Granular Ball Twin Support Vector Machine With Pinball Loss Function

A. Quadir and M. Tanveer, *Senior Member, IEEE*

***Abstract*—Alzheimer's disease (AD) and Schizophrenia (SCZ) are prominent neurodegenerative conditions and leading causes of dementia, resulting in progressive cognitive decline and memory loss. Several studies reveal that early detection and intervention can slow the progression of AD and SCZ. Numerous machine learning algorithms including twin support vector machine (TSVM) have been proposed for the early diagnosis of AD and SCZ. However, TSVM grapples with significant challenges: 1) TSVM relies on the hinge loss function, resulting in susceptibility to noise and instability; 2) TSVM encounters challenges in effectively handling large datasets, attributed to its computational complexity and dependence on matrix inversions. Keeping in view the aforementioned challenges, in this article, we propose a novel granular ball twin support vector machine with pinball loss function (Pin-GBTSVM). Pin-GBTSVM employs granular balls, as opposed to individual data points, as inputs for constructing a classifier, while also leveraging the pinball loss function to attain a heightened level of noise insensitivity. The proposed Pin-GBTSVM persists in facing challenges associated with the absence of integration of the structural risk minimization (SRM) principle and the requirement for matrix inversions. We further propose a novel large-scale Pin-GBTSVM (Pin-LGBTSVM). Pin-LGBTSVM achieves two crucial objectives: 1) it eliminates the necessity for matrix inversions, streamlining the computational efficiency of Pin-GBTSVM; and 2) it integrates the SRM principle by incorporating regularization terms, effectively addressing the concern of overfitting. Experiments are conducted on University of California Irvine (UCI), knowledge extraction based on evolutionary learning (KEEL), and normally distributed clustered (NDC) benchmark datasets, where both the proposed Pin-GBTSVM and Pin-LGBTSVM models consistently outperform the baseline models in terms of generalization performance. Furthermore, we implemented the proposed Pin-GBTSVM and Pin-LGBTSVM models on SCZ and Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets, showcasing the model's efficacy in real-world applications.**

***Index Terms*—Granular computing, noise insensitivity, pinball loss, quantile distance, twin support vector machine (TSVM).**

I. Introduction

NEUROLOGICAL conditions including Alzheimer's disease (AD) and Schizophrenia (SCZ) are prominent concerns in today's global health initiatives [1], [2]. AD is a form of dementia that primarily impacts older individuals. According to the World Alzheimer's Report, an estimated 152 million individuals will develop AD by the year 2050 [3]. AD causes brain structural and functional changes. This process takes years, turning healthy people into Alzheimer's patients [4]. SCZ, on the other hand, is a serious mental disorder characterized by positive symptoms such as hallucinations, delusions, and disorganized speech, as well as negative symptoms such as flattened affect, alogia, and avolition. Additionally, language disturbances are common in individuals with SCZ [5]. The early detection and diagnosis of SCZ and AD pose significant challenges due to the presence of multiple comorbidities associated with the disorder. These factors can complicate the optimal management of patients and potentially restrict positive outcomes. Individuals with SCZ and AD are increasingly turning to social media platforms to share their mental health concerns, connect with others who have similar experiences, and seek social support [6]. The textual content shared on social media platforms presents new opportunities for enhancing our understanding of self-expressed SCZ and AD at both the individual and community levels. Recent studies have concentrated on analyzing language from social media posts [7], [8]. With the emergence of social media, individuals coping with mental disorders have discovered online mental health communities where they can share their experiences. Recognizing the critical need for early diagnosis of SCZ and AD, the utilization of machine learning and natural language processing techniques on multimedia data provides a novel approach for detecting various types of mental illnesses [9], [10].

Support vector machines (SVMs) [11] have evolved into a formidable tool for tackling classification problems. Inspired by statistical learning theory and the maximum margin principle, SVM endeavors to determine an optimal separating hyperplane through the solving of a quadratic programming problem (QPP). SVM has found broad application in numerous real-world problems, such as remote sensing [12], electroencephalogram signal classification [13], text categorization [14], and feature extraction [15]. The SVM exhibits superior performance as it adheres to the structural risk minimization (SRM) principle. However, it faces challenges in real-world applications due to its higher computational complexity. Mangasarian

Manuscript received 29 February 2024; revised 3 May 2024 and 28 May 2024; accepted 4 June 2024. This work was supported by the Science and Engineering Research Board (SERB) through the MTR/2021/000787 grant as part of the Mathematical Research Impact-Centric Support (MATRICS) scheme. *(Corresponding author: M. Tanveer.)*

The authors are with the Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore 453552, India (e-mail: mscphd2207141002@iiti.ac.in; mtanveer@iiti.ac.in).

This article has supplementary downloadable material available at https://doi.org/10.1109/TCSS.2024.3411395, provided by the authors.

Digital Object Identifier 10.1109/TCSS.2024.3411395

and Wild [16] proposed generalized eigenvalue proximal SVM (GEPSVM), to alleviate the complexity associated with SVM. Subsequently, Jayadeva et al. [17] proposed twin SVM (TSVM) for solving binary classification problems. TSVM generates hyperplanes in such a way that each hyperplane is proximal to one of the classes and lies at a distance of atleast one unit from the other class. TSVM tackles two smaller QPPs rather than addressing a single large QPP, resulting in a speed enhancement of approximately four times compared to the standard SVM. In recent years, various variants of TSVM have been introduced for both small and large datasets such as twin-bounded SVM (TBSVM) [18], intuitionistic fuzzy weighted LSTSVM (IFW-LSTSVM) [19], and multiview large margin distribution machine (MVLDM) [20].

Although TSVM has its merits, it is still plagued by issues such as sensitivity to noise and instability when subjected to resampling, primarily due to its hinge loss function. The pinball loss function utilizes a quantile metric to measure margin distances, with the goal of reducing sensitivity to noise and enhancing stability, particularly in scenarios involving resampling. However, the incorporation of the pinball loss function precipitates a diminution in sparsity, a pivotal constituent influencing the performance of SVM classification. To address this limitation, pinball loss SVM (Pin-SVM) is proposed in [21], to mitigate the impact of noise during the process of obtaining sparse solutions. The authors showed that Pin-SVM demonstrates greater stability in handling outliers and resampling challenges compared to hinge loss-based SVM. Both models exhibit similar training times. Tanveer et al. [22] introduced generalized TSVM incorporating the pinball loss function, termed Pin-GTSVM. This model is characterized by its insensitivity to noise and enhanced stability in resampling. The computational complexity of Pin-GTSVM does not experience an escalation despite the incorporation of the pinball loss function.

Granular computing stands out as a significant methodology for emulating human cognition in addressing intricate problems. Since its inception by Lin and Zadeh in 1996, a growing number of scholars have delved into the concept of information granules. Granular computing aligns seamlessly with human cognition, following a paradigm of transitioning from large to small and from coarse to fine [23]. Within the realm of granular computing, increased efficiency and enhanced robustness against noise are associated with coarser granularity [24], [25]. Conversely, finer granularity in granular computing corresponds to a more detailed description of the objects under consideration [26]. As a result, observing, and analyzing problems with coarser granularity demonstrates robustness to noise. In contrast, many machine learning models rely on pixels or data points for training, often operating at the lowest available resolution. This approach lacks the efficiency and scalability inherent in the adaptable granulating capabilities of the human brain. The brain's capacity to granulate information as required plays a pivotal role in its scalability and efficiency. This stands in stark contrast to the constraints imposed on machine learning models due to their reliance on pixel-level granularity. The innovative classifier, which integrates granular computing and SVM [27], [28], was devised to encompass the concept of granular balls. This classifier employs hyperballs to partition datasets into different granular ball sizes. As emphasized in [29], this approach, mimicking cognitive processes observed in the human brain, provides a scalable, dependable, and efficient solution within the domain of granular computing by introducing larger granularity sizes. Researchers consistently explore novel applications, refine existing methodologies [30], [31], [32], and bridge interdisciplinary gaps to harness the potential of granular computing across diverse domains [33].

To effectively address issues related to noise and outliers, an efficient approach known as granular ball SVM (GBSVM) [34] has been proposed. This method integrates the granular ball concept into the framework of SVM. In GBSVM, each individual finest granularity point is substituted with a granular ball characterized by a center and a radius. This replacement is a distinctive feature of GBSVM's approach to handling granularity. A substantial radius signifies a more extensive granularity, whereas a diminished value indicates a more intricate granularity in the representation. Inspired by the advantages of employing a granular approach to handle noise and outliers, coupled with the efficiency demonstrated by TSVM, the potential integration of these two concepts not only appears intriguing but also holds promise. Therefore, we propose a granular ball TSVM with pinball loss (Pin-GBTSVM). In Pin-GBTSVM, granular balls are characterized by their center and radius, which leads to enhanced computational efficiency and a significant reduction in the number of training instances. This reduction consequently contributes to a decrease in overall training time. The essence of a granular ball resides in its center, which encapsulates all the relevant information of the data points within the ball. Utilizing the pinball loss function with an insensitive zone not only enhances the robustness of models to noise and outliers but also preserves the sparsity of the proposed model. However, the proposed Pin-GBTSVM model incorporates matrix inversions that impede its scalability for large-scale datasets and lack the SRM principle, thereby increasing the risk of overfitting. To address this, we further proposed large-scale Pin-GBTSVM (Pin-LGBTSVM), to effectively handle extensive datasets. Pin-LGBTSVM incorporates a regularization term in its primal form, eliminating the necessity for matrix inversions and mitigating the risk of overfitting. The contributions of this article are as follows.

1) We propose a novel granular ball TSVM with pinball loss (Pin-GBTSVM). Pin-GBTSVM utilizes granular balls as input instead of individual data points for constructing optimal nonparallel separating hyperplanes. The incorporation of granular balls significantly reduces the training time and enhances the overall model performance.
2) Pin-GBTSVM exhibits insensitivity to noise, maintains the sparseness of samples, and demonstrates stability for resampling, attributed to the utilization of the pinball loss function.
3) We propose a novel large-scale Pin-GBTSVM (Pin-LGBTSVM) to tackle the challenges posed by handling large-scale datasets, addressing concerns associated with the proposed Pin-GBTSVM model. Pin-LGBTSVM endeavors to reduce the

structural risk inherent in its formulation by incorporating the SRM principle. The proposed Pin-LGBTSVM demonstrates efficiency, scalability for large datasets, adept handling of overfitting, robustness, and enhanced generalization performance.

4) We conducted experiments on 33 real-world UCI [35] and KEEL [36] datasets (with and without label noise), as well as large-scale NDC [38] datasets ranging from 10 k to 5 m. Numerical experiments and statistical analysis showcase the superiority of the proposed Pin-GBTSVM and Pin-LGBTSVM models compared to the baseline models.
5) The application of machine learning techniques on multimedia data presents a novel approach for detecting various types of mental illnesses. SCZ and AD, a mental disorder extensively discussed on social media platforms, are diagnosed using machine learning techniques applied to multimedia, T1-weighted MRI datasets. We employed the proposed Pin-GBTSVM and Pin-LGBTSVM models for diagnosing SCZ and AD. Experimental results illustrate the superiority of the proposed models when compared to the baseline models.

The remainder of this article is organized as follows. In Section II, an overview of related work is provided. Sections III and IV elaborate on the mathematical formulation of the proposed Pin-GBTSVM and Pin-LGBTSVM models. The time complexity of the proposed Pin-GBTSVM is given in Section V. In Section VI, we discuss the experimental results and conduct statistical analysis. The conclusion and avenues for future work are presented in Section VII.

II. Related Work

We delve into the granular ball computing method in this section, discussing the pinball loss function, and the mathematical formulation of GBSVM, and TSVM in Section S.I in the Supplementary material. Let $\mathscr{D} = \{(x_i, t_i), i = 1, 2, 3, \ldots, n\}$ be the training dataset where $t_i \in \{+1, -1\}$ represents the label of $x_i \in \mathbb{R}^{1\times m}$. Consider the input matrices $X_1 \in \mathbb{R}^{n_1\times m}$ and $X_2 \in \mathbb{R}^{n_2\times m}$, where n_1 (n_2) is the number of data samples belonging to the $+1$ (-1) class, such that the total number of data samples is $n = n_1 + n_2$, and m is the total number of attributes of each data sample. Let $G = \{GB_i,\ \ i = 1, 2, \ldots, p\} = \{((c_i, r_i), y_i),\ \ i = 1, 2, \ldots, p\}$ be the set of generated granular balls, where c_i signifies the center, r_i indicates the radius, and y_i is the label of the i^{th} granular ball. $C_1 \in \mathbb{R}^{p_1\times m}$ ($C_2 \in \mathbb{R}^{p_2\times m}$) encapsulate the center affiliated with the class $+1$ (-1), where p_1 and p_2 are the total number of granular ball centers of $+1$ (-1) class, respectively. The cumulative number of centers is $p_1 + p_2 = p$. $R_1 \in \mathbb{R}^{p_1\times 1}$ ($R_2 \in \mathbb{R}^{p_2\times 1}$) denote the vectors encompassing the radii of the associated generated granular balls for $+1$ (-1) class.

A. Granular Ball Computing

Granular ball computing is a significant data processing technique introduced by Xia et al. [27] to tackle the scalability challenges associated with high-dimensional data. The fundamental concept of granular ball computing revolves around employing a hyperball to encompass all or a portion of the sample space. Utilizing the "granular ball" as input to represent the sample space enables the acquisition of multigranularity learning characteristics and the precise characterization of the sample space. Let "c" denotes the center of gravity of the data points encompassed within a granular ball, and it is defined as $c = (1/l)\sum_{i=1}^{l} x_i$, where "$x_i$" and "$l$" represent a data point, and the total number of data points inside the granular ball, respectively. The radius "r" of the granular ball is calculated as $r = (1/l)\sum_{i=1}^{l} \|x_i - c\|$. The label of the granular ball is determined by choosing the label that is most frequently observed among the samples enclosed within the granular ball. To quantitatively assess the divided granular ball's mass, the notion of a "purity threshold" is introduced. This threshold pertains to the percentage of samples within a granular ball that share the same label, specifically the majority labels.

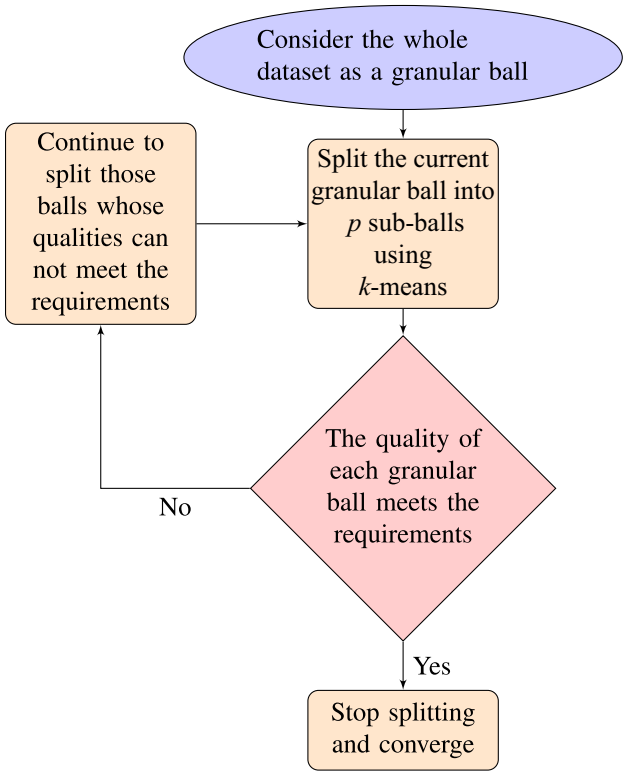

Fig. 1. Process of the granular ball generation.

Consider $\mathscr{D} = \{(x_i,\ t_i), i = 1, 2, 3, \ldots, n\}$, where x_i represents a data sample, and n represents the total number of samples in $\mathscr{D}$. The granular balls generated from the dataset $\mathscr{D}$ is denoted by GB_j $(j = 1, 2, 3, \ldots, p)$. Here, p represents the total number of granular balls generated from the dataset $\mathscr{D}$. The optimization problem of the granular-balls can be expressed as follows:

$$\begin{aligned} &\min\ \ \theta_1 \times \frac{n}{\sum_{i=1}^{p}|GB_i|} + \theta_2 \times p \\ &s.t.\ \ \text{purity}(GB_j) \geq \mathscr{T},\ \ j = 1, 2, \ldots, p, \end{aligned} \tag{1}$$

where $\mathscr{T}$ denotes the purity threshold, and θ_1 and θ_2 represent the weight coefficients.

III. Proposed Pin-GBTSVM

In this section, we introduce a novel Pin-GBTSVM to address the binary classification problem. We propose the use of granular balls that encompass either the complete sample space or a fraction of it during the training process. This method aids in separating nonparallel hyperplanes using granular balls and effectively classifying the original data points. In constructing optimal separating hyperplanes, our goal is to maximize the information from all training data points while minimizing the number of data points needed. To achieve this, we integrate both the centers and radii of all generated granular balls. Pin-GBTSVM exhibits insensitivity to noise, maintains the sparseness of samples, and demonstrates stability for resampling, attributed to the utilization of the pinball loss function. The optimization problem of linear Pin-GBTSVM is given as follows:

$$\begin{aligned}\min_{w_1,b_1}\ & \frac{1}{2}\|C_1w_1+e_1b_1\|^2+d_1e_2^T\zeta_2\\ s.t.\ & -(C_2w_1+e_2b_1)+\zeta_2\geq e_2+R_2,\\ & -(C_2w_1+e_2b_1)-\frac{\zeta_2}{\tau_2}\leq e_2+R_2,\end{aligned}\tag{2}$$

and

$$\begin{aligned}\min_{w_2,b_2}\ & \frac{1}{2}\|C_2w_2+e_2b_2\|^2+d_2e_1^T\zeta_1\\ s.t.\ & (C_1w_2+e_1b_2)+\zeta_1\geq e_1+R_1,\\ & (C_1w_2+e_1b_2)-\frac{\zeta_1}{\tau_1}\leq e_1+R_1.\end{aligned}\tag{3}$$

The Lagrangian function of (2) is

$$\begin{aligned}L=&\frac{1}{2}\|C_1w_1+e_1b_1\|^2+d_1e_2^T\zeta_2\\ &-\alpha_1^T(-(C_2w_1+e_2b_1)+\zeta_2-e_2-R_2)\\ &+\beta_1^T\left(-(C_2w_1+e_2b_1)-\frac{\zeta_2}{\tau_2}-e_2-R_2\right),\end{aligned}\tag{4}$$

where α_1 and β_1 are the vectors of Lagrangian multipliers. Applying the Karush–Kuhn–Tucker (K.K.T.) conditions, we obtain:

$$C_1^T(C_1w_1+e_1b_1)+C_2^T\alpha_1-C_2^T\beta_1=0,\tag{5}$$

$$e_1^T(C_1w_1+e_1b_1)+e_2^T\alpha_1-e_2^T\beta_1=0,\tag{6}$$

$$d_1e_2-\alpha_1-\frac{\beta_1}{\tau_2}=0.\tag{7}$$

Since $\alpha_1\geq 0$, and using the equation (7), we obtain as follows:

$$-\tau_2d_1e_2\leq(\alpha_1-\beta_1).\tag{8}$$

Combining (5) and (6) leads to

$$\begin{pmatrix}C_1^T\\ e_1^T\end{pmatrix}(C_1\ \ e_1)\begin{pmatrix}w_1\\ b_1\end{pmatrix}+\begin{pmatrix}C_2^T\\ e_2^T\end{pmatrix}(\alpha_1-\beta_1)=0.\tag{9}$$

Define $P=(C_1\ e_1)$, $Q=(C_2\ e_2)$ and $z_1=(w_1;\ b_1)$, then (9) can be rewritten as follows:

$$z_1=-(P^TP)^{-1}Q^T(\alpha_1-\beta_1).\tag{10}$$

Calculating the inverse of P^TP poses a significant challenge. Nevertheless, this problem can be tackled by introducing a regularization term represented as δI in (10), where I represent an identity matrix of suitable dimensions. Thus

$$z_1=-(P^TP+\delta I)^{-1}Q^T(\alpha_1-\beta_1).\tag{11}$$

Using (11) and the K.K.T. conditions mentioned above, we can derive the dual form of (2) as follows:

$$\begin{aligned}\max_{\alpha_1,\beta_1}\ & -\frac{1}{2}(\alpha_1-\beta_1)^TQ(P^TP+\delta I)^{-1}Q^T(\alpha_1-\beta_1)\\ & +(\alpha_1-\beta_1)^T(e_2+R_2)\\ s.t.\ & -\tau_2d_1e_2\leq(\alpha_1-\beta_1).\end{aligned}\tag{12}$$

Likewise, the Wolfe dual for (3) can be obtained as follows:

$$\begin{aligned}\max_{\alpha_2,\beta_2}\ & -\frac{1}{2}(\alpha_2-\beta_2)^TP(Q^TQ+\delta I)^{-1}P^T(\alpha_2-\beta_2)\\ & +(\alpha_2-\beta_2)^T(e_1+R_1)\\ s.t.\ & -\tau_1d_2e_1\leq(\alpha_2-\beta_2).\end{aligned}\tag{13}$$

Similarly, $z_2=(w_2;\ b_2)$ for the -1 class can be computed using the following equation:

$$z_2=(Q^TQ+\delta I)^{-1}P^T(\alpha_2-\beta_2).\tag{14}$$

After computing the optimal values of z_1 and z_2, the classification of a new input data point x into either the $+1$ or -1 class can be determined as follows:

$$\text{class}(i)=\arg\min_{i\in\{1,2\}}\frac{|w_i^Tx+b_i|}{\|w_i\|}.\tag{15}$$

The mathematical formulation of the Pin-GBTSVM model for the nonlinear kernel is outlined in Section S.II.A of the Supplementary material.

IV. Proposed Large Scale Pin-Gbtsvm (Pin-Lgbtsvm)

The efficiency of Pin-GBTSVM may decrease when dealing with large datasets due to its reliance on matrix inversion for solving the QPPs. Additionally, Pin-GBTSVM lacks the SRM principle. To adapt our proposed Pin-GBTSVM model for large-scale datasets, we strive to reformulate the problem to incorporate the advantages of both Pin-GTSVM [22] and TBSVM [18], thereby removing the requirement to compute massive matrix inverses. We propose the Pin-LGBTSVM model, which involves the inclusion of a regularization term and the incorporation of an equality constraint. Pin-LGBTSVM demonstrates insensitivity to noise, making it more stable in the context of resampling. Last, Pin-LGBTSVM inherently embodies the SRM principle, potentially enabling it to achieve enhanced classification accuracy. The linear optimization problem corresponding to the Pin-LGBTSVM is given as follows:

$$\min_{w_1,b_1,\eta_1,\xi_2}\ \frac{1}{2}d_3(\|w_1\|^2+b_1^2)+\frac{1}{2}\eta_1^T\eta_1+d_1e_2^T\zeta_2$$

$$\begin{aligned} s.t.\ \ & C_1 w_1 + e_1 b_1 = \eta_1, \\ & -(C_2 w_1 + e_2 b_1) + \zeta_2 \geq e_2 + R_2, \\ & -(C_2 w_1 + e_2 b_1) - \frac{\zeta_2}{\tau_2} \leq e_2 + R_2, \end{aligned} \tag{16}$$

and

$$\begin{aligned} \min_{w_2, b_2, \eta_2, \xi_1} \ & \frac{1}{2} d_4 (\|w_2\|^2 + b_2^2) + \frac{1}{2} \eta_2^T \eta_2 + d_2 e_1^T \zeta_1 \\ s.t.\ \ & C_2 w_2 + e_2 b_2 = \eta_2, \\ & (C_1 w_2 + e_1 b_2) + \zeta_1 \geq e_1 + R_1, \\ & (C_1 w_2 + e_1 b_2) - \frac{\zeta_1}{\tau_1} \leq e_1 + R_1, \end{aligned} \tag{17}$$

where ζ_1 and ζ_2 are slack variables, $d_1, d_2, d_3, d_4 (> 0)$ are penalty parameters, and e_2 and e_2 are vectors of one of appropriate dimensions. The dual form of equation (16) is presented as follows:

$$\begin{aligned} \min_{\mu_1, \beta_1} \ & \frac{1}{2} \begin{pmatrix} \mu_1^T & \beta_1^T \end{pmatrix} \tilde{P} \begin{pmatrix} \mu_1 \\ \beta_1 \end{pmatrix} - d_3 \beta_1^T (e_2 + R_2) \\ s.t.\ \ & -\tau_2 d_1 e_2 \leq \beta_1 \leq d_1 e_2, \\ \text{where } \ & \tilde{P} = \begin{pmatrix} C_1 C_1^T + d_3 I & C_1 C_2^T \\ C_2 C_1^T & C_2 C_2^T \end{pmatrix} + E. \end{aligned} \tag{18}$$

Here, the matrix E comprises entirely of ones, whereas I denotes the identity matrix of the appropriate dimension.

The dual formulation for (17) is expressed as follows:

$$\begin{aligned} \min_{\mu_2, \beta_2} \ & \frac{1}{2} \begin{pmatrix} \mu_2^T & \beta_2^T \end{pmatrix} \tilde{Q} \begin{pmatrix} \mu_2 \\ \beta_2 \end{pmatrix} - d_4 \beta_2^T (e_1 + R_1) \\ s.t.\ \ & -\tau_1 d_2 e_1 \leq \beta_2 \leq d_2 e_1, \\ \text{where } \ & \tilde{Q} = \begin{pmatrix} C_2 C_2^T + d_4 I & -C_2 C_1^T \\ -C_1 C_2^T & C_1 C_1^T \end{pmatrix} + E, \end{aligned} \tag{19}$$

and w_2 and b_2 can be given by

$$\begin{pmatrix} w_2 \\ b_2 \end{pmatrix} = \frac{1}{d_4} \begin{pmatrix} C_2^T & C_1^T \\ e_2^T & e_1^T \end{pmatrix} \begin{pmatrix} \mu_2 \\ \beta_2 \end{pmatrix}. \tag{29}$$

The pair of nonparallel hyperplanes $w_1^T x + b_1 = 0$, and $w_2^T x + b_2 = 0$ is determined by the optimal values of μ_1, β_1 and μ_2, β_2, respectively, where

$$w_1^* = -\frac{1}{d_3}(C_1^T \mu_1 + C_2^T \beta_1) \quad b_1^* = -\frac{1}{d_3}(e_1^T \mu_1 + e_2^T \beta_1)$$

$$w_2^* = \frac{1}{d_4}(C_2^T \mu_2 + C_1^T \beta_2) \quad b_2^* = \frac{1}{d_4}(e_1^T \mu_2 + e_2^T \beta_2).$$

The classification of a new input data point x into either the $+1$ or -1 class is then analogous to the decision function employed in the Pin-GBTSVM problem. The mathematical formulation of the Pin-LGBTSVM model for the linear and non-linear kernels are outlined in subsections S.II.B and S.II.C of the Supplementary material.

V. Time Complexity and Algorithm

Let $\mathscr{D}$ represent the training dataset having n samples. The time complexity of the k-means algorithm is $\mathcal{O}(nkt)$ [37], where the parameter k represents the number of clusters, and t denotes the number of iterations. Also, the time complexity of standard TSVM [17] is $\mathcal{O}(n^3/4)$. The Computational complexity of Pin-GTSVM [22] is similar to TSVM. We consider the training dataset $\mathscr{D}$ as the initial granular ball (GB) set. The set GB is divided into two granular balls using the 2-means clustering. During the initial phase of splitting, the computational complexity is $\mathcal{O}(2n)$. In the second phase, the two generated granular balls are further divided into four granular balls (if both the granular balls are impure), with a maximum computational complexity of $\mathcal{O}(2n)$, and so on. If there is a total t number of iterations, then the overall computational complexity of generating granular balls is (or less than) $\mathcal{O}(2nt)$. The complexity of the proposed Pin-GBTSVM model is (or less than) $\mathcal{O}(p^3/4) + \mathcal{O}(2nt)$, where p represents the total number of generated granular balls. Also, $\mathcal{O}(p^3) \ll \mathcal{O}(n^3)$ as $p \ll n$. Therefore, we compare our proposed Pin-GBTSVM model with the standard TSVM model given as follows: $\mathcal{O}(p^3/4) + \mathcal{O}(2nt) \ll \mathcal{O}(n^3/4)$. Therefore, the time complexity of the proposed Pin-GBTSVM model is much lower than TSVM, and hence, the proposed Pin-GBTSVM model is more efficient than the baseline models. This implies that in real-time applications, our Pin-GBTSVM model offers enhanced efficiency compared to baseline models, thereby facilitating quicker and more responsive processing of data, which is crucial for timely decision-making and application responsiveness. Comparison of the computational and space complexity of the proposed models along with the baseline models are shown in Table S.I in the Supplementary material. The algorithm of the proposed Pin-GBTSVM is briefly described in Algorithm 1.

VI. Experiments

In this section, we assess the efficacy of the proposed models, i.e., Pin-GBTSVM and Pin-LGBTSVM, along with the baseline models using publicly available UCI [35] and KEEL [36] benchmark datasets using different noise levels. Furthermore, we conducted experiments on datasets generated by the NDC Data Generator [38] and the results are discussed in Section S.V.B in the Supplementary material. Moreover, we deploy the proposed models on the AD dataset, accessible through the Alzheimer's Disease Neuroimaging Initiative (ADNI), and the SCZ dataset available from the Center for Biomedical Research Excellence (COBRE). In the Supplementary material, we perform several sensitivity analyses on different aspects of the proposed models. This involves examining how the parameters num and pur of the granular ball affect the proposed models in Section S.IV.B. We conduct sensitivity analyses on the hyperparameters τ_1 and τ_2 in Section S.IV.A. In Section S.IV.C, we analyze the relationship between the number of generated granular balls and the performance of the proposed models across different purities. The experimental setup is discussed in subsection S.V.A of the Supplementary material.

Algorithm 1 Proposed Pin-GBTSVM model.

Input: Training dataset $\mathscr{D}$, and the threshold purity $\mathscr{T}$.
Output: Pin-GBTSVM classifier.

1: Set the training dataset $\mathscr{D}$ as a granular ball GB, i.e., $GB = \mathscr{D}$ and set of granular balls, G, to be empty set, i.e., $G = \{\ \}$.
2: $Dummy = \{GB\}$.
3: $for\ i = 1 : |Dummy|$
4: $if\ pur(GB_i) < \mathscr{T}$
5: Split GB_i into GB_{i1} and GB_{i2}, using 2-means clustering algorithm.
6: $Dummy \leftarrow GB_{i1}, GB_{i2}$.
7: $end\ if$.
8: $else\ pur(GB_i) \geq \mathscr{T}$
9: Calculate the center $c_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_j$ of GB_i, where $x_j \in GB_i,\ j = 1, 2, \ldots, n_i$, and n_i is the number of training sample in GB_i.
10: Calculate the radius $r_i = \frac{1}{n_i}\sum_{j=1}^{n_i} |x_j - c_i|$ and the label y_i of GB_i, here y_i is assigned the label of majority class samples within GB_i.
11: Put $GB_i = \{((c_i, r_i), y_i)\}$ in G.
12: $end\ else$.
13: $end\ for$.
14: $if\ Dummy \neq \{\ \}$
15: Go to step 3 (for further splitting).
16: $end\ if$.
17: Set $G = \{GB_i,\ \ i = 1, 2, \ldots, p\} = \{((c_i, r_i), y_i),\ \ i = 1, 2, \ldots, p\}$, where c_i signifies the center, r_i indicates the radius, y_i is the label of GB_i and p is the number of generated granular balls.
18: Solve (12) and (13) to obtain the value of $\alpha - \beta$, where α and β are the Lagrangian multipliers.
19: Using (11) and (14), find optimal values of w_1, b_1, w_2, and b_2.
20: Testing sample is classified into class $+1$ or -1 using (15).

A. Experiments on Real World UCI and KEEL Datasets

In this section, a comprehensive analysis is presented, entailing a comparison between the proposed Pin-GBTSVM, and Pin-LGBTSVM models along with the baseline SVM [11], GBSVM [34], TSVM [17], Pin-GTSVM [22], and Wave-TSVM [39] models. The comparison is conducted across 33 UCI and KEEL benchmark datasets, both with and without label noise, to evaluate their performances. PSO algorithm is used to solve the optimization problem of GBSVM [34]. Our investigation covers a wide spectrum of scenarios, incorporating both linear and nonlinear cases. These scenarios undergo detailed numerical experimentation, ensuring a thorough analysis. The experimental outcomes are presented in Table S.VII of the Supplementary material. Table I shows the average accuracy (ACC) of the proposed models along with the baseline models. The average ACC at 0% level of noise of the proposed Pin-GBTSVM and Pin-LGBTSVM models are 88.06% and 85.74%, and the average ACC of baseline SVM, GBSVM, TSVM, Pin-GTSVM, and Wave-TSVM are 81.91%, 72.8%, 65.38%, 87.01%, and 85.19%, respectively. The average ACC of the proposed Pin-GBTSVM model surpasses that of the baseline models.

Although the average ACC can be swayed by outstanding performance in a single dataset that compensates for shortcomings across multiple datasets, it could potentially introduce bias into the measure. Hence, we utilize the ranking method to assess the effectiveness and evaluate the performance of the models. Within this ranking scheme [36], each model's assessment is contingent on its performance across individual datasets, the model with the poorest performance receives a higher rank, while the model demonstrating the best performance is ranked lower. Suppose that there are t models under evaluation across N datasets, and the rank of the jth model on the ith dataset is represented by r_j^i. Then the average rank of the model is calculated as $\mathscr{R}_j = (1/N)\sum_{i=1}^{N} r_j^i$. The average rank of the proposed Pin-GBTSVM and Pin-LGBTSVM along with baseline SVM, GBSVM, TSVM, Pin-GTSVM, and Wave-TSVM at 0% noise level are 1.97, 2.83, 4.33, 5.41, 6.76, 3.14, and 3.56, respectively. The proposed Pin-GBTSVM model achieved a lower average rank, securing the first position among the competing models. Meanwhile, the proposed Pin-LGBTSVM attained the second position. A lower rank signifies better performance of the model in comparison to baseline models. Now, we proceed with conducting statistical tests to ascertain the significance of the obtained results. We execute the Friedman test [40] to assess whether there are statistically significant differences among the models. Under the null hypothesis, the presumption is that all the models demonstrate an equal average rank, implying equivalent performance. The Friedman test follows the chi-squared distribution χ_F^2 with $(t-1)$ degrees of freedom (d.o.f.) and is given by $\chi_F^2 = (12N/t(t+1))\left[\sum_j \mathscr{R}_j^2 - (t(t+1)^2/4)\right]$. The F_F statistic is given by, $F_F = ((N-1)\chi_F^2/(N(t-1)-\chi_F^2))$, where the distribution of F_F has $(t-1)$ and $(N-1)\times(t-1)$ d.o.f. For $t = 7$ and $N = 33$, we get $\chi_F^2 = 114.116$ and $F_F = 43.53$. At 5% level of significance from the statistical F-distribution table, $F_F(6, 192) = 2.1460$. The null hypothesis is rejected as $43.53 > 2.1460$. Thus, the Friedman test detects a significant difference among the compared models. Next, we utilize the Nemenyi post hoc test to investigate the pairwise differences between the models. The critical difference (C.D.) is calculated as C.D. $= q_\alpha\sqrt{(t(t+1)/6N)}$, here q_α represents the critical value for the two-tailed Nemenyi test from the distribution table. According to the statistical F-distribution table, q_α is determined to be 2.949 at a 5% significance level. The C.D. is calculated as 1.568. The models are considered significantly different if there is a gap equal to or greater than the C.D. in their average ranks. The average rank differences of the proposed Pin-GBTSVM and Pin-LGBTSVM models with baseline SVM, GBSVM, TSVM, Pin-GTSVM, and Wave-TSVM models are $(2.36, 1.50)$, $(3.44, 2.58)$, $(4.79, 3.93)$, $(1.17, 0.31)$, and $(1.59, 0.73)$, respectively. C.D. is exceeded by the observed differences. Thus, as per the Nemenyi post hoc test, the proposed Pin-GBTSVM model exhibits significant distinctions from SVM, GBSVM, TSVM, and Wave-TSVM except

TABLE I
AVERAGE ACC AND RANK OF THE PROPOSED MODELS AND BASELINE MODELS ON UCI AND KEEL DATASETS WITH LINEAR KERNEL

Dataset	Noise	SVM [11]	GBSVM [34]	TSVM [17]	Pin-GTSVM [22]	Wave-TSVM [39]	Pin-GBTSVM†	Pin-LGBTSVM†
Average ACC	0%	81.91	72.8	65.38	*87.01*	85.19	**88.06**	85.74
	5%	81.82	76.21	82.22	85.18	84.49	**86.85**	*85.79*
	10%	82.62	74.53	84.04	84.71	83.94	**86.59**	*85.55*
	15%	80.81	73.61	80.95	84.72	83.08	**87.15**	*85.01*
	20%	81.98	74.19	82.11	83.02	82.01	*84.06*	**84.35**
Average rank	0%	4.33	5.41	6.76	3.14	3.56	**1.97**	*2.83*
	5%	4.8	5.61	4.77	3.68	3.92	**2.27**	*2.94*
	10%	4.53	5.92	3.95	3.89	4.14	**2.53**	*3.03*
	15%	4.26	5.86	4.39	3.56	4.42	**2.23**	*3.27*
	20%	4.21	5.61	4.24	3.88	4.33	**2.79**	*2.94*

†represents the proposed models.
The boldface and italics indicate the best and second-best models, respectively.

for Pin-GTSVM. Also, the proposed Pin-LGBTSVM model exhibits significant distinctions with GBSVM and TSVM. Consequently, both proposed models exhibit superior performance compared to baseline models, except for Pin-LGBTSVM when compared with SVM, Pin-GTSVM, Wave-TSVM, and Pin-GBTSVM when compared with Pin-GTSVM. However, the proposed Pin-LGBTSVM model surpasses SVM, Pin-GTSVM, and Wave-TSVM in terms of average rank.

The win-tie-loss test for the proposed models is discussed in subsection S.V.D and the experiments on real-world UCI and KEEL datasets for nonlinear kernel is discussed in the subsection S.V.B in the Supplementary material.

B. Experiments on Real World UCI and KEEL Datasets With Added Label Noise

To verify the effectiveness and resilience to noise of the proposed Pin-GBTSVM and Pin-LGBTSVM models, we conduct experiments on the same UCI and KEEL benchmark datasets, as outlined in Section VI-A, introducing label noise for comparative analysis. We introduce label noise at varying levels, specifically 5%, 10%, 15%, and 20%, into each dataset. The comparative experimental outcomes of the proposed Pin-GBTSVM and Pin-LGBTSVM models along with the baseline models for both linear and nonlinear cases, are presented in Table S.VII and Table S.VIII of the Supplementary material. The average ACC and average rank of the models employing linear kernel are displayed in Table I, across varying percentages of noise labels. Throughout the evaluations, both the proposed Pin-GBTSVM and Pin-LGBTSVM models consistently demonstrated superior performance compared to the baseline models. The proposed Pin-GBTSVM exhibits a notably higher average ACC in comparison to the baseline models, showcasing an improvement of up to 2% when escalating the label noise from 5% to 20% for the linear kernel. Similarly, our proposed Pin-LGBTSVM has a better average ACC than the baseline models with linear kernel. The average ACC of the proposed Pin-LGBTSVM model at noise levels of 5%, 10%, 15%, and 20% are 85.79%, 85.55%, 85.01%, and 84.35%, respectively, which is higher than all baseline models. It indicates that both the Pin-GBTSVM and Pin-LGBTSVM exhibit enhanced robustness compared to the other models in the comparison. For the nonlinear case, the average ACC of the proposed Pin-GBTSVM and Pin-LGBTSVM models are higher than the baseline models. This heightened resilience is attributed to the granular ball's coarser granularity, which serves to alleviate the influence of label noise points contained within it. The assignment of a label to a granular ball is primarily dictated by the prevailing label encapsulated within it. The existence of label noise points associated with minority labels within the granular ball does not markedly impact the determination of the granular ball's assigned label. This approach notably strengthens the models' resilience against the contamination of label noise. The consistent superiority of the proposed Pin-GBTSVM and Pin-LGBTSVM models over the baseline models underscores their potential effectiveness in real-world scenarios, particularly those where noise is frequently encountered in datasets.

C. Evaluation on SCZ Dataset

SCZ is a debilitating mental disorder that ranks among the foremost contributors to disability on a global scale. However, a significant number of SCZ cases remain untreated due to factors such as failure to diagnose, self-denial, and social stigma. With the emergence of social media, individuals grappling with SCZ are utilizing these platforms to share their mental health challenges and seek support and treatment options. Furthermore, machine learning approaches are being increasingly employed from multimedia and T1-weighted MRI datasets for diagnoses of SCZ. As an application, we employed the proposed Pin-GBTSVM and Pin-LGBTSVM models for the diagnosis of SCZ. The data utilized in this study is acquired from the COBRE[1]. The dataset comprises 74 subjects in a healthy control group (35.8 ± 11.5 years old, range $18 - 65$ years) and 72 subjects diagnosed with SCZ (38.1 ± 13.9 years old, range $18 - 65$ years). The CAT12 package[2] is implemented in the Statistical Parametric Mapping (SPM) toolbox version 12[3] is employed for image processing. The 3-D T1-weighted MRI scans underwent parcellation into white matter (WM), gray matter (GM), and cerebrospinal fluid, as well as segmentation of

[1]http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html
[2]http://www.neuro.uni-jena.de/cat/
[3]https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED MODELS AGAINST BASELINE MODELS ON THE SCZ DATASET

	SVM [11] (ACC(%), Seny.) (Spey., Pren.)	GBSVM [34] (ACC(%), Seny.) (Spey., Pren.)	TSVM [17] (ACC(%), Seny.) (Spey., Pren.)	Pin-GTSVM [22] (ACC(%), Seny.) (Spey., Precision)	Wave-TSVM [39] (ACC(%), Seny.) (Spey., Pren.)	Pin-GBTSVM† (ACC(%), Seny.) (Spey., Pren.)	Pin-LGBTSVM† (ACC(%),Seny.) (Spey., Pren.)
Dataset	(62.56, 55.56)	(60.42, 57.41)	(65.87, 54.39)	(*67.78*, 52.90)	(66.58, 49.72)	(**69.90**, 70.18)	(67.76, 69.34)
	(65.52, 65.27)	(77.39, 58.33)	(78.24, 63.16)	(78.24, 63.16)	(75.71, 59.52)	(76.11, 71.07)	(84.87, 73.85)

†represents the proposed models.
The boldface and italics indicate the best and second-best models, respectively, in terms of ACC.
Seny., Spey., and Pren. denote the Sensitivity, Specificity, and Precision, respectively.

TABLE III
PERFORMANCE COMPARISON OF THE PROPOSED MODELS AGAINST BASELINE MODELS ON THE ADNI DATASETS

Dataset	SVM [11] (ACC(%), Seny.) (Spey., Pren.)	GBSVM [34] (ACC(%), Seny.) (Spey., Pren.)	TSVM [17] (ACC(%), Seny.) (Spey., Pren.)	Pin-GTSVM [22] (ACC(%), Seny.) (Spey., Precision)	Wave-TSVM [39] (ACC(%), Seny.) (Spey., Pren.)	Pin-GBTSVM† (ACC(%), Seny.) (Spey., Pren.)	Pin-LGBTSVM† (ACC(%), Seny.) (Spey., Pren.)
CN_versus_AD	(64.4, 29.51)	(70.34, 32.79)	(45.6, 24.59)	(68.4, 32.79)	(73.27, 19.67)	(*74*, 49.18)	(**78.4**, 39.34)
	(88.7, 58.06)	(89.57, 62.5)	(89.57, 55.56)	(85.22, 54.05)	(92.17, 57.14)	(80, 56.6)	(89.57, 66.67)
CN_versus_MCI	(58.51, 85.96)	(66.44, 80.7)	(64.36, 73.68)	(*66.94*, 84.21)	(65.89, 64.91)	(65.74, 73.68)	(**67.55**, 80.7)
	(89.71, 87.5)	(86.76, 83.64)	(80.88, 76.36)	(92.65, 90.57)	(91.18, 86.05)	(92.65, 89.36)	(89.71, 86.79)
MCI_versus_AD	(56.25, 86.78)	(*69.16*, 86.78)	(64.66, 86.78)	(65.91, 82.64)	(68.77, 86.78)	(67.05, 87.6)	(**70.45**, 87.6)
	(47.76, 75)	(49.25, 75.54)	(50.75, 76.09)	(55.22, 76.92)	(49.25, 75.54)	(38.81, 72.11)	(52.24, 76.81)
Average ACC	59.72	68.65	58.21	67.08	69.31	68.93	72.13

†represents the proposed models.
The boldface and italics indicate the best and second-best models, respectively, in terms of ACC.
Seny., Spey., and Pren. denote the Sensitivity, Specificity, and Precision, respectively.

skull, scalp, and air cavities. Using the high-dimensional diffeomorphic anatomic registration through the exponentiated Lie algebra algorithm (DARTEL), the GM images were normalized to the Montreal Neurological Institute (MNI) space. The smoothed GM images were generated using an 8-mm full-width-half-maximum Gaussian kernel. Given the relatively small size of the dataset, we adopted a $70:30$ ratio for dividing the data into training and testing sets. To demonstrate the application of the proposed Pin-GBTSVM and Pin-LGBTSVM models, we assessed their effectiveness in diagnosing SCZ. From Table II, the proposed Pin-GBTSVM and Pin-LGBTSVM models achieved ACC of 69.90%, and 67.76% and the baseline SVM, GBSVM, TSVM, Pin-GTSVM, and Wave-TSVM are 62.56%, 60.42%, 65.87%, 67.78%, and 66.58%, respectively. This suggests that our proposed models exhibit better generalization ability compared to the baseline models. The sensitivity and precision of the proposed Pin-GBTSVM and Pin-LGBTSVM models are $(70.18, 71.07)$ and $(69.34, 73.85)$, respectively. The sensitivity and precision of the proposed Pin-GBTSVM and Pin-LGBTSVM models achieved the top position. Therefore, the sensitivity and precision of the proposed models confirm the superiority of the proposed models. The specificity of the proposed Pin-GBTSVM and Pin-LGBTSVM models are 71.11 and 84.87, respectively. The specificity of Pin-LGBTSVM secured the first position and Pin-GBTSVM secured the fourth position. Thus, the specificity of the attests to the superiority of the proposed Pin-LGBTSVM model.

D. Evaluation on ADNI Datasets

AD is a progressive neurological disorder that detrimentally affects memory and cognitive functions. AD commonly begins with mild cognitive impairment (MCI). Notably, not all MCI patients progress to AD [41]. Presently, the exact causes of AD remain incompletely understood. However, precise identification and diagnosis of AD are crucial for providing effective patient treatment, especially in the early stages. We utilize scans from the ADNI dataset to train the proposed Pin-GBTSVM and Pin-LGBTSVM models, which can be accessed at adni.loni.usc.edu. The ADNI project, spearheaded by Michael W. Weiner in 2003, is focused on investigating a range of neuroimaging methods, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and other diagnostic assessments for AD, particularly during the MCI stage. The feature extraction pipeline utilized in this study follows the same approach as described in [42]. The dataset encompasses three scenarios: control normal (CN) versus AD, CN versus MCI, and MCI versus AD. The performance metrics for AD diagnosis, comparing the proposed Pin-GBTSVM and Pin-LGBTSVM models along with the baseline models, are detailed in Table III. The proposed Pin-LGBTSVM model attained the highest position with an average ACC of 72.13%, followed by the baseline Wave-TSVM model in the second position with an average ACC of 69.31%, and the proposed Pin-GBTSVM model in the third position with an average ACC of 68.93%. On the other hand, the baseline models, comprising SVM, GBSVM, TSVM, and Pin-GTSVM, exhibited lower average accuracies of 59.72%, 68.65%, 58.21%, and 67.08%, respectively. The proposed Pin-LGBTSVM model achieves the highest ACC of 78..40% for the CN versus AD case, followed by the proposed Pin-GBTSVM model with an ACC of 74.49%. For the CN versus MCI case, the proposed Pin-LGBTSVM model demonstrates superior performance, achieving an average ACC

of 67.55%. For the MCI versus AD case, the proposed Pin-LGBTSVM stands out as the most accurate classifier, achieving an ACC of 70.45%. Therefore, the proposed Pin-LGBTSVM consistently demonstrates superior performance by achieving high ACC across various cases, establishing its prominence among the models. The overall results emphasize the efficacy of the proposed models in differentiating between various cognitive states. Furthermore, to visualize the comparison between the proposed Pin-GBTSVM and Pin-LGBTSVM models and the baseline models in terms of specificity, sensitivity, and precision, the generated bar graph is depicted in Figure S.3 of the Supplementary material. The performance of the proposed Pin-GBTSVM and Pin-LGBTSVM models closely aligns with that of the baseline models.

VII. Conclusion

In this article, we proposed a novel granular ball TSVM with pinball loss (Pin-GBTSVM) for the classification problems. Pin-GBTSVM utilizes granular balls, rather than individual data points, as inputs for constructing a classifier. Simultaneously, it employs the pinball loss function to achieve an enhanced level of noise insensitivity. However, Pin-GBTSVM still grapples with challenges, arising from the lack of integration of the SRM principle and the need for matrix inversions. To tackle these challenges, we again proposed a novel large-scale Pin-GBTSVM (Pin-LGBTSVM). Pin-LGBTSVM lies in its ability to circumvent matrix inversions, rendering it well suited for large-scale problems. Additionally, it effectively addresses concerns related to overfitting. To showcase the effectiveness, robustness, scalability, and efficiency of our proposed Pin-GBTSVM and Pin-LGBTSVM models, we conducted an extensive series of experiments in KEEL and UCI datasets with and without label noise and performed rigorous statistical analyses. The proposed Pin-GBTSVM and Pin-LGBTSVM models showcase exceptional generalization performance on UCI and KEEL datasets. To demonstrate the scalability of large-scale datasets, we conducted experiments on the NDC dataset. The results indicate that the proposed models are hundred to thousand times faster than the compared models. Moreover, machine learning techniques are increasingly applied to diagnose SCZ and AD using multimedia and T1-weighted MRI datasets. As an application, we utilized Pin-GBTSVM and Pin-LGBTSVM models, which demonstrated superior performance in diagnosing SCZ and AD. While our proposed models shown exceptional performance in binary classification problems, their evaluation in multiclass problems has not been conducted. A critical avenue for future research would involve adapting these models to be suitable for multiclass problems. Also one can incorporate the robust loss function, L_1 norm of regularization term, and intuitionistic fuzzy membership scheme in the proposed models to develop a more robust model.

Data Availability Statement

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. The source code of the proposed Pin-GBTSVM and Pin-LGBTSVM models is available at https://github.com/mtanveer1/Pin-GBTSVM.

Acknowledgment

The dataset utilized in this study was acquired with financial assistance from the ADNI, supported by the Department of Defense's ADNI award W81XWH-12-2-0012 and the National Institutes of Health's U01 AG024904 grant. ADNI receives funding from the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, as well as generous contributions from the following sources: Araclon Biotech; AbbVie, Alzheimer's Association; Cogstate; Eisai Inc.; BioClinica, Inc.; Alzheimer's Drug Discovery Foundation; EuroImmun; Biogen; Bristol–Myers Squibb Company; CereSpir, Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann–La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Takeda Pharmaceutical Company; Piramal Imaging; Pfizer Inc.; Servier; and Transition Therapeutics. Funding to support ADNI clinical sites in Canada is provided by the Canadian Institutes of Health Research. Private sector contributions are administered by the Foundation for the National Institutes of Health (www.fnih.org). The Northern California Institute for Research and Education acts as the grantee organization, with the study coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data dissemination is overseen by the Laboratory for Neuroimaging at the University of Southern California.

References

[1] M. Tanveer et al., "Ensemble deep learning for Alzheimer's disease characterization and estimation," *Nature Mental Health*, pp. 1–13, May 2024, doi: 10.1038/s44220-024-00237-x.

[2] S. Verma, T. Goel, M. Tanveer, W. Ding, R. Sharma, and R. Murugan, "Machine learning techniques for the Schizophrenia diagnosis: A comprehensive review and future research directions," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 5, pp. 4795–4807, 2023.

[3] J. Zeisel, K. Bennett, and R. Fleming, "World Alzheimer Report 2020: Design, dignity, dementia: Dementia-related design and the built environment," *Alzheimer's Dementia*, vol. 1, 2020, Art. no. 248.

[4] T. Wang, R. G. Qiu, and M. Yu, "Predictive modeling of the progression of Alzheimer's disease with recurrent neural networks," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 9161.

[5] C. M. Corcoran et al., "Prediction of psychosis across protocols and risk cohorts using automated language analysis," *World Psychiatry*, vol. 17, no. 1, pp. 67–75, 2018.

[6] G. J. Johnson and P. J. Ambrose, "Neo-tribes: The power and potential of online communities in health care," *Commun. ACM*, vol. 49, no. 1, pp. 107–113, 2006.

[7] A. A. Tahami Monfared, Y. Stern, S. Doogan, M. Irizarry, and Q. Zhang, "Stakeholder insights in Alzheimer's disease: Natural language processing of social media conversations," *J. Alzheimer's Disease*, vol. 89, no. 2, pp. 695–708, 2022.

[8] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: Observational study," *J. Med. Internet Res.*, vol. 22, no. 10, 2020, Art. no. e22635.

[9] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: A critical review," *NPJ Digit. Med.*, vol. 3, no. 1, 2020, Art. no. 43.

[10] R. Thorstad and P. Wolff, "Predicting future mental illness from social media: A big-data approach," *Behav. Res. Methods*, vol. 51, pp. 1586–1600, Apr. 2019.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995.

[12] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 1007–1011, 2005.

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

[14] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. Eur. Conf. Mach. Learn.,* Berlin, Heidelberg, Germany: Springer, 1998, pp. 137–142.

[15] Y. Li and C. Guan, "Joint feature re-extraction and classification using an iterative semi-supervised support vector machine algorithm," *Mach. Learn.*, vol. 71, pp. 33–53, Nov. 2008.

[16] 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.

[17] 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.

[18] 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.

[19] M. Tanveer, M. 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.

[20] K. Hu, Y. Xiao, W. Zheng, W. Zhu, and C.-H. Hsu, "Multiview large margin distribution machine," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2023.3349142.

[21] X. Huang, L. Shi, and J. A. Suykens, "Support vector machine classifier with pinball loss," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 984–997, May 2014.

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

[23] H. Guo, L. Wang, X. Liu, and W. Pedrycz, "Trend-based granular representation of time series and its application in clustering," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9101–9110, Sep. 2022.

[24] D. Cheng, S. Liu, S. Xia, and G. Wang, "Granular-ball computing-based manifold clustering algorithms for ultra-scalable data," *Expert Syst. Appl.*, vol. 247, 2024, Art. no. 123313.

[25] Q. Xie et al., "GBG++: A fast and stable granular ball generation method for classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 2, pp. 2022–2036, Apr. 2024.

[26] M. Sewwandi, Y. Li, and J. Zhang, "Automated granule discovery in continuous data for feature selection," *Inf. Sci.*, vol. 578, pp. 323–343, Nov. 2021.

[27] S. Xia, Y. Liu, X. Ding, G. Wang, H. Yu, and Y. Luo, "Granular ball computing classifiers for efficient, scalable and robust learning," *Inf. Sci.*, vol. 483, pp. 136–152, May 2019.

[28] W. Wenjian, G. Husheng, J. Yuanfeng, and B. Jingye, "Granular support vector machine based on mixed measure," *Neurocomputing*, vol. 101, pp. 116–128, Feb. 2013.

[29] S. Xia, S. Zheng, G. Wang, X. Gao, and B. Wang, "Granular ball sampling for noisy label classification or imbalanced classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 2144–2155, Apr. 2023.

[30] G. Zhang, Q. Liu, and L. JI, "Multi-granularity sentiment classification method based on sequential three-way decisions," *Comput. Sci.*, vol. 45, no. 12, pp. 153–159, 2018.

[31] D.-B. Bu, S. Bai, and G.-J. Li, "Principle of granularity in clustering and classification," *Chin. J. Comput.*, vol. 25, no. 8, pp. 810–816, 2002.

[32] D. Miao and S. Hu, "Uncertainty analysis based on granular computing," *J. Northwest Univ*, vol. 4, pp. 44–47, 2019.

[33] J. Liang, J. Wang, G. Yu, S. Xia, and G. Wang, "Multi-granularity causal structure learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 12, 2024, pp. 13727–13735.

[34] S. Xia, G. Wang, X. Gao, and X. Peng, "GBSVM: Granular-ball support vector machine," 2014, *arXiv:2210.03120*.

[35] D. Dua and C. Graff, "Automobile." UCI machine learning repository. [Online]. Available: http://Archive.ics.uci.edu/ml

[36] 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 Log. Soft Comput.*, vol. 17, pp. 255–287, Jan. 2015.

[37] Y. Zhou, H. Yu, and X. Cai, "A novel k-means algorithm for clustering and outlier detection," in *Proc. 2nd Int. Conf. Future Inf. Technol. Manage. Eng.,* Piscataway, NJ, USA: IEEE Press, 2009, pp. 476–480.

[38] D. R. Musicant, "NDC: Normally distributed clustered datasets." University of Wisconsin, Madison. [Online]. Available: www.cs.wisc.edu/dmi/svm/ndc/

[39] M. Akhtar, M. Tanveer, M. Arshad, and Alzheimer's Disease Neuroimaging Initiative, "Advancing supervised learning with the wave loss function: A robust and smooth approach," *Pattern Recognit.*, 2024, Art. no. 110637, doi: 10.1016/j.patcog.2024.110637.

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

[41] C. Davatzikos, S. M. Resnick, X. Wu, P. Parmpi, and C. M. Clark, "Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI," *Neuroimage*, vol. 41, no. 4, pp. 1220–1227, 2008.

[42] B. Richhariya, M. Tanveer, and Alzheimer's Disease Neuroimaging, "An efficient angle-based universum least squares twin support vector machine for classification," *ACM Trans. Internet Technol.*, vol. 21, no. 3, pp. 1–24, 2021.