm 0.01$	$0.69 \pm 0.01$	$0.69 \pm 0.01$
ecoli-0-	0.40, 0.51 $0.88 \pm 0.07$	0.55, 0.58 $0.9 \pm 0.08$	0.30, 0.39 $0.9 \pm 0.07$	0.37, 0.0 $0.9 \pm 0.08$	0.37, 0.39 $0.9 \pm 0.08$	0.38, 0.0 $0.89 \pm 0.08$	0.38, 0.01 $0.89 \pm 0.08$	0.39, 0.01 $0.91 \pm 0.06$	0.33, 0.01 $0.91 \pm 0.06$
1 vs 2-3-5	0.00 ± 0.07	0.5 ± 0.00	0.5 ± 0.07	0.0 ± 0.00	0.5 ± 0.00	0.05 ± 0.00	0.05 ± 0.00	0.51 ± 0.00	0.01 ± 0.00
	0.75, 0.77	0.79, 0.81	0.77, 0.78	0.74, 0.76	0.73, 0.74	0.71, 0.72	0.72, 0.74	0.75, 0.76	0.74, 0.76
ecoli-0-1_vs_5	$0.94\pm0.11$	$0.97\pm0.05$	$0.96\pm0.05$	$0.94\pm0.11$	$0.91\pm0.11$	$0.9\pm0.11$	$0.91\pm0.11$	$0.93\pm0.11$	$0.93\pm0.11$
	0.87, 0.88	0.89, 0.89	0.86, 0.87	0.82, 0.83	0.75, 0.76	0.74, 0.75	0.75, 0.76	0.78, 0.8	0.78, 0.8
ecoli-0-1-4-	$0.93 \pm 0.11$	$0.93 \pm 0.11$	$0.95 \pm 0.06$	$0.92 \pm 0.11$	$0.92 \pm 0.12$	$0.91 \pm 0.11$	$0.92 \pm 0.11$	$0.92 \pm 0.11$	$0.92 \pm 0.11$
6_VS_5	0.74.0.76	0 75 0 77	0.76.0.78	0.68 0.71	0.67.07	0.64.0.67	0.66.0.69	0.65.0.68	0.65.0.68
ecoli-0-1-4-	$0.9 \pm 0.04$	$0.9 \pm 0.04$	$0.9 \pm 0.05$	$0.89 \pm 0.06$	0.07, 0.7 $0.9 \pm 0.05$	$0.89 \pm 0.07$	0.00, 0.05 $0.9 \pm 0.07$	$0.88 \pm 0.1$	0.03, 0.00 $0.89 \pm 0.11$
7_vs_2-3-5-6									
	0.72, 0.74	0.69, 0.7	0.71, 0.73	0.67, 0.69	0.73, 0.74	0.71, 0.72	0.77, 0.78	0.72, 0.73	0.73, 0.74
ecoli-0-1-4-	$0.92\pm0.05$	$0.93\pm0.05$	$0.93\pm0.05$	$0.93\pm0.05$	$0.93\pm0.06$	$0.93\pm0.06$	$0.93\pm0.06$	$0.91\pm0.08$	$0.91\pm0.07$
7_vs_5-6									
	0.67, 0.69	0.71, 0.74	0.72, 0.74	0.69, 0.71	0.7, 0.72	0.71, 0.74	0.7, 0.72	0.71, 0.73	0.69, 0.71
4 vs 5	$0.94 \pm 0.11$	$0.94 \pm 0.11$	$0.99 \pm 0.01$	$0.96 \pm 0.06$	$0.94 \pm 0.07$	$0.94 \pm 0.07$	$0.91 \pm 0.11$	$0.94 \pm 0.07$	$0.91 \pm 0.11$
4_V3_5	0.85.0.85	0.85.0.85	0.89. 0.9	0.85, 0.86	0.84. 0.84	0.84, 0.84	0.8. 0.81	0.84. 0.84	0.79. 0.8
ecoli-0-3-4-	$0.92 \pm 0.11$	$0.91 \pm 0.11$	$0.96\pm0.06$	$0.94 \pm 0.07$	$0.94\pm0.07$	$0.94 \pm 0.07$	$0.96 \pm 0.05$	$0.97 \pm 0.06$	$0.96\pm0.07$
6_vs_5									
	0.84, 0.85	0.82, 0.83	0.88, 0.89	0.88, 0.88	0.86, 0.87	0.86, 0.87	0.87, 0.88	0.91, 0.91	0.89, 0.89
ecoli-0-3-4-	$0.87 \pm 0.1$	$0.89 \pm 0.07$	$0.93 \pm 0.08$	$0.88 \pm 0.12$	$0.88 \pm 0.11$	$0.89 \pm 0.11$	$0.89 \pm 0.11$	$0.89 \pm 0.1$	$0.89 \pm 0.1$
7_VS_5-6	0.68.0.60	07072	0.74.0.76	0.65 0.67	0.66.0.68	0.68 0.7	071072	07071	07071
ecoli-0-4-	0.03, 0.03 $0.84 \pm 0.11$	0.7, 0.72 $0.9 \pm 0.11$	0.74, 0.70 $0.95 \pm 0.07$	0.03, 0.07 $0.94 \pm 0.08$	0.00, 0.00 $0.94 \pm 0.08$	0.03, 0.7 $0.94 \pm 0.08$	$0.93 \pm 0.08$	$0.93 \pm 0.08$	$0.93 \pm 0.08$
6_vs_5									
	0.76, 0.76	0.86, 0.86	0.92, 0.92	0.9, 0.91	0.87, 0.87	0.86, 0.87	0.84, 0.85	0.81, 0.82	0.81, 0.82
ecoli-0-6-	$0.91\pm0.1$	$0.89\pm0.11$	$0.91\pm0.13$	$0.91\pm0.12$	$0.89\pm0.11$	$0.84\pm0.11$	$0.87\pm0.09$	$0.87\pm0.09$	$0.83\pm0.08$
7_vs_3-5									
acali 0 6	0.75, 0.78	0.71, 0.73	0.72, 0.74	0.71, 0.73	0.68, 0.71	0.62, 0.65	0.65, 0.67	0.69, 0.71	0.63, 0.65
7 vs 5	$0.89 \pm 0.07$	$0.9 \pm 0.07$	$0.92 \pm 0.07$	$0.9 \pm 0.07$	$0.92 \pm 0.07$	$0.89 \pm 0.07$	$0.91 \pm 0.08$	$0.91 \pm 0.07$	$0.88 \pm 0.00$
/	0.71.0.73	0.73.0.74	0.75. 0.77	0.73, 0.74	0.71.0.74	0.7.0.72	0.71.0.73	0.69. 0.72	0.65.0.68
ecoli0137vs26	$0.9\pm0.07$	$0.91\pm0.06$	$0.91\pm0.06$	$0.94\pm0.04$	$0.94\pm0.04$	$0.94\pm0.03$	$0.94\pm0.03$	$0.94\pm0.03$	$0.94\pm0.03$
	0.86, 0.86	0.87, 0.87	0.87, 0.87	0.9, 0.9	0.9, 0.9	0.89, 0.89	0.89, 0.89	0.89, 0.89	0.88, 0.89
ecoli01vs5	$0.97\pm0.05$	$0.97\pm0.05$	$1\pm0.01$	$1\pm0.01$	$1\pm0.01$	$1\pm0.01$	$1\pm0.01$	$1\pm0.01$	$0.99\pm0.01$
	0.95, 0.95	0.95, 0.95	0.98, 0.98	0.98, 0.98	0.98, 0.98	0.96, 0.96	0.96, 0.96	0.96, 0.96	0.94, 0.94
ecoliz	$0.85 \pm 0.02$	$0.87 \pm 0.05$ 0.66 0.68	$0.89 \pm 0.04$	$0.92 \pm 0.03$	$0.91 \pm 0.03$	$0.91 \pm 0.04$	$0.91 \pm 0.03$ 0.63 0.67	$0.91 \pm 0.03$ 0.63 0.67	$0.91 \pm 0.03$
ecoli3	0.03, 0.00 $0.85 \pm 0.02$	0.00, 0.00 $0.87 \pm 0.05$	0.03, 0.07 $0.89 \pm 0.04$	0.07, 0.7 $0.92 \pm 0.03$	0.03, 0.03 $0.91 \pm 0.03$	0.04, 0.03 $0.91 \pm 0.04$	0.03, 0.07 $0.91 \pm 0.03$	0.03, 0.07 $0.91 \pm 0.03$	0.04, 0.00 $0.91 \pm 0.03$
ccons	0.65, 0.66	0.66, 0.68	0.65, 0.67	0.67, 0.7	0.65, 0.69	0.64, 0.68	0.63, 0.67	0.63, 0.67	0.64, 0.68
ecoli4	$0.94\pm0.07$	$0.94\pm0.07$	$0.98\pm0.01$	$0.97\pm0.02$	$0.96\pm0.02$	$0.94\pm0.02$	$0.92\pm0.02$	$0.9\pm0.02$	$0.89 \pm 0.02$
	0.86, 0.87	0.8, 0.81	0.79, 0.81	0.69, 0.73	0.6, 0.65	0.52, 0.59	0.44, 0.53	0.39, 0.49	0.36, 0.47
glass4	$0.95 \pm 0.07$	$0.96 \pm 0.04$	$0.95 \pm 0.06$	$0.95 \pm 0.06$	$0.95 \pm 0.06$	$0.95 \pm 0.06$	$0.95 \pm 0.06$	$0.95 \pm 0.06$	$0.95 \pm 0.06$
heart stat	0.83, 0.85	0.7, 0.74	0.64, 0.69 0.78 $\pm$ 0.05	0.64, 0.69	0.64, 0.69	0.62, 0.67 0.82 $\pm$ 0.02	0.62, 0.67 0.83 $\pm$ 0.04	0.62, 0.67 0.82 $\pm$ 0.05	0.62, 0.67 0.82 $\pm$ 0.04
ncai t-stat	$0.74 \pm 0.03$ 0.67 0.69	$0.77 \pm 0.02$ 0.72 0.73	$0.78 \pm 0.03$ 0.74 0.75	$0.0 \pm 0.03$ 0.76 0.77	$0.81 \pm 0.04$ 0.77 0.78	$0.02 \pm 0.03$	$0.03 \pm 0.04$ 0.81 0.81	$0.02 \pm 0.03$ 0.79 0.79	$0.02 \pm 0.04$ 0.79 0.8
iono	$0.91 \pm 0.04$	$0.93 \pm 0.03$	$0.94 \pm 0.03$	$0.93 \pm 0.03$	$0.93 \pm 0.03$	$0.94 \pm 0.03$	$0.94 \pm 0.02$	$0.94 \pm 0.02$	$0.94 \pm 0.02$
-	0.89, 0.89	0.91, 0.91	0.92, 0.92	0.91, 0.91	0.91, 0.91	0.91, 0.91	0.91, 0.91	0.91, 0.91	0.91, 0.91
new-thyroid1	$\textbf{0.8} \pm \textbf{0.09}$	$0.9\pm0.12$	$0.94\pm0.09$	$1\pm0$	$1\pm0.01$	$0.99\pm0.01$	$0.99\pm0.01$	$0.99\pm0.01$	$\textbf{0.98} \pm \textbf{0.02}$
	0.74, 0.77	0.87, 0.89	0.93, 0.94	1,1	0.99, 0.99	0.96, 0.96	0.95, 0.95	0.94, 0.94	0.91, 0.92
pima	$0.7 \pm 0.03$	$0.71 \pm 0.02$	$0.73 \pm 0.05$	$0.71 \pm 0.05$	$0.7 \pm 0.06$	$0.66 \pm 0.04$	$0.66 \pm 0.04$	$0.68 \pm 0.05$	$0.69 \pm 0.05$
vehicle1	0.58, 0.6 0.77 $\pm$ 0.05	0.03, 0.03	0.00, 0.07	0.00, 0.68	0.04, 0.07	0.01, 0.05	0.01, 0.05	0.62, 0.66 0.85 $\pm$ 0.02	0.03, 0.00 0.85 ± 0.02
venicie i	$0.77 \pm 0.03$ 0.65 0.65	$0.6 \pm 0.02$ 0.68 0.68	$0.02 \pm 0.02$ 0 7 0 71	$0.04 \pm 0.01$ 0.72 0.73	$0.03 \pm 0.02$ 0.73 0.74	$0.03 \pm 0.03$ 0.73 0.74	$0.04 \pm 0.02$ 0.72 0.73	$0.03 \pm 0.02$ 0.72 0.74	$0.03 \pm 0.03$ 0.73 0.74
		,	,			· · · - , - · · ·	,	···	

(continued on next page)

Table 8 (continued).

$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	Datasets	2	3	4	5	6	7	8	9	10
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD	AUC±STD
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,	F-Measure,
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean	G-Mean
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	vehicle2	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$	$0.97\pm0.02$
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		0.96, 0.96	0.96, 0.96	0.96, 0.96	0.96, 0.96	0.96, 0.96	0.96, 0.96	0.96, 0.96	0.97, 0.97	0.96, 0.97
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	votes	$0.95\pm0.01$	$0.96\pm0.02$	$0.96\pm0.03$	$0.96\pm0.03$	$0.96\pm0.03$	$0.96\pm0.02$	$0.96\pm0.02$	$0.96\pm0.02$	$0.97\pm0.02$
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		0.93, 0.93	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.95, 0.95	0.96, 0.96
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	vowel	$0.99\pm0.01$	$0.99\pm0.01$	$0.99\pm0.01$	$0.99\pm0.01$	$0.99\pm0.01$	$0.99\pm0.01$	$1\pm0$	$1\pm0$	$1\pm0$
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		0.99, 0.99	0.99, 0.99	0.98, 0.98	0.99, 0.99	0.98, 0.98	0.98, 0.98	0.98, 0.98	0.98, 0.98	0.99, 0.99
$\begin{array}{c} 0.5, 0.51 \\ yeast-0-2-5- \\ 6\_vs\_3-7-8-9 \\ yeast-0-2-5- \\ yeast-0-2-5- \\ yeast-0-2-5- \\ yeast-0-2-5- \end{array} \\ \begin{array}{c} 0.5, 0.51 \\ 0.53, 0.54 \\ 0.74\pm 0.02 \\ 0.58, 0.6 \\ 0.9\pm 0.04 \\ 0.92\pm 0.04 \\ 0.91\pm 0.04 \\ 0.91\pm$	wpbc	$0.67 \pm 0.1$	$0.7\pm0.08$	$0.68\pm0.07$	$0.69\pm0.07$	$0.69\pm0.07$	$0.7\pm0.07$	$0.69\pm0.08$	$0.67\pm0.07$	$0.65\pm0.07$
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		0.5, 0.51	0.53, 0.54	0.51, 0.53	0.52, 0.54	0.52, 0.54	0.53, 0.54	0.51, 0.53	0.49, 0.5	0.47, 0.49
	yeast-0-2-5-	$0.68\pm0.04$	$0.74\pm0.02$	$0.77\pm0.05$	$0.79\pm0.03$	$0.81\pm0.02$	$0.81\pm0.02$	$0.81\pm0.02$	$0.82\pm0.02$	$0.82\pm0.02$
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	6_vs_3-7-8-9									
$yeast-0-2-5- 0.9 \pm 0.04 0.9 \pm 0.04 0.9 \pm 0.04 0.9 \pm 0.04 0.92 \pm 0.04 0.92 \pm 0.04 0.92 \pm 0.04 0.92 \pm 0.04 0.91 \pm $		0.5, 0.53	0.58, 0.6	0.62, 0.63	0.63, 0.64	0.65, 0.65	0.64, 0.64	0.62, 0.62	0.64, 0.64	0.63, 0.64
	yeast-0-2-5-	$0.9\pm0.04$	$0.9\pm0.04$	$0.9\pm0.04$	$0.92\pm0.04$	$0.92\pm0.04$	$0.92\pm0.04$	$0.92\pm0.04$	$0.91\pm0.04$	$0.91\pm0.04$
7-9_vs_3-6-8	7-9_vs_3-6-8									
0.77, 0.78 $0.77, 0.78$ $0.78, 0.78$ $0.79, 0.79$ $0.78, 0.79$ $0.78, 0.79$ $0.78, 0.79$ $0.77, 0.78$ $0.74, 0.75$ $0.73, 0.74$		0.77, 0.78	0.77, 0.78	0.78, 0.78	0.79, 0.79	0.78, 0.79	0.78, 0.79	0.77, 0.78	0.74, 0.75	0.73, 0.74
$yeast-0.5-6- 0.79 \pm 0.13  0.82 \pm 0.12  0.82 \pm 0.1  0.81 \pm 0.09  0.83 \pm 0.11  0.82 \pm 0.1  0.79 \pm 0.08  0.78 \pm 0.08  0.79 \pm 0.09  0.79 \pm 0.09  0.79 \pm 0.09  0.79 \pm 0.01  0.79  0.79  0.7$	yeast-0-5-6-	$0.79\pm0.13$	$0.82\pm0.12$	$0.82\pm0.1$	$0.81\pm0.09$	$0.83\pm0.11$	$0.82\pm0.1$	$0.79\pm0.08$	$0.78\pm0.08$	$0.79\pm0.09$
7-9_vs_4	7-9_vs_4									
$0.59, 0.6 \qquad 0.58, 0.59 \qquad 0.56, 0.58 \qquad 0.53, 0.56 \qquad 0.57, 0.59 \qquad 0.51, 0.54 \qquad 0.47, 0.5 \qquad 0.44, 0.48 \qquad 0.45, 0.49$		0.59, 0.6	0.58, 0.59	0.56, 0.58	0.53, 0.56	0.57, 0.59	0.51, 0.54	0.47, 0.5	0.44, 0.48	0.45, 0.49
$yeast-2_vs_4  0.88 \pm 0.05  0.88 \pm 0.04  0.88 \pm 0.04  0.88 \pm 0.04  0.89 \pm 0.03  0.88 \pm 0.03  0.88 \pm 0.02  0.88 \pm 0.03  0.87 \pm 0.04  0.88 \pm 0.04 $	yeast-2_vs_4	$0.88\pm0.05$	$0.88\pm0.04$	$0.88\pm0.04$	$0.88\pm0.04$	$0.89\pm0.03$	$0.88\pm0.03$	$0.89\pm0.02$	$0.88\pm0.03$	$0.87\pm0.04$
0.81, 0.81 0.8, 0.8 0.78, 0.78 0.78, 0.79 0.77, 0.77 0.75, 0.75 0.73, 0.73 0.71, 0.72 0.71, 0.71		0.81, 0.81	0.8, 0.8	0.78, 0.78	0.78, 0.79	0.77, 0.77	0.75, 0.75	0.73, 0.73	0.71, 0.72	0.71, 0.71
yeast2vs8 $0.77 \pm 0.16$ $0.77 \pm 0.16$ $0.8 \pm 0.14$ $0.77 \pm 0.16$	yeast2vs8	$0.77\pm0.16$	$0.77\pm0.16$	$0.8\pm0.14$	$0.77\pm0.16$	$0.77\pm0.16$	$0.77\pm0.16$	$0.77\pm0.16$	$0.77\pm0.16$	$0.77\pm0.16$
0.65, 0.69 $0.65, 0.69$ $0.7, 0.73$ $0.63, 0.66$ $0.65, 0.69$ $0.63, 0.66$ $0.65, 0.69$ $0.65, 0.69$ $0.63, 0.69$		0.65, 0.69	0.65, 0.69	0.7, 0.73	0.63, 0.66	0.65, 0.69	0.63, 0.66	0.65, 0.69	0.65, 0.69	0.63, 0.66
yeast $0.9 \pm 0.02$ $0.91 \pm 0.02$ $0.9 \pm 0.02$ $0.91 \pm 0.02$ $0.91 \pm 0.02$ $0.92 \pm 0.02$ $0.92 \pm 0.02$ $0.93 \pm 0.02$ $0.93 \pm 0.03$ $0.93 \pm 0.02$	yeast3	$0.9\pm0.02$	$0.91\pm0.02$	$0.9\pm0.02$	$0.91\pm0.02$	$0.92\pm0.02$	$0.92\pm0.02$	$0.93\pm0.02$	$0.93\pm0.03$	$0.93\pm0.02$
0.77, 0.77 $0.77, 0.77$ $0.75, 0.76$ $0.75, 0.76$ $0.74, 0.76$ $0.75, 0.76$ $0.74, 0.76$ $0.74, 0.76$ $0.74, 0.75$ $0.73, 0.75$		0.77, 0.77	0.77, 0.77	0.75, 0.76	0.75, 0.76	0.74, 0.76	0.75, 0.76	0.74, 0.76	0.74, 0.75	0.73, 0.75
$yeast5 \qquad 0.95 \pm 0.03 \qquad 0.98 \pm 0.01 \qquad 0.98 $	yeast5	$0.95\pm0.03$	$0.98\pm0.01$							
0.6, 0.64 0.6, 0.66 0.59, 0.65 0.62, 0.67 0.61, 0.66 0.6, 0.66 0.61, 0.67 0.6, 0.66 0.6, 0.65		0.6, 0.64	0.6, 0.66	0.59, 0.65	0.62, 0.67	0.61, 0.66	0.6, 0.66	0.61, 0.67	0.6, 0.66	0.6, 0.65



**Fig. 2.** Comparison of the classifiers based on the Nemenyi test. The classifiers which are connected are not significantly different while as the disconnected classifiers have significant difference between them.

Furthermore, we use Win–Tie–Loss sign test to evaluate the statistical significance of the models. In this test, under null hypothesis two models are equivalent if each model wins on approximately N/2 datasets out of the N datasets. At 5% level of significance, the two models are significantly different if one of models wins approximately on  $N/2 + 1.96\sqrt{N}/2$ . Also, if there is even number of ties between the two models then the number of ties are evenly distributed among them otherwise we ignore one and distribute the rest among the given classifiers.

Tables 5 and 6 gives the significant difference between the classifiers based on the pairwise win-tie-loss sign test on KEEL and biomedical datasets, respectively. It is evident that the proposed KWRUTSVM-CIL model is significantly superior compared to the existing models in KEEL and Biomedical datasets.

The nonparametric Wilcoxon signed-rank test [86] is employed to compare the performance of the given models. This test calculates  $d_i$ , that is the difference between the accuracy of two compared models corresponding to the *i*th dataset out of *N* datasets. The ascending order of absolute differences is ranked wherein the average ranks are given in tie cases. R + (R-) represents the sum of positive (negative) ranks, respectively. The sufficient difference between R+ and R- demonstrate that the compared models reject the null hypothesis. The null hypothesis is rejected if the *p*-value for this test is smaller than 0.05. Therefore, one can see from the Table 7 that the proposed KWRUTSVM-CIL outperforms the compared models, i.e., TWSVM,

UTSVM, RUTSVM, MVKWELM, MVCSKELM, KW-SMOTE-SVM and RFLSTSVM-CIL.

From the above analysis, superiority of the proposed KWRUTSM-CIL classifier in comparison with the existing models is evident.

The generalization of the proposed KWRUTSM-CIL is influenced by the hyperparameters. Fig. 1 gives the parameter sensitivity of the proposed KWRUTSM-CIL classifier. Thus, the model hyperparameters should be selected carefully for the optimal performance of the proposed KWRUTSM-CIL model.

#### 4.2. Biomedical data

For evaluating the applicability of the classifiers in real world scenarios, we use Alzheimer's disease and breast cancer datasets.

Table 9 gives the details of the biomedical data. The performance of the models on biomedical datasets is presented in Table 10. In PT\_vs\_PC subjects, TA\_vs\_LC subjects and ADN\_vs\_DC subjects the proposed model showed the average accuracy of 0.68, 0.76 and 0.89, respectively which is better in comparison with the existing classifiers. Thus, the proposed KWRUTSVM-CIL model shows better or competitive performance compared to the baseline models. It is evident that the proposed KWRUTSVM-CIL classifier obtained highest average classification performance in terms of accuracy. Also, the average rank is lowest i.e. the proposed model KWRUTSVM-CIL wins in most of the datasets. Hence, better generalization performance is demonstrated by the proposed KWRUTSVM-CIL model which shows its applicability in biomedical domain.

#### 5. Influence of the *K* nearest neighbours

Here, we investigate the influence of *K* nearest neighbours on the performance of the proposed KWRUTSM-CIL classifier. Here, we varied the parameter *K* from 2 to 10 with step size 1. Table 8 shows the performance of the model with varying number of *K* nearest neighbours. It is clear from the given table that *K* needs to be selected carefully for optimal performance.

#### Table 9 Biomedical dataset specifications

	Subclasses	Sample Size	Features
	Control Normal (CN)	228	91
Alzheimer's disease dataset	Mild Cognitive Impairment (MCI)	398	91
	Alzheimer's disease (AD)	187	91
	Adenosis (ADN)	106	768
Alzheimer's disease dataset Breast cancer dataset	Ductal Carcinoma (DC)	208	768
	Fibroadenema (FA)	237	768
	Mucinous Carcinoma (MC)	169	768
Breast cancer dataset	Phyllodes Tumour (PT)	115	768
	Tubular Adenoma (TA)	130	768
lzheimer's disease dataset	Lobular Carcinoma (LC)	137	768
	Papillary Carcinoma (PC)	138	768

#### Table 10

Classification accuracy of TSVM, UTSVM, RUTSVM-CIL and proposed KWRUTSVM-CIL models on biomedical datasets with Gaussian kernel.

$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	M-
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	,
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	2
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	
$ \begin{array}{c} \mbox{0.53, 0.53} & 0.66, 0.68 & 0.39, 0.42 & 0.39, 0.4 & 0.63, 0.67 & 0.51, 0.52 & 0.5, 0.51 & 0.54, 0.54 \\ \mbox{ADN_vs_MC} & 0.63, 0.67 & 0.60, 0.62 \pm 0.03 & 0.61 \pm 0.06 & 0.58 \pm 0.08 & 0.63 \pm 0.04 & 0.65 \pm 0.09 & 0.62 \pm 0.06 & 0.64 \pm 0.04 \\ 0.63, 0.67 & 0.69, 0.73 & 0.58, 0.6 & 0.55, 0.57 & 0.62, 0.66 & 0.53, 0.53 & 0.59, 0.61 & 0.52, 0.66 \\ \mbox{ADN_vs_PC} & 0.69, \pm 0.08 & 0.68 \pm 0.03 & 0.68 \pm 0.08 & 0.49 \pm 0.06 & 0.61 & 0.56 \pm 0.08 & 0.68 \pm 0.08 & 0.73 \pm 0.07 \\ 0.69, 0.71 & 0.69, 0.7 & 0.7, 0.73 & 0.54, 0.56 & 0.61, 0.62 & 0.52, 0.52 & 0.68, 0.7 & 0.71, 0.71 \\ \mbox{CN_vs_AD} & 0.89 \pm 0.05 & 0.88 \pm 0.02 & 0.88 \pm 0.04 & 0.86 \pm 0.04 & 0.87 \pm 0.03 & 0.89 \pm 0.05 & 0.87 \pm 0.05 \\ 0.88, 0.88 & 0.87, 0.87 & 0.87, 0.87 & 0.84, 0.84 & 0.86 & 0.85, 0.85 & 0.87, 0.88 & 0.86, 0.86 \\ \mbox{OLVs_MCI} & 0.69 \pm 0.03 & 0.69 \pm 0.03 & 0.69 \pm 0.01 & 0.67 \pm 0.01 & 0.69 \pm 0.02 & 0.69 \pm 0.03 & 0.69 \pm 0.03 & 0.7 \pm 0.03 \\ \mbox{OLNs_MCI} & 0.63, 0.64 & 0.63, 0.65 & 0.62, 0.63 & 0.61, 0.62 & 0.64, 0.66 & 0.62, 0.63 & 0.62, 0.63 & 0.64, 0.66 \\ \mbox{FA_vs_DC} & 0.81, 0.81 & 0.81, 0.81 & 0.8, 0.88 & 0.79, 0.79 & 0.82, 0.82 & 0.82, 0.82 & 0.81, 0.81 & 0.82, 0.82 \\ \mbox{OLNs_MCI} & 0.54, 0.54 & 0.52, 0.52 & 0.66, 65 & 0.58, 0.6 & 0.52, 0.52 & 0.63 \pm 0.05 & 0.65 \pm 0.03 \\ \mbox{OLNs_MCI} & 0.54, 0.54 & 0.52, 0.52 & 0.66, 65 & 0.58, 0.6 & 0.52, 0.52 & 0.46, 0.46 & 0.5, 0.51 & 0.55, 0.55 \\ \mbox{FA_vs_MC} & 0.55 \pm 0.02 & 0.58 \pm 0.03 & 0.58 \pm 0.05 & 0.69 \pm 0.06 & 0.51 \pm 0.05 & 0.57 \pm 0.07 & 0.58 \pm 0.03 \\ \mbox{OLNs_MCI} & 0.55 \pm 0.02 & 0.58 \pm 0.07 & 0.58, 0.6 & 0.52, 0.55 & 0.4, 0.44 & 0.59 \pm 0.06 & 0.51 \pm 0.05 & 0.57 \pm 0.07 & 0.58 \pm 0.04 \\ \mbox{OLNs_MCI} & 0.55 \pm 0.02 & 0.58 \pm 0.07 & 0.58 \pm 0.05 & 0.68 \pm 0.05 & 0.69 \pm 0.06 & 0.51 \pm 0.05 & 0.57 \pm 0.07 & 0.58 \pm 0.04 \\ \mbox{OLNs_MCI} & 0.54, 0.55 & 0.47, 0.48 & 0.37, 0.37 & 0.35, 0.39 & 0.4, 0.41 & 0.54, 0.55 & 0.54, 0.55 \\ \mbox{OLNs_MCI} & 0.54, 0.55 & 0.47, 0.48 & 0.37 \pm 0.04 & 0.58 \pm 0.04 & 0.54 \pm 0.04 & 0.55 \pm 0.06 & 0.66 \pm 0.05 & 0.66 \pm 0.05 & 0.66 \pm 0.05 & 0.66 \pm 0.05 & 0.6$	9
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	4
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	
$ \begin{array}{c} \mbox{0.69} \ 0.71 & 0.69 \ 0.7 & 0.7, 0.73 & 0.54 \ 0.56 & 0.61 \ 0.62 & 0.52 \ 0.52 & 0.68 \ 0.7 & 0.71 \ 0.71 $	7
$ \begin{array}{c} \mathrm{CN}\_\mathrm{vs}\_\mathrm{AD} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{AD} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{AC} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{MCI} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{LC} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{CN} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{LC} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{CN} \\ \mathrm{CN} \\ \mathrm{CN}\_\mathrm{vs}\_\mathrm{CN} \\ \mathrm{CN}$	
$ \begin{array}{c} \mbox{Res} & 0.88, 0.88 & 0.87, 0.87 & 0.87, 0.87 & 0.84, 0.84 & 0.86, 0.86 & 0.85, 0.85 & 0.87, 0.88 & 0.86, 0.86 \\ \mbox{CN\_vs\_MCI} & 0.69 \pm 0.06 & 0.69 \pm 0.03 & 0.69 \pm 0.01 & 0.69 \pm 0.02 & 0.69 \pm 0.03 & 0.69 \pm 0.03 & 0.69 \pm 0.03 \\ \mbox{OR} & 0.63, 0.64 & 0.63, 0.65 & 0.62, 0.63 & 0.61, 0.62 & 0.64, 0.66 & 0.62, 0.63 & 0.65, 0.51 & 0.55, 0.55 & 0.54, 0.54 & 0.52, 0.52 & 0.46, 0.46 & 0.5, 0.51 & 0.55, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.54 & 0.55 & 0.54, 0.55 & 0.55 & 0.54, 0.64 & 0.55, 0.55 & 0.54, 0.55 & 0.57 \pm 0.07 & 0.58 \pm 0.07 & 0.58 \pm 0.04 & 0.56 \pm 0.07 & 0.51 \pm 0.08 & 0.58 \pm 0.04 & 0.58 \pm 0.04 & 0.65 \pm 0.02 & 0.54, 0.55 & 0.59, 0.6 & 0.54, 0.55 & 0.58, 0.04 & 0.68 \pm 0.05 & 0.69 \pm 0.04 & 0.69 \pm 0.05 & 0.68 \pm 0.05 & 0.68 \pm 0.05 & 0.68 \pm 0.05 & 0.69 \pm 0.04 & 0.58 \pm 0.08 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 & 0.57, 0.58 & 0.68 \pm 0.05 & 0.58, 0.08 & 0.54, 0.55 & 0.57, 0.58 & 0.68 \pm 0.05 & 0.68 \pm 0.07 & 0.66 \pm 0.07 & 0.64 \pm 0.07 & 0.64 \pm 0.02 & 0.54 \pm 0.04 & 0.65 \pm 0.07 & 0.66 \pm 0.07 & 0.64 \pm 0.02 & 0.54 \pm 0.04 & 0.65 \pm 0.07 & 0.66 \pm 0.0$	5
$ \begin{array}{c} \mathrm{CN}\_vs\_MCI & 0.69\pm0.06 & 0.69\pm0.03 & 0.69\pm0.01 & 0.67\pm0.01 & 0.69\pm0.02 & 0.69\pm0.03 & 0.69\pm0.03 & 0.7\pm0.03 \\ 0.63, 0.64 & 0.63, 0.65 & 0.62, 0.63 & 0.62, 0.63 & 0.62, 0.63 & 0.64, 0.66 \\ 0.62, 0.63 & 0.62, 0.63 & 0.62, 0.63 & 0.62, 0.63 & 0.62, 0.63 \\ 0.81\pm0.01 & 0.82\pm0.03 & 0.8\pm0.03 & 0.81\pm0.02 & 0.83\pm0.01 & 0.82\pm0.02 & 0.81\pm0.03 & 0.82\pm0.02 \\ 0.81, 0.81 & 0.81, 0.81 & 0.8, 0.8 & 0.79, 0.79 & 0.82 & 0.82 & 0.82 & 0.81\pm0.03 & 0.65\pm0.03 \\ 0.54, 0.54 & 0.52, 0.52 & 0.56, 0.56 & 0.58, 0.6 & 0.52, 0.52 & 0.46, 0.46 & 0.5, 0.51 & 0.55, 0.55 \\ \mathrm{FA}\_vs\_MC & 0.55\pm0.02 & 0.58\pm0.08 & 0.55\pm0.03 & 0.58\pm0.06 & 0.51\pm0.05 & 0.57\pm0.07 & 0.58\pm0.07 \\ 0.36, 0.4 & 0.48, 0.48 & 0.39, 0.41 & 0.46, 0.47 & 0.5, 0.5 & 0.4, 0.4 & 0.48, 0.49 & 0.52, 0.52 \\ \mathrm{FA}\_vs\_PC & 0.63\pm0.05 & 0.63\pm0.04 & 0.55\pm0.07 & 0.51\pm0.08 & 0.58\pm0.04 & 0.58\pm0.04 & 0.65\pm0.03 & 0.65\pm0.03 \\ 0.54, 0.54 & 0.54, 0.55 & 0.47, 0.48 & 0.39, 0.41 & 0.46, 0.47 & 0.5, 0.5 & 0.4, 0.41 & 0.54, 0.55 & 0.54, 0.55 \\ \mathrm{FA}\_vs\_PC & 0.63\pm0.05 & 0.63\pm0.04 & 0.59\pm0.06 & 0.54, 0.55 & 0.54, 0.55 & 0.54, 0.55 \\ \mathrm{MC}\_vs\_AD & 0.68\pm0.06 & 0.68\pm0.05 & 0.69\pm0.04 & 0.69\pm0.05 & 0.68\pm0.05 & 0.69\pm0.05 & 0.69\pm0.04 \\ 0.58, 0.59 & 0.57, 0.57 & 0.59, 0.6 & 0.59, 0.6 & 0.54, 0.55 & 0.58, 0.59 & 0.6, 0.62 & 0.59, 0.6 \\ \mathrm{PT}\_vs\_DC & 0.85\pm0.04 & 0.86\pm0.04 & 0.87\pm0.05 & 0.88\pm0.05 & 0.88\pm0.05 & 0.89\pm0.05 & 0.89\pm0.05 \\ 0.81, 0.81 & 0.81, 0.82 & 0.82, 0.83 & 0.82, 0.82 & 0.84, 0.84 & 0.84, 0.84 & 0.82, 0.83 & 0.84, 0.84 \\ \mathrm{PT}\_vs\_MC & 0.65\pm0.04 & 0.65\pm0.03 & 0.59\pm0.06 & 0.61\pm0.03 & 0.54\pm0.05 & 0.54, 0.55 & 0.57, 0.58 \\ \mathrm{PT}\_vs\_DC & 0.65\pm0.04 & 0.65\pm0.03 & 0.58\pm0.06 & 0.61\pm0.03 & 0.54\pm0.05 & 0.54\pm0.05 & 0.58\pm0.07 & 0.61\pm0.03 & 0.54\pm0.05 & 0.54\pm0.05 & 0.58\pm0.06 & 0.61\pm0.03 & 0.54\pm0.05 & 0.54\pm0.05 & 0.54, 0.55 & 0.57, 0.58 \\ \mathrm{PT}\_vs\_MC & 0.65\pm0.04 & 0.65\pm0.03 & 0.59\pm0.04 & 0.61\pm0.03 & 0.54\pm0.03 & 0.6\pm0.07 & 0.61\pm0.07 & 0.61\pm0.07 & 0.55\pm0.04 & 0.65\pm0.04 $	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	2
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	3
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	7
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	2
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	4
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	5
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	7
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	5
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	4
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	4
$TA_v s_M C \qquad 0.63 \pm 0.06 \qquad 0.63 \pm 0.05 \qquad 0.6 \pm 0.03 \qquad 0.59 \pm 0.02 \qquad 0.62 \pm 0.04 \qquad 0.57 \pm 0.04 \qquad \textbf{0.65} \pm 0.07 \qquad 0.64 \pm 0.03 \qquad 0.59 \pm 0.02 \qquad 0.62 \pm 0.04 \qquad 0.57 \pm 0.04 \qquad \textbf{0.65} \pm 0.07 \qquad 0.64 \pm 0.03 \qquad 0.57 \pm 0.04 \qquad \textbf{0.65} \pm 0.07 \qquad 0.57 \pm 0.04 \qquad \textbf{0.65} \pm 0.07 \qquad 0.57 \pm 0.04 \qquad \textbf{0.57} \pm 0.07 \qquad $	
	3
0.58, 0.58         0.58, 0.58         0.51, 0.52         0.45, 0.47         0.41, 0.48         0.5, 0.5         0.61, 0.61         0.59, 0.59	
Average AUC         0.6988         0.7052         0.6877         0.6579         0.6867         0.6547         0.7001 <b>0.716</b>	
Average Rank         4.3235         3.4706         5.1176         6.7059         4.1176         6.2353         4.3235 <b>1.7059</b>	
Overall         [1,0,1]         [2,0,0]         [1,0,2]         [0,0,6]         [2,0,1]         [0,0,6]         [2,0,1]         [9,0,0]	

#### 6. Conclusion

In this paper, we proposed a novel *K*-nearest neighbour weighted reduced universum twin support vector machines for class imbalance learning (KWRUTSVM-CIL). To create a balance in the classes, the proposed KWRUTSVM-CIL model uses universum data. Also, the proposed model exploits the local neighbourhood information via incorporation of weight matrix in its objective function. To exploit the inter-class information, the weight vector is incorporated in the constraints of the corresponding objective functions. Different from TSVM, UTSVM and RUTSVM-CIL, we

added an extra regularization term in the objective function of the proposed KWRUTSVM-CIL formulation for maximizing the margin and ensuring that the matrices in the Wolfe dual of the proposed KWRUTSVM-CIL formulation are positive definite and structural risk minimization principle is implemented which is the marrow of statistical learning. Combination of undersampling with oversampling using universum data leads to the improved results in the class imbalance problems. The efficacy of the proposed KWRUTSVM-CIL model is demonstrated by the experimental results and the statistical analysis on benchmark KEEL and UCI datasets. As an application, we used the proposed

Knowledge-Based Systems 245 (2022) 108578

KWRUTSVM-CIL classifier for the diagnosis of diseases in biomedical domain like Alzheimer's disease and breast cancer disease. The proposed model showed highest average accuracy and lowest rank in comparison with the baseline classifiers in biomedical datasets. In future, one can explore the selection of universum data as it impacts the performance of the model.

#### **CRediT** authorship contribution statement

**M.A. Ganaie:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **M. Tanveer:** Conceptualization, Methodology, Validation, Supervision, Funding acquisition, Project administration, Resources, Writing – review & editing.

#### **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This work is funded by Government of India, Science and Engineering Research Board, India (SERB) under Ramanujan Fellowship Scheme, Grant No. SB/S2/RJN-001/2016, Department of Science and Technology under Interdisciplinary Cyber Physical Systems (ICPS) Scheme grant no. DST/ICPS/CPS-Individual/2018/276 and National Supercomputing Mission under DST and Miety, Govt. of India under Grant No. DST/NSM/R&D\_HPC\_Appl/2021/03.29. We acknowledge Council of Scientific & Industrial Research (CSIR), India, New Delhi, INDIA for Extra Mural Research (EMR) grant, under grant no. 22(0751)/17/EMR-II. We are grateful for the facilities and support by Indian Institute of Technology Indore, India. The data used in this study was funded by the ADNI (National Institutes of Health Grant U01 AG024904), and DOD ADNI (Department of Defence award number W81XWH-12-2-0012) whose funding is provided by National Institute on Ageing, United States, the National Institute of Biomedical Imaging and Bioengineering, United States, and through generous contributions from the following: F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc., Switzerland; Alzheimer's Drug Discovery Foundation, United States; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company, United States; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; EuroImmun; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; AbbVie, Alzheimer's Association; Fujirebio; GE Healthcare; IXICO Ltd.; Johnson & Johnson Pharmaceutical Research & Development LLC., United States; Eli Lilly and Company, United States; Lumosity; Lundbeck; Neurotrack Technologies; Takeda Pharmaceutical Company, Japan; Novartis Pharmaceuticals Corporation, United States; Pfizer Inc., United States; Piramal Imaging; Servier; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Foundation for the National Institutes of Health (www.fnih.org) provided private sector contributions. 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, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
- [2] X. Zhang, D. Jiang, T. Han, N. Wang, W. Yang, Y. Yang, Rotating machinery fault diagnosis for imbalanced data based on fast clustering algorithm and support vector machine, J. Sensors 2017 (2017).
- [3] B. Richhariya, M. Tanveer, EEG signal classification using universum support vector machine, Expert Syst. Appl. 106 (2018) 169–182.
- [4] B. Richhariya, M. Tanveer, A. Rashid, A.D.N. Initiative, Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE), Biomed. Signal Process. Control 59 (2020) 101903.
- [5] L. Shen, H. Wang, L. Da Xu, X. Ma, S. Chaudhry, W. He, Identity management based on PCA and SVM, Inf. Syst. Front. 18 (4) (2016) 711–716.
- [6] B. Richhariya, D. Gupta, Facial expression recognition using iterative universum twin support vector machine, Appl. Soft Comput. 76 (2019) 53–67.
- [7] R. Yuan, Z. Li, X. Guan, L. Xu, An SVM-based machine learning method for accurate internet traffic classification, Inf. Syst. Front. 12 (2) (2010) 149–156.
- [8] X. Wang, X. Chen, Z. Bi, Support vector machine and ROC curves for modeling of aircraft fuel consumption, J. Manag. Anal. 2 (1) (2015) 22–34.
- [9] D. Ding, F. He, L. Yuan, Z. Pan, L. Wang, M. Ros, The first step towards intelligent wire arc additive manufacturing: An automatic bead modelling system using machine learning through industrial information integration, J. Ind. Inf. Integr. 23 (2021) 100218.
- [10] Jayadeva, R. Khemchandani, S. Chandra, Twin support vector machines for pattern classification, IEEE Trans. Pattern Anal. Mach. Intell. 29 (5) (2007) 905–910.
- [11] M.A. Kumar, M. Gopal, Least squares twin support vector machines for pattern classification, Expert Syst. Appl. 36 (4) (2009) 7535–7543.
- [12] X. Peng, A v-twin support vector machine (v-TSVM) classifier and its geometric algorithms, Inform. Sci. 180 (20) (2010) 3863–3875.
- [13] C. Wang, Q. Ye, P. Luo, N. Ye, L. Fu, Robust capped  $L_1$ -norm twin support vector machine, Neural Netw. 114 (2019) 47–59.
- [14] H. Yan, Q. Ye, T. Zhang, D.-J. Yu, X. Yuan, Y. Xu, L. Fu, Least squares twin bounded support vector machines based on  $L_1$ -norm distance metric for classification, Pattern Recognit. 74 (2018) 434–447.
- [15] H. Zhao, Q. Ye, M.A. Naiem, L. Fu, Robust  $L_{2,1}$  -norm distance enhanced multi-weight vector projection support vector machine, IEEE Access 7 (2018) 3275–3286.
- [16] R. Rezvani-KhorashadiZadeh, M. Reza, WS-TWSVM: weighted structural twin support vector machine by local and global information, in: 2015 5th International Conference on Computer and Knowledge Engineering, ICCKE, IEEE, 2015, pp. 170–175.
- [17] S. Sun, X. Xie, C. Dong, Multiview learning with generalized eigenvalue proximal support vector machines, IEEE Trans. Cybern. 49 (2) (2018) 688–697.
- [18] X. Xie, Regularized multi-view least squares twin support vector machines, Appl. Intell. 48 (9) (2018) 3108–3115.
- [19] S. Ding, X. Zhang, Y. An, Y. Xue, Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification, Pattern Recognit. 67 (2017) 32–46.
- [20] J. Weston, R. Collobert, F. Sinz, L. Bottou, V. Vapnik, Inference with the universum, in: Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 1009–1016.
- [21] Z. Qi, Y. Tian, Y. Shi, Twin support vector machine with universum data, Neural Netw. 36 (2012) 112–119.
- [22] Y. Xu, M. Chen, G. Li, Least squares twin support vector machine with universum data for classification, Internat. J. Systems Sci. 47 (15) (2016) 3637–3645.
- [23] B. Richhariya, M. Tanveer, A fuzzy universum least squares twin support vector machine (FULSTSVM), Neural Comput. Appl. (2021) 1–12, http: //dx.doi.org/10.1007/s00521-021-05721-4.
- [24] F.H. Sinz, O. Chapelle, A. Agarwal, B. Schölkopf, An analysis of inference with the universum, in: NIPS, Vol. 7, 2007, p. 1.
- [25] D. Tomar, S. Agarwal, Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes, Adv. Artif. Neural Syst. 2015 (2015).
- [26] R. Batuwita, V. Palade, FSVM-CIL: fuzzy support vector machines for class imbalance learning, IEEE Trans. Fuzzy Syst. 18 (3) (2010) 558–571.
- [27] K. Li, H. Ma, A fuzzy twin support vector machine algorithm, Int. J. Appl. Innov. Eng. Manag. (IJAIEM) 2 (3) (2013) 459-465.
- [28] B.X. Wang, N. Japkowicz, Boosting support vector machines for imbalanced data sets, Knowl. Inf. Syst. 25 (1) (2010) 1–20.
- [29] Q. Fan, Z. Wang, D. Li, D. Gao, H. Zha, Entropy-based fuzzy support vector machine for imbalanced datasets, Knowl.-Based Syst. 115 (2017) 87–99.

- [30] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artificial Intelligence Res. 16 (2002) 321–357.
- [31] X.-Y. Liu, J. Wu, Z.-H. Zhou, Exploratory undersampling for class-imbalance learning, IEEE Trans. Syst. Man Cybern. B 39 (2) (2008) 539–550.
- [32] D.-J. Yu, J. Hu, Z.-M. Tang, H.-B. Shen, J. Yang, J.-Y. Yang, Improving protein-ATP binding residues prediction by boosting SVMs with random under-sampling, Neurocomputing 104 (2013) 180–190.
- [33] J. Mathew, C.K. Pang, M. Luo, W.H. Leong, Classification of imbalanced data by oversampling in kernel space of support vector machines, IEEE Trans. Neural Netw. Learn. Syst. 29 (9) (2017) 4065–4076.
- [34] B.S. Raghuwanshi, S. Shukla, Minimum variance-embedded kernelized extension of extreme learning machine for imbalance learning, Pattern Recognit. (2021) 108069.
- [35] Y. Tang, Y.-Q. Zhang, N.V. Chawla, S. Krasser, SVMS modeling for highly imbalanced classification, IEEE Trans. Syst. Man Cybern. B 39 (1) (2008) 281–288.
- [36] B. Richhariya, M. Tanveer, A robust fuzzy least squares twin support vector machine for class imbalance learning, Appl. Soft Comput. 71 (2018) 418–432.
- [37] M. Tanveer, A. Sharma, P.N. Suganthan, General twin support vector machine with pinball loss function, Inform. Sci. 494 (2019) 311–327.
- [38] M.A. Ganaie, M. Tanveer, Robust general twin support vector machine with pinball loss function, in: Machine Learning for Intelligent Multimedia Analytics, Springer, 2021, pp. 103–125.
- [39] M. Tanveer, A. Tiwari, R. Choudhary, M.A. Ganaie, Large-scale pinball twin support vector machines, Mach. Learn. (2021) 1–24, http://dx.doi.org/10. 1007/s10994-021-06061-z.
- [40] Y. Xu, Z. Yang, Y. Zhang, X. Pan, L. Wang, A maximum margin and minimum volume hyper-spheres machine with pinball loss for imbalanced data classification, Knowl.-Based Syst. 95 (2016) 75–85.
- [41] Y. Xu, Y. Zhang, J. Zhao, Z. Yang, X. Pan, KNN-based maximum margin and minimum volume hyper-sphere machine for imbalanced data classification, Int. J. Mach. Learn. Cybern. 10 (2) (2019) 357–368.
- [42] M.A. Ganaie, M. Tanveer, I. Beheshti, Brain age prediction using improved twin SVR, Neural Comput. Appl. (2022) 1–11, http://dx.doi.org/10.1007/ s00521-021-06518-1.
- [43] M.A. Ganaie, M. Tanveer, I. Beheshti, Brain age prediction with improved least squares twin SVR, IEEE J. Biomed. Health Inf. (2022) http://dx.doi.org/ 10.1109/JBHI.2022.3147524.
- [44] I. Beheshti, M.A. Ganaie, V. Paliwal, A. Rastogi, I. Razzak, M. Tanveer, Predicting brain age using machine learning algorithms: A comprehensive evaluation, IEEE J. Biomed. Health Inf. (2021) http://dx.doi.org/10.1109/ JBHI.2021.3083187.
- [45] Y.-J. Lee, O.L. Mangasarian, RSVM: Reduced support vector machines, in: Proceedings of the 2001 SIAM International Conference on Data Mining, SIAM, 2001, pp. 1–17.
- [46] M. Singh, J. Chadha, P. Ahuja, S. Chandra, Reduced twin support vector regression, Neurocomputing 74 (9) (2011) 1474–1477.
- [47] Y.-J. Lee, S.-Y. Huang, Reduced support vector machines: A statistical theory, IEEE Trans. Neural Netw. 18 (1) (2007) 1–13.
- [48] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) 21–27.
- [49] Q. Ye, C. Zhao, S. Gao, H. Zheng, Weighted twin support vector machines with local information and its application, Neural Netw. 35 (2012) 31–39.
- [50] Y. Xu, J. Yu, Y. Zhang, KNN-based weighted rough ν-twin support vector machine, Knowl.-Based Syst. 71 (2014) 303–313.
- [51] Y. Xu, L. Wang, K-nearest neighbor-based weighted twin support vector regression, Appl. Intell. 41 (1) (2014) 299–309.
- [52] Y. Xu, K-nearest neighbor-based weighted multi-class twin support vector machine, Neurocomputing 205 (2016) 430–438.
- [53] X. Pan, Y. Luo, Y. Xu, K-nearest neighbor based structural twin support vector machine, Knowl.-Based Syst. 88 (2015) 34–44.
- [54] M. Tanveer, A. Sharma, P.N. Suganthan, Least squares KNN-based weighted multiclass twin SVM, Neurocomputing 459 (2021) 454–464.
- [55] M.A. Ganaie, M. Hu, A.K. Malik, M. Tanveer, P.N. Suganthan, Ensemble deep learning: A review, 2021, arXiv preprint arXiv:2104.02395.
- [56] M.A. Ganaie, M. Tanveer, P.N. Suganthan, Oblique decision tree ensemble via twin bounded SVM, Expert Syst. Appl. 143 (2020) 113072.
- [57] M.A. Ganaie, M. Tanveer, LSTSVM classifier with enhanced features from pre-trained functional link network, Appl. Soft Comput. 93 (2020) 106305.
- [58] M. Tanveer, M.A. Ganaie, P.N. Suganthan, Ensemble of classification models with weighted functional link network, Appl. Soft Comput. 107 (2021) 107322.
- [59] M. Tanveer, T. Rajani, R. Rastogi, Y.H. Shao, M.A. Ganaie, Comprehensive review on twin support vector machines, Ann. Oper. Res. (2022) http: //dx.doi.org/10.1007/s10479-022-04575-w.

- [60] B. Richhariya, M. Tanveer, An efficient angle-based universum least squares twin support vector machine for classification, ACM Trans. Internet Technol. 21 (3) (2021) http://dx.doi.org/10.1145/3387131.
- [61] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
- [62] H. Parvin, B. Minaei-Bidgoli, H. Alizadeh, Detection of cancer patients using an innovative method for learning at imbalanced datasets, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2011, pp. 376–381.
- [63] S. Wang, X. Yao, Using class imbalance learning for software defect prediction, IEEE Trans. Reliab. 62 (2) (2013) 434–443.
- [64] M. Buda, A. Maki, M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Netw. 106 (2018) 249–259.
- [65] M. Kubat, R.C. Holte, S. Matwin, Machine learning for the detection of oil spills in satellite radar images, Mach. Learn. 30 (2) (1998) 195–215.
- [66] B. Krawczyk, M. Galar, Ł. Jeleń, F. Herrera, Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy, Appl. Soft Comput. 38 (2016) 714–726.
- [67] M.A. Ganaie, M. Tanveer, A.D.N. Initiative, Fuzzy least squares projection twin support vector machines for class imbalance learning, Appl. Soft Comput. 113 (2021) 107933.
- [68] S. Das, S. Datta, B.B. Chaudhuri, Handling data irregularities in classification: Foundations, trends, and future challenges, Pattern Recognit. 81 (2018) 674–693.
- [69] K. Yang, Z. Yu, C.P. Chen, W. Cao, H.-S. Wong, J. You, G. Han, Progressive hybrid classifier ensemble for imbalanced data, IEEE Trans. Syst. Man Cybern.: Syst. (2021).
- [70] M. Koziarski, Radial-based undersampling for imbalanced data classification, Pattern Recognit. 102 (2020) 107262.
- [71] B. Richhariya, M. Tanveer, A reduced universum twin support vector machine for class imbalance learning, Pattern Recognit. 102 (2020) 107150.
- [72] I. Nekooeimehr, S.K. Lai-Yuen, Adaptive semi-unsupervised weighted oversampling (a-SUWO) for imbalanced datasets, Expert Syst. Appl. 46 (2016) 405–416.
- [73] Z.-H. Zhou, X.-Y. Liu, Training cost-sensitive neural networks with methods addressing the class imbalance problem, IEEE Trans. Knowl. Data Eng. 18 (1) (2005) 63–77.
- [74] X. Yang, Q. Song, Y. Wang, A weighted support vector machine for data classification, Int. J. Pattern Recognit. Artif. Intell. 21 (05) (2007) 961–976.
- [75] Y.-H. Shao, W.-J. Chen, J.-J. Zhang, Z. Wang, N.-Y. Deng, An efficient weighted Lagrangian twin support vector machine for imbalanced data classification, Pattern Recognit. 47 (9) (2014) 3158–3167.
- [76] S. Rezvani, X. Wang, Class imbalance learning using fuzzy ART and intuitionistic fuzzy twin support vector machines, Inform. Sci. 578 (2021) 659–682.
- [77] C. Jimenez-Castaño, A. Alvarez-Meza, A. Orozco-Gutierrez, Enhanced automatic twin support vector machine for imbalanced data classification, Pattern Recognit. 107 (2020) 107442.
- [78] D. Dua, C. Graff, UCI machine learning repository, 2017, University of California, Irvine, School of Information and Computer Sciences, URL http: //archive.ics.uci.edu/ml.
- [79] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework, J. Mult.-Valued Logic Soft Comput. 17 (2011).
- [80] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2018) 880–893.
- [81] R.Y. Lo, W.J. Jagust, Predicting missing biomarker data in a longitudinal study of Alzheimer disease, Neurology 78 (18) (2012) 1376–1382.
- [82] E. Westman, J.-S. Muehlboeck, A. Simmons, Combining MRI and CSF measures for classification of alzheimer's disease and prediction of mild cognitive impairment conversion, Neuroimage 62 (1) (2012) 229–238.
- [83] M. Reuter, N.J. Schmansky, H.D. Rosas, B. Fischl, Within-subject template estimation for unbiased longitudinal image analysis, Neuroimage 61 (4) (2012) 1402–1418.
- [84] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (7) (2015) 1455–1462.
- [85] C. Gautam, P.K. Mishra, A. Tiwari, B. Richhariya, H.M. Pandey, S. Wang, M. Tanveer, A.D.N. Initiative, Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data, Neural Netw. 123 (2020) 191–216.
- [86] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.