Regularized robust fuzzy least squares twin support vector machine for class imbalance learning

M.A. Ganaie
Discipline of Mathematics,
Indian Institute of Technology Indore
Simrol, Indore, 453552, India
phd1901141006@iiti.ac.in

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

P.N. Suganthan
School of Electrical & Electronic Engineering,
Nanyang Technological University, Singapore
epnsugan@ntu.edu.sg

***Abstract*—Twin support vector machines (TWSVM) have been successfully applied to the classification problems. TWSVM is computationally efficient model of support vector machines (SVM). However, in real world classification problems issues of class imbalance and noise provide great challenges. Due to this, models lead to the inaccurate classification either due to higher tendency towards the majority class or due to the presence of noise. We provide an improved version of robust fuzzy least squares twin support vector machine (RFLSTSVM) known as regularized robust fuzzy least squares twin support vector machine (RRFLSTSVM) to handle the imbalance problem. The advantage of RRFLSTSVM over RFLSTSVM is that the proposed RRFLSTSVM implements the structural risk minimization principle by the introduction of regularization term in the primal formulation of the objective functions. This modification leads to the improved classification as it embodies the marrow of statistical learning theory. The proposed RRFLSTSVM doesn't require any extra assumption as the matrices resulting in the dual are positive definite. However, RFLSTSVM is based on the assumption that the inverse of the matrices resulting in the dual always exist as the matrices are positive semi-definite. To subsidize the effects of class imbalance and noise, the data samples are assigned weights via fuzzy membership function. The fuzzy membership function incorporates the imbalance ratio knowledge and assigns appropriate weights to the data samples. Unlike TWSVM which solves a pair of quadratic programming problem (QPP), the proposed RRFLSTSVM method solves a pair of system of linear equations and hence is computationally efficient. Experimental and statistical analysis show the efficacy of the proposed RRFLSTSVM method.**

***Index Terms*—Support vector machines, Twin support vector machines, Fuzzy Membership, class imbalance.**

I. Introduction

With the successful application of support vector machines (SVM) [1], [2] to the classification problems, SVM has been applied across different applications like face detection [3], [4], facial expression recognition [5], speaker identification [6], intrusion detection system [7] and so on. The performance of the SVM is better as it implements the structural risk minimization principle, however, suffers in real world application due to higher computational complexity. To improve the computational complexity of the SVM, twin SVM (TWSVM) [8] was proposed. Unlike SVM, TWSVM constructs the pair of non-parallel hyperplanes by optimizing the pair of quadratic programming problems (QPPs) in such a way that the optimal hyperplanes are proximal to the corresponding class. To reduce the complexity further, least squares twin SVM (LSTSVM) [9] solved a pair of system of linear equations instead of QPPs hence lead to faster computation. Robust and sparse linear programming TWSVM [10], [11] introduced sparseness and efficient angle based universum least squares twin support vector machine for pattern classification [12] used universum data for better generalization. TWSVM models have also been extended to multiclass problems [13] and regression problems [14]–[16]. In [17] unconstrained optimization problem was formulated which shows better generalization than TWSVM model.

The issues of class imbalance and noise are prominent in real world applications. Multiple domains like detection of faults [18], detection of defective modules in software [19] and so on mostly suffer due to the imbalance in the number of samples used for training the models and hence the classification modes are prone to be biased towards the majority class. To subsidize outlier effect of samples in the noisy data, fuzzy membership functions have been incorporated in SVM model to handle these problems in different domains like bankruptcy problem [20] and object tracking [21]. Distance based fuzzy membership function were used in fuzzy support vector machines [22] wherein the samples are weighted via its distance from the class centroid. The outlier effect is reduced as they are assigned smaller weights by the fuzzy membership function. Fuzzy least squares SVM [23] was proposed to tackle the problems in multi-class domain. Assuming that the sample is a member of both the classes with different proportion of membership weights, bilateral weighted fuzzy SVM [24] and proximal bilateral weighted fuzzy SVMs [25] were proposed. Fuzzy SVM [26] used within class structure based fuzzy membership and while as fuzzy SVM [27] maximized the partition index. To handle the multilabel classification problems, fuzzy SVMs for multilabel classification [28] was modeled. Robust energy based least squares twin SVMs (RELS-TSVM) [29] reduced noise via energy parameters and emerged as best classifier in recent evaluation [30]. Sparse and noise insensitive models [31]–[34] were introduced based on pinball loss function. TWSVM has also been used in ensemble methods to further improve the generalization ability [35].

In class imbalance learning, SVM classifiers are biased towards the majority class due to presence of majority of samples of a particular class and fewer samples of the other class which leads to more misclassifications in minority class. To overcome this problem, different approaches have been followed in the literature. FSVM-CIL [36] used different parameters and fuzzy membership functions, one-class SVM [37] used conformal kernel, boosting based SVM [38], scaling kernel-based SVM [39], hybrid sampling based SVM [40], weighted least squares projection twin SVMs with local information [41], fuzzy least squares twin SVMs [42], two-norm squared fuzzy based least squares twin parametric-margin SVM [43], fuzzy total margin based SVM [44] and maximum margin twin spheres SVM [45] for imbalanced data classification. Recently, entropy-based fuzzy SVM [46] assigned weights using information entropy. The down side of this model is that higher weights are assigned to those samples of majority which are outliers hence outliers are emphasized resulting in lower performance. Twin SVM models based on information entropy to handle the class imbalance problems have also been proposed [47]–[49]. The universum based approach, known as reduced universum twin SVM for class imbalance learning [50], used universum data points to handle the class imbalance problem. Robust fuzzy least squares twin SVMs (RFLSTSVM) for class imbalance learning [51] introduced a new membership function which takes imbalance ratio of the data samples into consideration. RFLSTSVM implements the empirical risk minimization principle and is based on the assumption that the inverse of the matrices resulting from the dual formulation always exist as the matrices are positive semi-definite. To overcome these issues in RFLSTSVM, we propose regularized robust fuzzy least squares twin SVMs (RRFLSTSVM) for class imbalance learning by introducing the regularization term in the primal formulation of the RFLSTSVM. The advantages of the proposed RRFLSTSVM over RFLSTSVM are:

- The proposed RRFLSTSVM implements structural risk minimization principle while as RFLSTSVM minimizes the empirical risk.
- The proposed RRFLSTSVM doesn't require any extra assumption as the matrices resulting in the dual of the proposed RRFLSTSVM are positive definite. However, RFLSTSVM is based on the assumption that the inverse of the matrices resulting in the dual always exist as the matrices are positive semi-definite.

The paper outline is given as: brief introduction is given in Section I, Section II gives related work, and the proposed work is discussed in Section III. Experimental evaluation and discussion are given in V and concluding remarks are given in Section VI.

In this paper, all vectors are column vectors unless transposed to a row vector. The vector of ones with appropriate dimensions is given by $e_i, i = 1, 2$. Consider a binary classification problem, with the minority class samples termed as positive samples $A \in \mathbb{R}^{m_1 \times n}$, majority class samples termed as negative samples $B \in \mathbb{R}^{m_2 \times n}$ of the training set, n is the dimension of each sample and $m_1 + m_2$ are the total number of samples in training set. Imbalance ratio (IR) is defined as

$$IR = \frac{\text{Number of negative samples}}{\text{Number of positive samples}}. \tag{1}$$

II. Related Work

Here, we will give the formulation of baseline methods- twin SVM (TWSVM) [8], least squares twin SVM (LSTSVM) [9], fuzzy twin SVMs (FTWSVM) [36] and robust fuzzzy least squares twin SVMs (RFLSTSVM) [51].

A. TWSVM

The primal formulation of the TWSVM [8] is given as:

$$\begin{aligned} \underset{w_1,b_1}{min} \quad & \frac{1}{2}\|K(A,D^t)w_1 + eb_1\|^2 + c_1 e^t \xi_1 \\ s.t. \quad & -(K(B,D^t)w_1 + eb_1) + \xi_1 \geq e, \ \xi_1 \geq 0e \end{aligned} \tag{2}$$

and

$$\begin{aligned} \underset{w_2,b_2}{min} \quad & \frac{1}{2}\|K(B,D^t)w_2 + eb_2\|^2 + c_2 e^t \xi_2 \\ s.t. \quad & (K(A,D^t)w_2 + eb_2) + \xi_2 \geq e, \ \xi_2 \geq 0e. \end{aligned} \tag{3}$$

The dual of quadratic programming problems (QPPs) (2) and (3) in terms of Lagrange multipliers α and β are given as follows:

$$\begin{aligned} \underset{\alpha}{Max} \quad & e^t\alpha - \frac{1}{2}\alpha^t G(H^tH)^{-1}G^t\alpha \\ s.t. \quad & 0e \leq \alpha \leq c_1 e, \end{aligned} \tag{4}$$

where $G = [K(B,D^t) \quad e]$ and $H = [K(A,D^t) \quad e]$, and

$$\begin{aligned} \underset{\beta}{Max} \quad & e^t\beta - \frac{1}{2}\beta^t P(Q^tQ)^{-1}P^t\beta \\ s.t. \quad & 0e \leq \beta \leq c_2 e, \end{aligned} \tag{5}$$

where $Q = [K(B,D^t) \quad e]$ and $P = [K(A,D^t) \quad e]$. After solving (4) and (5), the optimal hyperplanes are given as :

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(H^tH)^{-1}G^t\alpha, \tag{6}$$

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (Q^tQ)^{-1}P^t\beta. \tag{7}$$

B. LSTSVM

Least squares twin SVM (LSTSVM) [9] solves a system of linear equations instead of QPPs, hence this leads to the high computational efficiency.

The objective function of the LSTSVM are given as:

$$\begin{aligned} \underset{w_1,b_1}{min} \quad & \frac{1}{2}\|K(A,D^t)w_1 + e_1b_1\|^2 + \frac{c_1}{2}\xi_1^t\xi_1 \\ s.t. \quad & -(K(B,D^t)w_1 + e_2b_1) + \xi_1 = e_2, \end{aligned} \tag{8}$$

and

$$\begin{aligned} \underset{w_2,b_2}{min} \quad & \frac{1}{2}\|(K(B,D^t)w_2 + e_2b_2)\|^2 + \frac{c_2}{2}\xi_2^t\xi_2 \\ s.t. \quad & K(A,D^t)w_2 + e_1b_2 + \xi_2 = e_1. \end{aligned} \tag{9}$$

Substituting the equality constraints in the corresponding objective function, we have

$$\begin{aligned} \min_{w_1,b_1} \quad & \frac{1}{2}\|(K(A,D^t)w_1+e_1b_1)\|^2 \\ & +\frac{c_1}{2}\|K(B,D^t)w_1+e_2b_1+e_2\|^2. \end{aligned} \tag{10}$$

Take the gradient of QPP (10) w.r.t. w_1 and b_1 and set it to zero, and writing in matrix notation, we get

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(F^tF+\frac{1}{c_1}E^tE)^{-1}F^te_2, \tag{11}$$

where $E=[K(A,D^t) \quad e_1]$ and $F=[K(B,D^t) \quad e_2]$.

With the similar approach followed for solving QPP (9)

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (E^tE+\frac{1}{c_2}F^tF)^{-1}E^te_1. \tag{12}$$

C. FTWSVM

Here, the distance based fuzzy membership functions [36] is used in TWSVM [8] formulation. The primal formulation of the FTWSVM are given as

$$\begin{aligned} \min_{w_1,b_1,\xi_1} \quad & \frac{1}{2}\|K(A,D^t)w_1+e_1b_1\|^2+c_1s_2^t\xi_1 \\ s.t. \quad & -(K(B,D^t)w_1+e_2b_1)+\xi_1 \geq e_2, \xi_1 \geq 0 \end{aligned} \tag{13}$$

and

$$\begin{aligned} \min_{w_2,b_2,\xi_2} \quad & \frac{1}{2}\|K(B,D^t)w_2+e_2b_2\|^2+c_2s_1^t\xi_2 \\ s.t. \quad & K(A,D^t)w_2+e_1b_2+\xi_2 \geq e_1, \xi_2 \geq 0, \end{aligned} \tag{14}$$

where slack variables are given as ξ_1,ξ_2, penalty parameters as c_1,c_2 and s_1,s_2 denote fuzzy membership weights. Here, fuzzy membership function assigns weights based on the distance of each sample from the class centroid. If d denotes the sample distance from the class centroid and δ is a small positive integer then the fuzzy membership function is given as:

$$fmem = 1-\frac{d}{max(d)+\delta}. \tag{15}$$

Following the same procedure as in TWSVM for solving the objective functions (13) and (14), we have the dual as

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\alpha T(S^tS)T^t\alpha - e_2^t\alpha \\ s.t. \quad & 0 \leq \alpha \leq \alpha s_2c_1 \end{aligned} \tag{16}$$

and

$$\begin{aligned} \min_{\beta} \quad & \frac{1}{2}\beta S(T^tT)S^t\beta - e_1^t\beta \\ s.t. \quad & 0 \leq \beta \leq \beta s_1c_2, \end{aligned} \tag{17}$$

where $T=[K(B,D^t) \quad e_2]$ and $S=[K(A,D^t) \quad e_1]$.

The optimal hyperplanes are given as $\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(S^tS+\delta I)^{-1}T^t\alpha$ and $\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (T^tT+\delta I)^{-1}S^t\beta$, where δ is a small value to avoid the ill conditioning of the matrices.

D. RFLSTSVM

The objective functions of the robust fuzzy least squares twin SVM (RFLSTSVM) [51] are as follows:

$$\begin{aligned} \min_{w_1,b_1,\xi_1} \quad & \frac{1}{2}\|K(A,D^t)w_1+e_1b_1\|^2+\frac{c_1}{2}(s_2\xi_1)^t(s_2\xi_1) \\ s.t. \quad & -(K(B,D^t)w_1+e_2b_1)+\xi_1 = e_2, \end{aligned} \tag{18}$$

and

$$\begin{aligned} \min_{w_2,b_2,\xi_2} \quad & \frac{1}{2}\|K(B,D^t)w_2+e_2b_2\|^2+\frac{c_2}{2}(s_1\xi_2)^t(s_1\xi_2) \\ s.t. \quad & K(A,D^t)w_2+e_1b_2+\xi_2 = e_1, \end{aligned} \tag{19}$$

where A and B denote classes of the minority (class-1) and majority (class-2), respectively, s_1 and s_2 denote fuzzy membership functions and ξ_1,ξ_2 are the slack variables.

The function for assigning the fuzzy weights to the samples is given as

$$fmem = \begin{cases} 1, & \text{for positive class data points,} \\ z+z(\frac{exp(c_0(\frac{d_1-d_2}{d}-\frac{d_2}{r_2}))-exp(-2c_0))}{exp(c_0)-exp(-2c_0)}), & \text{otherwise.} \end{cases} \tag{20}$$

Here, $z=\frac{1}{1+IR}$, IR is imbalance ratio, d_1,d_2 represents the Euclidean distance of the data samples from the positive and negative class, respectively, d is distance between the class centroids, and negative class maximum distance from centroid is given by r_2, c_0 is the exponential scale of the membership function.

Following the similar approach as in LSTSVM for solving the QPPs (18) and (19), the optimal hyperplanes are given as

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(T^tT+\frac{1}{c_1}R^tR)^{-1}T^ts_2e_1, \tag{21}$$

where $R=[A \quad e_1]$ and $T=[s_2B_2 \quad s_2e_2]$ and

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (R^TR+\frac{1}{c_2}T^tT)^{-1}R^ts_1e_1, \tag{22}$$

where $R=[s_1A \quad s_1e_1]$ and $T=[B \quad e_2]$.

III. Proposed regularized Robust fuzzy least squares twin SVMs (RRFLSTSVM)

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

A. Linear RRFLSTSVM

The objective function of the proposed RRFLSTSVM for linear case is given as:

$$\begin{aligned} \min_{w_1,b_1} \quad & \frac{c_3}{2}(\|w_1\|^2+b_1^2)+\frac{1}{2}\eta_1^t\eta_1+\frac{c_1}{2}(s_2\xi_2)^t(s_2\xi_2) \\ s.t. \quad & Aw_1+e_2b_1=\eta_1, \\ & -(Bw_1+e_1b_1)+\xi_2=e_1, \end{aligned} \tag{23}$$

and

$$\min_{w_2,b_2} \quad \frac{c_4}{2}(\|w_2\|^2+b_2^2)+\frac{1}{2}\eta_2^t\eta_2+\frac{c_2}{2}(s_1\xi_1)^t(s_1\xi_1)$$
$$s.t. \quad Bw_2+e_1b_2=\eta_2,$$
$$(Aw_2+e_2b_2)+\xi_1=e_2, \tag{24}$$

where A, B are the matrices of minority and majority class and m_1, m_2 are the dimensions of fuzzy membership vectors s_2 and s_1, respectively.

Substituting the constraints of (23) in its objective function, we have

$$\min_{w_1,b_1} \quad \frac{c_3}{2}(\|w_1\|^2+b_1^2)+\frac{1}{2}\|Aw_1+e_2b_1\|^2 + \frac{c_1}{2}\|s_2((Bw_1+e_1b_1)+e_1)\|^2. \tag{25}$$

Setting the gradient of (25) w.r.t. w_1 and b_1 to zero, and writing in matrix notation, we get

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -\left[T^tT+\tfrac{1}{c_1}R^tR+\tfrac{c_3}{c_1}I\right]^{-1}(T^ts_2e_1). \tag{26}$$

where $R=[A \quad e_2]$ and $T=[s_2B \quad s_2e_1]$.

Similarly, the solution of QPP (24) is given as:

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = [R^tR+\frac{1}{c_2}T^tT+\frac{c_4}{c_2}I]^{-1}(R^ts_1e_2), \tag{27}$$

where $R=[s_1A \quad s_1e_2]$ and $T=[B \quad e_1]$.

B. Non-Linear RRFLSTSVM

The objective function of the proposed RRFLSTSVM for non-linear case is given as:

$$\min_{w_1,b_1} \quad \frac{c_3}{2}(\|w_1\|^2+b_1^2)+\frac{1}{2}\eta_1^t\eta+\frac{c_1}{2}(s_2\xi_2)^t(s_2\xi_2) \tag{28}$$
$$s.t. \quad K(A,D^t)w_1+e_2b_1=\eta_1, \tag{29}$$
$$-(K(B,D^t)w_2+e_1b_1)+\xi_2=e_1 \tag{30}$$

and

$$\min_{w_2,b_2} \quad \frac{c_4}{2}(\|w_2\|^2+b_2^2)+\frac{1}{2}\eta_2^t\eta+\frac{c_2}{2}(s_1\xi_1)^t(s_1\xi_1) \tag{31}$$
$$s.t. \quad K(B,D^t)w_2+e_1b_2=\eta_2, \tag{32}$$
$$K(A,D^t)w_2+e_2b_1+\xi_1=e_2, \tag{33}$$

where A, B denote the matrices of minority and majority class and m_1, m_2 are the dimensions of fuzzy membership vectors s_2 and s_1, respectively. Also, $D=[A;B], K(A,D^t)$ and $K(B,D^t)$ are the kernel matrices.

On the similar lines to linear case, we can obtain the following

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -\left[T^tT+\tfrac{1}{c_1}R^tR+\tfrac{c_3}{c_1}I\right]^{-1}(T^ts_2e_1). \tag{34}$$

where $R=[K(A,D^t) \quad e_2]$ and $T=[s_2K(B,D^t) \quad s_2e_1]$ and the second optimal hyperplane as follows

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = \left[R^tR+\tfrac{1}{c_2}T^tT+\tfrac{c_4}{c_2}I\right]^{-1}(R^ts_1e_2), \tag{35}$$

where $T=[K(B,D^t) \quad e_1]$ and $R=[s_1K(A,D^t) \quad s_1e_2]$.

Note that in both linear and non-linear cases of the proposed RRFLSTSVM, samples are weighted by the fuzzy membership function given in (20).

To reduce the inverse computation time, we use Sherman–Morrison–Woodbury (SMW) [52] formula.

Classification of the test sample $x \in R^n$ is based on the minimum perpendicular distance of the sample from the optimal hyperplanes $K(x^t,D^t)w_1+b_1=0$ and $K(x^t,D^t)w_2+b_2=0$.

It should be noted that adding the regularization term makes both the matrices $\left[T^tT+\frac{1}{c_1}R^tR+\frac{c_3}{c_1}I\right]^{-1}$ and $\left[R^tR+\frac{1}{c_2}T^tT+\frac{c_4}{c_2}I\right]^{-1}$ as positive definite, hence, the proposed RRFLSTSVM method is more robust and stable as compared to RFLSTSVM and LSTSVM methods.

IV. Computational Complexity

The time complexity of TWSVM is of the order of $O(2\times(\frac{m}{2})^3)$ for a balanced dataset of size m. In the formulation of LSTSVM, two matrix inversions are computed of the size $(m+1)$ where $m=m_1+m_2$ is the size of the training set. To reduce the computation further, SMW [52] is used which requires inversions of matrices smaller than $(m+1)$. The size of the matrices involved in the optimization problem of RFLSTSVM model is same as that LSTSVM model. However, additional complexity is involved for the computation of fuzzy weights. The complexity of fuzzy membership of RFLSTSVM model is $O(m_1)$ where m_1 is size of negative class samples as the samples of positive class are assigned weights as 1.

The computational complexity of the proposed RRFLSTSVM is similar to RFLSTSVM model as the size of the matrices involved in the optimization problem of the proposed RRFLSTSVM model is similar to that of RFLSTSVM model.

V. Numerical Experiments

In this subsection, we evaluate the proposed RRFLSTSVM method with the baseline methods i.e. TWSVM [8], LSTSVM [9], FTWSVM [36], RFLSTSVM [51] on various imbalanced datasets based on the accuracy and training time. The details of the datasets [53], [54] are given in Table-I. In Table-I, first column gives the dataset name, training and testing dataset sizes are given in second column, imbalance ratio in the whole dataset is given in third column and imbalance ratio of the training dataset is given in fourth column. For example, $ecoli-0-1-vs-5$ is divided into training and testing set of sizes 120×6 each, with the imbalance ratio (IR) equal to 11 in whole dataset and 16.429 in the training set.

In the given experiments, we used five-fold cross validation to evaluate the performance of the given baseline methods and the proposed RRFLSTSVM method. All the experiments were performed in Matlab R2017a on the machine Intel(R) core(TM) $i7-6700$ processor with 8GB RAM. We employed non-linear kernel (Gaussian kernel) $K(x_i,x_j)=-\|x_i-x_j\|^2/\mu^2$ where $x_i,x_j\in\mathbb{R}^m$ and μ is the kernel parameter in the given experiments. We used grid search method to obtain the optimal parameters. The optimal parameters

TABLE I: Dataset details

Datasets	(Train-size,Test-size)	IR(All)	IR(Train)
ecoli-0-1-vs-5	$(120 \times 6, 120 \times 6)$	11	16.1429
ecoli-0-1-4-7-vs-5-6	$(150 \times 6, 182 \times 6)$	12.28	10.5385
ecoli-0-2-3-4-vs-5	$(100 \times 7, 102 \times 7)$	9.1	6.69231
ecoli-0-2-6-7-vs-3-5	$(110 \times 7, 114 \times 7)$	9.18182	7.46154
ecoli-0-4-6-vs-5	$(100 \times 6, 103 \times 6)$	9.15	11.5
ecoli-0-6-7-vs-3-5	$(110 \times 7, 112 \times 7)$	9.09091	12.75
segment0	$(500 \times 19, 1808 \times 19)$	6.0152	6.04225
heart-stat	$(130 \times 13, 140 \times 13)$	0.8	0.830986
ripley	$(600 \times 2, 650 \times 2)$	1	1.07612
shuttle-c0-vs-c4	$(900 \times 9, 929 \times 9)$	13.8699	13.5161
ecoli-0-1-4-6-vs-5	$(150 \times 6, 130 \times 6)$	13	9.71429
vowel	$(500 \times 10, 488 \times 10)$	9.97778	9.86957
brwisconsin	$(300 \times 9, 383 \times 9)$	0.538288	0.875
vehicle2	$(400 \times 18, 446 \times 18)$	2.88073	2.47826
shuttle-6-vs-2-3	$(100 \times 9, 130 \times 9)$	22	19
pima	$(300 \times 8, 468 \times 8)$	1.86567	1.63158
new-thyroid1	$(100 \times 5, 115 \times 5)$	5.14286	3.54545
yeast1	$(500 \times 8, 2468 \times 8)$	2.45921	2.96825
segment0	$(500 \times 19, 1808 \times 19)$	6.0152	6.04225
ecoli0137vs26	$(180 \times 7, 131 \times 7)$	4.75926	4.625
cleve	$(150 \times 13, 147 \times 13)$	1.16788	1.23881
votes	$(200 \times 16, 235 \times 16)$	1.58929	1.5974
aus	$(300 \times 14, 390 \times 14)$	1.24756	1.12766
iono	$(200 \times 33, 151 \times 33)$	0.56	0.834862

are chosen from the following range of parameters $c_1 = [10^{-5}, ..., 10^5], c_3 = [10^{-5}, ..., 10^5], \mu = [2^{-5}, ..., 2^5], c_0 = [0.5, 1, 1.5, 2, 2.5]$. The parameters for TWSVM, LSTSVM, FTWSVM , RFLSTSVM and proposed RFLSTSVM are set $c_1 = c_2$. Also, for proposed RFLSTSVM we used $c_3 = c_4$ for evaluation.

A. Results and Discussion

The evaluation of the baseline methods and the proposed RRFLSTSVM method is performed on the datasets [53], [54]. The results obtained are given in Table-II. One can see from the given table that the proposed method achieved highest accuracy and lowest rank.

We perform statistical analysis to evaluate the effectiveness of the proposed RRFLSTSVM method. The average ranks of the given baseline methods and the proposed RRFLSTSVM method are presented in Table-II. We use Friedman test and Nemenyi post hoc test [55] to evaluate the performance of 5 methods on 24 datasets. Under null hypothesis, all the given models are assumed to be equal. The Friedman statistic χ^2 is given as follows:

$$\chi^2 = \frac{12N}{K(K+1)} \left[\sum_{j=1}^{K} R_j^2 - \frac{K(K+1)^2}{4} \right], \tag{36}$$

where K is the number of models evaluated on N datasets and

$$F_F = \frac{(N-1)\chi^2}{N \times (K-1) - \chi^2}, \tag{37}$$

where F_F is distributed with $F((K-1), (N-1)(K-1)) = F(4, 92) =$ degrees of freedom with 5 methods and 24 datasets. After calculation, we get $\chi^2 = 13.7385$ and $F_F = 3.8412$. At 5% level of significance, the critical value of $F(4, 92) = 2.465$. Since $3.8412 > 2.465$, hence we reject the null hypothesis. We use Nemenyi post hoc test to evaluate the methods in pairwise. To show that significant difference exists between the two methods, the average ranks of the two methods must differ by at least the critical difference (cd) given as

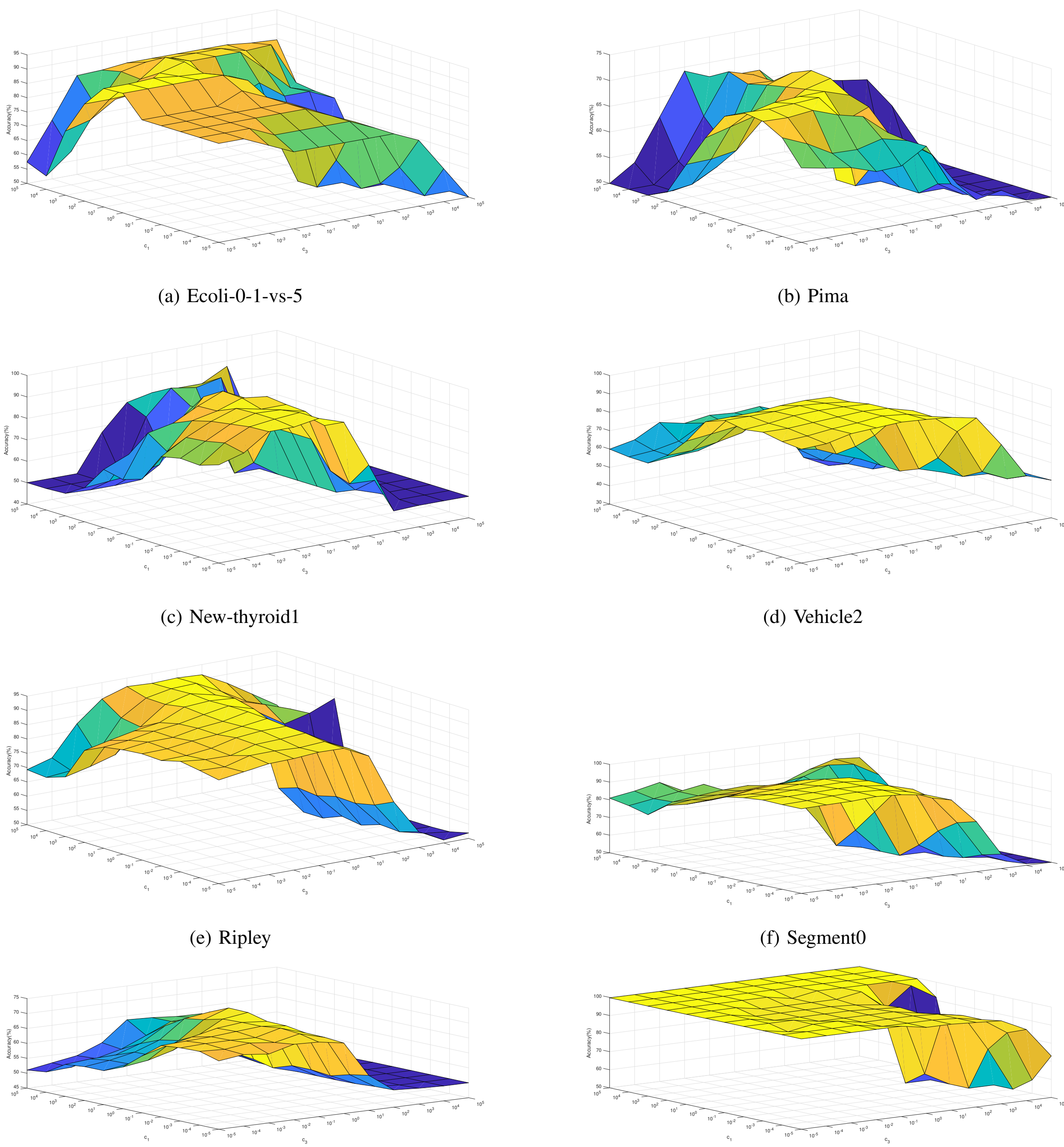
$$\text{cd} = q_\alpha \sqrt{\frac{K(K+1)}{6N}}. \tag{38}$$

At 5% level of significance, $q_{0.05} = 2.728$ for the evaluation of 5 methods. After calculation, we have $cd = 1.2452$. The difference of average ranks of pair of methods (RFLSTSVM, LSTSVM)= 1.4166, (RFLSTSVM, FTWSVM)= 1.5 which is greater than $cd = 1.2452$ hence, proposed RRFLSTSVM is better than LSTSVM and FTWSVM methods. However, Nemenyi test fails to detect the significant difference between the RRFLSTSVM, TWSVM and RRFLSTSVM, FLSTSVM. But one can see from Table-II that the proposed RRFLSTSVM achieved better performance and lower rank as compared to the given baseline methods.

Figure 1 shows the effect of parameters c_1 and c_3 on the performance of the proposed RRFLSTSVM method. In Figures (1a), (1b), (1c) and (1e), one can see that the performance is better in the middle range of c_1 and c_3 parameters. In Figure (1d), one can see that the performance of the proposed RRFLSTSVM method is lower at higher values of c_1 and c_3 and is higher in the middle range of c_1 parameters. In Figure (1h), the performance decreases at higher values of c_3 and lower values of c_1 after a certain range. Hence, given the effect of parameters the model parameters need to be chosen carefully to get the optimal performance.

VI. Conclusion and Future Work

To summarise the paper, we proposed regularized robust fuzzy least squares twin SVM (RRFLSTSVM) to handle the imbalance problem in classification tasks. In the proposed RRFLSTSVM method, regularization term is incorporated in the primal formulation of the objective function to implement the structural risk minimization principle. The proposed RRFLSTSVM method is not based on any assumptions as the matrices resulting from the dual formulation are positive definite. Hence, the proposed RRFLSTSVM model is more robust and stable as compared to RFLSTSVM method. The proposed RRFLSTSVM method is more efficient as compared to the TWSVM, the former solves the system of equations while as the latter solves a pair of quadratic programming problems for obtaining the optimal hyperplanes. The performance of the proposed RRFLSTSVM method is evaluated on multiple datasets. From the given results, one can see that the proposed RRFLSTSVM method achieved better generalization and lower rank. The statistical analysis further validate the efficiency of the proposed RRFLSTSVM method. In future, we would like to extend the work to multiclass datasets with class imbalance problems. Furthermore, solving the optimization problems more efficiently is another future directions. One can also focus on extending this work to large scale problems.



(a) Ecoli-0-1-vs-5

(b) Pima

(c) New-thyroid1

(d) Vehicle2

(e) Ripley

(f) Segment0

(g) Yeast1

(h) Shuttle-c0-vs-c4

Fig. 1: Impact of varying the parameters c_1 and c_3 on the proposed RRFLSTSVM method.

TABLE II: Comparison of classification models on multiple datasets based on non-linear kernel (Gaussian-kernel)

Datasets	TWSVM (Accuracy,Time) (c_1, μ)	LSTSVM (Accuracy,Time) (c_1, μ)	FTWSVM (Accuracy,Time) (c_1, μ)	RFLSTSVM (Accuracy,Time) (c_0, c_1, μ)	Proposed RRFLSTSVM (Accuracy,Time) (c_0, c_1, c_3, μ)
ecoli-0-1-vs-5	89.1667,0.0126 (0.00001, 0.03125)	89.1667, 0.0014 (0.00001, 0.03125)	76.9231, 0.0179 (0.00001, 16)	84.6154, 0.0107 (0.5, 0.001, 32)	91.4809, 0.0119 (2.5, 0.00001, 10, 32)
ecoli-0-1-4-7-vs-5-6	93.4066,0.0128 (0.00001, 0.03125)	93.4066, 0.0028 (0.00001, 0.03125)	83.0392, 0.0249 (0.01, 32)	90.7843, 0.0239 (0.5, 0.0001, 32)	90.7843, 0.0169 (0.5, 0.0001, 0.00001, 32)
ecoli-0-2-3-4-vs-5	93.1373,0.008 (0.00001, 0.03125)	93.1373, 0.0012 (0.00001, 0.03125)	100, 0.0141 (0.1, 32)	98.9474, 0.0114 (1.5, 0.0001, 32)	93.6842, 0.0086 (0.5, 0.00001, 10, 32)
ecoli-0-2-6-7-vs-3-5	92.1053,0.0122 (0.00001, 0.03125)	92.1053, 0.0015 (0.00001, 0.03125)	61.1111, 0.0157 (0.001, 32)	66.1905, 0.0098 (0.5, 0.1, 32)	86.9841, 0.0095 (1.5, 0.01, 1, 32)
ecoli-0-4-6-vs-5	88.3495,0.011 (0.00001, 0.03125)	88.3495, 0.0016 (0.00001, 0.03125)	75, 0.0149 (0.001, 32)	90.5678, 0.0092 (0.5, 0.00001, 32)	96.7033, 0.0099 (2, 0.00001, 10, 16)
ecoli-0-6-7-vs-3-5	87.5,0.0112 (0.00001, 0.03125)	87.5, 0.0018 (0.00001, 0.03125)	64.2857, 0.0161 (0.001, 16)	81.6327, 0.0094 (0.5, 0.0001, 32)	82.1429, 0.0095 (1, 0.001, 0.0001, 32)
segment0	91.0951,0.1464 (0.0001, 32)	91.0951, 0.0735 (0.01, 32)	97.8037, 0.2086 (100, 32)	98.902, 0.36 (0.5, 0.1, 32)	99.0958, 0.1888 (1, 0.1, 0.0001, 32)
heart-stat	55.7143,0.0211 (0.0001, 16)	55.7143, 0.002 (0.0001, 32)	68.4167, 0.0196 (0.1, 32)	64.4325, 0.0133 (0.5, 1, 32)	66.8188, 0.0149 (0.5, 0.00001, 0.001, 32)
ripley	90.9231,0.0922 (0.1, 0.03125)	89.2308, 0.0517 (0.01, 0.25)	90.9387, 0.2731 (0.1, 0.5)	91.9595, 0.2595 (1, 1, 0.5)	90.9909, 0.2569 (1, 10, 100, 0.125)
shuttle-c0-vs-c4	98.493,0.2203 (0.0001, 32)	98.1701, 0.1431 (0.01, 32)	97.541, 0.6835 (0.00001, 32)	99.9424, 0.6192 (1, 0.1, 32)	99.9424, 0.5932 (1, 0.1, 0.00001, 32)
ecoli-0-1-4-6-vs-5	95.3846,0.0125 (0.00001, 0.03125)	95.3846, 0.0021 (0.00001, 0.03125)	91.6667, 0.0254 (0.01, 32)	91.6667, 0.0173 (0.5, 0.01, 32)	95.9677, 0.0166 (2.5, 0.1, 10, 16)
vowel	97.1311,0.0715 (1, 32)	96.3115, 0.0272 (1, 2)	94.1032, 0.2012 (0.1, 32)	97.1744, 0.1837 (2, 0.1, 32)	97.0618, 0.1767 (2.5, 1, 0.0001, 32)
brwisconsin	98.1723,0.0244 (0.00001, 8)	97.6501, 0.0084 (0.0001, 4)	97.6046, 0.07 (10, 32)	97.3592, 0.0683 (2, 0.01, 16)	98.2626, 0.0712 (2, 0.1, 1, 16)
vehicle2	80.2691,0.0542 (0.0001, 32)	80.2691, 0.0264 (0.01, 32)	92.7666, 0.1315 (0.01, 32)	97.3294, 0.1195 (0.5, 0.1, 32)	94.8994, 0.1115 (2.5, 0.001, 0.0001, 32)
shuttle-6-vs-2-3	96.1538,0.0104 (0.00001, 0.03125)	96.1538, 0.0017 (0.00001, 0.03125)	60, 0.0146 (0.00001, 32)	60, 0.0112 (0.5, 0.00001, 32)	98.8, 0.0076 (1.5, 0.00001, 1, 32)
pima	67.9487,0.0371 (0.00001, 32)	67.5214, 0.0146 (0.001, 32)	67.8716, 0.0699 (0.00001, 32)	64.3478, 0.0625 (2.5, 0.001, 32)	68.2583, 0.0609 (0.5, 10, 100, 32)
new-thyroid1	98.2609,0.0185 (0.01, 32)	96.5217, 0.0016 (0.001, 16)	100, 0.0136 (0.00001, 32)	99.0196, 0.0083 (2, 0.1, 32)	99.0196, 0.0075 (2, 0.1, 0.0001, 32)
yeast1	74.8784,0.0598 (0.1, 0.25)	75.2026, 0.0337 (0.1, 1)	69.3355, 0.1882 (0.00001, 2)	69.8818, 0.1846 (1, 0.1, 2)	69.9169, 0.1832 (2.5, 0.001, 0.001, 0.25)
segment0	91.0951,0.1365 (0.0001, 32)	91.0951, 0.0723 (0.01, 32)	97.8037, 0.2128 (100, 32)	98.902, 0.2005 (0.5, 0.1, 32)	99.0958, 0.1809 (1, 0.1, 0.0001, 32)
ecoli0137vs26	94.6565,0.0335 (0.001, 0.125)	93.8931, 0.0032 (0.1, 1)	93.1818, 0.0315 (1, 1)	96.8098, 0.0228 (0.5, 0.1, 16)	95.4337, 0.0234 (1, 0.1, 100, 0.25)
cleve	76.8707,0.0192 (0.01, 32)	78.9116, 0.003 (0.1, 8)	82.5974, 0.023 (0.00001, 32)	79.1558, 0.0164 (2, 1, 16)	82.013, 0.0159 (0.5, 0.1, 0.001, 32)
votes	96.1702,0.0219 (0.01, 32)	95.3191, 0.0042 (0.0001, 8)	97.02, 0.0363 (10000, 32)	95.6311, 0.0294 (2, 1, 32)	95.6311, 0.0354 (2, 1, 0.00001, 32)
aus	84.6154,0.0235 (0.1, 32)	77.9487, 0.0104 (1, 32)	85.8514, 0.07 (0.1, 8)	86.7766, 0.0641 (1, 1, 16)	86.5426, 0.0618 (2.5, 1, 0.1, 4)
iono	92.053,0.0199 (0.1, 32)	92.7152, 0.005 (1000, 0.5)	84.9877, 0.0375 (0.1, 4)	84.8522, 0.0299 (2.5, 1, 8)	90.5665, 0.0317 (0.5, 1, 0.01, 4)
Average Accuracy	88.4813	88.0322	84.5771	86.9534	90.4207
Average Rank	3.0417	3.4583	3.5417	2.9167	2.0417

ACKNOWLEDGMENT

This work was supported by Science and Engineering Research Board (SERB) as Early Career Research Award grant no. ECR/2017/000053 and Council of Scientific & Industrial Research (CSIR), New Delhi, INDIA under Extra Mural Research (EMR) Scheme grant no. 22(0751)/17/EMR-II. We gratefully acknowledge the Indian Institute of Technology Indore for providing facilities and support.

REFERENCES

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

[2] S. R. Gunn, “Support vector machines for classification and regression,” *ISIS technical report*, vol. 14, no. 1, pp. 5–16, 1998.

[3] E. Osuna, R. Freund, and F. Girosit, “Training support vector machines: an application to face detection,” in *Proceedings of IEEE computer society conference on computer vision and pattern recognition*. IEEE, 1997, pp. 130–136.

[4] P. J. Phillips, “Support vector machines applied to face recognition,” in *Advances in Neural Information Processing Systems*, 1999, pp. 803–809.

[5] P. Michel and R. El Kaliouby, “Real time facial expression recognition in video using support vector machines,” in *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 2003, pp. 258–264.

[6] M. Schmidt and H. Gish, “Speaker identification via support vector classifiers,” in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 1. IEEE, 1996, pp. 105–108.

[7] L. Khan, M. Awad, and B. Thuraisingham, “A new intrusion detection

system using support vector machines and hierarchical clustering," *The VLDB journal*, vol. 16, no. 4, pp. 507–521, 2007.

[8] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 5, pp. 905–910, 2007.

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

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

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

[12] B. Richhariya and M. Tanveer, "An efficient angle based universum least squares twin support vector machine for pattern classification," *ACM Transactions on Internet Technology (TOIT)*, 2020.

[13] M. Tanveer, A. Sharma, and P. N. Suganthan, "Least squares KNN-based weighted multiclass twin SVM," *Neurocomputing*, 2020.

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

[15] M. Tanveer and K. Shubham, "A regularization on lagrangian twin support vector regression," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 3, pp. 807–821, 2017.

[16] S. Balasundaram and M. Tanveer, "On lagrangian twin support vector regression," *Neural Computing and Applications*, vol. 22, no. 1, pp. 257–267, 2013.

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

[18] X. Zhang, D. Jiang, T. Han, N. Wang, W. Yang, and Y. Yang, "Rotating machinery fault diagnosis for imbalanced data based on fast clustering algorithm and support vector machine," *Journal of Sensors*, vol. 2017, 2017.

[19] D. Tomar and S. Agarwal, "Prediction of defective software modules using class imbalance learning," *Applied Computational Intelligence and Soft Computing*, vol. 2016, p. 6, 2016.

[20] A. Chaudhuri and K. De, "Fuzzy support vector machine for bankruptcy prediction," *Applied Soft Computing*, vol. 11, no. 2, pp. 2472–2486, 2011.

[21] S. Zhang, S. Zhao, Y. Sui, and L. Zhang, "Single object tracking with fuzzy least squares support vector machine," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5723–5738, 2015.

[22] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE transactions on neural networks*, vol. 13, no. 2, pp. 464–471, 2002.

[23] D. Tsujinishi and S. Abe, "Fuzzy least squares support vector machines for multiclass problems," *Neural Networks*, vol. 16, no. 5-6, pp. 785–792, 2003.

[24] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 820–831, 2005.

[25] S. Balasundaram and M. Tanveer, "On proximal bilateral-weighted fuzzy support vector machine classifiers," *International Journal of Advanced Intelligence Paradigms*, vol. 4, no. 3-4, pp. 199–210, 2012.

[26] W. An and M. Liang, "Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises," *Neurocomputing*, vol. 110, pp. 101–110, 2013.

[27] Z. Wu, H. Zhang, and J. Liu, "A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method," *Neurocomputing*, vol. 125, pp. 119–124, 2014.

[28] S. Abe, "Fuzzy support vector machines for multilabel classification," *Pattern Recognition*, vol. 48, no. 6, pp. 2110–2117, 2015.

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

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

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

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

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

[34] M. Tanveer, S. Sharma, R. Rastogi, and P. Anand, "Sparse support vector machine with pinball loss," *Wiley Transactions on Emerging Telecommunications Technologies*.

[35] M. A. Ganaie, M. Tanveer, and P. N. Suganthan, "Oblique decision tree ensemble via twin bounded SVM," *Expert Systems with Applications*, vol. 143, p. 113072, 2020.

[36] R. Batuwita and V. Palade, "FSVM-CIL: fuzzy support vector machines for class imbalance learning," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 558–571, 2010.

[37] G. Cohen, M. Hilario, and C. Pellegrini, "One-class support vector machines with a conformal kernel. a case study in handling class imbalance," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2004, pp. 850–858.

[38] B. X. Wang and N. Japkowicz, "Boosting support vector machines for imbalanced data sets," *Knowledge and information systems*, vol. 25, no. 1, pp. 1–20, 2010.

[39] Y. Zhang, P. Fu, W. Liu, and G. Chen, "Imbalanced data classification based on scaling kernel-based support vector machine," *Neural Computing and Applications*, vol. 25, no. 3-4, pp. 927–935, 2014.

[40] Q. Wang, "A hybrid sampling SVM approach to imbalanced data classification," in *Abstract and Applied Analysis*, vol. 2014. Hindawi, 2014.

[41] X. Hua and S. Ding, "Weighted least squares projection twin support vector machines with local information," *Neurocomputing*, vol. 160, pp. 228–237, 2015.

[42] J. S. Sartakhti, H. Afrabandpey, and N. Ghadiri, "Fuzzy least squares twin support vector machines," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 402–409, 2019.

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

[44] H.-L. Dai, "Class imbalance learning via a fuzzy total margin based support vector machine," *Applied Soft Computing*, vol. 31, pp. 172–184, 2015.

[45] Y. Xu, "Maximum margin of twin spheres support vector machine for imbalanced data classification," *IEEE transactions on cybernetics*, vol. 47, no. 6, pp. 1540–1550, 2016.

[46] Q. Fan, Z. Wang, D. Li, D. Gao, and H. Zha, "Entropy-based fuzzy support vector machine for imbalanced datasets," *Knowledge-Based Systems*, vol. 115, pp. 87–99, 2017.

[47] D. Gupta, B. Richhariya, and P. Borah, "A fuzzy twin support vector machine based on information entropy for class imbalance learning," *Neural Computing and Applications*, vol. 31, no. 11, pp. 7153–7164, 2019.

[48] B. Richhariya and M. Tanveer, "A fuzzy universum support vector machine based on information entropy," in *Machine Intelligence and Signal Analysis*. Springer, 2019, pp. 569–582.

[49] D. Gupta and B. Richhariya, "Entropy based fuzzy least squares twin support vector machine for class imbalance learning," *Applied Intelligence*, vol. 48, no. 11, pp. 4212–4231, 2018.

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

[51] ——, "A robust fuzzy least squares twin support vector machine for class imbalance learning," *Applied Soft Computing*, vol. 71, pp. 418–432, 2018.

[52] G. Golub and C. Van Loan, "Matrix computations 3rd edition the john hopkins university press," *Baltimore, MD*, 1996.

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

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

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