Robust High-Dimensional Classification via Granular Balls-based $L_{0/1}$-SVM

1st Guangming Lang[a,b], 2nd Xinyu Du[a,b], 3rd Jie Zhou[a,b,*], 4th Qimei Xiao[a,b]

[a]School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha, Hunan 410114, P.R. China

[b]Hunan Provincial Key Laboratory of Mathematical Modeling and Analysis in Engineering, Changsha University of Science and Technology, Changsha, Hunan 410114, P.R. China

[*]Corresponding author E-mail address: jie_jpu@163.com

***Abstract*—High-dimensional data frequently exhibit feature redundancy, noise sensitivity, and complex underlying distributions, all of which severely impair classification performance. To address these challenges, we propose a novel classification framework termed granular ball $L_{0/1}$ support vector machine (GB-$L_{0/1}$SVM), which synergistically integrates principal component analysis (PCA), granular ball (GB) computing, and $L_{0/1}$ soft-margin SVM ($L_{0/1}$-SVM). In this paper, we apply PCA to reduce dimensionality and eliminate redundant features, and generate GBs in the transformed feature space to form compact and homogeneous information granules, thereby enhancing structural interpretability and robustness to noise. After that, GBs are fed into an $L_{0/1}$-SVM classifier to enable sparse and discriminative decision boundaries. Finally, we present an algorithm for the GB-$L_{0/1}$SVM classifier and conduct extensive experiments on six benchmark datasets. Experimental results demonstrate that GB-$L_{0/1}$SVM effectively bridges the gap between robust non-convex optimization and large-scale data mining, offering a superior trade-off between sparsity and predictive accuracy.**

***Index Terms*—Dimensionality Reduction, Granular Ball, $L_{0/1}$-SVM, Nonconvex Optimization.**

I. Introduction

Support Vector Machines (SVMs) [1] have long served as a cornerstone of supervised learning. They have been widely applied in bioinformatics, healthcare, image recognition, and other domains [2], [3]. In the Big Data era, real-world datasets increasingly exhibit the "3V" characteristics: high volume, high velocity, and high veracity [4], [5]. Conventional SVMs solve a quadratic programming problem with $O(n^3)$ time complexity [6], rendering them computationally prohibitive for large-scale data. Moreover, standard SVMs typically rely on unbounded loss functions, such as the squared hinge loss, which are inherently sensitive to outliers.

To alleviate computational bottlenecks, several optimization strategies have been proposed, including the sequential minimal optimization algorithm [7] and stochastic gradient descent [8]. Robustness has also been pursued through non-convex loss functions, including truncated Huber loss [9] and flexible pinball loss [10]. More recently, the $L_{0/1}$ soft-margin loss has attracted attention for its explicit control over solution sparsity and label noise tolerance. However, minimizing this loss function remains challenging: it is non-convex, discontinuous, and NP-hard, meaning that no known algorithm can find the global optimum in polynomial time. While solvers based on the ADMM, in which an iterative framework that breaks complex problems into smaller sub-problems [11], [12], improve practical efficiency, they remain infeasible for massive-scale datasets. Conventional iterative convex optimization solvers [13] are computationally expensive, as they involve a vast number of variables in the entire instance space, severely limiting scalability.

Granular Computing (GrC) provides a principled alternative paradigm for efficient learning by structuring data into information granules at multiple levels of abstraction [14]. Among existing GrC models, the granular ball (GB) framework [15] has attracted growing interest. Granular balls (GBs) adaptively cover the data space using hyperspherical granules: large GBs capture homogeneous regions, while small GBs preserve intricate decision boundaries. This representation achieves effective data compression and inherently suppresses local noise via purity constraints. Recent studies have further enhanced GB modeling and its integration with downstream learning tasks. In particular, improvements in GB generation strategies have strengthened stability and coverage quality under complex data distributions [16], [17]. Meanwhile, combining GBs with classifiers such as SVM and TWSVM has demonstrated notable performance gains in noisy label and imbalanced learning scenarios [18], [19]. Yet a critical limitation persists for high-dimensional data. GB generation depends heavily on distance metrics, which suffer from computational cost and theoretical instability under the curse of dimensionality [20].

To address the trade-off between computational efficiency and statistical robustness, this paper proposes granular ball $L_{0/1}$ support vector machine (GB-$L_{0/1}$SVM), a unified classification framework that integrates granular ball computing with sparse optimization. The methodology follows a rigorous coarse-to-fine paradigm: first, PCA reduces feature dimensionality to suppress feature noise [21]; second, GBs are constructed in the reduced dimensional space to achieve adaptive, topology aware coverage of the underlying data distribution; finally, classification is formulated as an $L_{0/1}$ loss function optimization problem over the GB representation.

Beyond computational efficiency and robustness, the proposed GB-$L_{0/1}$SVM shifts the learning paradigm from point-wise optimization to structural abstraction. By representing data as multi-scale granular balls, the model inherently filters

feature-level and instance-level noise before the optimization begins. The main contributions of this work are summarized as follows:

- A unified classification framework is proposed that sequentially integrates PCA-based dimensionality reduction, GB computing, and $L_{0/1}$ loss optimization. The framework effectively bridges feature compression and classifier construction, thereby mitigating the challenges arising from high dimensionality and large-scale datasets.
- Speedup of multiple orders of magnitude: Extensive experiments on large-scale benchmarks show that our method reduces training time from hours to sub-second levels, achieving speedups of three orders of magnitude.
- The proposed approach ensures robustness across three complementary levels. Specifically, PCA suppresses feature-level noise; GBs mitigate local instance noise via coarse-graining; and the $L_{0/1}$ loss explicitly bounds the influence of label corruption.

II. Preliminaries

This section outlines the theoretical foundations of the proposed framework: GB representation [15] and $L_{0/1}$ soft-margin SVM [12].

A. Granular Ball Representation

Granular ball computing provides a region level abstraction of local neighborhoods in the feature space. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top \in \mathbb{R}^{n\times m}$ denote the input data matrix, where each sample $\mathbf{x}_i$ is associated with a label $y_i \in \{-1, +1\}$. The set of GBs generated from the input dataset is denoted as $\mathcal{B} = \{\mathrm{GB}_1, \mathrm{GB}_2, \ldots, \mathrm{GB}_k\}$. For the j-th granular ball GB_j, let $\mathcal{S}_j \subseteq \{1, \ldots, n\}$ denote the index set of samples contained in the GB, where $|\mathcal{S}_j|$ indicates the number of objects. The granular ball GB_j is characterized by four quantities: its center $\mathbf{c}_j$, radius r_j, majority-vote label y_j, and label purity p_j:

$$\mathbf{c}_j = \frac{1}{|\mathcal{S}_j|}\sum_{i\in\mathcal{S}_j}\mathbf{x}_i, \qquad r_j = \frac{1}{|\mathcal{S}_j|}\sum_{i\in\mathcal{S}_j}\|\mathbf{x}_i - \mathbf{c}_j\|, \tag{1}$$

$$y_j = \arg\max_{l\in\{-1,+1\}}\sum_{i\in\mathcal{S}_j}\mathbb{I}(y_i = l), \quad p_j = \frac{1}{|\mathcal{S}_j|}\sum_{i\in\mathcal{S}_j}\mathbb{I}(y_i = y_j), \tag{2}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

B. $L_{0/1}$ Soft-Margin SVM

The $L_{0/1}$ Soft-Margin SVM is a robust SVM variant that explicitly controls misclassification by penalizing margin violations directly. Given a training sample $(\mathbf{x}_i, y_i)$ with $y_i \in \{-1, +1\}$, the margin violation is defined as $u_i = 1 - y_i(\mathbf{w}^\top\mathbf{x}_i + b)$. The $L_{0/1}$-SVM solves the following regularized empirical risk minimization problem:

$$\min_{\mathbf{w},b,\mathbf{u}} \frac{1}{2}\|\mathbf{w}\|^2 + C\|\mathbf{u}_+\|_0, \tag{3}$$

$$\text{s.t. } u_i = 1 - y_i(\mathbf{w}^\top\mathbf{x}_i + b), \quad i = 1, \ldots, n, \tag{4}$$

where $\|\mathbf{u}_+\|_0$ counts the number of samples with positive margin violations; $\mathbf{u}_+$ denotes the element-wise positive part, i.e., $(\mathbf{u}_+)_i = \max(0, u_i)$; the term $\frac{1}{2}\|\mathbf{w}\|^2$ implements structural risk minimization via geometric margin maximization, and the hyperparameter C balances model simplicity against the cardinality of the misclassified set.

III. Proposed Method

Although the $L_{0/1}$ loss provides the most direct misclassification penalty, its nonconvexity and discontinuity render instance optimization theoretically and practically intractable. Moreover, independent sample treatment introduces redundant constraints and local inconsistencies, while overlooking inherent homogeneity in local neighborhoods. To this end, we propose a unified framework that integrates PCA-based denoising, GB computing, and $L_{0/1}$-margin learning over compact GB representatives, thereby jointly suppressing label noise and drastically reducing the optimization scale.

A. Formulations

The proposed GB-$L_{0/1}$SVM learns a classifier from a compact set of GB representatives $\{(\mathbf{c}_i, r_i, y_i)\}_{i=1}^s$, where each GB summarizes a local region of the feature space via its centroid $\mathbf{c}_i$, radius r_i, and majority class label y_i. Under this representation, the classifier is obtained by solving:

$$\min_{\mathbf{w},b,\mathbf{u}} \frac{1}{2}\|\mathbf{w}\|^2 + C\|\mathbf{u}_+\|_0, \tag{5}$$

$$\text{s.t. } u_i = 1 - y_i(\mathbf{w}^\top\mathbf{c}_i + b) + \|\mathbf{w}\|r_i, \quad i = 1, \ldots, s. \tag{6}$$

In this formulation, $\mathbf{w}$ and b define the decision boundary; $\mathbf{u} \in \mathbb{R}^s$ collects the margin violation variables; and $C > 0$ trades off geometric margin maximization against misclassification tolerance.

The formulation operates natively at the GB level, reducing the constraint count from n to s ($s \ll n$). Each constraint incorporates a radius dependent correction term $\|\mathbf{w}\|r_i$, which bounds the maximal margin violation over all points in the i-th GB. As a result, the margin condition implicitly guarantees separation for every instance within the GB.

The overall framework of the proposed GB-$L_{0/1}$SVM is illustrated in Fig. 1. To generate the inputs for the optimization in Eqs. (5)–(6), GB representatives are constructed via a coarse-to-fine pipeline. First, PCA is applied to the input data matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$ to mitigate high-dimensional noise, yielding an orthogonal projection matrix $\mathbf{P} \in \mathbb{R}^{m\times d}$ and reduced features $\mathbf{Z} = \mathbf{XP}$. The dimension d is selected to retain a cumulative explained variance of 90%. Subsequently, using $\mathbf{Z}$, a collection of GBs $\mathcal{B} = \{\mathrm{GB}_1, \ldots, \mathrm{GB}_s\}$ is generated, each assigned a purity measure p_i. GBs with $p_i < 0.9$ undergo recursive bisecting k-means partitioning until all resulting GBs satisfy $p_i \geq 0.9$. By combining PCA-based manifold stabilization with this splitting, the resulting GB representation maintains structural consistency across varying data perturbations, ensuring that the centroids $\mathbf{c}_i$ reliably capture local class-homogeneity. This granular abstraction replaces thousands of individual samples with s GBs, providing a more transparent

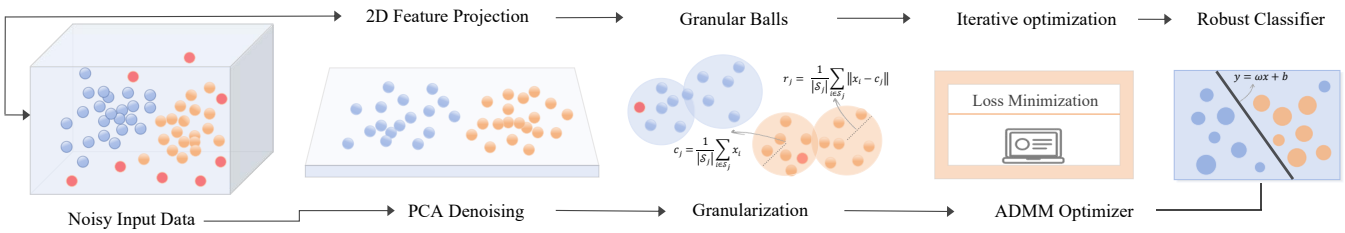

Fig. 1. The overall framework of the proposed GB-$L_{0/1}$SVM.

decision-making process than traditional point-wise models. Compared to existing granular ball computing frameworks, our approach further enhances model interpretability by identifying a significantly smaller number of support GBs through the sparsity-inducing $L_{0/1}$ penalty. This structural transparency, combined with the statistical stability afforded by PCA-denoising, ensures the trustworthiness of the classification results, as the decision boundary is anchored by robust GBs rather than being swayed by isolated outliers.

The constraint in Eq. (6) is theoretically grounded in enforcing the robust margin condition $y_i(\mathbf{w}^\top \mathbf{z} + b) \geq 1$ for all $\mathbf{z} \in \mathrm{GB}_i$. Leveraging the Lipschitz continuity of the linear functional $f(\mathbf{z}) = y_i(\mathbf{w}^\top \mathbf{z} + b)$, with Lipschitz constant $\|\mathbf{w}\|$, allows the infinite constraint set over GB_i to be equivalently replaced by the uniform bound

$$y_i(\mathbf{w}^\top \mathbf{c}_i + b) - \|\mathbf{w}\| r_i \geq 1. \tag{7}$$

Rearranging yields the margin violation variable $u_i = 1 - y_i(\mathbf{w}^\top \mathbf{c}_i + b) + \|\mathbf{w}\| r_i$.

The radius r_i is defined as the average distance rather than the maximum distance. This choice is deliberate: while the maximum distance is highly sensitive to boundary outliers, the average distance provides a robust statistical measure of the granule's scale, ensuring that the GB representation remains stable even in the presence of localized noise.

B. Optimization

To solve the nonconvex and nonsmooth problem in Eqs. (5)–(6), we develop an ADMM-based iterative scheme. It seeks a discriminative hyperplane leveraging the manifold structure extracted from PCA while simultaneously minimizing constraints via a compact set of granular balls. In practice, the solver efficiently identifies and focuses on the most critical "Support GBs" near the decision boundary, while ignoring "easy" granules that do not contribute to the margin definition. The detailed derivation is as follows.

Define the constraint residual:

$$s_i(\mathbf{w}, b) = 1 - y_i(\mathbf{w}^\top \mathbf{c}_i + b) + \|\mathbf{w}\| r_i, \tag{8}$$

and introduce auxiliary variables $u_i = s_i(\mathbf{w}, b)$. The augmented Lagrangian is given by:

$$\mathcal{L}_\rho(\mathbf{w}, b, \mathbf{u}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\|\mathbf{u}_+\|_0 + \sum_{i=1}^{s} \lambda_i (u_i - s_i(\mathbf{w}, b)) + \frac{\rho}{2} \sum_{i=1}^{s} (u_i - s_i(\mathbf{w}, b))^2. \tag{9}$$

At iteration t, the variables $(\mathbf{w}, b, \mathbf{u})$ are updated alternately.

1) Update of $\mathbf{u}$*:* With $(\mathbf{w}^{(t)}, b^{(t)}, \boldsymbol{\lambda}^{(t)})$ fixed, define:

$$q_i^{(t)} = s_i(\mathbf{w}^{(t)}, b^{(t)}) - \lambda_i^{(t)}/\rho. \tag{10}$$

The update of u_i admits the closed-form solution:

$$u_i^{(t+1)} = \begin{cases} 0, & 0 \leq q_i^{(t)} \leq \sqrt{2C/\rho}, \\ q_i^{(t)}, & \text{otherwise.} \end{cases} \tag{11}$$

2) Update of $\mathbf{w}$*:* The non-smooth term $\|\mathbf{w}\|$ is linearized at $\mathbf{w}^{(t)}$ using its subgradient:

$$\mathbf{g}^{(t)} = \frac{\mathbf{w}^{(t)}}{\|\mathbf{w}^{(t)}\|}, \quad \|\mathbf{w}^{(t)}\| > 0. \tag{12}$$

This yields the linearized constraint:

$$\tilde{s}_i(\mathbf{w}, b^{(t)}) = 1 - y_i b^{(t)} - \mathbf{w}^\top \tilde{\mathbf{x}}_i^{(t)}, \tag{13}$$

where

$$\tilde{\mathbf{x}}_i^{(t)} = y_i \mathbf{c}_i - r_i \mathbf{g}^{(t)}. \tag{14}$$

The resulting subproblem is convex quadratic and leads to the linear system:

$$\left(\mathbf{I} + \rho \sum_{i=1}^{s} \tilde{\mathbf{x}}_i^{(t)} (\tilde{\mathbf{x}}_i^{(t)})^\top \right) \mathbf{w}^{(t+1)} = \rho \sum_{i=1}^{s} v_i^{(t)} \tilde{\mathbf{x}}_i^{(t)}, \tag{15}$$

with $v_i^{(t)} = 1 - y_i b^{(t)} - u_i^{(t+1)} - \lambda_i^{(t)}/\rho$.

3) Update of b*:* The bias term is updated by setting $\partial \mathcal{L}_\rho / \partial b = 0$:

$$b^{(t+1)} = b^{(t)} + \frac{1}{\rho \sum_{i=1}^{s} y_i^2} \sum_{i=1}^{s} y_i \big(\rho (s_i - u_i^{(t+1)}) - \lambda_i^{(t)} \big). \tag{16}$$

4) Update of $\boldsymbol{\lambda}$*:* The dual variables are updated as:

$$\lambda_i^{(t+1)} = \lambda_i^{(t)} + \rho \big(u_i^{(t+1)} - s_i(\mathbf{w}^{(t+1)}, b^{(t+1)}) \big). \tag{17}$$

Algorithm 1 GB-$L_{0/1}$SVM classifier

Input: Data matrix $\mathbf{X}$, labels $\mathbf{y}$, PCA ratio η, purity T, parameters C, ρ, tol ϵ.
Output: Classifier parameters $\mathbf{w}, b$.
Compute PCA projection $\mathbf{P}$ retaining η fraction of variance; $\mathbf{Z} \leftarrow \mathbf{XP}$;
Let $d = \dim(\mathbf{Z})$;
Initialize $\mathcal{B} \leftarrow$ single GB containing all samples in $\mathbf{Z}$;
while $\exists \mathrm{GB} \in \mathcal{B}$ has purity $< T$ **do**
    Split it via bisecting K-means; update $\mathcal{B}$;
end while
Form granular ball set $\{(\mathbf{c}_i, r_i, y_i)\}_{i=1}^{s}$;
Initialize: $\mathbf{w}^{(0)} \in \mathbb{R}^d$, $b^{(0)} = 0, \mathbf{u}^{(0)} = \mathbf{0}, \boldsymbol{\lambda}^{(0)} = \mathbf{0}$;
for $t = 0, 1, 2, \ldots$ **do**
    Compute direction $\mathbf{g}^{(t)}$ and effective data $\tilde{\mathbf{x}}_i^{(t)}$;
    Update $\mathbf{u}^{(t+1)}$ via closed-form solution Eq. 11;
    Update $\mathbf{w}^{(t+1)}$ by solving the linear system Eq. 15;
    Update $b^{(t+1)}$ via Eq. 16;
    Update $\boldsymbol{\lambda}^{(t+1)}$ via Eq. 17;
    Calculate residuals $\delta_s^{(t+1)}, \delta_d^{(t+1)}$ via Eqs. (18)–(19);
    if $\delta_s^{(t+1)} \le \epsilon\sqrt{s}$ & $\delta_d^{(t+1)} \le \epsilon\sqrt{d}$ **then break**
    end if
end for
return $\mathbf{w}, b$.

5) Stopping Criteria: At iteration $t+1$, the primal residual $\delta_s^{(t+1)}$ quantifies violation of the primal equality constraints:

$$\delta_s^{(t+1)} = \left\| \mathbf{u}^{(t+1)} - \mathbf{s}(\mathbf{w}^{(t+1)}, b^{(t+1)}) \right\|_2. \tag{18}$$

The dual residual $\delta_d^{(t+1)}$ reflects stability of the auxiliary variable update:

$$\delta_d^{(t+1)} = \rho \left\| \mathbf{u}^{(t+1)} - \mathbf{u}^{(t)} \right\|_2. \tag{19}$$

Termination occurs when

$$\delta_s^{(t+1)} \le \epsilon\sqrt{s} \quad \text{and} \quad \delta_d^{(t+1)} \le \epsilon\sqrt{d}, \tag{20}$$

where $\epsilon = 10^{-3}$ is the absolute tolerance, s is the number of GBs, and d is the feature dimension.

The overall procedure is summarized in Algorithm 1.

C. Computational Complexity

Complexity analysis proceeds in three stages, with n denoting the number of samples, m the original feature dimension, d the reduced dimension, and s the number of GBs:

- Preprocessing: Covariance computation and eigendecomposition dominate, costing $\mathcal{O}(nm^2 + m^3)$.
- Granular ball construction: Recursive bisecting k-means costs $\mathcal{O}(nd \log s)$.
- ADMM optimization: Per iteration, updating $\mathbf{w}$ is performed via coefficient matrix assembly and linear solve, which costs $\mathcal{O}(sd^2 + d^3)$ as $\mathbf{g}^{(t)}$ varies across iterations. Updates for $\mathbf{u}$, b, and $\boldsymbol{\lambda}$ cost $\mathcal{O}(s)$. Over T iterations, total cost is $\mathcal{O}(T(sd^2 + d^3))$.

The overall complexity is:

$$\mathcal{O}(nm^2 + nd \log s + T(sd^2 + d^3)). \tag{21}$$

Because $s \ll n$ and $d \ll m$, the method achieves substantial computational savings. This decouples the costly iterative phase from sample size, ensuring scalability to large-scale data.

IV. Experiments

In this section, we present a comprehensive evaluation of the proposed method against representative baseline models: Ramp-SVM [22], FSVM [23], GBSVM [18], GBKNN [24], PCA-SVM and $L_{0/1}$-SVM [12]. Specifically, PCA-SVM refers to the standard SVM classifier [1] applied after PCA dimensionality reduction. All algorithms are implemented in Python 3.11 and executed on a desktop computer equipped with an Intel® Core™ i7-10700 CPU operating at 2.90 GHz and 16 GB of RAM.

A. Experimental Setup

To rigorously evaluate the effectiveness of GB-$L_{0/1}$SVM, comprehensive experiments were conducted on six benchmark datasets selected from the UCI repository [25] and the LIBSVM database [26]. These datasets cover sample sizes from 748 to over 65,000 and feature dimensions from 3 to 300, as summarized in Table I. All datasets were normalized to the $[0, 1]$ range, and a stratified five-fold cross-validation protocol was employed to ensure statistical robustness. Model performance was assessed using five standard metrics: accuracy (Acc), precision (Prec), recall (Rec), F1-score (F1), and training time (Time).

TABLE I
Statistical Characteristics of Benchmark Datasets

Dataset	Samples	Features
Transfusion (Transfusion)	748	4
Glioma grading clinical (Glioma)	839	23
RICE (RICE)	3,810	7
Htru2 (Htru2)	17,898	8
w6a (w6a)	49,749	300
Skin segmentation (Skin)	65,536	3

The comparative analysis includes an ablation study and a robustness evaluation against six baseline methods: Ramp-SVM, FSVM, GBSVM, GBKNN, PCA-SVM and $L_{0/1}$-SVM. These baselines are categorized into three groups: (i) robust non-convex SVM variants (Ramp-SVM, FSVM, and $L_{0/1}$-SVM) to assess noise-tolerance; (ii) GB computing paradigms (GBSVM and GBKNN) to evaluate the efficiency and representation of GB abstraction; and (iii) dimensionality reduction models (PCA-SVM) to isolate the contribution of feature preprocessing.

For model implementation, all hyperparameters were tuned via grid search on the training sets. The search ranges were determined based on a preliminary parameter sensitivity analysis to ensure that the chosen windows for C, ρ, the PCA retention ratio η, and the purity threshold T cover the most

TABLE II
THE OPTIMAL PARAMETERS OF SIX MODELS ON DATASETS WITHOUT NOISE.

Dataset	Ramp-SVM C, s	FSVM C	GBSVM C	PCA-SVM C	$L_{0/1}$SVM C, ρ	GB-$L_{0/1}$SVM C, ρ
Transfusion	$2^0, 2^1$	2^1	2^0	2^0	$2^{-3}, 2^{-2}$	$2^{-3}, 2^{-3}$
Glioma	$2^1, 2^3$	2^1	2^{-5}	2^{-3}	$2^{-4}, 2^{-2}$	$2^{-1}, 2^{-2}$
RICE	$2^0, 2^3$	2^{-1}	2^{-1}	2^{-3}	$2^{-3}, 2^{-2}$	$2^{-3}, 2^0$
Htru2	$2^{-2}, 2^0$	2^0	2^{-4}	2^{-1}	$2^4, 2^1$	$2^{-1}, 2^{-1}$
w6a	$2^{-1}, 2^0$	2^{-2}	2^1	2^{-3}	$2^1, 2^{-1}$	$2^{-5}, 2^{-3}$
Skin	$2^5, 2^3$	2^1	2^{-3}	2^{-2}	$2^3, 2^{-2}$	$2^{-1}, 2^{-2}$

TABLE III
THE RUNNING TIME (SECONDS) OF SEVEN MODELS ON DATASETS WITHOUT NOISE.

Dataset	Ramp-SVM	FSVM	GBSVM	GBKNN	PCA-SVM	$L_{0/1}$SVM	GB-$L_{0/1}$SVM
Transfusion	0.012	0.013	0.086	12.074	3.643	0.138	**0.001**
Glioma	0.012	0.041	0.237	37.592	3.725	0.127	**0.001**
RICE	0.858	0.044	158.100	273.739	1.224	0.093	**0.020**
Htru2	10.757	0.225	6392.138	1056.386	39.168	0.765	**0.170**
w6a	25.077	41.579	64132.058	3922.232	41340.337	16.389	**1.362**
Skin	10.549	12.478	464.714	156.402	4044.621	0.427	**0.001**
Ave Time	7.878	9.063	11857.889	909.738	7572.120	2.990	**0.259**
Ave Rank	3.50	3.00	6.17	6.83	5.00	2.00	**1.00**

TABLE IV
THE ACCURACY, PRECISION, AND F1-SCORE OF SEVEN MODELS ON DATASETS WITHOUT NOISE.

Dataset	Ramp-SVM Acc ± Sd (Prec, F1)	FSVM Acc ± Sd (Prec, F1)	GBSVM Acc ± Sd (Prec, F1)	GBKNN Acc ± Sd (Prec, F1)	PCA-SVM Acc ± Sd (Prec, F1)	$L_{0/1}$SVM Acc ± Sd (Prec, F1)	GB-$L_{0/1}$SVM Acc ± Sd (Prec, F1)
Transfusion	0.775 ± 0.012 (0.761, 0.768)	0.763 ± 0.011 (0.582,0.660)	0.773 ± 0.003 (0.777 , 0.775)	0.673 ± 0.045 (0.632 , 0.652)	0.633 ± 0.006 (0.770, 0.695)	0.785 ± 0.017 (0.617 , 0.691)	$\mathbf{0.793 \pm 0.015}$ $(\mathbf{0.792}, \mathbf{0.792})$
Glioma	0.854 ± 0.012 (0.871, 0.862)	0.852 ± 0.007 (0.882, 0.867)	0.804 ± 0.021 (0.827, 0.815)	0.821 ± 0.036 (0.821, 0.821)	0.848 ± 0.023 (0.878, 0.863)	0.844 ± 0.021 (0.862, 0.853)	$\mathbf{0.857 \pm 0.011}$ $(\mathbf{0.857}, \mathbf{0.857})$
RICE	0.918 ± 0.011 (0.918, 0.918)	0.917 ± 0.014 (0.917, 0.917)	0.911 ± 0.001 (0.913, 0.912)	0.903 ± 0.012 (0.903, 0.903)	0.920 ± 0.013 (0.926, 0.923)	0.905 ± 0.011 (0.906, 0.905)	$\mathbf{0.925 \pm 0.013}$ $(\mathbf{0.925}, \mathbf{0.925})$
Htru2	0.977 ± 0.007 (0.976, 0.976)	0.979 ± 0.009 (0.979, 0.979)	0.936 ± 0.033 (0.962, 0.949)	0.972 ± 0.048 (0.971, 0.971)	0.969 ± 0.018 (0.971, 0.970)	0.981 ± 0.026 (0.981, 0.981)	$\mathbf{0.997 \pm 0.011}$ $(\mathbf{0.997}, \mathbf{0.997})$
w6a	0.971 ± 0.007 (0.968, 0.969)	0.984 ± 0.007 (0.982, 0.984)	0.870 ± 0.012 (0.926, 0.897)	0.973 ± 0.027 (0.947, 0.960)	0.961 ± 0.018 (0.977, 0.969)	0.984 ± 0.031 (0.983, 0.983)	$\mathbf{0.986 \pm 0.011}$ $(\mathbf{0.986}, \mathbf{0.986})$
Skin	0.972 ± 0.010 (0.975, 0.973)	0.969 ± 0.012 (0.970, 0.969)	0.790 ± 0.005 (0.794, 0.792)	0.979 ± 0.021 (0.979, 0.979)	0.967 ± 0.012 (0.969, 0.968)	0.979 ± 0.023 (0.979, 0.979)	$\mathbf{0.979 \pm 0.013}$ $(\mathbf{0.979}, \mathbf{0.979})$
Avg Acc	0.911	0.911	0.847	0.887	0.883	0.913	**0.923**
Avg Rank	3.50	3.75	6.33	5.00	5.17	3.25	**1.00**

stable performance regions. Specifically, for GB-$L_{0/1}$SVM and $L_{0/1}$-SVM, we jointly optimized the regularization parameter $C \in \{2^i \mid i = -5, \ldots, 8\}$ and the ADMM penalty $\rho \in \{2^i \mid i = -5, \ldots, 7\}$. For GBSVM, PCA-SVM, and FSVM, C was tuned over the same range, while Ramp-SVM additionally required optimization of the truncation parameter s. For high-dimensional datasets, PCA retained at least 90% of the cumulative explained variance. The optimal configurations for each dataset are detailed in Table II.

B. Experimental Results on Datasets without Noise

1) Computational Efficiency: Training times are reported in Table III. All values reflect the average training time across five cross-validation folds, using optimal hyperparameters. GB-$L_{0/1}$SVM achieves the shortest training time on nearly all datasets, with the lowest overall mean.

The efficiency advantage grows markedly with dataset size n. On small datasets such as *Transfusion*, runtimes are comparable across methods. On large-scale datasets, however, baseline models suffer severe scalability bottlenecks. For example, on *Skin segmentation*, PCA-SVM requires 4,032s and GBSVM takes 462s. Meanwhile, GB-$L_{0/1}$SVM completes training in just 0.001s, a speedup of over three orders of magnitude. Standard $L_{0/1}$-SVM incurs substantial computational latency when dealing with high-dimensional datasets like *w6a*. In contrast, GB-$L_{0/1}$SVM maintains high efficiency, demonstrating that the GB representation effectively reduces the problem scale for the ADMM solver.

2) Classification Performance: Table IV presents accuracy, precision, and F1-score results across all benchmark datasets. Among the seven compared methods, GB-$L_{0/1}$SVM achieves both the highest average accuracy and the best average rank.

PCA-SVM compromises discriminative structure through

TABLE V
THE ACCURACY OF SEVEN MODELS ON DATASETS WITH DIFFERENT NOISE LEVELS.

Dataset	Noise	Ramp-SVM	FSVM	GBSVM	GBKNN	PCA-SVM	$L_{0/1}$SVM	GB-$L_{0/1}$SVM
Transfusion	0.05	0.732	0.730	**0.800**	0.653	0.627	0.785	0.793
	0.10	0.700	0.693	0.773	0.673	0.606	0.781	**0.800**
	0.15	0.669	0.665	0.753	0.534	0.593	0.779	**0.789**
	0.20	0.647	0.649	0.750	0.574	0.573	0.774	**0.787**
Glioma	0.05	0.826	0.820	0.800	0.785	0.826	0.856	**0.875**
	0.10	0.803	0.792	0.793	0.732	0.802	0.848	**0.863**
	0.15	0.757	0.753	0.753	0.691	0.760	0.835	**0.862**
	0.20	0.735	0.727	0.750	0.601	0.722	0.821	**0.855**
RICE	0.05	0.888	0.887	0.833	0.825	0.886	0.911	**0.923**
	0.10	0.848	0.846	0.900	0.783	0.843	0.908	**0.925**
	0.15	0.805	0.806	0.867	0.702	0.804	0.901	**0.925**
	0.20	0.765	0.766	0.767	0.673	0.763	0.911	**0.924**
Htru2	0.05	0.929	0.925	0.904	0.921	0.925	0.981	**0.986**
	0.10	0.882	0.874	0.942	0.858	0.874	0.979	**0.981**
	0.15	0.834	0.824	0.930	0.795	0.822	0.975	**0.979**
	0.20	0.785	0.776	0.839	0.730	0.770	0.969	**0.979**
w6a	0.05	0.975	0.976	0.800	0.927	0.907	0.971	**0.984**
	0.10	0.962	0.965	0.793	0.885	0.819	0.966	**0.982**
	0.15	0.948	0.952	0.753	0.829	0.786	0.960	**0.979**
	0.20	0.929	0.935	0.750	0.782	0.738	0.957	**0.975**
Skin	0.05	0.930	0.920	0.800	0.935	0.925	**0.978**	**0.978**
	0.10	0.882	0.875	0.793	0.858	0.880	**0.978**	**0.978**
	0.15	0.834	0.830	0.753	0.802	0.833	0.976	**0.980**
	0.20	0.787	0.784	0.750	0.734	0.785	0.975	**0.980**
Average	0.05	0.880	0.876	0.823	0.841	0.849	0.914	**0.923**
	0.10	0.846	0.841	0.832	0.798	0.804	0.910	**0.922**
	0.15	0.808	0.805	0.802	0.726	0.766	0.904	**0.919**
	0.20	0.775	0.773	0.768	0.675	0.709	0.901	**0.917**

global dimensionality reduction and consequently underperforms on complex datasets such as *Transfusion*. In contrast, GB-$L_{0/1}$SVM preserves local geometric coherence via GB representatives. This advantage is evident on *Htru2*, where GB-$L_{0/1}$SVM achieves an accuracy of 0.997 and surpassing all baseline methods.

On high-dimensional sparse datasets such as *w6a*, GB-$L_{0/1}$SVM also outperforms instance weighted SVM variants: it achieves an accuracy of 0.986, exceeding Ramp-SVM and FSVM. Critically, its consistently high F1-scores and low standard deviations across datasets demonstrate that GB fulfills a dual function: enhancing robustness to label noise while simultaneously acting as an effective geometric regularizer to improve generalization performance.

To assess the stability of GB-$L_{0/1}$SVM against its key hyperparameters, we conducted sensitivity tests for the PCA variance retention ratio $\eta \in [0.8, 0.98]$ and the purity threshold $T \in [0.7, 1.0]$. On the *RICE* and *Transfusion* datasets, the model's test accuracy remains remarkably stable within these broad ranges, with fluctuations typically below 1%. This stability indicates that while the initial parameter selection involves empirical ranges, the GB representation buffers the model against minor parameter variations, ensuring consistent performance across diverse data distributions.

C. Experimental Results on Datasets with Noise

Table V reports classification accuracy for all methods under symmetric 5%-20% label noise. Noise is introduced via label flipping.

As shown in Table V, GB-$L_{0/1}$SVM achieves the highest average accuracy at 20% noise and surpasses $L_{0/1}$-SVM while substantially outperforming standard baselines. Its performance curve remains consistently flat across all noise levels. This robustness is especially pronounced on large-scale datasets. On the *Skin* dataset, the accuracy of Ramp-SVM drops to 0.787 at 20% noise, while GB-$L_{0/1}$SVM maintains 0.980. On *w6a*, it achieves 0.975 versus Ramp-SVM's 0.929. These results confirm that GB enables stable pattern extraction in high-noise, large-scale settings.

PCA-SVM and GBKNN also degrade sharply: on the *Transfusion* dataset, PCA-SVM's classification accuracy declines from 0.627 under 5% label noise to 0.573 under 20% label noise; on the *RICE* dataset, its accuracy drops by over 10 percentage points. In contrast, GB-$L_{0/1}$SVM shows minimal decline. Even though $L_{0/1}$-SVM is competitive, it is instance-based and thus vulnerable to isolated mislabels. GB-$L_{0/1}$SVM mitigates this by grouping samples into GBs: local geometric consistency inherently suppresses label noise without explicit weighting.

D. Statistical Analysis

To statistically validate the superiority and robustness of the proposed GB-$L_{0/1}$SVM, we conduct the Friedman test and the Nemenyi post-hoc test across 30 experimental scenarios comprising 6 original datasets and 24 noisy variants with varying noise levels. The Friedman test yields a p-value of 1.98×10^{-23}, which is far below the significance level $\alpha = 0.05$. This result decisively rejects the null hypothesis, confirming substantial performance differences among the compared models. The Nemenyi post-hoc analysis further reveals that GB-$L_{0/1}$SVM significantly outperforms GBKNN, PCA-SVM, GBSVM, FSVM, and Ramp-SVM at the 95% confidence level. Collectively, these results provide rigorous statistical evidence that the joint incorporation of granular ball computing and PCA meaningfully improves the classification accuracy and label-noise robustness of the $L_{0/1}$-SVM framework.

V. Conclusion

This paper proposes GB-$L_{0/1}$SVM, which is a robust and computationally efficient classification framework that replaces instance level optimization with structure level GB computing. The method integrates PCA-based feature preprocessing, GB representation, and $L_{0/1}$ loss minimization, thereby simultaneously mitigating feature noise and instance label corruption. Extensive experiments on diverse benchmark datasets show that GB-$L_{0/1}$SVM delivers speedups of one to two orders of magnitude over existing robust SVM variants. Beyond performance metrics, the significance of this work lies in the structural abstraction of data. Our results confirm that learning from GBs rather than points is a more reliable paradigm for noise-prone environments.

Despite its empirical success, this study has limitations. First, the ADMM solver used for the non-convex $L_{0/1}$ loss lacks a formal global convergence guarantee, although its practical convergence is demonstrated across all benchmarks. Second, the current implementation is restricted to binary classification. Future extensions will focus on multi-class strategies and kernelized settings, and deriving formal generalization bounds to further solidify the theoretical foundation of granular computing in sparse optimization. Furthermore, we intend to establish rigorous theoretical convergence guarantees, adapt the GB computing strategy for realtime online learning on streaming data, and potentially offering a new path toward sustainable and robust large-scale machine learning.

References

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[2] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sept. 2020.

[3] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An overview on the advancements of support vector machine models in healthcare applications: a review," *Information*, vol. 15, no. 4, p. 235, Apr. 2024.

[4] A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015.

[5] H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee, "Learning from noisy labels with deep neural networks: a survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8135–8153, Nov. 2023.

[6] L. Bottou and C.-J. Lin, "Support vector machine solvers," in *Large Scale Kernel Machines*, L. Bottou et al., Eds. Cambridge, MA, USA: MIT Press, 2007, pp. 301–320.

[7] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA, USA: MIT Press, 1999, pp. 185–208.

[8] S. Shalev-Shwartz, Y. Singer, N. Srebro, and A. Cotter, "Pegasos: Primal estimated sub-gradient solver for SVM," *Math. Program.*, vol. 127, no. 1, pp. 3–30, Mar. 2011.

[9] H. Wang and Y.-H. Shao, "Fast truncated Huber loss SVM for large scale classification," *Knowl. Based Syst.*, vol. 260, p. 110074, Jan. 2023.

[10] A. Kumari, M. Akhtar, M. Tanveer, and M. Arshad, "Diagnosis of breast cancer using flexible pinball loss support vector machine," *Appl. Soft Comput.*, vol. 157, p. 111454, May 2024.

[11] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, Jan. 2019.

[12] H. Wang, Y. Shao, S. Zhou, C. Zhang, and N. Xiu, "Support vector machine classifier via $L_{0/1}$ soft-margin loss," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7253–7265, Oct. 2022.

[13] H. Wang and Y. Shao, "Fast generalized ramp loss support vector machine for pattern classification," *Pattern Recognit.*, vol. 146, p. 109987, Feb. 2024.

[14] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sept. 1997.

[15] 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.

[16] S. Xia, X. Dai, G. Wang, X. Gao, and E. Giem, "An efficient and adaptive granular ball generation method in classification problem," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 4, pp. 5319–5331, Apr. 2024.

[17] J. Pan, G. Lang, Q. Xiao, and T. Yang, "A framework of granular ball generation for classification via granularity tuning," *Appl. Intell.*, vol. 55, no. 1, p. 63, Dec. 2024.

[18] S. Xia, X. Lian, G. Wang, X. Gao, J. Chen, and X. Peng, "GBSVM: An efficient and robust support vector machine framework via granular ball computing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 5, pp. 9253–9267, May 2025.

[19] G. Lang, L. Zhao, D. Miao, and W. Ding, "Granular ball computing based fuzzy twin support vector machine for pattern classification," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 7, pp. 2148–2160, Jul. 2025.

[20] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Inf. Fusion*, vol. 59, pp. 44–58, Jul. 2020.

[21] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. R. Soc. A*, vol. 374, no. 2065, p. 20150202, Apr. 2016.

[22] J. P. Brooks, "Support vector machines with the ramp loss and the hard margin loss," *Oper. Res.*, vol. 59, no. 2, pp. 467–479, Mar. 2011.

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

[24] 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.

[25] D. Dua and C. Graff, "UCI machine learning repository," School of Information and Computer Science, University of California, Irvine, CA, USA, 2019. [Online]. Available: https://archive.ics.uci.edu/ml.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM Data Sets," 2011. [Online]. Available: https://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets/.