Class Probability and Generalized Bell Fuzzy Twin SVM for Imbalanced Data

Anuradha Kumari, *Graduate Student Member, IEEE*, M. Tanveer, *Senior Member, IEEE*, C. T. Lin, *Fellow, IEEE*, and the Alzheimer's Disease Neuroimaging Initiative

***Abstract*—The data mining community has a major challenge in classifying datasets with noise, outliers, and imbalanced classes. Twin support vector machine (TSVM) is a well-known plane-based learning technique for classification, however, it has poor performance on the aforementioned datasets. To address the issue, in this article, we propose a novel class probability and generalized bell fuzzy twin SVM for imbalanced data (CGFTSVM-ID). The proposed CGFTSVM-ID assigns membership value to the data points using a new membership function called class probability and generalized bell (CPGB) function. The membership function for the majority class is a combination of the generalized bell (gbell) function, class probability, and imbalance ratio. The gbell function suppresses the negative impact of outliers in the training data by assigning them less value. The less class probability of the majority class data points denotes their higher possibility to be noise. The imbalance ratio of the classes considered in the membership function tackles the imbalancing issue of the datasets. In order to ensure the importance of the minority class samples in model learning, relatively high memberships are assigned to them. Thus, the proposed CPGB function handles the class imbalance learning problem with noise and outliers. We employ successive overrelaxation technique to solve the proposed optimization problem. The extensive numerical experiments and statistical analysis carried out over imbalanced real-world UCI and KEEL datasets clearly reveal that the proposed CGFTSVM-ID has superior generalization performance in comparison to baseline models. Moreover, the experiments are also conducted on the publicly available ADNI dataset for Alzheimer's disease classification and the results demonstrate the superiority of the proposed CGFTSVM-ID.**

***Index Terms*—Alzheimer's disease (AD), generalized bell (gbell) function, intuitionistic fuzzy theory, twin support vector machine (TSVM).**

I. Introduction

SUPPORT vector machine (SVM) [1] is one of the rapidly expanding techniques of machine learning. It has wide applications in various domains such as tea identification [2], dendritic spine detection [3], EEG signal classification [4], and so on. Over many other classification paradigms, SVM has a strong mathematical background and leads to a unique optimal solution. The objective is to find the optimal hyperplane with a maximal margin between the two classes. The maximum margin can ensure a reduced Vapnik–Chervonenkis (VC) dimension, thus improving the generalization performance. However, SVM suffers from high time complexity [5].

Apart from SVM, various classifiers, such as generalized eigenvalue proximal SVM (GEPSVM) [6] and twin support vector machine (TSVM) [7], have been proposed that draw two nonparallel hyperplanes. TSVM solves two smaller quadratic programming problems (QPPs) as compared to SVM, which solves a large QPP. TSVM creates hyperplanes in such a way that each hyperplane is proximal to one of the classes and lies at a distance of at least one unit from the other class. TSVM is four times more efficient than SVM [7]. Unlike SVM, TSVM considers only empirical risk minimization [8] and does not consider structural risk minimization (SRM), which is addressed in twin bounded SVM (TBSVM) [9]. Various variants of TSVM have been presented in the last decade, which are briefly discussed in a comprehensive review of TSVM [10].

Learning from an imbalanced dataset and addressing the noise and outliers present in the training dataset are the two primary challenges related to classification problems. For the datasets having two classes, imbalanced datasets have one class in abundance (majority class) as compared to the other class (minority class). There are several real-world domains, such as face recognition [11], where an imbalanced dataset can be seen. In general, there are two techniques to deal with class imbalance learning (CIL) problems: data-level approach and algorithm-level approach. Data-level approaches involve preprocessing techniques to get balanced data such as undersampling

Manuscript received 14 December 2023; revised 31 January 2024; accepted 6 February 2024. Date of publication 19 February 2024; date of current version 3 May 2024. We express our gratitude to the Science and Engineering Research Board (SERB) for their support under the Mathematical Research Impact-Centric Support (MATRICS) scheme, under Grant MTR/2021/000787, and to the National Supercomputing Mission under the Department of Science and Technology (DST) and Ministry of Electronics and Information Technology (MeitY), Government of India, under Grant DST/NSM/R&D_HPC_Appl/2021/03.29 for funding this work. Ms. Anuradha Kumari (File no - 09/1022 (12437)/2021-EMR-I) would like to express her appreciation to the CSIR New Delhi, India, for the financial assistance provided as fellowship. Recommended by Associate Editor M. Pratama. *(Corresponding author: Mohammad Tanveer.)*

Anuradha Kumari and M. Tanveer are with the Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore 453552, India (e-mail: phd2101141007@iiti.ac.in; mtanveer@iiti.ac.in).

C. T. Lin is with the School of Computer Science, Human-Centric AI Centre, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: Chin-Teng.Lin@uts.edu.au).

The data for this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database available at adni.loni.usc.edu. While the ADNI investigators provided data and contributed to the design and implementation of the initiative, they were not involved in the analysis or writing of this report. For a comprehensive list of ADNI investigators, please refer to: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

The code for the proposed CGFTSVM-ID can be found on https://github.com/mtanveer1/CGFTSVM-ID.

This article has supplementary downloadable material available at https://doi.org/10.1109/TFUZZ.2024.3366936, provided by the authors.

Digital Object Identifier 10.1109/TFUZZ.2024.3366936

(neighborhood cleaning rule [12]) and oversampling (synthetic minority oversampling technique [13]). However, data-level approaches may discard valuable information (undersampling) or introduce noise (oversampling) [14]. The algorithmic-level approach modifies the algorithm. The TSVM algorithm performs well on a balanced dataset, however, when dealing with imbalanced datasets, TSVM considers all samples equal and overlooks the distinction between minority and majority classes, which might lead to the obtained classification boundary being biased in favor of the majority class [15]. Furthermore, TSVM provides equal attention to all the training samples, including potential outliers or class noise that can significantly skew the decision hyperplane [7].

To ensure good performance of a model for imbalance learning with noise and outliers, various SVM variants have been proposed where the contribution of input data depends on membership functions, which lessen the influence of outliers and noise on classification technique. Fuzzy SVM (FSVM) [16] is proposed to handle outliers and noise, where the assignment of fuzzy value depends on the distance from the class center. The assignment of membership to the input data is the key point in FSVM variants. Though fuzzy concept overcomes the effect of noise/outliers, it still suffers from CIL issue. To handle the class imbalance problem in the presence of outliers and noise, Batuwita and Palade [17] proposed FSVM for imbalance data (FSVM-CIL) with four different membership functions. In FSVM-CIL, the hyperplane-based membership function assumes the initially obtained hyperplane is an accurate prediction of the final hyperplane, which is not always the case. To overcome the limitations of the aforementioned model for CIL in the presence of noise/outliers effectively, Tao et al. [18] introduced affinity and class probability-based FSVM (ACFSVM) designed for imbalanced datasets, which uses support vector data description (SVDD) [19] and class probability to obtain the membership value of the data points. An alternative approach to tackle the challenge posed by outliers and noise in the data involves employing the large-scale pinball TSVM [20].

Similar to SVM, many twin variants have been proposed to overcome the impact of outliers and noise. Rezvani et al. [15] introduced intuitionistic fuzzy TSVM (IFTWSVM) by incorporating the intuitionistic fuzzy number (IFN) in the TSVM model. It assigns membership and nonmembership values to each sample, which is used to calculate the score values of the data points. Thus, IFTWSVM has diminished the effect of noise/outliers on classification problems. However, IFTWSVM encounters computational challenges for large datasets due to the computation of matrix inverses, making it intractable in such scenarios. Improved IFTWSVM (IIFTWSVM) [21] overcomes the requirement of matrix inversion and has resistance against noise and outliers. To deal with the CIL problem along with noise/outliers, IFTWSVM for imbalanced data (IFTWSVM-ID) [22] is proposed, where a weighting strategy is given to the data points to tackle the class imbalance issue. Further, to enhance the time complexity of IFTWSVM along with considering the local neighborhood information, Tanveer et al. [23] introduced intuitionistic fuzzy weighted least square TSVM. A large-scale fuzzy least square TSVM [24] is proposed to classify large-scale data. K-nearest neighborhood (KNN) weighted reduced universum twin SVM for CIL is introduced by Ganaie et al. [25], which used the available prior and local neighborhood information of the data in training of the model. Another method to address the effect of noise in the data is presented by Tian et al. [26] by building a semisupervised model. Later, the branch and bound method is updated in [27] and another effective model is constructed. Another variant that modified the concept of intuitionistic fuzzy in a better way is introduced by Zhang et al. [28]. Further, the idea of affinity and class probability is extended to twin variants, and fuzzy TSVM based on affinity and class probability (ACFTSVM) for CIL [29] is introduced.

To deal with the major concerns of noise, outliers and CIL issues, in this article, we propose a novel class probability and generalized bell fuzzy twin SVM for imbalanced data (CGFTSVM-ID). It assigns varying membership values to the training data points, determining their impact on the creation of nonparallel hyperplanes. In the proposed method, the generalized bell (gbell) function assigns the membership value to the majority class data points based on their location and reduces the effect of outliers in the classification process by assigning them a lesser value. To further reduce the noise impact, the class probability of the majority samples is calculated. By combining the gbell values, class probability of the majority samples, and the imbalance ratio of the datasets, we define a class probability and generalized bell (CPGB) membership function. The contribution of the proposed CGFTSVM-ID can be summarized as follows.

1) We present a new membership function called (CPGB to solve imbalance learning problems having noise and outliers, which assign weights to individual samples to decide their influence in the construction of the final optimal plane.
2) The proposed CGFTSVM-ID employs regularization term to follow the SRM principle, which avoids the overfitting problem.
3) The optimization problem of the proposed CGFTSVM-ID is solved using the successive overrelaxation (SOR) technique, which processes datasets effectively without residing in memory.
4) The extensive numerical experiments conducted over real-world UCI and KEEL datasets consistently show that the proposed CGFTSVM-ID outperforms the baseline models in terms of classification performance. The proposed model's effectiveness is further demonstrated on the ADNI dataset, demonstrating its real-world application in classifying Alzheimer's disease (AD).

The rest of this article is organized as follows. Section II discusses the related work. Section III provides a detailed explanation of the proposed work, while Section IV encompasses numerical experiments and the subsequent statistical tests. Section V contains the application of the proposed work. Finally, Section VI concludes this article.

II. Related Work

In this section, we discuss state-of-the-art algorithmic-level approaches to reduce the impact of noise and outliers in the data along with imbalanced learning techniques. Assume the set of

training samples denoted as $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where $x_k \in \mathbb{R}^m$ represents the feature vector and $y_k \in \{\pm 1\}$. Let T_1 and T_2 be the matrix containing positive and negative samples, respectively, of order $t_1 \times m$ and $t_2 \times m$. Presume the positive class constitutes the majority class, while the negative class serves as the minority class, and the imbalance ratio (ir) is expressed as the ratio between the cardinality of the majority class and the minority class. We represent the sample x_k belonging to the majority class as x_k^{maj}, and if it belongs to the minority class, denote it as x_k^{min}. Additionally, let θ represent the mapping from the feature space to a higher dimensional space.

A. Intuitionistic Fuzzy Scheme

The intuitionistic fuzzy [15] approach involves assigning an IFN to each training sample in the dataset based on their membership and nonmembership functions, respectively. To assess the presence of noise or outliers in the dataset, a score function is established utilizing the membership and nonmembership functions.

1) Membership Function: It is defined in terms of the distance between the class center and the training data points in the high-dimensional feature space. Mathematically,

$$\mu_k = \begin{cases} 1 - \frac{\|\theta(x_k) - C_1\|}{R_1 + \Delta}, & \text{if } y_k = 1, \\ 1 - \frac{\|\theta(x_k) - C_2\|}{R_2 + \Delta}, & \text{if } y_k = -1, \end{cases} \tag{1}$$

where Δ is a small positive number. $C_1 = \frac{1}{t_1} \sum_{y_k=1} \theta(x_k)$ and $C_2 = \frac{1}{t_2} \sum_{y_k=-1} \theta(x_k)$ are the centers of positive and negative classes, respectively. And, $R_1 = \max_{y_k=1} \|\theta(x_i) - C_1\|$ and $R_2 = \max_{y_k=-1} \|\theta(x_i) - C_2\|$ denote the radius of positive and negative classes, respectively.

2) Nonmembership Function: It depends on the proportion of dissimilar points to the total number of points in a neighborhood around each sample. It is defined as:

$$\nu_k = (1 - \mu_k)\zeta(x_k), \tag{2}$$

where

$$\zeta(x_k) = \frac{|x_l : \|\theta(x_k) - \theta(x_l)\| \leq \beta, y_k \neq y_l|}{|x_l : \|\theta(x_k) - \theta(x_l)\| \leq \beta|}.$$

Here, β corresponds to user-defined positive parameter.

Score function: The pair (μ_k, ν_k) for each training sample (x_k) is called IFN and is used to calculate score value of data points by the function given as:

$$s_k = \begin{cases} \mu_k, & \text{if } \nu_k = 0 \\ 0, & \text{if } \mu_k \leq \nu_k \\ \frac{1-\nu_k}{2-\mu_k-\nu_k}, & \text{otherwise.} \end{cases} \tag{3}$$

B. Intuitionistic Fuzzy TSVM

By incorporating the concept of IFN into the classical TSVM, Rezvani et al. [15] proposed the IFTWSVM, which is effective against noise and outliers. The kernel-generated surfaces for the nonlinear case are given by:

$$K(x, T^t)w_1 + b_1 = 0, \text{ and } K(x, T^t)w_2 + b_2 = 0$$

where $K(.,.)$ represents the kernel function; w_1, w_2 are weight vectors; $b_1\, b_2$ are bias terms and $T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$. The nonlinear optimization problems corresponding to IFTWSVM are given as:

$$\min_{w_1, b_1, \chi_2} \frac{1}{2}\|K(T_1, T^t)w_1 + e_1 b_1\|^2 + \frac{1}{2}\lambda_1 \|w_1\|^2 + \lambda_2 s_2^t \chi_2$$
$$\text{s.t. } -(K(T_2, T^t)w_1 + e_2 b_1) + \chi_2 \geq e_2, \chi_2 \geq 0 \tag{4}$$

and

$$\min_{w_2, b_2, \chi_1} \frac{1}{2}\|K(T_2, T^t)w_2 + e_2 b_2\|^2 + \frac{1}{2}\lambda_3 \|w_2\|^2 + \lambda_4 s_1^t \chi_1$$
$$\text{s.t. } (K(T_1, T^t)w_2 + e_1 b_2) + \chi_1 \geq e_1, \chi_1 \geq 0 \tag{5}$$

where λ_1, λ_2, λ_3, and λ_4 denote the positive hyperparameters. χ_1 and χ_2 correspond to the slack variables. s_1 and s_2 are score vectors calculated using (3) for positive and negative class data points, respectively. e_1 and e_2 are the column vectors of ones with appropriate dimensions.

C. Affinity and Class Probability Based Fuzzy Support Vector Machine for Imbalanced Datasets

To overcome the limitation of SVM in the presence of noise/outliers and CIL issue, Tao et al. [18] proposed ACFSVM, where they assigned membership values to the data points. Using the concept of SVDD [19], the affinity of the majority class is calculated, which identifies border samples (outliers) existing among the majority samples effectively. The KNN is employed to calculate the probability of data points in the majority class. Here, R denotes the center and radius of the training data points and d_k^{svdd} denotes the distance of each training data point (x_k) from the class center. The function to calculate affinity is written as:

$$\mu_{aff}(x_k) = \begin{cases} 0.8\left(\frac{d_k^{\text{svdd}} - \min(d_k^{\text{svdd}})}{\max(d_k^{\text{svdd}}) - \min(d_k^{\text{svdd}})}\right), & \text{if } d_k^{\text{svdd}} < \rho \times R, \\ 0.2\left(\exp(\beta(1 - \frac{d_k^{\text{svdd}}}{\rho \times R}))\right), & \text{if } d_k^{\text{svdd}} \geq \rho \times R, \end{cases} \tag{6}$$

where $\min(d_k^{\text{svdd}})$ and $\max(d_k^{\text{svdd}})$ denote the minimum and maximum distance of the data points from the class center, $\rho \in (0, 1]$ controls the size of outliers and the border samples and β is a positive weight decay parameter.

Using the concept of KNN, a probability value $(p(x_k))$ is assigned to each data point, which is calculated as the ratio between the same class data points among the k selected nearest neighborhood data points in the kernel space and k. Hence, ACFSVM calculates the final membership values as follows:

$$\mu(x_k) = \begin{cases} 1, & \text{if } x_k = x_k^{\text{min}}, \\ \mu_{\text{aff}}(x_k) \times p(x_k), & \text{if } x_k = x_k^{\text{maj}}. \end{cases} \tag{7}$$

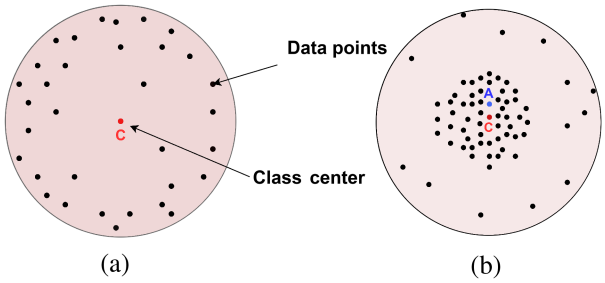

Fig. 1. Representation of a training sample. (a) Sample having majority of the data points lying on the boundary of class. (b) Sample having majority of the data points lying around the center of the class.

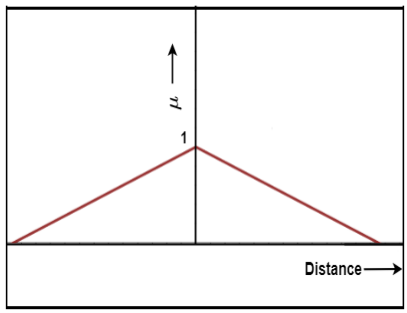

Fig. 2. Graphical illustration of membership function in (1).

D. Limitation of the Existing Membership Functions

Though the score function defined using the concept of IFN in (3) has resistance against noise and outliers, it has certain limitations, such as follows:

1) Assume the data has a distribution characterized by a significant concentration of the training data points along the boundary of the class as depicted in Fig. 1(a). Then, the majority of data points are assigned a membership that is close to zero according to (1), as visually depicted in Fig. 2. Consequently, their impact on the classification process is minimized.
2) The membership function of IFN decreases linearly with distance, i.e., the membership value decreases as the distance between data points and the corresponding class center increases. Though the points lying near its class center belong to the dense region of its class distribution [e.g., data point A in Fig. 1(b)], these are given a lower membership value compared to the class center (C). Consequently, their significance in the classification process is reduced.

The membership function discussed in ACFSVM in (7) is an effective way to deal with noise, outliers, and imbalanced data simultaneously, however, it has the following limitation.

1) The computation of affinity in (6) involves solving QPP associated with the SVDD technique, which has high complexity. Furthermore, the determination of class probability entails selecting the k nearest neighbors for each training data point, a factor that significantly contributes to the overall increase in time complexity.

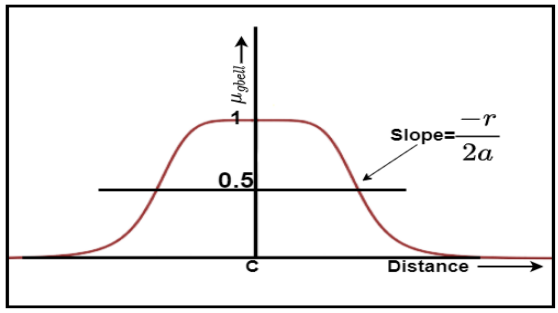

Fig. 3. Graphical representation of gbell membership function of (8).

III. Proposed Work

In the preceding section, we discussed the limitations of the two well-known weighting assignment techniques for noise and outliers. To address these limitations and draw inspiration from the aforementioned existing weighting techniques, we propose the CPGB membership function. By amalgamating the CPGB membership function with TSVM, we present a novel framework called CGFTSVM-ID.

A. Weight Assignment to the Training Data Points

In this, section, we discuss the strategy of assigning weights to the training data points.

1) Determination of Membership Value: The gbell function is a well-known fuzzy set membership function, frequently employed in fuzzy logic and fuzzy systems. It enables more adaptable and versatile fuzzy modeling and is an extension of the bell-shaped membership function [30]. In the proposed CGFTSVM-ID, it assigns membership value to the data points depending on their location in higher dimensional space. The assigned membership value is used to identify the outliers and border samples present in the majority class. Formally, the gbell function [30] is expressed as follows:

$$\mu_{\text{gbell}}(x_k) = \begin{cases} \frac{1}{1+\left(\frac{\|\theta(x_k)-C_1\|}{a}\right)^{2r}}, & \text{if } y_k = 1, \\ \frac{1}{1+\left(\frac{\|\theta(x_k)-C_2\|}{a}\right)^{2r}}, & \text{if } y_k = -1. \end{cases} \tag{8}$$

Here, C_1 and C_2 correspond to the center of the positive class and negative class, respectively, defined as $C_1 = \frac{1}{t_1}\sum_{y_k=1}\theta(x_k)$ and $C_2 = \frac{1}{t_2}\sum_{y_k=-1}\theta(x_k)$. a and r correspond to membership function positive parameters responsible for the width and steepness of the function (8), respectively. The gbell function is illustrated in Fig. 3.

2) Determination of Class Probability: The impact of noise during model training is mitigated by the class probability value assigned to the training data points. This value, computed for each sample, is determined by the ratio of the number of samples belonging to the same class within a predefined neighborhood to the total number of data points in that neighborhood:

$$p(x_k) = \frac{|\{x_j : \|\theta(x_j) - \theta(x_k)\| \le \delta, y_j = y_k\}|}{|x_j : \|\theta(x_j) - \theta(x_k)\| \le \delta|}, \tag{9}$$

where δ is a fixed positive real number. The greater value of $p(x_k)$ implies the larger belongingness of x_k to its own class. Thus, less probability value of noise leads to a reduced effect on the construction of hyperplanes. In (9), we have

$$(\|\theta(x) - \theta(\tilde{x})\|)^2 = K(x,x) - 2K(x,\tilde{x}) + K(\tilde{x},\tilde{x}).$$

3) Proposed CPGB Membership Function: Combining the calculated gbell function values of each sample using (8) with the probability value calculated using (9) and the imbalance ratio (ir) of classes, we define CPGB membership function as follows:

$$\mu_g(x_k) = \begin{cases} 1, & \text{if } x_k = x_k^{\min} \\ \mu_{\text{gbell}}(x_k) \times p(x_k) \times \frac{1}{\text{ir}}, & \text{if } x_k = x_k^{\text{maj}}. \end{cases} \quad (10)$$

Theorem 1: For all the positive (negative) data points and the gbell function parameters C_1 (C_2), $a > 0$ and $r > 0$, the proposed CPGB membership function lies between 0 and 1. In other words, $0 \le \mu_g(x_k) \le 1$, $k = 1, 2, \ldots, n$.

Proof: For the minority (negative) class sample, the proof holds. For majority class samples,

$$\|\theta(x_k) - C_1\| \ge 0 \quad \forall\, x_k \in \text{majority (positive) class}$$

$$\Rightarrow \left(\frac{\|\theta(x_k) - C_1\|}{a}\right)^{2r} \ge 0 \quad \forall\, C_1,\ a > 0 \text{ and } r > 0.$$

$$\Rightarrow 1 + \left(\frac{\|\theta(x_k) - C_1\|}{a}\right)^{2r} \ge 1 \Rightarrow \frac{1}{1 + \left(\frac{\|\theta(x_k) - C_1\|}{a}\right)^{2r}} \le 1.$$

$$\text{Also, } \frac{1}{1 + \left(\frac{\|\theta(x_k) - C_1\|}{a}\right)^{2r}} > 0.$$

Hence, $\forall\ x_k \in$ majority class, $0 < \mu_{\text{gbell}}(x_k) \le 1$. Also, $0 \le p(x_k) \le 1$ and $0 < \frac{1}{\text{ir}} \le 1$. Thus, the value of the proposed CPGB membership function lies between 0 and 1. ■

The proposed CPGB function effectively overcomes the limitations discussed in Section II-D in the following way.

1) Unlike IFN, where membership value decreases linearly from the center, the proposed CPGB function assigns membership value depending on the parameter 'a' of (8) that decides the width up to which the training data points should be assigned a high membership value so that the classification performance improves. Thus, for the data distribution having majority samples at the boundary of the class, membership value varies with that of 'a' leading to an optimal classifier.
2) It can be clearly seen from Fig. 3 that, up to the distance 'a' from the class center, the assigned membership value is high depending on the steepness parameter 'r' of the gbell function. Thus, the membership value of all the training points lying in the dense region of the class distribution varies depending on 'r' and may take an equivalent value to that of the class center. Hence, leading to an improved contribution of such points in the classification process.
3) The proposed CPGB function does not require the solution of the other QPP of SVDD. Also, we are considering the class probability of the training data points in some fixed radii (δ) instead of KNN. Hence, the calculation of the proposed weighting technique has better efficiency with respect to (wrt) time than the membership function of ACFSVM.

B. Proposed CGFTSVM-ID: Linear Case

By integrating the above proposed CPGB membership function (10) with the TSVM model, we propose a novel CGFTSVM-ID. The equations of nonparallel hyperplanes for positive and negative classes are given as:

$$w_1^t x + b_1 = 0 \text{ and } w_2^t x + b_2 = 0 \quad (11)$$

respectively. The optimization problems for the proposed CGFTSVM-ID are expressed as:

$$\min_{w_1,b_1,\chi_2} \frac{1}{2}\|T_1 w_1 + e_1 b_1\|^2 + \frac{1}{2}\lambda_1(\|w_1\|^2 + b_1^2) + \lambda_2 \mu_{g_2}^t \chi_2$$
$$\text{s.t. } -(T_2 w_1 + e_2 b_1) + \chi_2 \ge e_2, \chi_2 \ge 0 \quad (12)$$

and

$$\min_{w_2,b_2,\chi_1} \frac{1}{2}\|(T_2 w_2 + e_2 b_2)\|^2 + \frac{1}{2}\lambda_3(\|w_2\|^2 + b_2^2) + \lambda_4 \mu_{g_1}^t \chi_1$$
$$\text{s.t. } (T_1 w_2 + e_1 b_2) + \chi_1 \ge e_1, \chi_1 \ge 0, \quad (13)$$

where λ_1, λ_2, λ_3, and λ_4 are the positive hyperparameters, and e_1, e_2 are the column vectors. $\mu_{g_1} \in \mathbb{R}^{t_1}$ and $\mu_{g_2} \in \mathbb{R}^{t_2}$ are the membership values of positive and negative classes; χ_1 and χ_2 are the slack variables. The first term in QPP (12) minimizes the distance of the positive class data points from the positive class hyperplane, and the second term corresponds to the regularization term, which considers the SRM principle. The third term minimizes the total penalty imposed on negative class data points, which are at a distance less than 1 from the positive class hyperplane. A similar explanation follows for the QPP (13). The Lagrangian corresponding to QPP (12) having α_1 and β_1 as the positive Lagrange multipliers, is written as:

$$\mathscr{L}(w_1, b_1, \chi_2, \alpha_1, \beta_1) = \frac{1}{2}\|T_1 w_1 + e_1 b_1\|^2 + \frac{1}{2}\lambda_1(\|w_1\|^2 + b_1^2)$$
$$+ \lambda_2 \mu_{g2}^t \chi_2 + \alpha_1^t(e_2 - \chi_2 + (T_2 w_1 + e_2 b_1)) - \beta_1^t \chi_2. \quad (14)$$

From the Karush Kuhn Tucker (K.K.T.) conditions, differentiate the Lagrangian (14) wrt w_1, b_1, and χ_2, we obtain:

$$\frac{\partial \mathscr{L}}{\partial w_1} = T_1^t(T_1 w_1 + e_1 b_1) + \lambda_1 w_1 + T_2^t \alpha_1 = 0, \quad (15)$$

$$\frac{\partial \mathscr{L}}{\partial b_1} = e_1^t(T_1 w_1 + e_1 b_1) + \lambda_1 b_1 + e_2^t \alpha_1 = 0, \quad (16)$$

$$\frac{\partial \mathscr{L}}{\partial \chi_2} = \lambda_2 \mu_{g_2} - \alpha_1 - \beta_1 = 0. \quad (17)$$

Combining (15) and (16), we get

$$\left(\begin{bmatrix} T_1^t \\ e_1^t \end{bmatrix}\begin{bmatrix} T_1 & e_1 \end{bmatrix} + \lambda_1 I\right)\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} T_2^t \\ e_2^t \end{bmatrix}\alpha_1 = 0. \quad (18)$$

Assuming $P = \begin{bmatrix} T_1 & e_1 \end{bmatrix}$, $Q = \begin{bmatrix} T_2 & e_2 \end{bmatrix}$, (18) reduces to

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(P^t P + \lambda_1 I)^{-1} Q^t \alpha_1. \tag{19}$$

Similarly, for the QPP (13) having α_2 and β_2 as the positive Lagrange multipliers, we get the relation:

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (Q^t Q + \lambda_3 I)^{-1} P^t \alpha_2. \tag{20}$$

The Wolfe dual for the primal problems (12) and (13) can be expressed as:

$$\min_{\alpha_1} \frac{1}{2}\alpha_1^t Q(P^t P + \lambda_1 I)^{-1} Q^t \alpha_1 - e_2^t \alpha_1$$
$$\text{s.t. } 0 \le \alpha_1 \le \lambda_2 \mu_{g_2} \tag{21}$$

and

$$\min_{\alpha_2} \frac{1}{2}\alpha_2^t P(Q^t Q + \lambda_3 I)^{-1} P^t \alpha_2 - e_1^t \alpha_2$$
$$\text{s.t. } 0 \le \alpha_2 \le \lambda_4 \mu_{g_1}, \tag{22}$$

respectively. A new data point x is allocated to a class based on the function:

$$f(x) = \text{argmin}\left\{ \frac{|w_1^t x + b_1|}{\|w_1\|}, \frac{|w_2^t x + b_2|}{\|w_2\|} \right\}. \tag{23}$$

C. Proposed CGFTSVM-ID: Nonlinear Case

For the nonlinear case, the kernel-generated nonparallel surfaces are given by the following equations:

$$K(x, T^t)w_1 + b_1 = 0, \quad K(x, T^t)w_2 + b_2 = 0, \tag{24}$$

where $K(.,.)$ represents the kernel function. The nonlinear optimization problems corresponding to the proposed CGFTSVM-ID are defined as:

$$\min_{w_1, b_1, \chi_2} \frac{1}{2}\|K(T_1, T^t)w_1 + e_1 b_1\|^2 + \frac{1}{2}\lambda_1(\|w_1\|^2 + b_1^2) + \lambda_2 \mu_{g_2}^t \chi_2$$
$$\text{s.t. } -(K(T_2, T^t)w_1 + e_2 b_1) + \chi_2 \ge e_2, \chi_2 \ge 0 \tag{25}$$

and

$$\min_{w_2, b_2, \chi_1} \frac{1}{2}\|K(T_2, T^t)w_2 + e_2 b_2\|^2 + \frac{1}{2}\lambda_3(\|w_2\|^2 + b_2^2) + \lambda_4 \mu_{g_1}^t \chi_1$$
$$\text{s.t. } (K(T_1, T^t)w_2 + e_1 b_2) + \chi_1 \ge e_1, \chi_1 \ge 0, \tag{26}$$

where the terms have the same notion as discussed in the linear case. Solving as the linear case by using K.K.T. conditions, we obtain the dual for primal problems (25) and (26) as:

$$\min_{\alpha_1} \frac{1}{2}\alpha_1^t Q(P^t P + \lambda_1 I)^{-1} Q^t \alpha_1 - e_2^t \alpha_1$$
$$\text{s.t. } 0 \le \alpha_1 \le \lambda_2 \mu_{g_2} \tag{27}$$

and

$$\min_{\alpha_2} \frac{1}{2}\alpha_2^t P(Q^t Q + \lambda_3 I)^{-1} P^t \alpha_2 - e_1^t \alpha_2$$
$$\text{s.t. } 0 \le \alpha_2 \le \lambda_4 \mu_{g_1} \tag{28}$$

respectively, where $P = [K(T_1, T^t) \;\; e_1]$, $Q = [K(T_2, T^t) \; e_2]$.

A new data point (x) is labeled as positive or negative depending on the function:

$$f(x) = \text{argmin}\left\{ \frac{|K(x, T^t)w_1 + b_1|}{\sqrt{w_1^t K(T, T^t)w_1}}, \frac{|K(x, T^t)w_2 + b_2|}{\sqrt{w_2^t K(T, T^t)w_2}} \right\}. \tag{29}$$

D. Time Complexity

Consider a binary class classification scenario with a total of n samples, where t_1 samples represent the positive class and t_2 samples represent the negative class. Assuming the positive class as the majority class, the imbalance ratio (ir)$= \frac{t_1}{t_2}$. The time complexity of the proposed CGFTSVM-ID is determined by two primary components. First, it involves calculating the CPGB membership value for the data points, which comprises the calculation of the gbell function value [see (8)] and the class probability [see (9)] for the majority class samples. The complexity of calculating gbell value would be O (t_1), as it entails a fixed number of arithmetic operations. Similarly, the complexity for calculating the class probability is equal to O(t_1) [15]. Second, the computational complexity of SVM is O(n^3), which implies the complexity of TSVM is O(t_1^3)+O(t_2^3) [7], where $n = t_1 + t_2$. Therefore, solving the obtained QPPs involves a computational complexity of $(\text{ir}^3 + 1)$ O(t_2^3). Consequently, the overall time complexity of the proposed CGFTSVM-ID is $(\text{ir}^3 + 1)\text{O}(t_2^3)$ + O(t_1) + O(t_1).

IV. Numerical Experiments

In this section, we delve into the numerical experiments and present the corresponding results for the proposed CGFTSVM-ID along with state-of-the-art baseline models, which are TSVM [7], IFTWSVM [15], IFTWSVM-ID [22], intuitionistic fuzzy TSVMs with the insensitive pinball loss (pin-IFTWSVM) [31], IIFTWSVM [21], ACFSVM [18], and ACFTSVM [29].

A. Experimental Setup

All the experiments are carried out in MATLAB 2022b on a desktop PC with 11th Gen Intel(R) Core(TM) i7-11700 @ 2.50 GHz processor and 16 GB RAM. 70% of each dataset is used for training and remaining dataset is used for testing. For the selection of the optimal parameters, we used grid search with 10-fold cross-validation. For brevity and increased computational efficiency, we assumed $\lambda_1 = \lambda_3$ and $\lambda_2 = \lambda_4$. All the datasets are normalized using the Z-score normalization technique. To implement the nonlinear case, we used the Gaussian kernel $(\exp(\|x_i - x_j\|^2/2\sigma^2))$. The optimization problems of the proposed CGFTSVM-ID are solved using the SOR technique. The

Algorithm 1: Algorithm CGFTSVM-ID.

Input: Training data point $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, testing data, hyperparameters λ_1, λ_2, λ_3, λ_4, membership function parameters a and r.

Output: Labels of testing dataset.

1) Determine the gbell membership value of training data points using
$$\mu_{gbell}(x_k) = \begin{cases} \frac{1}{1+\left(\frac{\|\theta(x_k)-C_1\|}{a}\right)^{2r}}, & \text{if } y_k = 1, \\ \frac{1}{1+\left(\frac{\|\theta(x_k)-C_2\|}{a}\right)^{2r}}, & \text{if } y_k = -1. \end{cases}.$$
2) Obtain class probability of training data using $p(x_k) = \frac{|\{x_j : \|\theta(x_j)-\theta(x_k)\| \le \delta, y_j = y_k\}|}{|x_j : \|\theta(x_j)-\theta(x_k)\| \le \delta|}$.
3) Calculate the CPGB membership value of the training data points following
$$\mu_g(x_k) = \begin{cases} 1, & \text{if } x_k = x_k^{\min}, \\ \mu_{\text{gbell}}(x_k) \times p(x_k) \times \frac{1}{\text{ir}}, & \text{if } x_k = x_k^{\text{maj}}. \end{cases}$$
4) Integrating the assigned membership value, obtain α_1, α_2 by solving the optimization problems (21) and (22) for the linear case, and (27), (28) for the nonlinear case.
5) Obtain $\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(P^tP + \lambda_1 I)^{-1} Q^t \alpha_1$ and $\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (Q^tQ + \lambda_3 I)^{-1} P^t \alpha_2$.
6) Determine labels of test data point using $f(x) = \text{argmin}\left\{ \frac{|w_1^t x + b_1|}{\|w_1\|}, \frac{|w_2^t x + b_2|}{\|w_2\|} \right\}$ for linear case and $f(x) = \text{argmin}\left\{ \frac{|K(x,T^t)w_1 + b_1|}{\sqrt{w_1^t K(T,T^t) w_1}}, \frac{|K(x,T^t)w_2 + b_2|}{\sqrt{w_2^t K(T,T^t) w_2}} \right\}$ for the nonlinear case.

relaxation factor (ω) in the SOR method is typically constrained to the interval $(0, 2)$ [32]. A larger ω accelerates convergence, but it can introduce instability. On the contrary, a smaller ω results in slower convergence. To strike a balance and achieve optimal convergence with stable results, we empirically selected ω as 0.5. For better comparison, we solved the optimization problems of the compared twin variants using SOR, except for ACFSVM, which is solved using the MATLAB quadprog toolbox. Different parameters of the optimization problems vary in the following way: λ_1 and λ_2 in the twin variant models (and λ in ACFSVM) are chosen from range $\{10^{-5}, \ldots, 10^5\}$; the Gaussian kernel parameter σ is chosen from the range $\{2^{-5}, \ldots, 2^5\}$. The values of ϵ and τ for pin-IFTWSVM are taken the same as in [31]. The parameters of ACFSVM except for λ and σ are taken the same as in [18]. The gbell membership function parameter 'a' takes the range $[1/2, 5/8, 3/4, 7/8, 1]*$(radius of positive/negative class). We considered the proposed CPGB membership function steepness parameter 'r' same as the value of 'a.' We considered δ as the maximum radius encompassing positive and negative classes, ensuring a broader neighborhood coverage. The class probability, influenced by this neighborhood information, extends across a wide area around each data point. Moreover, choosing δ as the maximum radii of the positive and negative classes dynamically adapts δ according to the spread or variance of the data. This can help the model adjust to the scale of the dataset, potentially improving generalization performance. The performance indicators for the experiments are area under ROC (AUC), sensitivity (Sens.) specificity (spec.), G-mean, and F-measure, which are defined as follows:

$$\text{Sens.} = \frac{\text{tp}}{\text{tp} + \text{fn}}, \tag{30}$$

$$\text{Spec.} = \frac{\text{tn}}{\text{tn} + \text{fp}}, \tag{31}$$

$$\text{AUC} = \frac{1 + \text{tpr} - \text{fpr}}{2}, \tag{32}$$

$$\text{G-mean} = \sqrt{\text{Sens.} \times \text{Spec.}}, \tag{33}$$

$$\text{F-measure} = \frac{2 \times \text{Sens.} \times \text{Prec.}}{\text{Sens.} + \text{Prec.}}, \tag{34}$$

where tp, fp, tn, and fn denote the true positive, false positive, true negative, and false negative samples, respectively. tpr, fpr, and Prec. denote the true positive rate, false positive rate, and precision calculated as:

$$\text{tpr} = \frac{\text{tp}}{\text{tp} + \text{fn}}, \ \text{fpr} = \frac{\text{fp}}{\text{fp} + \text{tn}} \text{ and Prec.} = \frac{\text{tp}}{\text{tp} + \text{fp}},$$

respectively. To carry out the experiments, we have considered the synthetic KEEL datasets and several real-world KEEL and UCI [33] datasets. For the application of the proposed CGFTSVM-ID, we employed it on the publicly available AD neuroimaging initiative (ADNI) dataset for the classification of AD.

B. Synthetic Datasets

To analyze the performance of the proposed CGFTSVM-ID with the baseline models, we employed experiments on the synthetic datasets publicly available on the imbalance KEEL dataset repository. The details of the used synthetic datasets are given in Table SI of the supplementary file. The performance of the models on the binary class synthetic datasets is depicted in Table SII of the supplementary file. The proposed CGFTSVM-ID exhibits the highest AUC values for datasets 04clover5z-600-5-0-BI, 04clover5z-600-5-30-BI, 04clover5z-600-5-60-BI, 04clover5z-800-7-0-BI, and 04clover5z-800-7-50-BI. While baseline models outperform the proposed CGFTSVM-ID in the remaining datasets, the average AUC of the proposed CGFTSVM-ID is 0.8289 with an average rank of 1.7, which is superior to those of the baseline models. This clearly establishes the superiority of CGFTSVM-ID over the baseline models. Table SIII of the supplementary file comprises the optimal parameters of the models corresponding to the maximum AUC across each synthetic dataset. For better visualization, we have compared the AUC values using bar graph, which is depicted in Figure S1 of the supplementary file.

TABLE I
AVERAGE AUC, AVERAGE RANK, AVERAGE G-MEAN, AVERAGE F-MEASURE, AVERAGE SENSITIVITY, AVERAGE SPECIFICITY, AND AVERAGE TRAINING TIME OF THE PROPOSED CGFTSVM-ID AND BASELINE MODELS ON KEEL AND UCI DATASETS

Dataset	metric	TSVM [7]	IFTWSVM [15]	IFTWSVM-ID [22]	pin-IFTWSVM [31]	IIFTWSVM [21]	ACFSVM [18]	ACFTSVM [29]	proposed CGFTSVM-ID
KEEL	Average AUC	0.7754	0.7769	0.7641	0.7728	0.7022	0.7694	0.7914	**0.8344**
	Average Rank	4.56	4.2	4.82	4.44	6.77	4.88	4.24	**2.09**
	Avergae G-mean	0.7347	0.7593	0.732	0.7601	0.613	0.7318	0.7406	**0.831**
	Average F-measure	0.5859	0.5763	0.5483	0.5732	0.4583	0.5686	0.5773	**0.6357**
	Average Sens.	0.6605	0.6869	0.6627	0.6893	0.6479	0.7215	0.6731	**0.7999**
	Average Spec.	0.8904	0.8724	0.8671	0.8675	0.7672	0.8245	**0.9111**	0.8766
	Average time (seconds)	0.1198	0.1484	0.1514	0.1749	0.5667	1.0902	0.7462	0.1247
UCI	Average AUC	0.7892	0.7889	0.7849	0.7764	0.7186	0.762	0.7998	**0.8098**
	Average Rank	4.06	4.11	4.67	4.72	6.83	6.03	3.56	**2.03**
	Avergae G-mean	0.7805	0.7783	0.7766	0.7675	0.6931	0.7471	0.7921	**0.8048**
	Average F-measure	0.7558	0.7573	0.7537	0.7422	0.6693	0.733	0.762	**0.7768**
	Average Sens.	0.7873	0.8084	0.7918	0.7734	0.6886	0.7803	0.814	**0.8153**
	Average Spec.	0.791	0.7693	0.7781	0.7794	0.7487	0.7436	0.7857	**0.8043**
	Average time (seconds)	0.02	0.0266	0.0252	0.1352	0.032	0.1752	0.1255	0.0227

The bold values represent the best values.

C. Real-World Datasets

To further examine the performance of the proposed CGFTSVM-ID in comparison to the state-of-the-art methods TSVM [7], IFTWSVM [15], IFTWSVM-ID [22], pin-IFTWSVM [31], IIFTWSVM [21], ACFSVM [18], and ACFTSVM [29], we performed experiments on 51 real-world KEEL and UCI datasets. The details of the datasets are mentioned in Table SIV of the supplementary file. First, we will discuss the results of the KEEL datasets, followed by the results over UCI datasets.

The detailed performance of 33 real-world KEEL datasets (with an imbalance ratio in the range of 1–85.88) in terms of AUC, training time (in seconds), sensitivity, and specificity are depicted in Table SV of the supplementary file. For the datasets, namely, crossplane150, ecoli0267vs35, ecoli0347vs56, ecoli067vs35, aus, checkerboard, monk2, monk3, sonar, vowel, wpbc, yeast2vs8, abalone-17vs78910, abalone-20vs8910, kr-vs-k-zero_vs_eight, shuttle-2vs5, winequality-red-8vs67, winequality-red-8vs6, winequality-white-39vs5, and winequality-white-3vs7, the proposed CGFTSVM-ID has the highest AUC value. Other datasets have a greater AUC for baseline models, nevertheless, the average AUC of the proposed CGFTSVM-ID is 0.8344, which is greater than that of baseline models. ACFTSVM has the second greater value, which is 0.7914. The average AUC of the remaining models is as follows: 0.7769, 0.7754, 0.7728, 0.7694, 0.7641, and 0.7022 for IFTWSVM, TSVM, pin-IFTWSVM, ACFSVM, IFTWSVM-ID, and IIFTWSVM, respectively. Hence, CGFTSVM-ID has better classification performance over KEEL datasets than baseline models. However, the average AUC value is not reliable enough to depict the superiority of a model, as the lower AUC value of a dataset may be compensated by the high AUC value of another dataset. To overcome this, we assigned ranks to all the models over each dataset and calculated the average ranks of the models. The assignment of rank is done via: the best-performing model is assigned the lowest rank and the worst-performing model is assigned the highest rank over each dataset. The proposed CGFTSVM-ID has the lowest average rank of 2.09, followed by IFTWSVM with an average rank of 4.2. The ranks of the other baseline models are the following: 4.24, 4.44, 4.56, 4.82, 4.88, and 6.77 for the models ACFTSVM, pin-IFTWSVM, TSVM, IFTWSVM-ID, ACFSVM, and IIFTWSVM, respectively. The average metric values across the KEEL datasets of the proposed CGFTSVM-ID and baseline models are presented in Table I. A comparison of the count of wins, ties, and losses for the proposed CGFTSVM-ID wrt the baseline models over KEEL datasets is shown in the first row of Table II, which is always greater for the proposed CGFTSVM-ID. Thus, the average rank and win, tie, loss statistics show the superiority of the proposed CGFTSVM-ID over the baseline models. Table SVI of the supplementary file denotes the G-mean and F-measure values of the proposed CGFTSVM-ID and baseline models over KEEL datasets. Table SVII of the supplementary file showcases the optimal parameters corresponding to the maximum AUC value. Section 2B of the supplementary material explains the effect on the AUC values of randomly selected KEEL data on varying different parameters. Figure S2 pictorially illustrates the effect of parameters on AUC values.

The performances on 18 UCI datasets are shown in Table SVIII of the supplementary file and Table I depicts the average AUC and rank over the 18 datasets. The average AUC values of the baseline models over UCI datasets are 0.7998, 0.7892, 0.7889, 0.7849, 0.7764, 0.762, 0.7186, for ACFTSVM, TSVM, IFTWSVM, IFTWSVM-ID, pin-IFTWSVM, ACFSVM, and IIFTWSVM, respectively, whereas the proposed CGFTSVM-ID has an average AUC value of 0.8098. Hence, CGFTSVM-ID has better classification performance over UCI datasets than baseline models. The mean rank of the proposed CGFTSVM-ID is 2.03, which is the least, followed by ACFTSVM with a value of 3.56. The average rank of TSVM, IFTWSVM, IFTWSVM-ID, pin-IFTWSVM, IIFTWSVM, and ACFSVM are 4.06, 4.11, 4.67, 4.72, 6.83, and 6.03, respectively. Table SIX of the supplementary file consists of the optimal parameters corresponding to maximum AUC values over UCI datasets. Figure S3 of the supplementary file depicts the pictorial comparison of the sensitivity metric of the baseline models with the proposed CGFTSVM-ID on UCI datasets. Through the last row of Table II, we demonstrate the count of wins, ties, and loses of the proposed CGFTSVM-ID wrt the baseline models over UCI datasets.

In order to experimentally verify the limitation of membership of ACFSVM [18] discussed in Section II-D wrt the proposed CPGB function, we carried out numerical experiments on the

TABLE II
PERFORMANCE EVALUATION OF THE PROPOSED CGFTSVM-ID AND BASELINE MODELS IN TERMS OF WIN-TIE-LOSS; THE NOTATION [WIN TIE LOSS] INDICATES THE COUNT OF WINS, TIES, AND LOSSES OF THE PROPOSED CGFTSVM-ID OVER THE RESPECTIVE COLUMN MODEL

Dataset	TSVM [7]	IFTWSVM [15]	IFTWSVM-ID [22]	pin-IFTWSVM [31]	IIFTWSVM [21]	ACFSVM [18]	ACFTSVM [29]
KEEL	[26 4 3]	[22 6 5]	[25 4 4]	[26 6 1]	[31 0 2]	[29 3 1]	[26 3 4]
UCI	[13 1 3]	[16 0 2]	[17 1 0]	[16 1 1]	[17 1 0]	[16 1 1]	[9 2 7]

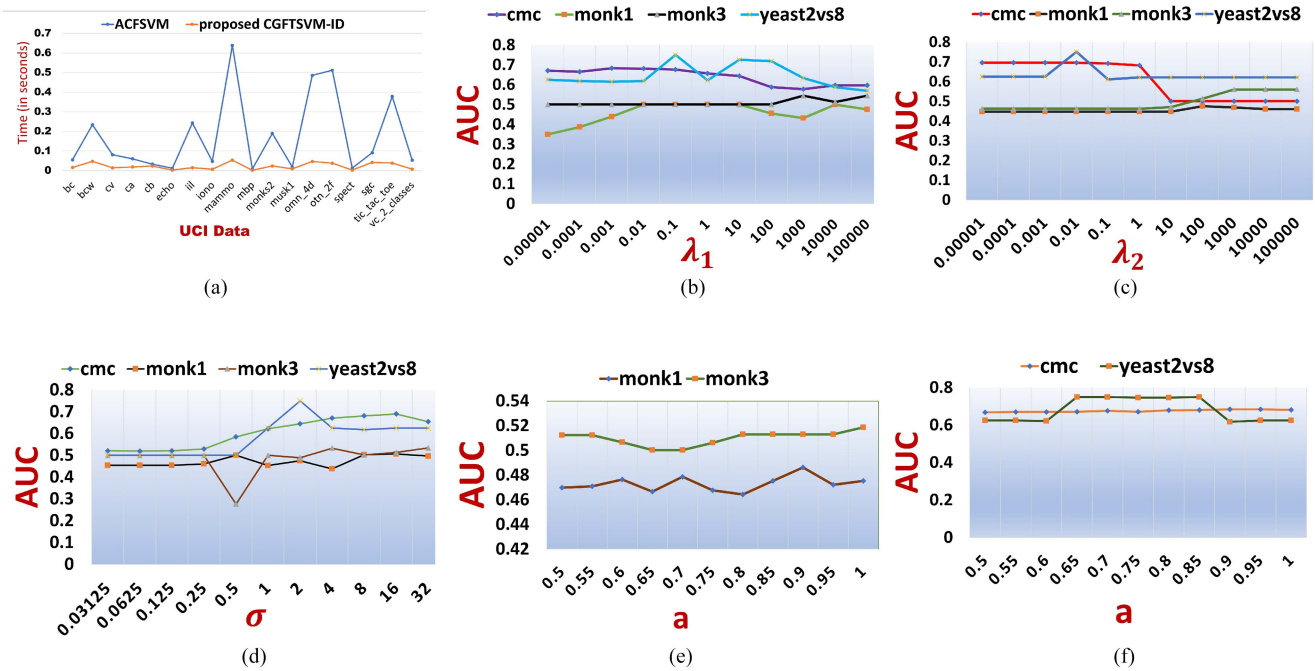

Fig. 4. (a) Training time of ACFSVM and proposed CGFTSVM-ID. (b)–(f) Effect of parameters on AUC values of the proposed CGFTSVM-ID over KEEL datasets cmc, monks1, monks3, and Yeast2vs8.

UCI datasets with their optimal parameters for the ACFSVM and proposed CGFTSVM-ID. The training time (in seconds) obtained for each dataset is demonstrated in Fig. 4(a), which clearly depicts that the proposed model is more efficient in terms of training time than ACFSVM.

D. Statistical Analysis Over Real-World KEEL Datasets

In this section, we statistically analyze the performance of the models over KEEL datasets by conducting the Friedman, Nemenyi-posthoc, and Wilcoxon signed rank test [34].

The null hypothesis of the Friedman test hypothesizes that the models are equivalent and possess identical average ranks. Friedman statistic is calculated as: $\chi_F^2 = \frac{12D}{l(l+1)}\left[\sum_k r_k^2 - \frac{l(l+1)^2}{4}\right]$ and has the chi-squared (χ_F^2) distribution having $l-1$ degrees of freedom. Here, r_k represents the average rank of the kth model, D and l correspond to the count of datasets and models, respectively. To refrain from the undesirably conservative nature of Friedman statistic, F_F statistic is calculated as: $F_F = \frac{(D-1)\chi_F^2}{D(l-1)-\chi_F^2}$, which follows F-distribution with degrees of freedom $(l-1, (l-1)(D-1))$. In our case, we have $(l=)8$ models executed over $(D=)33$ real-world KEEL datasets, thus, $\chi_F^2 = 62.5493$ and $F_F = 11.8823$ with degree of freedom 7 and (7,224), respectively. At 5% level of significance, $F_{(7,224)} = 2.0506$, which is less than calculated F_F, hence the null hypothesis is rejected and models under comparison are not equivalent.

To further strengthen the statistical analysis, we conduct the Nemenyi-posthoc test which includes pairwise comparison of the models. According to this test, if the difference in mean ranks between the two models exceeds the critical difference (Cdiff), they are deemed to be significantly different. The Cdiff is calculated as: Cdiff $= q_\alpha\sqrt{\frac{l(l+1)}{6D}}$, where $q_\alpha = 3.031$ for 8 classifiers [34]. The calculated value of Cdiff with $\alpha = 0.05$ is 1.8278. Table III first row demonstrates the significant difference between the proposed CGFTSVM-ID wrt the baseline models. It is clear from Table III that all the baseline models are significantly different from the proposed CGFTSVM-ID.

Further, we also conduct the Wilcoxon-signed rank test, a pairwise test with the null hypothesis assumption of the equivalence of the two models. It ranks the differences in the performances of the two models depending on their magnitude. Let R^+ and R^- denote the sum of positive and negative ranks, respectively. Table IV, first part, depicts the Wilcoxon signed rank tests of the performance of the proposed and baseline algorithms over KEEL datasets. If the p-value of an algorithm is less than 0.05,

TABLE III
NEMENYI POSTHOC SIGNIFICANT DIFFERENCE BETWEEN THE BASELINE COLUMN MODELS AND THE PROPOSED CGFTSVM-ID ACROSS UCI AND KEEL DATASETS

Dataset	Significance	TSVM [7]	IFTWSVM [15]	IFTWSVM-ID [22]	pin-IFTWSVM [31]	IIFTWSVM [21]	ACFSVM [18]	ACFTSVM [29]
KEEL	proposed CGFTSVM-ID	Yes	Yes	Yes	Yes	Yes	Yes	Yes
UCI	proposed CGFTSVM-ID	No	No	Yes	Yes	Yes	Yes	No

TABLE IV
WILCOXON-SIGNED RANK TEST OF PROPOSED CGFTSVM-ID WRT BASELINE MODELS ON KEEL AND UCI DATASETS

Model	R^+	R^-	p-value	Hypothesis
KEEL dataset				
TSVM[7]	406.5	28.5	< 0.00001	Rejected
IFTWSVM [15]	341	37	0.00026	Rejected
IFTWSVM-ID [22]	415	20	< 0.00001	Rejected
pin-IFTWSVM [31]	371	7	< 0.00001	Rejected
IIFTWSVM [21]	556	5	< 0.00001	Rejected
ACFSVM [18]	456	9	< 0.00001	Rejected
ACFTSVM [29]	400	65	0.00056	Rejected
UCI dataset				
TSVM [7]	131	22	0.00988	Rejected
IFTWSVM [15]	157	14	0.00188	Rejected
IFTWSVM-ID [22]	153	0	0.0003	Rejected
pin-IFTWSVM [31]	148	5	0.00072	Rejected
IIFTWSVM [21]	153	0	0.0003	Rejected
ACFSVM [18]	146	7	0.001	Rejected
ACFTSVM [29]	84	52	0.40654	Not rejected

then the null hypothesis assumption is rejected. Thus, first part of Table IV shows the baseline models are not equivalent to the proposed CGFTSVM-ID.

E. Statistical Analysis Over Real-World UCI Datasets

To further demonstrate the superiority of the proposed CGFTSVM-ID model, we perform statistical analysis on the results over UCI datasets. On simple calculations for $D = 18$ and $l = 8$, we get Friedman statistics $\chi_F^2 = 45.8019$ and $F_F = 9.7089$ with degree of freedom 7 and (7,119), respectively. At 5% level of significance $F_{(7,119)} = 2.1713$ which is less than calculated F_F, thus, the null hypothesis assumption is rejected leading to nonequivalence of the models. Now, we employ the pairwise Nemenyi-posthoc test and the Wilcoxon test. The calculated Cdiff value for $\alpha = 0.05$ is 2.0874. Table III, second row, shows all the models, except TSVM, IFTWSVM, and ACFTSVM are significantly different from the proposed CGFTSVM-ID. However, other statistical analysis and numerical experiments demonstrate the superiority of the proposed CGFTSVM-ID wrt TSVM, IFTWSVM, and ACFTSVM. Table IV, second part, contains the results of the Wilcoxon test over UCI datasets, which clearly demonstrate that the proposed CGFTSVM-ID and baseline models, except ACFTSVM, are not equivalent. The nonequivalence of ACFTSVM with the proposed CGFTSVM-ID is shown by other tests.

Hence, through the numerical experiments and statistical analysis of real-world datasets, we conclude that the proposed CGFTSVM-ID model is superior to the baseline models TSVM, IFTWSVM, IFTWSVM-ID, pin-IFTWSVM, IIFTWSVM, ACFSVM, and ACFTSVM.

F. Influence of Parameters on the Performance of the Proposed CGFTSVM-ID

The nonlinear case of the proposed CGFTSVM-ID has four parameters that have to be prespecified: the hyperparameters λ_1, λ_2, the membership function parameter a, and the Gaussian kernel parameter σ. To demonstrate the effect of each parameter on the generalization performance of the proposed CGFTSVM-ID, we carried out experiments on four randomly selected KEEL datasets—cmc, monks1, monks3, and yeast2vs8. While carrying out the investigations, we implemented grid search on the respective parameters whose effect is to be determined and the other parameters are fixed at their optimal values. The values of parameters λ_1, λ_2, and σ are same as given in experimental setup and $a \in [0.5, 0.55. \ldots, 0.95, 1]$. Fig. 4(b)–(f) represents the effect of parameters on AUC values, which clearly demonstrate that the parameters significantly effect the AUC values of the proposed CGFTSVM-ID. Thus, the generalization performance of the proposed model is sensitive to the choice of optimal parameters.

G. Choice of Steepness Parameter r in the Proposed CPGB Membership Function

To determine the optimal steepness parameter (r) of the proposed CGFTSVM-ID, we conducted numerical experiments on several real-world KEEL datasets, including bupa, crossplane150, ecoli0147vs2356, ecoli0234vs5, ecoli0267vs35, ecoli0347vs56, ecoli067vs35, ecoli01vs5, and glass5. We systematically tuned the parameter within the range of $[a, a/2, a/4, a/8]$. For most datasets, the optimal value for the steepness parameter (r) is found to be 'a,' except bupa, where the optimal value is determined as '$a/4$.' Importantly, the variation in the steepness parameter of bupa from 'a' to '$a/4$' did not lead to any significant change in the testing AUC value. Consequently, for the sake of consistency, we conducted all numerical experiments using a fixed value of $r = a$.

V. APPLICATION

In this section, we examine the application of the proposed CGFTSVM-ID in the real world by conducting numerical experiments on ADNI datasets,[1] which is publicly available. It was launched in 2003 by its Principal Investigator Michael W. Weiner to analyze different neuroimaging techniques for the diagnosis of AD from mild cognitive impairment (MCI). From the ADNI database, we downloaded and processed 150 T1-weighted images using the technique outlined in [35]. Subsequently, we employed our proposed CGFTSVM-ID and baseline models for the classification tasks involving AD versus MCI subjects, AD

[1][Online]. Available: www.adni-info.org

TABLE V
AUC VALUE WITH TRAINING TIME, SENSITIVITY, AND SPECIFICITY OF THE MODELS ON ADNI DATASETS

Dataset (ir) (samples,features)	TSVM [7] (AUC,time(seconds)) (Sens.,Spec.)	IFTWSVM [15] (AUC,time(seconds)) (Sens.,Spec.)	IFTWSVM-ID [22] (AUC,time(seconds)) (Sens.,Spec.)	pin-IFTWSVM [31] (AUC,time(seconds)) (Sens.,Spec.)	IIFTWSVM [21] (AUC,time(seconds)) (Sens.,Spec.)	ACFSVM [18] (AUC,time(seconds)) (Sens.,Spec.)	ACFTSVM [] (AUC,time(seconds)) (Sens.,Spec.)	proposed CGFTSVM-ID (AUC,time(seconds)) (Sens.,Spec.)
CN_vs_AD (1.22) (415, 91)	(0.8736, 0.0206) (0.8431, 0.9041)	(0.8834, 0.0444) (0.8627, 0.9041)	(0.8834, 0.0615) (0.8627, 0.9041)	(0.901, 0.0412) (0.8431, 0.9589)	(0.8033, 0.067) (0.9216, 0.6849)	(0.802, 0.053) (0.6863, 0.9178)	(**0.9128**, 0.0797) (0.9216, 0.90411)	(0.8962, 0.0244) (0.902, 0.8904)
CN_vs_MCI (1.75) (626, 91)	(0.6219, 0.0158) (0.8504, 0.3934)	(0.673, 0.029) (0.7559, 0.5902)	(0.7006, 0.0359) (0.811, 0.5902)	(0.6621, 0.0371) (0.4882, 0.8361)	(0.6461, 0.0988) (0.4724, 0.8197)	(0.6939, 0.0423) (0.5354, 0.8525)	(**0.7071**, 0.1037) (0.6929, 0.721311)	(0.6776, 0.029) (0.7323, 0.623)
MCI_vs_AD (2.13) (585, 91)	(0.6587, 0.0293) (0.5538, 0.7636)	(0.6615, 0.0209) (0.5231, 0.8)	(0.615, 0.0264) (0.3846, 0.8455)	(0.6462, 0.0271) (0.4923, 0.8)	(0.6455, 0.0441) (0.4, 0.8909)	(0.6182, 0.0416) (0.4, 0.8364)	(0.6178, 0.0849) (0.5538, 0.681818)	(**0.6801**, 0.0229) (0.5692, 0.7909)
Average AUC	0.7181	0.7393	0.733	0.7364	0.6983	0.7047	0.7459	**0.7513**
Average Rank	5.67	3.83	4.83	4	6.33	5.67	3	**2.67**

The bold values represent the best values.

subjects versus control normal (CN) subjects, and MCI versus CN subjects.

A. Experimental Discussion on AD Dataset

The numerical experiments are employed on the proposed CGFTSVM-ID and baseline models over ADNI datasets with their performance shown in Table V. For the CN versus AD case, ACFTSVM is the best-performing model with AUC 0.9128, pin-IFTWSVM with AUC 0.901 is the second-best model, followed by the proposed CGFTSVM-ID with AUC 0.8962. For the CN versus MCI classification, the proposed model lies at the fourth number in performance having AUC 0.6776 with ACFTSVM, IFTWSVM-ID, and ACFSVM being at the first, second and third number having AUC 0.7071, 0.7006, and 0.6939, respectively. The third and most important classification of MCI versus AD case, the proposed CGFTSVM-ID is the best performing with AUC value 0.6801 followed by baseline models in the order IFTWSVM, TSVM, pin-IFTWSVM, IIFTWSVM, ACFSVM, ACFTSVM, and IFTWSVM-ID with AUC values 0.6615, 0.6587, 0.6462, 0.6455, 0.6182, 0.6178, and 0.615, respectively. The average AUC of the proposed CGFTSVM-ID is 0.7513 and average rank for the proposed CGFTSVM-ID is 2.67, which are the best among the models. Tables SX and XI of the supplementary file consist of metric G-mean, F-measure, and optimal parameters corresponding to the maximum AUC over the ADNI dataset. For ease of visualization, the AUC values of the models are shown in Figure S4 of the supplementary file.

VI. CONCLUSION

In this article, a novel model termed CGFTSVM-ID is proposed to reduce the effect of noise/outliers and imbalanced data simultaneously. In CGFTSVM-ID, we employed a novel CPGB membership function according to which the majority class data points are assigned weights by combining the gbell function, class probability, and imbalance ratio. The gbell function effectively identified the potential outliers within the majority class. To further mitigate the impact of class noise and ID, we calculated the class probability of each majority class sample and considered the imbalance ratio. To ensure the significance of the minority class samples, they were assigned relatively high membership values. The experimental results and statistical analysis of the proposed CGFTSVM-ID with the baseline models on classifying real-world KEEL and UCI datasets clearly demonstrated the superior performance of the proposed model. The proposed CGFTSVM-ID is also applied to ADNI datasets to demonstrate its real-world applications for classifying AD subjects from CN and MCI. The proposed CGFTSVM-ID proved to be the best classifier for the diagnosis of MCI versus AD case, which is challenging to classify according to the literature. Nonetheless, it is important to highlight that the proposed CGFTSVM-ID is specifically tailored for binary class datasets. Furthermore, its applicability is limited when dealing with large-scale datasets due to the computational intensity associated with matrix inversion. Hence, there is potential for extending its functionality to accommodate multiclass classification tasks. Moreover, deep learning models had shown efficiency in feature extraction and dimension reduction, ultimately contributing to improved classification performance. Therefore, it could be worthwhile to explore the extension of the proposed CGFTSVM-ID to its deep variant.

ACKNOWLEDGMENT

The collection and dissemination of data for this investigation were funded by the Alzheimers Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Piramal Imaging; Servier; AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Johnson & Johnson Pharmaceutical Research & Development LLC.; Araclon Biotech; BioClinica, Inc.; Biogen; IXICO Ltd.; Bristol-Myers Squibb Company; Pharmaceuticals Corporation; CereSpir, Inc.; Cogstate; Eisai Inc.; Novartis Pfizer Inc.; Elan Pharmaceuticals, Inc.; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Takeda Pharmaceutical Company; Eli Lilly and Company; EuroImmun; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; and Transition Therapeutics. The Foundation for the National Institutes of Health (www.fnih.org) plays a facilitative role in securing private sector contributions. The Canadian Institutes of Health Research provide support for ADNI clinical locations in Canada. The Northern California Institute for Research and Education is the grantee agency, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. The dissemination of ADNI data is handled by the Center for Neuro Imaging at the University of Southern California.

REFERENCES

[1] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1999.

[2] S. Wang, X. Yang, Y. Zhang, P. Phillips, J. Yang, and T.-F. Yuan, "Identification of green, oolong and black teas in China via wavelet packet entropy and fuzzy support vector machine," *Entropy*, vol. 17, no. 10, pp. 6663–6682, 2015.

[3] S. Wang, Y. Li, Y. Shao, C. Cattani, Y. Zhang, and S. Du, "Detection of dendritic spines using wavelet packet entropy and fuzzy support vector machine," *CNS Neurological Disord.-Drug Targets (Formerly Curr. Drug Targets-CNS Neurological Disorders)*, vol. 16, no. 2, pp. 116–121, 2017.

[4] B. Richhariya and M. Tanveer, "EEG signal classification using universum support vector machine," *Expert Syst. Appl.*, vol. 106, pp. 169–182, 2018.

[5] E. Osuna and F. Girosi, "Reducing the run-time complexity of support vector machines," in *Proc. Int. Conf. Pattern Recognit.*, Citeseer, 1998, pp. 687–695.

[6] O. L. Mangasarian and E. W. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 69–74, Jan. 2006.

[7] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905–910, May 2007.

[8] C. Zhang, Y. Tian, and N. Deng, "The new interpretation of support vector machines on statistical learning theory," *Sci. China Ser. A: Math.*, vol. 53, no. 1, pp. 151–164, 2010.

[9] Y.-H. Shao, C.-H. Zhang, X.-B. Wang, and N.-Y. Deng, "Improvements on twin support vector machines," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 962–968, Jun. 2011.

[10] M. Tanveer, T. Rajani, R. Rastogi, Y.-H. Shao, and M. A. Ganaie, "Comprehensive review on twin support vector machines," *Ann. Operations Res.*, pp. 1–46, 2022, doi: 10.1007/s10479-022-04575-w.

[11] M. Hanmandlu, "A new entropy function and a classifier for thermal face recognition," *Eng. Appl. Artif. Intell.*, vol. 36, pp. 269–286, 2014.

[12] J. Laurikkala, "Improving identification of difficult small classes by balancing class distribution," in *Proc. Artif. Intell. Med.: 8th Conf. Europe*, 2001, pp. 63–66.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[14] S. Rezvani and X. Wang, "A broad review on class imbalance learning techniques," *Appl. Soft Comput.*, vcol. 143, 2023, Art. no. 110415. https://doi.org/10.1016/j.asoc.2023.110415

[15] S. Rezvani, X. Wang, and F. Pourpanah, "Intuitionistic fuzzy twin support vector machines," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 11, pp. 2140–2151, Nov. 2019.

[16] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 464–471, Mar. 2002.

[17] R. Batuwita and V. Palade, "FSVM-CIL: Fuzzy support vector machines for class imbalance learning," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 558–571, Jun. 2010.

[18] X. Tao et al., "Affinity and class probability-based fuzzy support vector machine for imbalanced data sets," *Neural Netw.*, vol. 122, pp. 289–307, 2020.

[19] D. M. Tax and R. P. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, pp. 45–66, 2004.

[20] M. Tanveer, A. Tiwari, R. Choudhary, and M. Ganaie, "Large-scale pinball twin support vector machines," *Mach. Learn.*, vol. 111, pp. 3525–3548, 2021, doi: 10.1007/s10994-021-06061-z.

[21] M. A. Ganaie, A. Kumari, A. K. Malik, and M. Tanveer, "EEG signal classification using improved intuitionistic fuzzy twin support vector machines," *Neural Comput. Appl.*, vol. 36, pp. 163–179, 2022, doi: 10.1007/s00521-022-07655-x.

[22] S. Rezvani and X. Wang, "Intuitionistic fuzzy twin support vector machines for imbalanced data," *Neurocomputing*, vol. 507, pp. 16–25, 2022.

[23] M. Tanveer, M. A. Ganaie, A. Bhattacharjee, and C. T. Lin, "Intuitionistic fuzzy weighted least squares twin SVMs," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4400–4409, Jul. 2023, doi: 10.1109/TCYB.2022.3165879.

[24] M. Ganaie, M. Tanveer, and C.-T. Lin, "Large-scale fuzzy least squares twin SVMs for class imbalance learning," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 11, pp. 4815–4827, Nov. 2022.

[25] M. Ganaie, M. Tanveer, and A. D. N. Initiative, "KNN weighted reduced universum twin SVM for class imbalance learning," *Knowl.-Based Syst.*, vol. 245, 2022, Art. no. 108578.

[26] Y. Tian, M. Sun, Z. Deng, J. Luo, and Y. Li, "A new fuzzy set and nonkernel SVM approach for mislabeled binary classification with applications," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1536–1545, Dec. 2017.

[27] Y. Tian, Z. Deng, J. Luo, and Y. Li, "An intuitionistic fuzzy set based S^3VM model for binary classification with mislabeled information," *Fuzzy Optim. Decis. Mak.*, vol. 17, pp. 475–494, 2018.

[28] L. Zhang, Q. Jin, S. Fan, and D. Liu, "A novel dual-center based intuitionistic fuzzy twin bounded large margin distribution machines," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 9, pp. 3121–3134, Sep. 2023, doi: 10.1109/TFUZZ.2023.3245215.

[29] B. B. Hazarika, D. Gupta, and P. Borah, "Fuzzy twin support vector machine based on affinity and class probability for class imbalance learning," *Knowl. Inf. Syst.*, vol. 65, pp. 5259–5288, 2023.

[30] A. Dorzhigulov and A. P. James, "Generalized bell-shaped membership function generation circuit for memristive neural networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2019, pp. 1–5.

[31] Z. Liang and L. Zhang, "Intuitionistic fuzzy twin support vector machines with the insensitive pinball loss," *Appl. Soft Comput.*, vol. 115, 2022, Art. no. 108231.

[32] O. L. Mangasarian and D. R. Musicant, "Successive overrelaxation for support vector machines," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1032–1037, Sep. 1999.

[33] D. Dua and C. Graff, "UCI machine learning repository," 2017.

[34] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.

[35] B. Richhariya, M. Tanveer, A. H. Rashid, and A. D. N. Initiative, "Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE)," *Biomed. Signal Process. Control*, vol. 59, 2020, Art. no. 101903.