Enhancing Imbalance Learning: A Novel Slack-Factor Fuzzy SVM Approach

M. Tanveer* *Senior Member, IEEE,*, Anushka Tiwari, C.T. Lin, *Fellow, IEEE*

***Abstract*—In real-world scenarios, class imbalanced datasets are prevalent, posing challenges for algorithms like the Support Vector Machine (SVM) in effectively handling imbalance, noise, and outliers. Fuzzy Support Vector Machines (FSVMs) have emerged as a solution, leveraging varying fuzzy memberships across samples to tackle class imbalances. However, the efficacy of FSVMs is often hindered by the fuzzy membership function's sensitivity to imbalanced datasets, resulting in inaccurate assessments of sample importance and subsequently impacting FSVM performance. Addressing this challenge, a novel approach termed Slack-Factor-Based Fuzzy Support Vector Machine (SFFSVM) has been introduced. SFFSVM enhances traditional FSVMs by incorporating slack factors, which govern the relationship between optimal and estimated hyperplanes. By adjusting fuzzy membership based on the slack factor, SFFSVM effectively assigns higher membership values to misclassified majority class samples with excessively high slack factor values, thereby rectifying misclassifications induced by the hyperplane obtained via Different Error Cost (DEC). To further refine this methodology, we propose an Enhanced Slack-Factor-Based Fuzzy Support Vector Machine (ISFFSVM), integrating a novel parameter known as the location parameter. The distinguishing feature of ISFFSVM lies in constraining the DEC hyperplane from extending beyond a specified position where the slack factor scores of majority class observations approach the location parameter value. Consequently, the ISFFSVM model achieves superior classification accuracy for minority class samples compared to the SFFSVM model. Extensive experimentation on real-world KEEL datasets validates the superiority of the proposed ISFFSVM model over the baseline classifier, showcasing its efficacy in handling class imbalance and improving classification performance.**

***Index Terms*—Fuzzy membership (FM), support vector machine (SVM), different error cost (DEC), slack-factor-based fuzzy support vector machine (SFFSVM).**

I. Introduction

THE support vector machine (SVM) [1] is the algorithm that maximizes the margin between two hyperplanes leading to solving a convex quadratic programming problem. It uses a structural risk minimization (SRM) [2] technique that deals with overfitting [3] by summing up the square of the norm of all the weights. SVM performs inappropriately with imbalanced datasets (where a class's sample size is significantly higher than that of other classes) because it gets biased towards the majority class and treats the minority class as noise. Therefore, SVM is not an efficient algorithm for class imbalance learning [4] commonly exists in real world such as credit scoring [5], breast cancer malignancy [6], fraud detection [7], and so forth. Two types of methods are available in the literature to deal with imbalanced datasets: data-level and algorithm-level [8].

Data-level approaches are preprocessing techniques that make the number of samples per class identical. Many data-level methods [9] are available for class imbalanced problems, such as oversampling [10], undersampling [11], resampling [12] with different ratios, and so forth. In oversampling, we add more samples to the minority class; in undersampling, we remove some data samples from the majority class. Another preprocessing technique, *i.e.,* algorithm-level methods, aims to keep the dataset constant and alter the training algorithm. Many algorithm-level techniques, such as boundary-shifting methods [13], cost-sensitive learning [14], threshold adjustment strategy [15], scaling kernel-based [16], and so forth, are available to lessen the influence of imbalance. Cost-sensitive learning is one of the most powerful techniques, providing different weights to the different data samples. For example, Veropoulos et al. [17] came up with the Different Error Cost (DEC) Model in which the imbalance ratio (the proportion of the sizes of the samples from the positive and negative classes, denoted as IR) is the amount by which the data points from the minority class are weighted more than those from the majority class. It has been seen that the DEC model performs better than SVM but both models are sensitive to noise and outliers.

Fuzzy Support Vector Machine (FSVM) [18], a variant of SVM, reduces the issue of noise and outliers by assigning slighter weights to noise and outliers compared to other samples. FSVM initially uses a previously defined fuzzy membership (FM) function to determine the FM values of samples. The impact of noise and outliers is then eliminated when the decision plane is generated using these FM values in SVM. In the classical FSVM-CIL [19], three types of tactics are applied to describe FM functions, *i.e.,* the separation between data points and the obtained decision boundary, the separation between data points and the pre-estimated decision boundary, and the separation between data points and their own class center. The FM functions based on these three strategies have an issue of misclassification by the approximated and computed decision hyperplane. For instance, in Fig.1 (a), points A and B are equally far from the ideal hyperplane but A is more crucial in constructing the hyperplane than B [20]. Furthermore, the class imbalance is another factor for the incorrect display of the importance of the samples by FM values.

To overcome this issue, a new model named slack-factor-based fuzzy support vector machine (SFFSVM) [20] has been proposed recently. The slack factor of the samples is taken

*Corresponding Author
M. Tanveer and Anushka Tiwari are with the Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore, 453552, India (e-mail: mtanveer@iiti.ac.in, anushkatiwari9911@gmail.com).
Chin-Teng Lin is with School of Computer Science, Human-Centric AI Centre, University of Technology Sydney, Australia (e-mail: Chin-Teng.Lin@uts.edu.au).

into account in the SFFSVM model to define a new FM function. The main purpose of slack factor values is to move the hyperplane M_{dec} obtained by DEC model to the right side to reach the optimal hyperplane M^* (as depicted in Fig.1 (b)). The more the slack factor value of the sample, the higher chances of the sample being considered an outlier or noise and misclassified. Based on this observation, a new FM function is defined in the SFFSVM model so that the samples with higher slack factor values get lower FM values.

The SFFSVM model also has a drawback while defining FM values of the majority class samples given in Eq.(5). FM function defined for the majority class samples allocates high FM values to the samples, misclassified by the DEC model hyperplane, with a slack factor value less than 2. However, 2 is not always the correct choice because while shifting M_{dec} to the right side, we have to ensure that correctly classified minority class samples do not get misclassified. To overcome this issue with the SFFSVM model, we propose an improved slack-factor-based fuzzy support vector machine (ISFFSVM) model which introduces a new parameter, known as the location parameter, denoted by a. The advantage of introducing the location parameter a is that the number of minority class samples that will be misclassified after shifting M_{dec} hyperplane depends on the parameter a. Consequently, the proposed ISFFSVM model is more efficient for class imbalance learning as it attempts to correctly classify a larger number of minority samples. Moreover, Fig.3 also demonstrates that the decision boundary of the proposed ISFFSVM model is superior to the SFFSVM model in classifying minority class samples.

With the following nine models, we have compared our results: (1) Ensemble-based category: Hashing-based Undersampling Ensemble (HUE) [21]; (2) Fuzzy-based category: FSVM based on centered kernel alignment (CKA-FSVM) [22] and Fuzzy SVM for class imbalance learning [19] (FSVM-CIL: FSVM-CIL-exp and FSVM-CIL-lin); (3) Other category: Different Error Cost (DEC) [17] and Complement Naive Bayes (CNB) [23]; (4) Oversampling-based category: Polynom fit SMOTE (PF-SMOTE) [24], SMOTE-Tomeklinks [25] and MWMOTE [26]. The thorough experimental findings show that the proposed ISFFSVM model outperforms the above models.

The remaining sections of the paper are organized as follows: The approaches for class imbalance learning available in the literature are discussed in Section 2. The proposed method is explained in Section 3. In Section 4, experimental comparisons are thoroughly presented and Section 5 includes conclusion with future work.

II. Related Work

This section briefly explains the related work on the imbalanced datasets. Over the past few decades, numerous algorithm-level techniques have been proposed. Some of them are given below:

A. Different Error Cost (DEC)

In DEC [17] model, a variant of SVM, an imbalance ratio (IR) has been introduced to deal with class imbalance problems. The minority class samples' misclassification cost is IR in the DEC. It is better than SVM on the imbalanced dataset due to the involvement of the IR. The optimization problem of DEC is as follows:

$$\begin{aligned} \min_{w,b,\xi_i} \ & \frac{1}{2}\left\|w\right\|^2 + \zeta^+ \sum_{x\in X^+}^{|X^+|} \xi_i + \zeta^- \sum_{x\in X^-}^{|X^-|} \xi_i \qquad (1)\\ \text{s.t.} \ & y_i(w^T x_i + b) \geq 1-\xi_i, i = 1,2...,N,\\ & \xi_i \geq 0, i=1,2,\ldots,N, \end{aligned}$$

where $\zeta^- = \zeta$ and $\zeta^+ = \zeta * IR$.
DEC model is increasing the membership of the minority class by multiplying the corresponding slack factors with IR. In this way, it assigns different FM values to the majority and minority class samples. However, like SVM, DEC model is sensitive to noise and outliers as it cannot distinguish them properly.

B. Slack Factor Based Fuzzy Support Vector Machine (SFFSVM)

We discuss the formulation of SFFSVM [20] in this section. SFFSVM defines slack-factor-based FM which helps to define different FM values for the misclassified and correctly classified samples based on their importance. In this way, it reduces the impact of imbalance and noise or outliers on the FM values.

$$\begin{aligned} \min_{w,b,\xi_i} \ & \frac{1}{2}\|w\|^2 + \zeta \sum_{i=1}^{N} \xi_i \qquad (2)\\ \text{s.t.} \ & y_i(w^T x_i + b) \geq 1-\xi_i, i=1,2,\ldots,N,\\ & \xi_i \geq 0, i=1,2,...,N. \end{aligned}$$

The decision boundary $M_{svm} = \{x|\tilde{w}^T x + \tilde{b} \cap x \in \mathbb{R}^d\}$ can be constructed by optimizing Eq.(2) on the data set $X = \{(x_i, y_i)\}_{i=1}^N$, where $\tilde{w}$ and $\tilde{b}$ represent the best possible outcome. After obtaining M_{svm} on X, we determine the slack factor value of a sample x_i using the hinge loss function

$$\xi_i = max(0, 1 - y_i(w^T x_i + b)).$$

A slack-factor-based FM is defined below based on the observation that the likelihood of misclassifying the related sample increases with the size of the slack factor values.

$$\psi_{x_i} = \begin{cases} 1, & x_i \in T, \\ e^{(-\mu\xi_i)}, & x_i \in F. \end{cases} \qquad (3)$$

Here,

$$\begin{aligned} T &= \{x_i | x_i \in X \cap (0 \leq \xi_i < 1)\}, \\ F &= \{x_i | x_i \in X \cap (\xi_i \geq 1)\}. \end{aligned}$$

μ is the smoothness parameter which determines how smoothly the DEC hyperplane moves.

SFFSVM defines a new FM function based on the above slack-factor-based FM function. It uses DEC to find the slack factor values and then define FM values for majority and minority data samples. Conveniently, suppose that M^* and

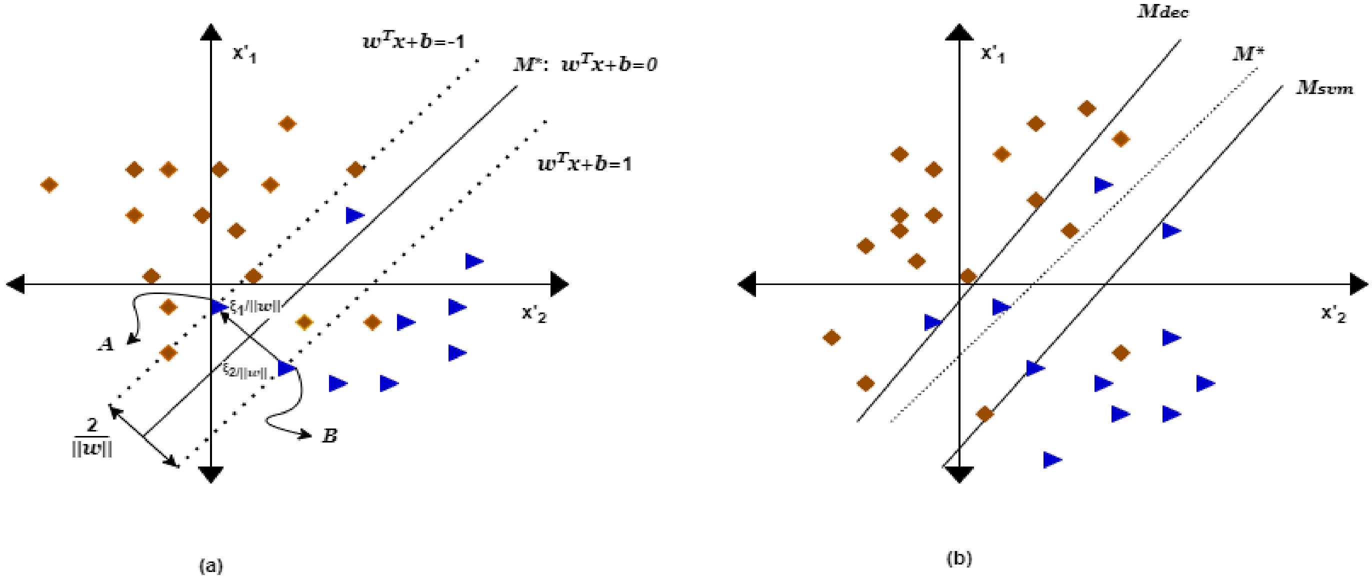


Fig. 1: Decision boundaries of SVM and DEC on an imbalanced dataset. M^* indicates optimal decision boundary; M_{dec} indicates the decision boundary obtained by DEC; M_{svm} indicates the decision boundary obtained by SVM.

M_{dec} denote the most ideal decision hyperplane on $X = \{X^+ \cup X^-\}$ and a decision hyperplane derived by instructing DEC on X, respectively. Depending on the values of ξ^+(or ξ^-) of M_{dec} can divide X$^+$(X$^-$) into $T^+(T^-)$ and $F^+(F^-)$. Note that, we apply superscripts "-" and "+" to represent the majority and minority class samples, respectively. For instance, T^+ denotes the collection of minority class samples correctly classified with a slack factor greater than or equal to zero and less than one.

The following FM values are set for the minority class samples:

$$\psi^+_{x_i} = \begin{cases} 2/(e^{\mu\xi_i}+1), & x_i \in T^+, \\ 0, & x_i \in F^+. \end{cases} \tag{4}$$

Eq.(4) depicts that the FM value for correctly classified minority class samples decreases exponentially as the slack factor value increases. At the same time, FM value for wrongly classified minority class samples is zero.

The following FM values are set for the majority class samples:

$$\psi^-_{x_i} = \begin{cases} e^{-\mu\xi_i}, & x_i \in \{x | x \in F^- \cap \xi(x) \geq 2\}, \\ 1, & x_i \in \{x | x \in F^- \cup T^- \cap \xi(x) < 2\}. \end{cases} \tag{5}$$

Here, the FM value of the majority class samples having a slack factor less than 2 is one, and for the samples with a slack factor greater than or equal to 2, FM value decreases exponentially.

Implementation of the above memberships in the DEC model can be expressed as follows:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + \zeta^+ \sum_{x\in X^+}^{|X^+|} \psi^+_{x_i}\xi_i + \zeta^- \sum_{x \in X^-}^{|X^-|} \psi^-_{x_i}\xi_i \tag{6}$$
$$\text{s.t. } y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, 2..., N,$$
$$\xi_i \geq 0, i = 1, 2, \ldots, N,$$

where $\zeta^- = \zeta$ and $\zeta^+ = \zeta * IR$.

SFFSVM considers the concept of slack factor, which helps to resolve the issue of assigning equal membership to points A and B in Fig.1 (a). Moreover, slack-factor-based FM functions are less affected by class imbalance problems because it employs DEC to find the FM values.

III. Proposed Work

FSVM-CIL has an issue of misclassification by the approximated and computed decision hyperplane. For instance, in Fig.1 (a), points A and B are equally far from the ideal hyperplane M^*, however, A is more crucial in constructing the hyperplane than B. Furthermore, the class imbalance is another factor for the incorrect display of the importance of the samples by FM values. To overcome this issue, a new slack-factor-based FM function has been proposed in the SFFSVM model. In this model, all the majority class samples wrongly classified by the DEC hyperplane, i.e., they are on the right side of the DEC hyperplane and with slack factor value less than 2, are allocated one membership value. In other words, majority class samples between the decision boundary obtained by DEC and the supporting hyperplane $w^T_{dec}x + b_{dec} = 1$ are given equal membership as the majority class samples which are correctly classified by the DEC hyperplane.

In the proposed ISFFSVM model, we introduce a new parameter a given in Eq.(9), termed as the location parameter. Unlike the SFFSVM, we assign one membership value to only those majority class samples with slack factor value less than a and greater than zero. For instance, in Fig.2, suppose that points A, B, and C have slack factor values less than a ($a < 2$) and slack factor values of points E and G are greater than or equal to a but less than 2. According to the formula given in Eq.(9), points A, B, and C will be assigned one membership value, and points E and G will be assigned a membership value less than one. In this way, it is highly likely that correctly classified minority points D and F will not be misclassified after shifting DEC hyperplane M_{dec} to the right because we are not giving high membership to E and G samples. Thus, more minority samples will be correctly classified which is our priority in class imbalance learning. However, as per the SFFSVM model, E and G will also be assigned one membership value. Consequently, there is a very high chance that minority samples D and F will be misclassified which is contrary to our objective. Also, we can verify through Fig.3 that, in comparison to SFFSVM model, the proposed ISFFSVM model successfully classifies a larger number of minority samples.

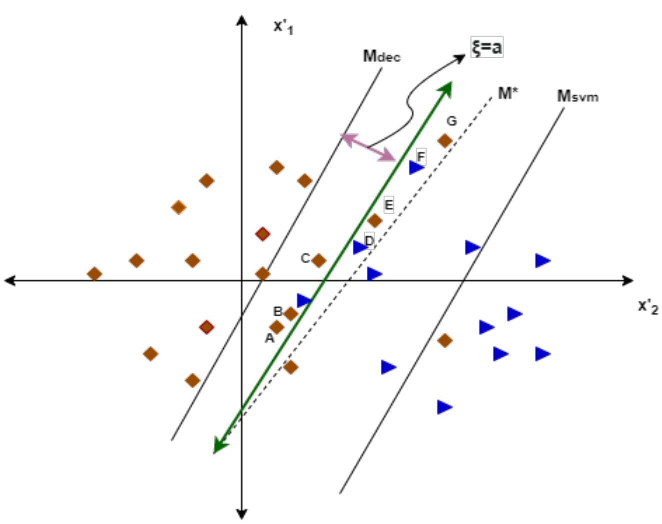


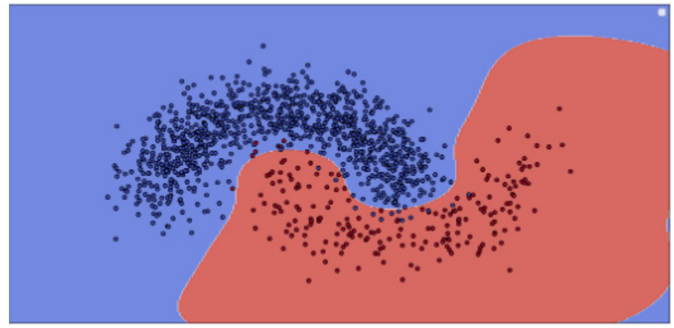
Fig. 2: Pictorial representation of the location parameter a.

The following DEC model is employed to obtain the FM values:

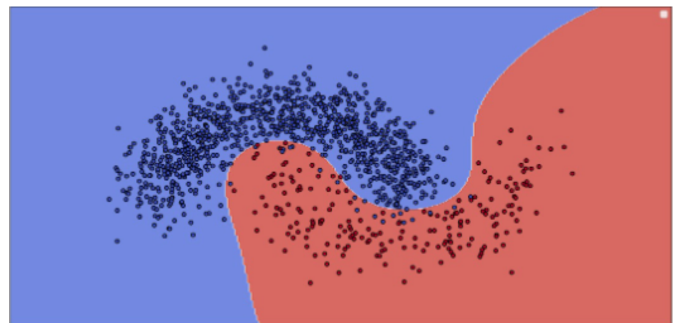
$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + \zeta^+ \sum_{x\in X^+}^{|X^+|} \xi_i + \zeta^- \sum_{x\in X^-}^{|X^-|} \xi_i \tag{7}$$
$$\text{s.t.} \;\; y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, 2..., N,$$
$$\xi_i \geq 0, i = 1, 2, ..., N.$$

FM values set for the minority class samples are as follows:

$$\psi^+_{x_i} = \begin{cases} 2/(e^{\mu\xi_i} + 1), & x_i \in T^+, \\ 0, & x_i \in F^+. \end{cases} \tag{8}$$



(a) Decision surface of SFFSVM



(b) Decision surface of ISFFSVM

Fig. 3: Decision hyperplanes of SFFSVM and ISFFSVM on dataset moon_1000_200_2 that contain $1,000$ majority class data points and 200 minority class data points (i.e., IR $= 5$).

FM values set for the majority class samples are as follows:

$$\psi^-_{x_i} = \begin{cases} e^{-\mu\xi_i}, & x_i \in \{x | x \in F^- \cap \xi(x) \geq a\}, \\ 1, & x_i \in \{x | x \in F^- \cup T^- \cap \xi(x) < a\}. \end{cases} \tag{9}$$

Implementation of the above membership function in the DEC model can be expressed as follows:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + \zeta^+ \sum_{x\in X^+}^{|X^+|} \psi^+_{x_i}\xi_i + \zeta^- \sum_{x\in X^-}^{|X^-|} \psi^-_{x_i}\xi_i \tag{10}$$
$$\text{s.t.} \;\; y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, 2..., N,$$
$$\xi_i \geq 0, i = 1, 2, ..., N,$$

where $\zeta^- = \zeta$ and $\zeta^+ = \zeta * IR$.

We are tuning the parameter a in the range $[1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2]$. It has been noted that 2 is not always the most suitable choice for the parameter a; values other than 2 are appearing as the best value for a. For instance, in Fig.3 (b), the decision boundary is drawn with a=1.3. It can be seen that the decision boundary of the proposed ISFFSVM model is better as compared to SFFSVM model. The proposed ISFFSVM model misclassifies a lesser number of minority samples than SFFSVM model; consequently, it deals with class imbalance more competently than SFFSVM model.

Unlike DEC, the proposed ISFFSVM model assigns various FM values to samples according to their significance.

Moreover, unlike SFFSVM model, it does not always take 2 as the value of the slack factor below which the majority class samples have been assigned FM value one. Due to the addition of location parameter a, the proposed ISFFSVM model provides more accurate FM values. When $a = 2$, the proposed ISFFSVM model becomes equivalent to SFFSVM model.

Properties of the proposed ISFFSVM model are as follows:

- DEC hyperplane cannot go beyond the position at which the ξ score of the majority class data points equal a.
- Majority class samples with ξ greater than or equal to a are abnormal samples and their FM values are less than zero.
- The number of majority class samples on the right side of the DEC hyperplane with FM value one depends on the parameter a.
- The number of misclassified minority class samples after shifting the M_{dec} hyperplane depends on the parameter a.

Algorithm 1 Improved Slack-Factor-Based Fuzzy Support Vector Machine

Let $X \in \mathbb{R}^{N\times m}$ be the given input dataset and Y contains the target labels.
Obtain optimal w and b using Eq.7.
for $i = 1, 2, \cdots, N$ **do**
  Calculate slack factor value for each x_i using the formula $\xi_i = max(0, 1 - y_i(w^T x_i + b))$.
  if $y_i = 1$ **then**
    if $\xi_i < 1$ put x_i into T^+, otherwise into F^+.
    Calculate FM value using Eq.(8).
  else
    if $\xi_i < 1$ put x_i into T^-, otherwise into F^-.
    Calculate FM value using Eq.(9).
  end if
end for
Apply obtained FM values in Eq.(10) and calculate the optimal w and b.

TABLE I: F1 score analysis on the low IR imbalanced datasets with the baseline models and the proposed ISFFSVM model.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Pima	65.1±3.86	66.81±3.75	66.03±3.33	62.39±4.84	62.21±5.41	62.43±4.25	58.09±5.77	64.79±3.78	53.66±4.4	61.93±12.29	63.75±4.43
Haberman	38.51±13.39	44.64±7.67	45.85±7.15	30.26±10.88	36.74±10.62	42.89±8.73	32.89±8.77	41.69±8.6	46.78±8.53	80.91±3.49	80.37±4.55
Vehicle0	94.58±3.08	93.07±3.38	92.59±3.34	93.73±2.94	94.11±2.79	94.35±2.45	94.66±2.52	89.95±2.64	53.87±3.97	94.69±2.46	95.03±93.57
Spect	50.59±10.88	56.54±12.19	54.76±11.24	52.45±11.33	49.59±12.12	46.83±12.14	51.29±10.11	51.54±9.81	22.09±9.49	57.24±13.26	60.98±12.5
Yeast1	60.57±4.96	57.8±4.14	56.68±3.72	60.38±5.75	51.93±5.51	55.18±6.41	48.02±5.48	55.02±3.57	50.78±4.8	55.21±9.57	56.24±9.09
Newthyroid1	94.73±5.95	94.76±6.1	97.27±4.11	91.99±8.32	94.74±7.28	94.43±5.75	96.41±4.24	93.13±6.45	90.5±6.3	94.8±5.71	95.86±4.73
Newthyroid2	95±5.54	94.35±8.06	94.08±7.04	91.56±7.51	95.17±5.39	94.4±7.5	96.25±5.15	92.86±6.5	89.47±6.49	93.07±5.59	94.44±5.65
Segment0	98.74±0.98	98.84±0.93	98.79±0.88	98.46±1.09	98.92±0.87	98.78±0.97	98.7±1.17	96.87±1.65	38.98±1.5	98.92±0.74	98.88±0.8
Glass6	79.04±12.2	73.78±11.57	68.44±10.7	78.76±13.07	81.31±11.75	82.5±10.41	81.93±10.03	78.41±9.31	72.46±8.87	83.56±10.17	86.67±13.25
Yeast3	74.85±4.16	71.7±4.87	71.03±4.69	78.03±3.74	68.28±9.48	71.23±9.94	69.22±10.15	72.01±3.77	61.01±4.34	76.74±4.37	76.1±6.1
Pageblocks0	83.25±2.17	79.19±2.21	80.22±2.4	83.89±2.26	84.83±2.66	86.51±1.97	82.75±2.07	80.73±2.05	61.77±4.03	83.93±1.99	84.5±2.05
Yeast05679vs4	45.2±9.48	42.23±7.78	43.26±8.34	49.79±14.92	40.42±12.64	39.27±13.36	41.34±13.36	42.97±5.78	43±9.61	45.13±9.95	47.33±9.93
Liver	60.23±6.56	61.03±8.73	62.12±6.72	55.77±8.17	60.64±7.3	58.72±6.74	59.12±7.65	60.95±5.88	61.04±5.06	62.65±5.94	62.76±6.94
Iris	91.78±5.86	92.98±5.44	92.41±5.71	92.24±4.85	92.6±4.84	91.75±5.25	91.97±5.7	93.22±4.64	64.78±4.97	99.01±2.94	99.61±1.3
Ecoli1	75.56±6.39	76.31±5.36	76.56±5.88	75.6±6.57	76.45±7.56	82.1±6.24	74.37±8.49	78.23±5.33	66.04±5.65	99.13±1.51	99.62±1.11
Mean	73.85	73.6	73.34	73.02	72.53	73.43	71.8	72.83	58.42	79.13	**80.14**

IV. Numerical Experiments

For the assessment of performance of the proposed ISFFSVM model and the baseline models, we choose real-world datasets given by Knowledge Extraction Based on Evolutionary Learning (KEEL) [27]. All datasets are divided into two groups: one with IR < 10 and another with IR $>= 10$. The experiments are performed on a machine with Python 3.7 on the system with 2 Intel Xeon processors, 128 GB of RAM, and 4 TB of secondary storage. The dataset is divided into 80:20 ratio for training and testing the models. We used the grid search method to tune the hyperparameters via a five-fold cross-validation method. To lessen the unpredictability of the studies, we run five-fold cross-validation tests ten times independently on each dataset.

Evaluation Metrics We choose the area under the Precision-Recall curves (AUC-PR), the Matthews correlation

TABLE II: MCC score analysis on the low IR imbalanced datasets with the baseline models and the proposed ISFFSVM model.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Pima	44.11±6.24	46.32±6.31	45.28±5.61	41.36±7.78	40.53±8.26	43.18±5.97	35.46±8.2	44.42±5.62	25.77±7.02	45.92±9.67	46.09±6.81
Haberman	24.46±12.41	23.18±10.35	23.41±10.25	18.44±12.36	13.19±10.92	26.48±11.26	9.18±7.71	17.16±11.45	27.27±11.44	12.21±11.43	13.34±11.32
Vehicle0	93±3.97	90.99±4.41	90.37±4.35	91.94±3.75	92.37±3.59	92.73±3.16	93.14±3.23	87.04±3.43	37.63±6.63	93.12±3.12	93.57±2.5
Spect	40.88±13.16	47.71±14.21	43.7±13.59	42.15±13.78	37.42±14.62	35.78±13.16	40.1±11.56	37.82±13.66	3.08±6.11	36.72±12.63	39.95±15.24
Yeast1	53±5.83	49.03±5.21	47.77±4.77	54.08±6.63	41.47±6.95	46.49±7.78	37.19±6.65	45.89±4.76	40.28±6.45	38.31±6.01	38.97±7.57
Newthyroid1	93.9±6.88	94±6.88	96.84±4.74	91.08±8.84	94.08±8.02	93.62±6.6	95.83±4.95	92.12±7.35	88.94±7.37	94.12±6.35	95.27±5.3
Newthyroid2	94.31±6.05	93.54±9.16	93.21±7.95	90.55±8.19	94.42±6.21	93.8±8.13	95.64±5.97	91.84±7.29	87.82±7.75	92.77±6.19	95.31±5.49
Segment0	98.53±1.13	98.66±1.08	98.6±1.02	98.21±1.26	98.75±1.01	98.59±1.11	98.5±1.34	96.37±1.92	31.39±2.61	98.75±0.86	98.7±0.93
Glass6	78.32±11.37	71.86±12.35	66.46±11.16	79.01±11.63	80.75±11.48	81.94±10.02	81.17±9.95	75.85±10.65	69.09±10.2	81.9±11.27	85.45±14.41
Yeast3	72.11±4.64	69.33±5.38	68.89±4.92	75.43±4.21	64.98±10.51	68.42±10.39	66.08±11.04	70.17±3.98	58.59±4.68	74.06±4.93	73.41±6.51
Pageblocks0	81.53±2.36	77.85±2.19	78.78±2.45	82.28±2.47	83.31±2.92	85±2.2	80.91±2.3	79.49±2.17	57.45±4.54	82.31±2.2	82.9±2.27
Yeast05679vs4	40.06±11.26	37.63±9.59	39.14±9.66	45.96±15.63	34.49±13.42	33.28±14.55	35.42±15.02	40.08±7.45	38.25±11.8	39.91±11.3	42.41±11.51
Liver	31.96±10.52	32.23±13.79	34.68±11	26.81±12.18	32.34±11.48	28.25±10.45	30.82±11.22	33.1±9.95	22.8±11.65	40.02±8.83	40.86±11.01
Iris	87.93±8.71	89.69±7.97	88.88±8.43	88.67±7.06	89.27±7.05	87.8±7.82	88.3±8.4	90.19±6.7	45.27±9.01	98.54±4.32	99.43±1.93
Ecoli1	67.93±8.61	69.24±7.21	69.6±7.96	68.07±8.75	69.25±9.95	76.82±8.09	66.58±11.07	71.8±7.09	55.94±7.87	99.05±1.65	99.58±1.21
Mean	66.8	66.08	65.71	66.27	64.44	66.15	63.62	64.89	45.97	68.52	**69.68**

TABLE III: AUC-PR score analysis on the low IR imbalanced datasets with the baseline models and the proposed ISFFSVM model.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Pima	65.56±5.78	66.67±6.42	64.19±5.36	67.71±5.77	61.89±6.97	66.58±5.77	59.3±6.36	67.45±4.79	48.97±6.25	71.24±5.94	71.69±5.5
Haberman	43.77±12.42	44.07±9.16	44.84±9.25	39.18±10.56	34.84±7.9	42.99±8.44	30.51±6.31	39.19±7.8	45.42±10.77	76.62±6.96	77.5±5.67
Vehicle0	99.06±0.81	98.72±1.12	98.63±1.33	98.89±0.85	98.88±0.91	99.1±0.66	99.25±0.58	95.41±2.59	59.86±7.24	99.08±0.8	99.16±0.83
Spect	59.35±13.42	62.07±13.71	53.64±12.45	61.24±13.15	52.72±13.35	51.89±12.36	55.19±13.65	60.33±12.25	25.22±5.59	66.87±9.88	69.08±10.64
Yeast1	61.9±5.87	61.7±6.7	60.39±6.64	62.47±6.73	52.86±7.86	56.86±7.51	46.88±5.91	62.76±5.28	47.87±6.55	58.34±4.64	58.07±5.21
Newthyroid1	98.52±3.5	99.62±1.28	99.46±1.85	98.12±3.68	98.99±2.35	99.1±1.86	99.46±1.51	98.33±3.26	98.67±2.12	99.58±1.24	99.68±1.13
Newthyroid2	99.21±1.69	98.71±3	98.65±2.51	98.52±2.62	98.89±1.92	99.06±2.23	99.06±2.52	98.51±2.41	97.76±3.25	99.92±1.38	99.76±0.77
Segment0	99.84±0.32	99.86±0.25	99.85±0.23	99.73±0.39	99.77±0.55	99.78±0.6	99.84±0.43	99.53±0.65	21.92±1.82	99.72±0.68	99.84±0.32
Glass6	94.52±5.26	94.13±6.33	93.74±6.08	95.3±5.14	94.22±5.38	92.61±6.33	95.02±5.44	92.22±7.58	82.01±13.84	88.9±10.41	94.49±7.14
Yeast3	79.85±5.75	80±6.64	81.21±6.45	80.44±6.47	71.64±10.43	76.39±10.88	72.58±12.26	81.98±4.47	51.54±7.94	81.45±5.97	81.15±5.36
Pageblocks0	84.96±3.62	82.44±3.18	82.42±3.67	87.32±3.11	86.33±3.62	88.87±2.46	83.63±3.17	91.76±1.97	64.41±6.31	82.89±3.79	82.01±4.49
Yeast05679vs4	46.49±14.65	43.29±12.61	51.19±15.53	52.3±15.04	39.95±13.99	37.44±11.7	43.14±14.64	55.88±12.38	37.81±10.36	47.06±11.8	50.52±13.39
Liver	67.48±7.62	67.78±7.57	68.8±7.03	61.5±7.59	66.69±7.47	64.3±7.68	65.98±6.51	66.93±6.82	58.03±9.61	67.31±6.99	69.45±7.58
Iris	98.12±3.59	98.74±2.26	98.3±2.38	98.15±2.09	98.27±2.53	98.57±1.75	98.21±2.19	96.77±3.82	37.06±6.96	95.13±1.94	98.37±11.47
Ecoli1	76.44±11.13	76.6±10.78	79.51±9.38	80.31±8.66	77.89±10.11	84.16±8.22	77.1±9.4	84.47±5.5	82.25±9	99.99±0.06	99.99±0.04
Mean	78.34	78.29	78.32	78.75	75.59	77.18	75.01	79.43	57.25	82.27	**83.38**

TABLE IV: F1 score analysis on the high IR imbalanced datasets with the baseline models and the proposed ISFFSVM model.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Abalone918	31.9±10.68	28.09±5.51	26.58±7.71	31.41±16.36	32.66±14.59	35.86±14.54	28.97±12.63	24.68±5	15.41±4.26	49.81±10.53	50.92±9.77
Dermatology6	62.5±21.8	71.18±20.21	63.22±21.99	38±25.21	40.99±25.32	46.36±27.96	51.33±27.89	96±5.39	91.22±8.02	96.19±7.33	97.14±21.81
Ecoli0137vs26	46.8±43.91	45.47±39.01	33.15±36.28	70±46.29	35.57±41.11	68.33±45.95	28.93±33.74	15.8±9.67	55.67±42	62±41.93	69.33±41.93
Ecoli4	81.89±12.09	81.76±13.01	67.61±14.93	86.76±10.91	79.98±12.75	80.51±12.93	78.06±14.57	56.8±11.75	48.05±9.54	77.95±13.47	78.56±16.64
Glass016vs2	35.51±22.74	37.87±19.09	36.07±16.94	14.53±23.44	41.26±22.71	44.42±29.5	37.1±25.36	31.91±10.22	10.53±8.35	37.55±19.5	40.28±20.2
Glass016vs5	67.82±27.7	36.14±21.9	43.88±25.7	53.8±33.23	60.6±32.55	58.27±30.04	55±31.62	46.36±10.85	20.61±13.13	70.28±24.72	74±27.82
Glass2	39.52±20.53	31.92±16.59	32.51±20.16	20.22±23.9	42.04±22.13	43.58±25.1	35.82±22.06	30.94±10.28	12.83±7.76	37.72±20.19	39.21±21
Glass4	72.8±24.15	61.81±21.97	66.83±20.52	69.03±26.18	72.91±22.02	81.01±20.41	87.99±15.53	55.33±11.83	36.65±11.86	69.8±22.25	73.81±20.62
Glass5	65.68±30.24	40.29±29.11	44.96±30.24	37±36.62	56±33.05	62.33±32.89	52.47±33.39	47.34±14.46	15.53±10.73	65.76±29.15	74.88±24.55
Shuttle6vs23	50.78±39.23	67.7±18.92	58.94±28.43	61.33±31.84	60±38.69	63.33±35.15	72.67±29.11	100±0	72.42±18.08	91.87±14.08	93.6±13.06
Shuttle-0vs4	98.6±1.54	97.58±1.69	98.32±1.51	98.42±1.84	98.43±1.73	98.32±1.57	97.88±1.8	100±0	99.26±1.17	98.91±1.25	99.1±1.4
WQ-R3vs5	10.1±14.56	8.35±10.75	8.26±10.35	0±0	2.8±11.26	0±0	6.39±15.44	7.64±4.07	6.4±2.47	9.88±16.18	8.92±18.39
WQ-R4	16.21±8.87	14.53±4.67	15.34±5.84	17.09±9.85	14.07±8.12	12.74±8.42	15.99±8.24	13.63±2.63	11.46±2.01	16.8±9.46	15.27±8.02
WQ-R8vs67	10.67±11.8	8.61±3.08	6.07±6.15	3.84±9.87	7.01±10.67	9.18±11.88	12.56±12.71	11.57±2.71	9.81±2.76	6.97±11.43	8.43±12.62
WQ-R8vs6	14.34±12.97	13.37±3.91	14.51±11.87	1.14±5.66	22.79±22.06	13.5±14.62	18.03±16.31	17.57±4.45	15.72±3.91	12.23±15.39	13.33±13.95
WQ-W3vs7	29.65±17.84	15.4±6.33	11.93±3.09	0±0	1.82±7.49	0±0	1.83±7.39	13.18±3.82	12.92±5.09	20.29±21.34	22.52±23.88
WQ-W9vs4	62±49.03	53.33±44.54	44.03±47.27	36±48.49	28.67±43.13	32±47.12	40±49.49	19.29±10.75	14.81±9.89	60.4±35.36	64.47±35.53
Yeast2vs8	43.39±20.02	36.15±17.88	24.55±17.74	41.01±23.96	24.92±19.74	36.47±23.33	22.49±17.5	23.89±6.09	60.58±17.3	62.91±21.31	62.98±22.64
Yeast1vs7	28.08±12.34	32.01±7.82	24.1±7.5	34.8±16.69	20.19±11.37	21.37±15.32	25.36±13.83	27.29±5.06	29.83±7.86	16.12±18.07	18.29±18.76
Yeast4	33.92±10.67	24.24±5.08	23.22±4.23	35.68±6.55	28.08±8.93	30.25±11.52	29.32±10.96	23.43±2.74	28.83±4.9	31.69±11.3	31.69±11.14
Yeast5	68.13±9.49	54.92±11.26	54.28±10.87	68.84±9.36	66.56±10.47	67.24±12.41	69.15±9.94	58.09±7.38	36.65±4.21	67.29±9.93	67.41±10.43
Aba.17vs789A	30.34±6.61	22.42±3.58	24.18±4.41	29.07±9.24	29.3±11.12	22.83±10.51	26.59±10.71	20.44±2.69	9.1±1.41	29.55±6.18	30.63±6.06
Aba.19vsABCD	7.35±5.49	6.02±2.2	6.38±3.26	5.87±9.2	8.7±9.4	8.88±10.43	8.79±9.26	7.05±1.52	4.36±1.94	11.21±7.17	10.21±7.4
Aba.21vs8	41.56±24.09	39.88±15.17	31.99±17.8	53.29±23.37	33.03±25.51	44.89±22.5	33.41±22.98	27.13±7.15	13.86±3.63	53.42±23.99	55.95±19.5
Aba.3vsB	98.57±4.33	99.14±3.43	96.57±6.16	99.6±2.83	98.57±4.33	100±0	100±0	85.13±12.59	50.54±9.8	96.71±6.91	97.79±5.72
Poker89vs5	7.04±6.97	5.89±3.61	5.3±6.19	4.32±7.96	11.37±13.26	18.44±12.26	16.31±12.21	6.65±1.97	2.1±1.01	15.76±14.98	15.91±13.75
Poker8vs6	63.84±29.29	27.75±16.2	48.83±32.29	42±32.2	62.6±32.56	63.8±30.7	76.4±21.07	5.02±2.36	1.68±0.96	92.4±12.55	89.93±19.14
Poker9vs7	20.33±35.84	37.17±26.58	7.25±9.35	0±0	24±34.36	50.67±33.16	43.33±37.04	15.05±8.84	5.83±5.41	39.04±31.02	43.98±31
Yeast6	42.62±10.95	39.49±6.62	35.06±6.62	49.92±8.63	41.34±12.47	39.53±13.36	38.12±11.08	19.74±2.57	27.08±3.71	42.68±12.9	42.33±12.57
WQ-W39vs5	11.07±11.99	11.07±5.12	9.09±9.36	3.66±8.55	7.73±13.11	2.46±7.49	3.86±7.49	7.27±1.89	4.87±1.85	9.3±12.02	11.75±12.94
Vowel0	99.12±1.61	98.61±2.39	98.93±2.76	99.44±2.62	98.99±3.19	99.51±1.06	99.45±1.1	83.04±5.06	57.83±5.3	99.62±1.08	99.46±1.21
Car	98.1±3.38	94.35±6.48	92.56±8.76	95.33±6.46	98.42±2.73	99.05±1.71	98.06±3.26	82.98±5.2	29.7±2.58	90.64±6.33	90.53±5.65
Shuttlec2vsc4	68.8±39.1	76.6±31.08	81±31.77	76±43.14	72.67±40.37	90±30.3	74.47±32.17	100±0	54.33±43.49	73.67±38.26	80.8±37.39
Mean	47.24	43	40.47	41.74	43.15	46.82	45.03	38.8	29.29	52.01	**53.74**

TABLE V: MCC score analysis on the high IR imbalanced datasets with the baseline models and the proposed ISFFSVM model.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Abalone918	30.91±11.97	28.8±7.13	26.3±10.01	30.74±16.63	29.26±15.26	32.98±15.1	25.28±13.29	24.98±7.96	12.63±7.85	48.91±11.27	49.89±10.38
Dermatology6	66.64±18.79	74.25±16.97	66.24±21.35	43.19±26.67	46.26±26.1	50.94±27.85	55.2±27.7	95.96±5.44	91.28±7.84	96.31±7	97.19±5.67
Ecoli0137vs26	48.55±43.92	48.17±39.26	36.08±37	70±46.29	37.12±41.72	68.54±45.91	31.41±35.3	24.2±13.34	57.09±42.17	63.09±41.98	70.55±41.98
Ecoli4	81.5±12.63	82.07±12.77	67.7±14.74	86.73±11.02	80.13±12.39	80.55±12.98	77.78±15.04	58.24±11.88	50.22±9.75	77.87±13.79	78.5±16.97
Glass016vs5	68.57±27.97	38.79±21.33	46.62±25.1	54.82±33.91	60.94±33.08	59.68±30.65	55.64±32.27	50.99±9.56	19.31±15.12	35.21±20.98	37.03±21.28
Glass016vs2	32.92±23.52	35.53±21.09	32.73±19.03	13.79±23.19	38.28±23.33	42.79±29.88	34.31±25.67	30.92±14.27	4.45±8.38	70.91±25.07	74.53±27.99
Glass2	36.54±21.43	29.17±18.33	29.8±21.5	19.01±23.81	39.5±23.32	41.39±25.37	33.79±22.53	31.84±14.69	8.46±10.59	36.13±21.37	37.22±21.58
Glass4	73.31±23.96	63.33±21.06	67.24±20.84	70.38±25.56	74.11±21.27	81.57±19.98	88.45±14.79	56.27±11.98	35.32±14.96	69.73±22.46	73.91±20.52
Glass5	66.78±30.55	43.74±28.45	46.32±30.66	37.62±37.33	57.1±33.64	63.66±33.22	53.04±33.99	52.51±12.67	15.23±12.88	66.69±29.43	75.72±24.55
Shuttle6vs23	51.9±39.66	70.1±17.5	61.54±28.37	63.35±32.23	61.37±38.96	64.96±35.35	74.36±28.8	100±0	74.64±16.26	92.61±12.75	94.2±11.81
Shuttle-0vs4	98.52±1.61	97.46±1.75	98.22±1.59	98.35±1.91	98.35±1.81	98.22±1.65	97.77±1.89	100±0	99.21±1.24	98.85±1.32	99.05±1.47
WQ-R3vs5	10.85±15.52	9.62±12.35	8.97±12.19	0±0	2.77±11.12	0±0	6.44±15.42	12.96±8	10.55±6.34	10.37±17.8	9.53±19.25
WQ-R4	15.19±9.12	14.57±6.61	15.03±6.98	14.81±9.89	11.2±7.96	10.32±8.04	13.27±8.36	15.11±5.43	11.86±4.58	16.25±10.22	14.11±8.45
WQ-R8vs67	10.74±12.01	12.13±5.85	8.1±8.33	3.76±9.8	6.47±9.94	8.69±11.75	11.7±12.1	16.57±6.28	13.4±6.3	7.48±12.51	9.17±14.44
WQ-R8vs6	14.33±13.19	17.87±6.89	15.64±13.52	1.08±5.34	22.78±22.74	12.94±14.83	17.11±16.03	22.3±7.93	20.63±7.35	13.64±18.11	14.25±16.71
WQ-W3vs7	30.62±18.76	20.2±7.5	16.02±5.98	0±0	1.95±8.38	0±0	1.84±7.47	17.56±7.48	16.77±8.85	20.63±22.14	23.53±25.07
WQ-W9vs4	62±49.03	54.16±44.73	44.13±47.66	36±48.49	28.96±43.4	32±47.12	40±49.49	26.63±13.39	21.25±13.51	62.76±34.92	66.56±34.86
Yeast2vs8	44.49±21.08	36.83±19.29	23.31±19.17	44.38±24.79	24.13±20.18	38.54±24.58	20.86±17.28	25.2±8.24	61.3±16.98	64.91±20.38	64.62±22.16
Yeast1vs7	24.6±13.18	30.49±10	21.07±10.13	34.5±17.51	14.8±11.46	16.8±15.15	20.38±14.88	27.19±7.76	27.08±10.5	16.68±19.89	17.03±18.68
Yeast4	33.37±11.13	29.28±5.62	27.82±5.57	36.23±7.59	26.07±9.71	28.12±12.11	27.19±11.63	30.17±3.79	32.07±6.01	31.57±11.93	30.78±11.35
Yeast5	68.07±9.73	58.28±9.73	57.52±9.73	68.73±9.68	66.11±10.87	66.88±12.68	68.82±10.16	62.16±5.96	44.68±3.64	67.45±10.04	67.54±10.79
Aba.17vs789A	32.44±7.45	28.96±4.42	30.66±5.77	27.46±9.63	28.03±11.7	21.05±10.6	24.98±10.68	27.57±4.48	11.53±3.86	31.4±7.43	32.23±7.22
Aba.19vsABCD	7.64±5.94	7.8±4.31	7.43±5.22	5.49±8.91	8±9.13	8.27±10.31	8.01±8.84	10.03±4.36	3.78±4.57	11.4±7.83	10.86±8.22
Aba.21vs8	43.23±24.54	42.74±15.25	35.98±18.13	54.76±23.7	34.08±26.62	45.33±23.06	33.67±23.81	32.58±8.12	20.36±7.54	53.82±24.28	56.29±19.64
Aba.3vsB	98.62±4.19	99.17±3.32	96.68±5.97	99.62±2.65	98.62±4.19	100±0	100±0	86.08±11.48	56.36±8.18	96.85±6.59	97.87±5.47
Poker89vs5	7.95±8.23	8.91±5.95	6.56±7.51	4.21±7.79	11.12±13.2	18.2±12.6	16.06±12.23	12.59±4.96	1.25±1.83	16.4±15.21	16.61±14.05
Poker8vs6	66.18±28.85	35.02±15.76	53.79±29.86	45.62±33.57	65.8±31.79	66.46±30.32	78.93±18.98	11.08±5.56	1.27±1.91	93.12±11.09	90.71±18.06
Poker9vs7	20.75±36.43	38.61±27.24	8.3±11.47	0±0	24.99±35.63	52.78±34.22	44.99±38.12	15.64±12.45	3.65±6.22	39.52±32.02	45.15±31.95
Yeast6	45.05±12.11	43.57±6.85	40.95±7.22	51.45±9.1	40.55±13.08	38.82±13.8	37.34±11.68	27.29±4.2	33.89±4.87	43.29±13.35	43.05±13.64
WQ-W39vs5	10.92±12.06	15.37±7.88	9.74±11.1	3.58±8.44	8.04±13.99	2.46±7.54	3.57±6.95	10.69±4.58	5.49±4.74	9.6±12.75	12.76±14.29
Vowel0	99.05±1.75	98.5±2.56	98.86±2.89	99.43±2.65	98.94±3.22	99.47±1.15	99.41±1.2	82.34±5.11	56.26±5.85	99.59±1.17	99.42±1.31
Car	98.09±3.37	94.41±6.35	92.79±8.22	95.38±6.27	98.39±2.81	99.04±1.73	98.05±3.27	83.6±4.71	37.66±2.45	90.51±6.43	90.35±5.79
Shuttlec2vsc4	69.84±38.81	78.1±30	81.79±31.26	76±43.14	73.42±40.03	90±30.3	76.32±30.82	100±0	55.12±43.69	74.41±38	81.1±37.19
Mean	47.58	45.03	41.82	42.14	42.96	46.71	44.73	41.86	30.73	52.36	**53.98**

TABLE VI: AUC-PR score analysis on the high IR imbalanced datasets with the baseline models and the proposed ISFFSVM model.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Abalone918	32.84±17.7	33.26±13.64	32.41±14.6	35.84±14.38	32.46±16.9	35.35±15.02	29.21±14	37.66±15.58	22.09±10.22	60.81±16.63	59.14±14.68
Dermatology6	86.4±34.06	93.83±23.29	100±0	100±0	100±0	100±0	100±0	99.66±1.35	100±0	100±0	100±0
Ecoli0137vs26	50.53±44.94	59.7±46.13	46.69±44.64	70.91±44.89	45.06±47.6	68.69±46.16	44.71±46.23	51±45.03	70.25±42.26	74.71±43.15	79.08±39.92
Ecoli4	87.79±13.49	92.37±11.16	80.68±14.67	93.58±10.17	85.53±13.02	88.71±12.13	84.99±13.12	84.46±15.79	37.2±18.46	88.39±10.42	90.22±12
Glass016vs5	82.22±19.77	76.72±24.52	75.68±25.85	75.32±26.86	77.44±22.97	77.29±25.93	77.92±26.02	91.3±16.94	13.34±4.15	34.47±21.51	37.69±21.51
Glass016vs2	36.11±23.19	45.97±26.61	35.91±20.52	28.56±21.26	41.23±23.83	46.69±27.17	40.06±25.46	36.13±21.21	16.57±13.9	75.71±27.28	83.5±26.03
Glass2	49.97±28	44.37±25.12	42.05±25.97	29.19±21.41	42.57±22.67	45.52±24.66	39.74±23.26	30.61±19.26	12.25±9.61	37.65±19.86	38.04±18.65
Glass4	89.23±16.74	87.51±16.22	85.25±17.86	89.71±15.42	90.28±15.33	93.33±10.57	93.51±10.67	83.55±14.56	28.38±17.76	82.22±19.16	82.41±20.92
Glass5	76.11±25.11	74.56±28.01	71.67±25.7	70.35±26.38	78.6±26.91	83.09±20.08	73.78±28.65	90.29±19.55	10.34±3.75	80.48±25.63	84.01±24.18
Shuttle6vs23	96.17±9.02	95.79±15.16	97±8.34	98.42±6.4	99.17±4.12	99.17±4.12	98.58±5.75	100±0	100±0	99.17±4.12	100±0
Shuttle-0vs4	99.99±0.02	99.99±0.04	99.98±0.05	99.99±0.04	100±0	99.99±0.02	99.97±0.05	100±0	100±0	100±0.01	100±0.01
WQ-R3vs5	9.19±12.11	9.15±12	7.31±7.92	14.76±16.82	8.21±11.91	4.91±4.69	8.33±13.32	18.87±22.32	26.13±22.89	20.89±26	21.42±26.76
WQ-R4	12.93±6.05	11.99±4.81	12.14±4.51	13.81±6.89	8.61±4.77	7.73±3.33	9.36±3.9	12.51±5.93	13.55±5.84	13.19±5.72	11.41±5.08
WQ-R8vs67	8.84±9.64	5.09±2.47	6.06±6.4	6.28±8	7.22±8.18	8.56±9.44	11.64±11.44	24.3±17.14	6.33±2.66	9.08±10.13	8.47±8.61
WQ-R8vs6	15.1±12.79	19.06±14.22	17.54±12.24	9.44±6.83	25.14±18.43	17.09±13.36	20.39±15.15	35.2±20.66	12.73±7.14	15.06±11.45	15.43±12.45
WQ-W3vs7	18.39±10.29	16.96±9.68	12.29±4.91	38.96±20.5	15.97±9.39	13.33±6.81	17.27±8.19	20.28±16.8	16.57±14.4	26.99±19.89	34.31±19.65
WQ-W9vs4	50.95±49.55	39.16±48.11	63.56±45.58	86.37±34.12	68.19±45.06	70.16±44.23	71.31±42.67	68.69±42.44	53.82±45.4	78.77±40.41	74.94±41
Yeast2vs8	53.87±22.11	48.82±24.3	31.91±17.52	55.17±23.01	24.5±17.09	46.69±22.67	21.49±17.55	58.26±20.01	56±18.09	59.55±18.17	59.75±23.42
Yeast1vs7	25.71±13.81	30.46±10.91	26.92±15.5	34.15±15.17	18.7±10.97	19.23±11.66	22.47±13.15	34.37±15.93	29.8±13.2	34.43±18.04	31.55±17.99
Yeast4	32.34±13.41	33.27±12.08	31.79±12.49	35.76±12.58	23.63±7.98	24.22±9.36	24.49±11.07	36.93±13.86	16.48±4.51	28.75±12.35	29.26±11.25
Yeast5	72.02±14.36	62.41±16.1	60.47±14.93	66.67±14.67	69.78±14.01	70.83±15.93	73.02±13.17	75.85±10.55	29.01±6.81	67.73±15.41	68.12±13.99
Aba.17vs789A	26.69±8.62	24.32±8.09	29.13±12.37	26.84±10.84	22.4±9.45	20.97±9.8	21.26±8.74	32.27±12.1	11.93±6.77	29.44±11.63	31.26±12.87
Aba.19vsABCD	6.55±8.76	7.3±6.77	10.82±9.99	5.21±5.63	7.11±5.82	8.08±7.93	5.99±5.94	5.79±4.97	5.56±4.89	7.73±5.45	8.28±6.28
Aba.21vs8	53.11±27.73	50.77±22.26	49.84±25.05	62.95±24.25	38.7±21.18	47.75±22.31	33.31±19.79	58.83±27.22	34.16±23.69	64.29±25.52	66.04±23.79
Aba.3vsB	100±0	100±0	100±0	100±0	100±0	100±0	100±0	99.75±1.77	95.48±6.71	100±0	100±0
Poker89vs5	5.11±3.71	3.76±3.92	5.47±4.38	5.73±3.9	10.67±8.74	12.96±10.04	14.81±10.17	14.68±11.68	2.59±4.74	14.62±13.22	14.4±12.67
Poker8vs6	67.42±24.75	60.38±26.96	59.71±25.71	63.65±26.26	67.31±28.43	64.95±24.32	81.6±20.7	37.06±32.01	3.05±6.71	98.57±5.79	96.53±13.66
Poker9vs7	39.58±33.22	12.86±23.86	53.58±29.71	77.96±26.25	68.72±23.11	63.32±27.29	57.52±27.27	25.44±24.37	18.17±21.83	58.15±33.9	63.67±31.56
Yeast6	47.17±18.78	52.04±13.19	49.05±16.96	58.25±16.49	34.71±13.85	36.55±14.75	34.39±13.14	60.22±17.37	16.86±4.35	48.08±17.37	48.42±14.9
WQ-W39vs5	10.35±9.43	13.25±10.47	13.34±10.78	13.55±9.76	9.93±9.01	6.49±5.4	8.98±8.21	16.89±12.56	10.86±10.55	12.48±14.84	12.46±13.28
Vowel0	99.98±0.08	99.97±0.2	99.99±0.04	99.98±0.14	99.98±0.12	100±0	99.99±0.06	95.06±3.84	79.47±8.85	100±0	100±0
Car	99.88±0.5	99.67±1.07	99.44±1.22	99.52±1.29	99.94±0.32	99.96±0.22	99.93±0.33	100±0	64.87±10.82	97.25±3.5	96.69±3.91
Shuttlec2vsc4	87.33±29.37	95.08±19.7	93.83±21.15	95.33±18.68	89.33±26.73	95.5±17.99	95.17±19.35	100±0	77.33±34.99	95.5±17.99	97±14.85
Mean	52.42	51.51	51.56	56.43	51.85	53.85	51.98	55.63	36.1	58	**58.89**

TABLE VII: Average F1-Score and Average Rank Test.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]	ISFFSVM
Average F1 Score (IR$<$ 10)	73.85	73.6	73.34	73.02	72.53	73.42	71.8	72.82	58.42	79.13	80.14
Average F1 Score (IR$>=$ 10)	47.24	43	40.47	41.74	43.15	46.82	45.03	38.8	29.29	52.01	53.74
Average Rank (IR$<$ 10)	5.73	5.6	5.53	6.73	6.37	6.53	7.2	6.93	9.53	3.43	2.4
Average Rank (IR$>=$ 10)	4.45	6.65	7.71	6.52	6.17	5.12	5.53	7.52	8.7	4.23	3.41

TABLE VIII: Pairwise Win-Tie-Loss.

	DEC [17]	FSVMCIL$_{exp}$ [19]	FSVMCIL$_{lin}$ [19]	CKAFSVM [22]	MWMOTE [26]	PFSMOTE [24]	SMOTETL [25]	HUE [21]	CNB [23]	SFFSVM [20]
ISFFSVM (IR$<$ 10)	[12,0,3]	[13,0,2]	[12,0,3]	[12,0,3]	[12,0,3]	[14,0,1]	[13,0,2]	[14,0,1]	[15,0,0]	[12,0,3]
ISFFSVM (IR$>=$ 10)	[20,0,13]	[27,0,6]	[28,0,5]	[24,0,9]	[26,0,7]	[20,0,13]	[24,0,9]	[27,0,6]	[29,0,4]	[26,0,7]

coefficient (MCC), and the F1-measure (F1) as the performance measures due to the fact that accuracy cannot accurately assess the differences between the prior approaches in the class imbalance [28, 29, 30]. To calculate these measures, we must determine the confusion matrix, which indicates how many samples a classifier properly or erroneously classifies for each class. We can easily obtain the above metrics after computing the confusion matrix by the following formulas:

$$\text{Recall (R)} = \frac{T_P}{T_P + F_N}, \tag{11}$$

$$\text{Precision (P)} = \frac{T_P}{T_P + F_P}, \tag{12}$$

$$\text{F1} = \frac{2 \times R \times P}{R + P}, \tag{13}$$

$$\text{MCC} = \frac{T_P \times T_N - F_P \times F_N}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}}, \tag{14}$$

where T_N denotes the number of correctly predicted negative data points, T_P denotes the number of correctly predicted positive data points by the model, F_N denotes false negative data points which are incorrectly classified by the model, and F_P denotes false positive data points which the model incorrectly classifies.

We follow different metrics based on analysis and statistical tests [31] for an overall comparison of the models

Average F1 score: In comparison to the baseline models, the proposed ISFFSVM model's average F1 score is larger. From Table.VII, the proposed ISFFSVM model has a higher F1 score than SFFSVM model by 1.01 % on low IR datasets and 1.73 % on high IR datasets.

Ranking Scheme: The average F1 score may be an inaccurate measure as better performance in one dataset may compensate for worse performance in others. To assess the model's performance while keeping this flaw in mind, we rank it on each dataset. In this ranking scheme, every model is scored on each dataset with the model that performed worse receiving a higher rank and the model that performed better receiving a lower rank. Let D number of datasets be used to evaluate l models, and let Δ_m^d represent the rank of the m^{th} model on the d^{th} dataset. The m^{th} algorithm's average rank is then calculated as follows:

$$R_m = \frac{\sum_{d=1}^{D} \Delta_m^d}{D}. \tag{15}$$

The proposed ISFFSVM model has a much lower rank than the competing models. It attains 2.4 rank on the low IR datasets and 3.41 rank on the high IR datasets.

Friedman Test [31]: We statistically assess the models using the Friedman test. Assuming they perform similarly, the models' average rank is equal under the null hypothesis. The Friedman test follows the chi-squared distribution (χ_F^2) with $l-1$ degrees of freedom (d.o.f.), where l is the number of models being compared. Now,

$$\chi_F^2 = \frac{12D}{l(l+1)}\left(\sum_{m=1}^{l} R_m^2 - \frac{l(l+1)^2}{4}\right), \tag{16}$$

$$F_F = \frac{(D-1)\chi_F^2}{D(l-1) - \chi_F^2}, \tag{17}$$

where the distribution of F_F has $(l-1)$ and $(l-1)(D-1)$ d.o.f. For low IR datasets ($l = 11$ and $D = 15$), we get χ_F^2=48.75 and F_F=6.74 and for high IR datasets ($l = 11$ and $D = 33$), we get χ_F^2=79.39 and F_F=10.14 . According to the statistical F-distribution table, F_F(10,140)=1.9 and F_F(10,330)=1.86 for low and high IR datasets, respectively at the 5% level of significance. We reject the null hypothesis due to the fact that $6.74 > 1.9$ and $10.14 > 1.86$. As a result, the models differ significantly on high IR datasets as well as on low IR datasets. We examine whether there is a significant difference between the models using the Nemenyi post hoc test.

Nemenyi post hoc test: The critical difference (C.D.) is given by

$$\text{C.D.} = q_\alpha \left(\sqrt{\frac{l(l+1)}{6D}}\right), \tag{18}$$

where $q_\alpha\,(\alpha = 0.05)$ is the critical value for the two-tailed Nemenyi test from the distribution table. After calculation, we get C.D. as 3.90 and 7.20 for low and high IR datasets, respectively. If the difference between average ranks of the two models is more than C.D., then the models are considered to be significantly different. The average rank differences on the low IR datasets between the proposed ISFFSVM model and PF-SMOTE [24], CKA-FSVM [22], MWMOTE [26], SMOTE-TL [25], HUE [21], and CNB [23] models are $4.33, 4.17, 4.13, 4.8, 4.53, 7.13$, respectively. All the differences are more than the C.D. value. Therefore, according to Nemenyi post hoc test, the proposed ISFFSVM model differs significantly from the existing baseline models. On high IR datasets, its average rank does not significantly differ from the baseline models, but it is clear from Table VII that it outperforms the baseline models in terms of average ranks.

Win-tie-loss-sign test: Win-tie-loss is a common statistical test in research and data analysis to determine if there is a significant difference between the outcomes of two or more models. Under the null hypothesis, two models are performing equally in the win-tie-loss sign test which means each model succeeds on $D/2$ out of D datasets. The minimum number of wins required for two models to be considered significantly different is $D/2 + 1.96\sqrt{D}/2$. In the event of a tie, the score is evenly split between the models under consideration. For low IR ($D = 15$), $D/2 + 1.96\sqrt{D}/2 = 11.3$ and for high IR ($D = 33$), $D/2 + 1.96\sqrt{D}/2 = 22.13$. The pairwise examination of the models is provided in Table VIII using straightforward computations. In our case, if any of the two models perform better on at least 12 datasets in the case of low IR, they are significantly different; if they perform better on at least 23 datasets in the case of high IR, they are considerably different. From Table VIII, it is evident that the proposed ISFFSVM model differs significantly from DEC, FSVM-CIL-exp, PF-SMOTE, SMOTE-TL, FSVM-CIL-lin, CKA-FSVM, MWMOTE, HUE, CNB, and SFFSVM in the case of low IR datasets and from FSVM-CIL-exp, SMOTE-TL, HUE, FSVM-CIL-lin, CKA-FSVM, MWMOTE, CNB, and SFFSVM in the case of high IR datasets. It has been determined that the proposed ISFFSVM model has emerged victorious, thereby, substantiating its superiority over other baseline models.

Parameter Analysis: There are four parameters in the proposed ISFFSVM model including the kernel width parameter γ, regularization parameter ζ, the smoothing parameter μ, and the location parameter a. The hyperparameters corresponding to different models are selected from the range: $\zeta = \left[2^{-i}, |i = -5, -4, ..., 11\right], \gamma = \left[2^{i}|i = -10, -9, ..., 3\right]$. Also, we set a range of $\mu = [0.5, 1, 1.5, 2, 2.5]$ and $a = [1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2]$. Here, regularization and kernel width parameters are the same as in the SVM model based on the RBF Kernel function [32], the smoothing parameter is the same as in the SFFSVM, and their role is the same as in the SVM and SFFSVM. For further research, please refer [33]. The range of location parameter a introduced in the ISFFSVM is $[1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2]$. It adjusts the distance the DEC hyperplane will move to the right side to reach the optimal hyperplane M^*. From Fig.4, it is clear that setting the value of a to 1.1 is most likely to occur. The value 1.1 for a has been encountered 1803 times in 2400 instances, the value 2 has been encountered 105 times, and the rest of the values fall in between these two values.

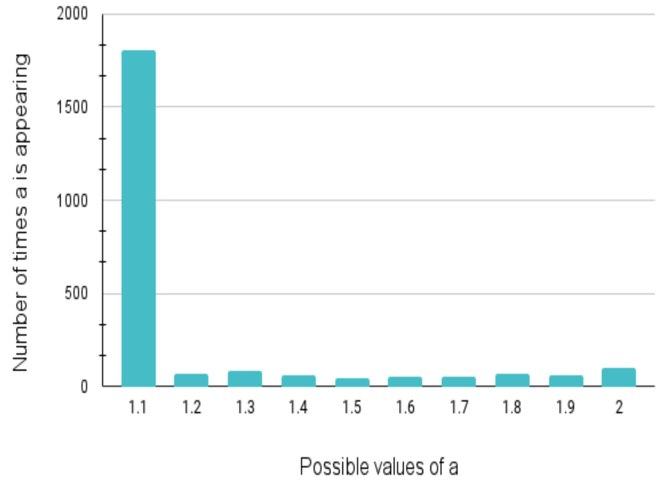


Fig. 4: Analysis of values attained by location parameter a.

V. Conclusion

In this paper, we have come up with an improvement in the existing slack-factor-based fuzzy membership function. In the SFFSVM model, wrongly classified majority class samples by the hyperplane obtained by DEC (M_{dec}) with slack factor value less than 2 are assigned one membership value. However, 2 is not always the correct choice due to the fact that while shifting M_{dec}, we make sure that correctly classified minority samples do not get misclassified. To solve this issue, we come up with an improvement in the fuzzy membership function for majority class samples in the SFFSVM model. We introduced a location parameter a in our improved slack-factor-based fuzzy support vector machine (ISFFSVM) model. Now, only those wrongly classified majority class samples with slack factor value less than a are assigned one membership value. The proposed ISFFSVM's key benefit is that it will result in a reduction in minority samples being misclassified. On the basis of the $F1$-measure (F1), Mathews correlation coefficient (MCC), and area under precision-recall curves (AUC-PR), the proposed ISFFSVM model performs more efficiently compared to other models. Considering that the proposed ISFFSVM model is a SVM variant, it can be challenging to handle large-scale imbalanced datasets. Therefore, in the future, the proposed startegy can be applied with an algorithm that can address the issue of imbalance. Moreover, it can also be employed in twin support vector machines (TWSVM).

References

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[2] M. Sewell, "Structural risk minimization," *Unpublished Ph. D. Dissertation, University College London, Department of Computer Science*, 2008.

[3] L. Baumes, J. Serra, P. Serna, and A. Corma, "Support vector machines for predictive modeling in heterogeneous catalysis: a comprehensive introduction and overfitting investigation based on two real applications," *Journal of Combinatorial Chemistry*, vol. 8, no. 4, pp. 583–596, 2006.

[4] S. M. Abd Elrahman and A. Abraham, "A review of class imbalance problem," *Journal of Network and Innovative Computing*, vol. 1, no. 2013, pp. 332–340, 2013.

[5] K. Kennedy, "Credit scoring using machine learning," 2013.

[6] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*. IEEE, 2018, pp. 1–4.

[7] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 631–641, 2019.

[8] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[9] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, 2018.

[10] J. Mathew, C. K. Pang, M. Luo, and W. H. Leong, "Classification of imbalanced data by oversampling in kernel space of support vector machines," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4065–4076, 2017.

[11] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2020, pp. 243–248.

[12] A. More, "Survey of keel techniques for improving classification performance in unbalanced datasets," *ArXiv:1608.06048*, 2016.

[13] Z. A. Huang, Y. Sang, Y. Sun, and J. Lv, "Neural network with absent minority class samples and boundary shifting for imbalanced data classification," *Neural Computing and Applications*, pp. 1–17, 2023.

[14] N. Thai-Nghe, Z. Gantner, and L. Schmidt-Thieme, "Cost-sensitive learning methods for imbalanced data," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2010, pp. 1–8.

[15] H. Yu, C. Mu, C. Sun, W. Yang, X. Yang, and X. Zuo, "Support vector machine-based optimized decision threshold adjustment strategy for classifying imbalanced data," *Knowledge-Based Systems*, vol. 76, pp. 67–78, 2015.

[16] 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, pp. 927–935, 2014.

[17] K. Veropoulos, C. Campbell, and N. Cristianini, "Controlling the sensitivity of support vector machines," in *Proceedings of the International Joint Conference on AI*, vol. 55. Stockholm, 1999, p. 60.

[18] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.

[19] 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.

[20] J. Ren, Y. Wang, and X. Deng, "Slack-factor-based fuzzy support vector machine for class imbalance problems," *ACM Transactions on Knowledge Discovery from Data*, 2023.

[21] W. W. Ng, S. Xu, J. Zhang, X. Tian, T. Rong, and S. Kwong, "Hashing-based undersampling ensemble for imbalanced pattern classification problems," *IEEE Transactions on Cybernetics*, vol. 52, no. 2, pp. 1269–1279, 2020.

[22] T. Wang, Y. Qiu, and J. Hua, "Centered kernel alignment inspired fuzzy support vector machine," *Fuzzy Sets and Systems*, vol. 394, pp. 110–123, 2020.

[23] J. D. Rennie, "Tackling the poor assumptions of naive bayes text classification. machine learning," *ICML-2003, Washington DC*, 2003.

[24] S. Gazzah and N. E. B. Amara, "New oversampling approaches based on polynomial fitting for imbalanced data sets," in *2008 the Eighth IAPR International Workshop on Document Analysis Systems*. IEEE, 2008, pp. 677–684.

[25] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations NewsLetter*, vol. 6, no. 1, pp. 20–29, 2004.

[26] S. Barua, M. M. Islam, X. Yao, and K. Murase, "MWMOTE–majority weighted minority oversampling technique for imbalanced data set learning," *IEEE Transactions on knowledge and Data Engineering*, vol. 26, no. 2, pp. 405–425, 2012.

[27] J. Derrac, S. Garcia, L. Sanchez, and F. Herrera, "Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Mult. Valued Logic Soft Comput*, vol. 17, 2015.

[28] G. He, H. Han, and W. Wang, "An over-sampling expert system for learing from imbalanced data sets," in *2005 International Conference on Neural Networks and Brain*, vol. 1. IEEE, 2005, pp. 537–541.

[29] J. Wang, Y. Qian, and F. Li, "Learning with mitigating random consistency from the accuracy measure," *Machine Learning*, vol. 109, pp. 2247–2281, 2020.

[30] J. Wang, Y. Qian, F. Li, J. Liang, and Q. Zhang, "Generalization performance of pure accuracy and its application in selective ensemble learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[31] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[32] S. Han, C. Qubo, and H. Meng, "Parameter selection in svm with rbf kernel function," in *World Automation Congress 2012*. IEEE, 2012, pp. 1–4.

[33] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.