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 (RFLSTSV) known as regularized robust fuzzy least squares twin support vector machine (RRFLSTSV) to handle the imbalance problem. The advantage of RRFLSTSV over RFLSTSV is that the proposed RRFLSTSV 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 narrow of statistical learning theory. The proposed RRFLSTSV doesn’t require any extra assumption as the matrices resulting in the dual are positive definite. However, RFLSTSV 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 RRFLSTSV method solves a pair of system of linear equations and hence is computationally efficient. Experimental and statistical analysis show the efficacy of the proposed RRFLSTSV 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 (LSTSV) [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 univerism least squares twin support vector machine for pattern classification [12] used univerism 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 RFLSTSVM 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 1, 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 fuzzy least squares twin SVMs (RFLSTSVM) [51].

#### A. TWSVM

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

\[
\begin{align*}
\min_{w_1, b_1} & \quad \frac{1}{2} \| K(A, D^t)w_1 + e b_1 \|^2 + c_1 e^t \xi_1 \\
\text{s.t.} \quad & - (K(B, D^t)w_1 + e b_1) + \xi_1 \geq e, \; \xi_1 \geq 0 e
\end{align*}
\]

and

\[
\begin{align*}
\min_{w_2, b_2} & \quad \frac{1}{2} \| K(B, D^t)w_2 + e b_2 \|^2 + c_2 e^t \xi_2 \\
\text{s.t.} \quad & (K(A, D^t)w_2 + e b_2) + \xi_2 \geq e, \; \xi_2 \geq 0 e. \tag{3}
\end{align*}
\]

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

\[
\begin{align*}
\max_\alpha & \quad e^t \alpha - \frac{1}{2} \alpha^t G (H^t H)^{-1} G^t \alpha \\
\text{s.t.} \quad & 0 e \leq \alpha \leq c_1 e,
\end{align*}
\]

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

\[
\begin{align*}
\max_\beta & \quad e^t \beta - \frac{1}{2} \beta^t P (Q^t Q)^{-1} P^t \beta \\
\text{s.t.} \quad & 0 e \leq \beta \leq c_2 e,
\end{align*}
\]

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

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

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

#### 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:

\[
\begin{align*}
\min_{w_1, b_1} & \quad \frac{1}{2} \| K(A, D^t)w_1 + e_1 b_1 \|^2 + \frac{c_1}{2} \xi_1^t \xi_1 \\
\text{s.t.} \quad & - (K(B, D^t)w_1 + e_1 b_1) + \xi_1 = e_2,
\end{align*}
\]

and

\[
\begin{align*}
\min_{w_2, b_2} & \quad \frac{1}{2} \| (K(B, D^t)w_2 + e_2 b_2) \|^2 + \frac{c_2}{2} \xi_2^t \xi_2 \\
\text{s.t.} \quad & K(A, D^t)w_2 + e_1 b_2 + \xi_2 = e_1.
\end{align*}
\]
and as:

positive integer then the fuzzy membership function is given as:

where slack variables are given as:

where \(w\) and \(b\) and set it to zero, and writing in matrix notation, we get:

Following the same procedure as in TWSVM for solving the QPP (9)

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:

The objective function of the proposed RRFLSTSVM for solving the QPPs (18) and (19), the optimal hyperplanes are given as:

D. RFLST SVM

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

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 \(\xi_1, \xi_2\) are the slack variables.

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

Here, \(z = \frac{1}{IR}\), \(IR\) 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:

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

The optimal hyperplanes are given as:

where \(\delta\) is a small value to avoid the ill conditioning of the matrices.
and
\[
\min_{\eta_1, \eta_2} \frac{c_1}{2}(\|w_1\|^2 + b_1^2) + \frac{1}{2} \eta_1^2 \eta_2 + \frac{c_2}{2} (s_1 \xi_1)^t (s_1 \xi_1)
\]
\[\text{s.t. } (A \eta_1 + e_2 b_1)^t + \xi_1 = e_2, \]

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
\[
\min_{w_1, b_1} \frac{c_1}{2}(\|w_1\|^2 + b_1^2) + \frac{1}{2} \|Aw_1 + e_2 b_1\|^2
\]
\[+ \frac{c_1}{2} \|s_2((Bw_1 + e_1 b_1) + e_1)\|^2. \tag{25}\]

Setting the gradient of (25) w.r.t. \( w_1 \) and \( b_1 \) to zero, and writing in matrix notation, we get
\[
\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = - \left[ T^T + \frac{1}{c_1} R^T R + \frac{c_1}{c_2} I \right]^{-1} (T^T s_2 e_1). \tag{26}\]

where \( R = [A, e_2] \) and \( T = [s_2 B, s_2 e_1] \).

Similarly, the solution of QP (24) is given as:
\[
\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = \left[ R^T R + \frac{1}{c_2} T^T T + \frac{c_1}{c_2} I \right]^{-1} (R^T s_1 e_2), \tag{27}\]

where \( R = [s_1 A, s_1 e_1] \) and \( T = [B, e_1] \).

### B. Non-Linear RRFLSTSVM

The objective function of the proposed RRFLSTSVM for non-linear case is given as:
\[
\min_{w_1, b_1} \frac{c_1}{2}(\|w_1\|^2 + b_1^2) + \frac{1}{2} \eta_1^2 \eta_2 + \frac{c_1}{2} (s_2 \xi_2)^t (s_2 \xi_2) \tag{28}\]
\[\text{s.t. } K(A, D^t) w_1 + e_2 b_1 = \eta_1, \tag{29}\]
\[− (K(B, D^t) w_2 + e_1 b_1) + \xi_1 = e_1 \tag{30}\]

and
\[
\min_{w_2, b_2} \frac{c_2}{2}(\|w_2\|^2 + b_2^2) + \frac{1}{2} \eta_2^2 \eta_2 + \frac{c_2}{2} (s_1 \xi_1)^t (s_1 \xi_1) \tag{31}\]
\[\text{s.t. } K(B, D^t) w_2 + e_1 b_2 = \eta_2, \tag{32}\]
\[K(A, D^t) w_2 + e_2 b_1 + \xi_2 = e_2 \tag{33}\]

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
\[
\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = - \left[ T^T + \frac{1}{c_1} R^T R + \frac{c_1}{c_2} I \right]^{-1} (T^t s_2 e_1). \tag{34}\]

where \( R = [K(A, D^t), e_2] \) and \( T = [s_2 K(B, D^t), s_2 e_1] \) and the second optimal hyperplane as follows
\[
\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = \left[ R^T R + \frac{1}{c_2} T^T T + \frac{c_1}{c_2} I \right]^{-1} (R^t s_1 e_2), \tag{35}\]

where \( T = [K(B, D^t) e_1] \) and \( R = [s_1 K(A, D^t) s_1 e_2] \).

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) 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 \([T^T T + \frac{1}{c_1} R^T R + \frac{c_1}{c_2} I], [R^T R + \frac{1}{c_2} T^T T + \frac{c_1}{c_2} I]^{-1}\) 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 of 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\(-0.1 - 0.5\) is divided into training and testing set of sizes 120 \(\times\) 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 \(x_i, x_j \in R^n\) and \(\mu\) 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

<table>
<thead>
<tr>
<th>Datasets</th>
<th>(Train-size, Test-size)</th>
<th>IR(All)</th>
<th>IR(Train)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ecoli-0-1-vs-5</td>
<td>(120 x 6, 120 x 6)</td>
<td>11</td>
<td>16.1429</td>
</tr>
<tr>
<td>ecoli-0-1-4-vs-5-6</td>
<td>(150 x 6, 182 x 6)</td>
<td>12.28</td>
<td>10.5385</td>
</tr>
<tr>
<td>ecoli-0-2-3-4-vs-5</td>
<td>(100 x 7, 102 x 7)</td>
<td>9.1</td>
<td>6.69221</td>
</tr>
<tr>
<td>ecoli-0-2-6-7-vs-3-5</td>
<td>(110 x 7, 114 x 7)</td>
<td>9.1812</td>
<td>7.46154</td>
</tr>
<tr>
<td>ecoli-0-4-6-vs-5</td>
<td>(100 x 6, 103 x 6)</td>
<td>9.15</td>
<td>11.75</td>
</tr>
<tr>
<td>ecoli-0-6-7-vs-3-5</td>
<td>(110 x 7, 112 x 7)</td>
<td>9.09091</td>
<td>12.75</td>
</tr>
<tr>
<td>segment0</td>
<td>(500 x 19, 1808 x 19)</td>
<td>6.0152</td>
<td>6.04225</td>
</tr>
<tr>
<td>heart-stat</td>
<td>(130 x 13, 140 x 13)</td>
<td>0.8</td>
<td>0.830986</td>
</tr>
<tr>
<td>ripley</td>
<td>(600 x 2, 650 x 2)</td>
<td>1</td>
<td>1.07612</td>
</tr>
<tr>
<td>shuttle-cb-vs-c4</td>
<td>(900 x 9, 929 x 9)</td>
<td>13.8699</td>
<td>13.5161</td>
</tr>
<tr>
<td>ecoli-0-1-4-6-vs-5</td>
<td>(150 x 6, 130 x 6)</td>
<td>13</td>
<td>9.71429</td>
</tr>
<tr>
<td>brwiscinsin</td>
<td>(300 x 9, 383 x 9)</td>
<td>0.538288</td>
<td>0.875</td>
</tr>
<tr>
<td>vehicle2</td>
<td>(400 x 18, 446 x 18)</td>
<td>2.88073</td>
<td>2.47826</td>
</tr>
<tr>
<td>shuttle-6-vs-2-3</td>
<td>(100 x 9, 130 x 9)</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>pima</td>
<td>(300 x 8, 468 x 8)</td>
<td>1.86567</td>
<td>1.63158</td>
</tr>
<tr>
<td>new-thyroid1</td>
<td>(100 x 5, 115 x 5)</td>
<td>5.14286</td>
<td>3.54545</td>
</tr>
<tr>
<td>yeast1</td>
<td>(500 x 8, 2468 x 8)</td>
<td>2.45921</td>
<td>2.96825</td>
</tr>
<tr>
<td>segment0</td>
<td>(500 x 19, 1808 x 19)</td>
<td>6.0152</td>
<td>6.04225</td>
</tr>
<tr>
<td>ecoli0137vs26</td>
<td>(180 x 7, 131 x 7)</td>
<td>4.75926</td>
<td>4.625</td>
</tr>
<tr>
<td>cleave</td>
<td>(150 x 13, 147 x 13)</td>
<td>1.16788</td>
<td>1.23881</td>
</tr>
<tr>
<td>votes</td>
<td>(200 x 16, 235 x 16)</td>
<td>1.58929</td>
<td>1.5974</td>
</tr>
<tr>
<td>aus</td>
<td>(300 x 14, 390 x 14)</td>
<td>1.24756</td>
<td>1.12766</td>
</tr>
<tr>
<td>iono</td>
<td>(200 x 33, 151 x 33)</td>
<td>0.56</td>
<td>0.834862</td>
</tr>
</tbody>
</table>

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, LST SVM, FTWSVM, RFLST SVM and proposed RFLST SVM are set $c_1 = c_2$. Also, for proposed RFLST SVM we used $c_3 = c_4$ for evaluation.

A. Results and Discussion

The evaluation of the baseline methods and the proposed RFLST SVM 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 RFLST SVM method. The average ranks of the given baseline methods and the proposed RFLST SVM 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 $\chi^2$ is given as follows:

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

(36)

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

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

(37)

where $F_F$ is distributed as $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

$$cd = q_0 \sqrt{\frac{K(K+1)}{6N}}.$$  

(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 (RFLST SVM, LST SVM)=$ 1.4166$, (RFLST SVM, FTWSVM)=$1.5$ which is greater than $cd = 1.2452$ hence, proposed RFLST SVM is better than LST SVM and FTWSVM methods. However, Nemenyi test fails to detect the significant difference between the RFLST SVM, TWSVM and RFLST SVM, FLST SVM. But one can see from Table-II that the proposed RFLST SVM 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 RFLST SVM 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 RFLST SVM 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 (RFLST SVM) to handle the imbalance problem in classification tasks. In the proposed RFLST SVM method, regularization term is incorporated in the primal formulation of the objective function to implement the structural risk minimization principle. The proposed RFLST SVM method is not based on any assumptions as the matrices resulting from the dual formulation are positive definite. Hence, the proposed RFLST SVM model is more robust and stable as compared to RFLST SVM method. The proposed RFLST SVM 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 RFLST SVM method is evaluated on multiple datasets. From the given results, one can see that the proposed RFLST SVM method achieved better generalization and lower rank. The statistical analysis further validate the efficiency of the proposed RFLST SVM 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 direction. One can also focus on extending this work to large scale problems.
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)

<table>
<thead>
<tr>
<th>Datasets</th>
<th>TWSVM (Accuracy,Time)</th>
<th>LSVM (Accuracy,Time)</th>
<th>PTWSVM (Accuracy,Time)</th>
<th>RFLSTVM (Accuracy,Time)</th>
<th>Proposed RFLSTVM (Accuracy,Time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ecoli-0-1-vs-5</td>
<td>90.9167, 0.0176</td>
<td>90.7656, 0.0185</td>
<td>90.9167, 0.0176</td>
<td>90.9167, 0.0176</td>
<td>90.9167, 0.0176</td>
</tr>
<tr>
<td>ecoli-0-1-4-7-vs-5-6</td>
<td>93.4066, 0.0176</td>
<td>93.4066, 0.0176</td>
<td>93.4066, 0.0176</td>
<td>93.4066, 0.0176</td>
<td>93.4066, 0.0176</td>
</tr>
<tr>
<td>ecoli-0-2-3-4-vs-5</td>
<td>93.1373, 0.008</td>
<td>93.1373, 0.008</td>
<td>93.1373, 0.008</td>
<td>93.1373, 0.008</td>
<td>93.1373, 0.008</td>
</tr>
<tr>
<td>ecoli-0-2-6-7-vs-3-5</td>
<td>92.1053, 0.0122</td>
<td>92.1053, 0.0122</td>
<td>92.1053, 0.0122</td>
<td>92.1053, 0.0122</td>
<td>92.1053, 0.0122</td>
</tr>
<tr>
<td>ecoli-0-4-6-vs-5</td>
<td>88.3495, 0.011</td>
<td>88.3495, 0.011</td>
<td>88.3495, 0.011</td>
<td>88.3495, 0.011</td>
<td>88.3495, 0.011</td>
</tr>
<tr>
<td>ecoli-0-6-7-vs-3-5</td>
<td>87.5, 0.0112</td>
<td>87.5, 0.0112</td>
<td>87.5, 0.0112</td>
<td>87.5, 0.0112</td>
<td>87.5, 0.0112</td>
</tr>
<tr>
<td>segment0</td>
<td>91.0951, 0.1464</td>
<td>91.0951, 0.1464</td>
<td>91.0951, 0.1464</td>
<td>91.0951, 0.1464</td>
<td>91.0951, 0.1464</td>
</tr>
<tr>
<td>heart-stat</td>
<td>55.7143, 0.0211</td>
<td>55.7143, 0.0211</td>
<td>68.4167, 0.0196</td>
<td>68.4167, 0.0196</td>
<td>70.9509, 0.0195</td>
</tr>
<tr>
<td>riple</td>
<td>90.9231, 0.0122</td>
<td>90.9231, 0.0122</td>
<td>90.9231, 0.0122</td>
<td>90.9231, 0.0122</td>
<td>90.9231, 0.0122</td>
</tr>
<tr>
<td>shuttle-c0-cv-c4</td>
<td>98.4939, 0.203</td>
<td>98.4939, 0.203</td>
<td>98.4939, 0.203</td>
<td>98.4939, 0.203</td>
<td>98.4939, 0.203</td>
</tr>
<tr>
<td>ecoli-0-1-4-6-vs-5</td>
<td>95.3846, 0.0125</td>
<td>95.3846, 0.0125</td>
<td>95.3846, 0.0125</td>
<td>95.3846, 0.0125</td>
<td>95.3846, 0.0125</td>
</tr>
<tr>
<td>vowel</td>
<td>97.1311, 0.0175</td>
<td>97.1311, 0.0175</td>
<td>97.1311, 0.0175</td>
<td>97.1311, 0.0175</td>
<td>97.1311, 0.0175</td>
</tr>
<tr>
<td>brwiscouness</td>
<td>98.1723, 0.0244</td>
<td>98.1723, 0.0244</td>
<td>98.1723, 0.0244</td>
<td>98.1723, 0.0244</td>
<td>98.1723, 0.0244</td>
</tr>
<tr>
<td>vehicle2</td>
<td>80.2691, 0.0542</td>
<td>80.2691, 0.0542</td>
<td>80.2691, 0.0542</td>
<td>80.2691, 0.0542</td>
<td>80.2691, 0.0542</td>
</tr>
<tr>
<td>shuttle-6-vs-2-3</td>
<td>96.1538, 0.0104</td>
<td>96.1538, 0.0104</td>
<td>96.1538, 0.0104</td>
<td>96.1538, 0.0104</td>
<td>96.1538, 0.0104</td>
</tr>
<tr>
<td>pima</td>
<td>67.9457, 0.0831</td>
<td>67.9457, 0.0831</td>
<td>67.9457, 0.0831</td>
<td>67.9457, 0.0831</td>
<td>67.9457, 0.0831</td>
</tr>
<tr>
<td>new-thyroid</td>
<td>98.2609, 0.0185</td>
<td>98.2609, 0.0185</td>
<td>98.2609, 0.0185</td>
<td>98.2609, 0.0185</td>
<td>98.2609, 0.0185</td>
</tr>
<tr>
<td>yeast1</td>
<td>74.8784, 0.0598</td>
<td>74.8784, 0.0598</td>
<td>74.8784, 0.0598</td>
<td>74.8784, 0.0598</td>
<td>74.8784, 0.0598</td>
</tr>
<tr>
<td>segment0</td>
<td>91.0951, 0.1365</td>
<td>91.0951, 0.1365</td>
<td>91.0951, 0.1365</td>
<td>91.0951, 0.1365</td>
<td>91.0951, 0.1365</td>
</tr>
<tr>
<td>ecoli1037vs26</td>
<td>94.6565, 0.0335</td>
<td>94.6565, 0.0335</td>
<td>94.6565, 0.0335</td>
<td>94.6565, 0.0335</td>
<td>94.6565, 0.0335</td>
</tr>
<tr>
<td>cleve</td>
<td>76.8707, 0.0192</td>
<td>76.8707, 0.0192</td>
<td>76.8707, 0.0192</td>
<td>76.8707, 0.0192</td>
<td>76.8707, 0.0192</td>
</tr>
<tr>
<td>votes</td>
<td>96.1702, 0.0219</td>
<td>96.1702, 0.0219</td>
<td>96.1702, 0.0219</td>
<td>96.1702, 0.0219</td>
<td>96.1702, 0.0219</td>
</tr>
<tr>
<td>aus</td>
<td>84.6514, 0.0235</td>
<td>84.6514, 0.0235</td>
<td>84.6514, 0.0235</td>
<td>84.6514, 0.0235</td>
<td>84.6514, 0.0235</td>
</tr>
<tr>
<td>iono</td>
<td>92.0531, 0.0191</td>
<td>92.0531, 0.0191</td>
<td>92.0531, 0.0191</td>
<td>92.0531, 0.0191</td>
<td>92.0531, 0.0191</td>
</tr>
</tbody>
</table>

Average Accuracy: 88.4813
Average Rank: 3.0417

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


