Shahid Bahonar University of Kerman
Journal of Mahani Mathematical Research
Print ISSN: 2251-7952 Online ISSN: 2645-4505

# ROBUST SUPPORT VECTOR MACHINES UNDER LABEL NOISE: AN EFFICIENT SIGNED REFORMULATION APPROACH

E. Ahmadi and S. Fallahi ✉

Article type: Research Article
(Received: 02 May 2026, Received in revised form 27 July 2026)
(Accepted: 23 August 2026, Published Online: 23 August 2026)

Abstract. This paper presents a new mixed integer optimization reformulation for robust binary support vector machines under symmetric label noise. Recent work has addressed this issue through mixed integer programming formulations that incorporate label uncertainty via binary decision variables. Representative approaches, such as CDRESVM and DRESVM, introduce mechanisms for label correction in the dual formulation. Despite their modeling flexibility, these methods depend on Big-$M$ constraints and auxiliary variables, which often result in weak continuous relaxations, numerical instability, and increased computational burden. To address these shortcomings, a new formulation, termed Signed DRESVM (SD-DRESVM), is developed. The model is based on a signed decomposition in which each dual variable is expressed as the difference of two nonnegative components, governed by a single binary variable. This construction eliminates the Big-$M$ parameters and auxiliary coupling constraints required by CDRESVM, yielding a more compact mixed-integer optimization formulation with a simpler constraint structure while preserving its label correction mechanism. Experimental results on sixteen benchmark datasets with symmetric label noise demonstrate that the proposed formulation consistently achieves classification performance that is statistically comparable to the strongest existing robust SVM formulations while substantially improving computational efficiency. These results indicate that the proposed signed reformulation preserves predictive robustness while providing a more compact and computationally attractive optimization model.

*Keywords*: Robust support vector machine, label noise, mixed integer optimization, signed decomposition.
*2020 MSC*: 90C90, 62H30, 68T05.

## 1. Introduction

Support vector machines (SVMs) constitute a central methodology in supervised learning, supported by a well-established theoretical foundation, a convex optimization framework, and strong generalization properties [16, 23]. The use of kernel functions enables SVMs to address high-dimensional and nonlinear

✉ s.fallahi@kazerunsfu.ac.ir, ORCID: 0000-0003-4661-5345
https://doi.org/10.22103/jmmr.2026.27092.1962
Publisher: Shahid Bahonar University of Kerman
How to cite: E. Ahmadi, S. Fallahi, *Robust support vector machines under label noise: an efficient signed reformulation approach*, J. Mahani Math. Res. 2027; 16(1): 267-288.

classification tasks effectively, which has contributed to their broad adoption across diverse application domains [5, 10–12, 18]. Nevertheless, standard SVM formulations are known to be sensitive to imperfections in the training data, particularly when label noise is present [13]. Since the hinge loss treats all observations as equally reliable, even moderate mislabeling can distort the decision boundary and lead to a noticeable decline in predictive performance.

Learning in the presence of label noise has therefore been studied extensively. Such noise may arise from annotation errors, ambiguous class definitions, or adversarial manipulation, and can adversely affect both accuracy and generalization [4, 17, 21, 26]. The presence of corrupted labels necessitates mechanisms that explicitly account for uncertainty during training. Existing approaches broadly fall into three categories: modifications of the loss function, sample filtering or reweighting strategies, and optimization-based formulations designed to incorporate robustness [6, 7, 22, 24].

Recent years have witnessed substantial advances in robust support vector machines, driven by the development of novel loss functions and algorithmic frameworks specifically designed to mitigate the adverse effects of label noise and outliers. The eagle loss function, which assigns lower loss values to outliers and noisy samples while preserving discriminative power for boundary instances, has been incorporated into the Eagle SVM formulation [19]. Similarly, the RoBoSS (Robust, Bounded, Sparse, and Smooth) loss function has been integrated into the SVM framework and evaluated on 88 benchmark datasets, demonstrating superior generalization performance under label noise and outlier contamination [2]. Nonconvex loss functions have also attracted considerable attention: the capped asymmetric elastic net (CaEN) loss and the Huberized kernel based (HK) loss have been combined within a fused robust geometric nonparallel SVM (FRGNHSVM), achieving more than 5% improvement in average prediction accuracy on label contaminated datasets [9]. Beyond loss function design, alternative robust SVM variants have emerged, including the confident learning based SVM (CL-SVM), which identifies and removes noisy labels by ranking optimized confidence values and maintains nearly 90% accuracy under noise ratios up to 35% [28]. The granular ball twin SVM (GBTSVM) and its least squares variant (LS-GBTSVM) have demonstrated robustness against noise and outliers through a data grouping mechanism that takes granular balls rather than individual data points as inputs [15]. Furthermore, the Flexi Fuzz least squares SVM introduces a flexible weighting mechanism that integrates class probability and imbalance ratio considerations, showing robust performance on biomedical applications [1]. These developments collectively illustrate the diversity of strategies ranging from loss function engineering to data driven sample weighting that have been pursued to enhance SVM robustness under label noise.

Beyond the SVM specific literature, learning from noisy labels has emerged as a fundamental challenge across the broader machine learning community, with recent surveys systematically categorizing approaches for managing label noise in

both classical and deep learning settings [20, 25]. In the deep learning domain, where noisy labels severely degrade generalization performance, methods include loss correction techniques such as forward and backward correction, sample selection strategies like co-teaching and MentorNet, and semi-supervised approaches including DivideMix [20]. The challenge of label noise is further compounded by class imbalance, instance dependent noise, and long tailed data distributions, leading to increasingly sophisticated frameworks that address these intertwined difficulties [8]. Meanwhile, mixed integer programming (MIP) has emerged as a powerful paradigm for embedding robustness directly into the optimization formulation of classifiers. Recent advances have explored robust optimization frameworks for nonparallel SVM through uncertainty set based formulations using hyper rectangular and hyper-ellipsoidal sets [27], and adjustable robust optimization models have been developed for SVM under feature uncertainty [14]. However, existing MIP based formulations typically rely on Big-M constraints to linearize interactions between continuous and binary variables, which introduce well documented limitations including weak continuous relaxations, numerical instability, and increased computational burden.

Early contributions include robust classification models based on optimization principles [3] and SVM based relabeling approaches such as RESVM and CRESVM [4], which jointly determine classifier parameters and labeling decisions. These formulations demonstrate that embedding relabeling within the optimization problem can enhance robustness in noisy settings.

More recent work has considered dual-based formulations, including DRESVM and CDRESVM [17], where label correction is incorporated directly into the dual representation of the SVM. In these models, the effect of relabeling is reflected through modifications to the sign or magnitude of dual variables, resulting in more compact formulations compared to primal counterparts. This perspective not only reduces the number of variables and constraints but also enables the construction of classifiers beyond those attainable with standard SVM formulations, thereby improving robustness under label noise.

Despite these advances, existing MIP-based formulations typically rely on auxiliary variables and Big-$M$ constraints to linearize interactions between continuous and binary decisions. Although widely used, such techniques introduce several well-known limitations. In particular, Big-$M$ constraints often lead to weak continuous relaxations, which in turn degrade lower bounds and increase the computational effort required by branch-and-bound algorithms. Furthermore, large constants may cause numerical instability, while auxiliary variables increase both the dimensionality and structural complexity of the model.

These observations suggest that the principal difficulty lies not in representing label uncertainty itself, but in how the associated logical structure is incorporated into the optimization problem. Standard linearization approaches tend to obscure this structure, motivating alternative formulations that capture it more directly while maintaining tractability.

Unlike many recent studies that improve robustness through new loss functions, sample reweighting, or data cleaning strategies, the contribution of this work is fundamentally different. Rather than modifying the learning objective, we revisit the underlying mixed integer optimization formulation of robust SVMs. Specifically, we develop a signed reformulation that replaces the conventional Big-$M$ modeling strategy used in existing robust SVM formulations with a compact representation based on signed variable decomposition. Consequently, the proposed model preserves the original classification objective while improving the mathematical structure of the optimization problem and eliminating the need for a user defined Big-$M$ parameter.

In this work, a new formulation for robust SVMs, referred to as *Signed DRESVM* (SD-DRESVM), is introduced. The approach replaces the conventional Big-$M$-based framework with a structure-oriented representation based on signed variable decomposition. Each dual variable is expressed as the difference of two nonnegative components, with a single binary variable determining which component is active. This construction embeds the disjunctive nature of label correction within the variable domain and removes the need for auxiliary variables and Big-$M$ constraints. From an optimization standpoint, the proposed formulation offers several advantages. The absence of Big-$M$ parameters eliminates the dependence on artificially selected constants and removes the associated logical coupling constraints. The resulting formulation consists only of bound constraints and a single linear equality, leading to a structurally simpler mixed-integer optimization model. The proposed approach is evaluated through computational experiments on a diverse set of benchmark datasets under varying levels of label noise. The results indicate that SD-DRESVM attains predictive performance comparable to, and in several cases exceeding, that of existing robust SVM formulations. At the same time, it demonstrates improved computational efficiency and more consistent runtime behavior.

The main contributions of this work are summarized as follows:

- A novel mixed integer optimization reformulation for robust support vector machines based on signed variable decomposition.
- A compact and structurally simplified mixed-integer optimization model that eliminates Big-$M$ parameters and auxiliary coupling variables, resulting in a structurally simpler mixed-integer optimization model with reduced formulation complexity.
- An empirical study demonstrating competitive predictive performance alongside reduced computational effort relative to existing methods.

The remainder of the paper is organized as follows. Section 2 reviews related robust SVM formulations and highlights their structural limitations. Section 3 introduces the proposed SD-DRESVM model. Computational results are presented in Section 4, and concluding remarks are provided in Section 5.

## 2. Baseline Robust SVM Formulations

In this section, we review representative mixed integer optimization models for robust support vector machines that explicitly incorporate label uncertainty into the learning process. These formulations extend the classical dual SVM by embedding discrete decision variables, enabling the model to adjust or reinterpret potentially corrupted labels during training. Among the existing approaches, we focus on two influential formulations, namely CDRESVM and DRESVM [17], which serve as primary benchmarks for the proposed method.

### 2.1. CDRESVM Formulation.

The CDRESVM model introduces robustness by augmenting the dual SVM formulation with binary variables that regulate the admissible range of each dual variable. Through this mechanism, the model captures the possibility that an observed label may be unreliable, allowing the optimization process to adapt the contribution of each training sample accordingly.

Consider a training set $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$. Let $H \in \mathbb{R}^{n\times n}$ denote the kernel matrix defined by $H_{ij} = y_i y_j K(x_i, x_j)$, with $K(\cdot, \cdot)$ being a positive semidefinite kernel.

The CDRESVM model can be formulated as the following mixed integer quadratic program:

$$\min_{\alpha,\xi,\theta} \quad \frac{1}{2}\alpha^\top H\alpha - \mathbf{1}^\top\alpha + C_2\sum_{i=1}^n \xi_i$$

subject to:

$$\begin{gathered}
\sum_{i=1}^n y_i\alpha_i = 0,\\
-M(1-\theta_i) - C_1\frac{(\theta_i - y_i)}{2} \le \alpha_i \le M(1-\theta_i) + C_1\frac{(\theta_i + y_i)}{2}, \quad \forall i,\\
-M\theta_i - C_1\frac{(1+\theta_i + y_i)}{2} \le \alpha_i \le M\theta_i + C_1\frac{(1-\theta_i - y_i)}{2}, \quad \forall i,\\
-C_1\xi_i \le \alpha_i \le C_1(1-\xi_i), \quad \forall i,\\
\xi_i, \theta_i \in \{0, 1\}, \quad \forall i.
\end{gathered}$$

Within this formulation, the binary variables $\theta_i$ and $\xi_i$ play complementary roles in shaping the feasible region of each dual variable $\alpha_i$. The variable $\theta_i$ governs consistency with the observed class label, while $\xi_i$ determines whether the corresponding sample contributes in its original or reversed form. Through the interaction of these variables, the model implicitly encodes label correction in the dual space.

Despite its modeling flexibility, the formulation relies extensively on Big-$M$ constraints to enforce logical relationships between variables. This dependence often results in weak continuous relaxations and sensitivity to the magnitude

of the parameter $M$, which can adversely affect both numerical stability and computational performance.

2.2. **DRESVM Formulation.** The DRESVM formulation adopts an alternative strategy by introducing auxiliary variables that explicitly modify the dual representation. Rather than directly constraining the sign or range of the dual variables, this model constructs a transformed variable that captures the effective contribution of each sample under potential label flipping.

Let $\{(x_i, y_i)\}_{i=1}^n$ denote the training data. In addition to the standard dual variables $\alpha_i$, the model introduces auxiliary continuous variables $\beta_i \geq 0$ and binary variables $\xi_i \in \{0, 1\}$. The effective dual variable is defined as

$$\tilde{\alpha}_i = \alpha_i - 2\beta_i,$$

which determines the role of each observation in the objective function.

The DRESVM formulation is given by:

$$\min_{\alpha,\beta,\xi} \quad \frac{1}{2}(\alpha - 2\beta)^\top H(\alpha - 2\beta) - \mathbf{1}^\top \alpha + C_2 \sum_{i=1}^{n} \xi_i$$

subject to:

$$\begin{aligned} \sum_{i\in I^+} (\alpha_i - 2\beta_i) - \sum_{i\in I^-} (\alpha_i - 2\beta_i) &= 0, \\ \alpha_i - M(1-\xi_i) \leq \beta_i \leq \alpha_i + M(1-\xi_i), &\quad \forall i, \\ -M\xi_i \leq \beta_i \leq M\xi_i, &\quad \forall i, \\ 0 \leq \alpha_i \leq C_1, &\quad \forall i, \\ \xi_i \in \{0,1\}, &\quad \forall i. \end{aligned}$$

The coupling between $\alpha_i$, $\beta_i$, and $\xi_i$ governs whether a data point contributes positively or negatively in the dual formulation. Specifically, when $\xi_i = 0$, the auxiliary variable is forced to zero, yielding $\tilde{\alpha}_i = \alpha_i$ and recovering the standard SVM behavior. Conversely, when $\xi_i = 1$, the constraints enforce $\beta_i = \alpha_i$, resulting in $\tilde{\alpha}_i = -\alpha_i$, which corresponds to reversing the effect of the sample.

This construction provides an explicit and interpretable mechanism for label correction. However, similar to CDRESVM, the formulation depends on Big-$M$ bounds to impose these logical conditions. As a consequence, the model inherits the associated drawbacks, including weak relaxations and increased computational burden. Additionally, the introduction of auxiliary variables $\beta_i$ expands the dimensionality of the problem, which can limit scalability, particularly in large or high-dimensional datasets.

2.3. **Computational Limitations.** While both CDRESVM and DRESVM provide flexible frameworks for handling label noise, their computational performance is constrained by several structural issues. First, both models rely

extensively on Big-$M$ constraints to encode logical implications between binary and continuous variables. These constraints may weaken the linear programming relaxation because the introduction of large Big-$M$ coefficients enlarges the feasible region of the relaxed problem. Consequently, weaker lower bounds, larger branch-and-bound trees, and reduced numerical stability may arise, particularly when conservative values of $M$ are employed. Second, the formulations introduce additional variables per sample. In CDRESVM, binary variables $\theta_i$ and $\xi_i$ increase combinatorial complexity, while in DRESVM, the auxiliary variables $\beta_i$ significantly increase the number of continuous variables. Third, the presence of large constants $M$ can lead to numerical instability, further degrading solver performance. Taken together, these limitations highlight the need for alternative reformulations that eliminate or reduce dependence on Big-$M$ constraints while providing more compact mathematical representations. Addressing these issues is central to improving numerical stability and computational efficiency and directly motivates the reformulation proposed in the next section.

## 3. Signed DRESVM: A Structure-Based Reformulation

Existing robust SVM formulations, including CDRESVM and DRESVM, account for label uncertainty through interactions between continuous dual variables and binary indicators. These interactions are typically enforced via auxiliary variables and Big-$M$ constraints, which serve to linearize implicit bilinear relationships. While such formulations offer considerable modeling flexibility, they are associated with well-documented computational challenges, notably weak continuous relaxations, potential numerical instability, and increased model size due to additional variables and coupling constraints.

To address these issues, a reformulation based on a *signed decomposition* of the dual variables is proposed. The approach is designed to capture the inherent disjunctive structure of label correction directly, rather than relying on constraint-based linearization. Table 1 summarizes the notation used throughout the proposed formulation.

The key idea of the proposed formulation is to represent each dual variable as the difference between two nonnegative components,

$$\alpha_i = \alpha_i^+ - \alpha_i^-, \qquad \alpha_i^+, \alpha_i^- \geq 0, \tag{1}$$

where $\alpha_i^+$ and $\alpha_i^-$ correspond to contributions aligned with the original and flipped labels, respectively. A binary variable $\xi_i \in \{0, 1\}$ is introduced to ensure that at most one of these components is active:

$$\alpha_i^+ \leq C_1(1 - \xi_i), \qquad \alpha_i^- \leq C_1\xi_i. \tag{2}$$

This formulation induces the feasible set

$$\alpha_i \in [0, C_1] \cup [-C_1, 0],$$

with the binary variable selecting the active interval. In contrast to Big-$M$ approaches, the disjunction is incorporated directly into the variable bounds, avoiding the introduction of artificial constants. It is important to emphasize that the proposed SD-DRESVM is derived as a Big-$M$-free reformulation of the CDRESVM optimization model. In CDRESVM, the optimization variable $\alpha_i$ is already represented as a signed decision variable whose sign is controlled through Big-$M$ constraints. To eliminate these constraints, each signed variable is expressed using the decomposition

$$\alpha_i = \alpha_i^+ - \alpha_i^-,$$

where $\alpha_i^+, \alpha_i^- \geq 0$ and at most one of them can be positive. Consequently, every occurrence of the signed variable $\alpha_i$ in the CDRESVM objective function is replaced directly by its signed decomposition. In particular, the linear term

$$-\sum_{i=1}^{n} \alpha_i$$

becomes

$$-\sum_{i=1}^{n} \left(\alpha_i^+ - \alpha_i^-\right),$$

while the quadratic term is obtained by substituting

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}^+ - \boldsymbol{\alpha}^-.$$

Table 1. Notation used in the proposed SD-DRESVM formulation.

| Symbol | Description |
|---|---|
| $l$ | Number of training samples |
| $x_i$ | Training feature vector |
| $y_i \in \{-1, +1\}$ | Class label |
| $K$ | Kernel matrix |
| $H$ | Label-weighted kernel matrix ($H = YKY$) |
| $C_1$ | Regularization parameter |
| $C_2$ | Label correction penalty parameter |
| $\alpha_i^+$ | Positive component of signed dual variable |
| $\alpha_i^-$ | Negative component of signed dual variable |
| $\alpha_i$ | Signed dual variable ($\alpha_i = \alpha_i^+ - \alpha_i^-$) |
| $\xi_i$ | Binary variable indicating label correction |

Accordingly, SD-DRESVM preserves the optimization objective of CDRESVM while replacing its Big-$M$-based sign representation with an equivalent signed-variable decomposition. Let $H = YKY \in \mathbb{R}^{l \times l}$ denote the kernel matrix. The resulting Signed DRESVM (SD-DRESVM) formulation is given by:

$$
\begin{aligned}
\min_{\alpha^+,\alpha^-,\xi} \quad & \frac{1}{2}(\alpha^+ - \alpha^-)^\top H(\alpha^+ - \alpha^-) - \sum_{i=1}^{l}(\alpha_i^+ - \alpha_i^-) + C_2 \sum_{i=1}^{l} \xi_i \\
\text{s.t.} \quad & \sum_{i=1}^{l} y_i(\alpha_i^+ - \alpha_i^-) = 0, \\
& \alpha_i^+ \leq C_1(1 - \xi_i), \quad \forall i, \\
& \alpha_i^- \leq C_1 \xi_i, \quad \forall i, \\
& \alpha_i^+ \geq 0, \quad \alpha_i^- \geq 0, \quad \forall i, \\
& \xi_i \in \{0, 1\}, \quad \forall i.
\end{aligned}
$$

The first constraint preserves the equality condition inherited from the dual SVM formulation. The second constraint restricts the positive component to be active only when no label correction is selected, whereas the third constraint activates the negative component only when a label correction is permitted. Together, these constraints guarantee that the signed variable automatically assumes the correct admissible range without introducing auxiliary Big-$M$ coefficients. Constraints (2) ensure that no more than one of $\alpha_i^+$ or $\alpha_i^-$ can take a positive value, thereby determining the sign of $\alpha_i$ implicitly. This eliminates the need for auxiliary variables and explicit bilinear constraints, while preserving the capacity to represent label correction within the dual formulation.

From an optimization perspective, the feasible region is notably simpler and more structured. The constraint set consists solely of bound constraints together with a single linear equality, and does not involve cross-variable coupling. As a result, the proposed formulation completely eliminates the dependence on Big-$M$ constants and their associated logical coupling constraints. Since the disjunctive structure is represented directly through variable bounds, the formulation avoids the need for an artificially selected Big-$M$ parameter.

The choice of the Big-$M$ parameter plays an important role in the computational behavior of classical mixed integer formulations. If $M$ is selected excessively large, the continuous relaxation generally becomes weaker, which may lead to poorer lower bounds and increased branch-and-bound effort. Conversely, selecting $M$ too small may inadvertently exclude feasible integer solutions. In contrast, the proposed SD-DRESVM formulation does not require any Big-$M$ parameter, thereby eliminating this formulation dependency and avoiding the need for problem-specific parameter selection.

In terms of dimensionality, the formulation introduces $2l$ continuous variables and $l$ binary variables, yielding a total of $3l$ variables. The constraint system comprises $4l$ bound constraints and one equality constraint, for a total of $4l + 1$ constraints. Although the overall size is comparable to that of

existing models, the absence of auxiliary constructs and the reduced interaction among constraints lead to a more compact and computationally efficient representation.

Beyond computational considerations, the signed decomposition provides a direct interpretation of robustness. The variables $\alpha_i^+$ and $\alpha_i^-$ distinguish between agreement and disagreement with the observed labels, allowing the model to represent corrected and uncorrected contributions in an explicit manner.

Overall, the SD-DRESVM formulation replaces conventional linearization techniques with a structure-based representation of the underlying disjunction. Overall, the SD-DRESVM formulation replaces conventional linearization techniques with a structure-based representation of the underlying disjunction. This yields a structurally simpler optimization model that avoids Big-$M$ parameters while retaining the flexibility required for robust classification in the presence of label noise.

## 4. **Experimental Results**

This section reports a comprehensive empirical evaluation of the proposed SD-DRESVM formulation relative to SVM, CDRESVM, and DRESVM. The presentation is organized in two stages. We first describe the experimental protocol and benchmark setting, and then examine predictive accuracy, statistical significance, and computational efficiency in detail. The experiments are carried out on a heterogeneous collection of benchmark binary classification datasets, summarized in Table 2. The selected instances cover a broad range of sample sizes and feature dimensions, including small-sample high-dimensional problems (e.g., *Leukemia*, *ALLAML*) together with larger-scale datasets (e.g., *Twonorm*). Such variation is important for assessing performance under distinct structural conditions, since the impact of label noise may differ substantially across problem classes.

To evaluate robustness against annotation errors, synthetic label noise is generated by independently flipping class labels at rates of $0\%, 10\%, 20\%, 30\%$, and $40\%$. This controlled contamination scheme permits a systematic examination of how predictive performance deteriorates as the proportion of corrupted labels increases. All experiments were conducted using a linear kernel. For each training fold, the kernel matrix was computed as

$$K = X_{\text{train}} X_{\text{train}}^\top,$$

and the corresponding matrix

$$H = YKY$$

was constructed exclusively from the training samples. Consequently, the kernel matrix was recomputed independently within every cross-validation fold, ensuring that no information from the test fold was used during model training.

Prior to model training, each dataset was standardized using zero-mean and

TABLE 2. Characteristics of datasets used in the computational experiments

| **Dataset** | **Samples (n)** | **Features (d)** | **Classes** |
|---|---|---|---|
| Fertility | 100 | 9 | 2 |
| Parkinsons | 195 | 22 | 2 |
| Pima Indians Diabetes | 768 | 8 | 2 |
| Heart Disease | 270 | 13 | 2 |
| Haberman Survival | 306 | 3 | 2 |
| Mammographic Mass (Ordinal) | 961 | 12 | 2 |
| Promoters | 106 | 57 | 2 |
| SPECT Heart | 267 | 22 | 2 |
| WPBC (Breast Cancer Prognostic) | 198 | 33 | 2 |
| Sonar | 208 | 60 | 2 |
| Vertebral Column | 310 | 6 | 2 |
| Hayes-Roth | 160 | 4 | 2 |
| Australian Credit | 690 | 14 | 2 |
| Twonorm | 7400 | 20 | 2 |
| Leukemia | 72 | 7070 | 2 |
| ALLAML | 72 | 7129 | 2 |

unit-variance normalization. The regularization parameters $C_1$ and $C_2$ were selected independently for each method using an exhaustive grid search over

$$\{2^{-3}, 2^{-2}, \ldots, 2^{3}\},$$

resulting in 49 candidate parameter pairs.
For the original DRESVM and CDRESVM formulations, the Big-$M$ constant was fixed at $M = 1000$ throughout all experiments.

Model performance is estimated using a repeated cross-validation protocol. For all datasets except *Twonorm*, we conduct 10 independent repetitions of 10-fold cross-validation. Owing to its larger scale, *Twonorm* is evaluated using a single 10-fold cross-validation run for computational efficiency. Reported values correspond to the mean classification accuracy and associated standard deviation computed over all folds and repetitions. All formulations are implemented in Python through the CVX framework and solved using MOSEK. The experiments are executed on a high-performance computing system equipped with 96 CPU cores, while the solver is restricted to a single thread in order to maintain a consistent and reproducible computational setting.

Table 5 presents the classification accuracy (mean $\pm$ standard deviation) obtained by all methods across the considered noise levels.

Across the full set of datasets, a consistent trend is observed: the predictive performance of the standard SVM declines markedly as the noise level increases, reaffirming its established sensitivity to mislabeled observations. By contrast,

Table 3. Average rank of models based on classification accuracy under different label noise levels

| Model | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | Overall |
|---|---|---|---|---|---|---|
| CDRESVM | 1.938 | 1.906 | 2.094 | 2.000 | 2.031 | 1.994 |
| DRESVM | 2.250 | 2.281 | 2.344 | 2.125 | 2.219 | 2.244 |
| SD-DRESVM | 1.812 | 1.812 | 1.562 | 1.875 | 1.750 | 1.762 |

Table 4. Pairwise Nemenyi post-hoc p-values.

| Comparison | p-value |
|---|---|
| CDRESVM vs DRESVM | 0.2538 |
| CDRESVM vs SD-DRESVM | 0.3091 |
| DRESVM vs SD-DRESVM | **0.0066** |

all robust formulations exhibit substantially greater resistance to label corruption, with a notably slower loss in accuracy. The proposed SD-DRESVM delivers results that are comparable to, and in many instances statistically indistinguishable from, those of CDRESVM. On several datasets, including *fertility*, *parkinsons*, and *spect*, the two methods attain identical or nearly identical accuracy across all contamination levels. More importantly, SD-DRESVM remains competitive in more demanding settings, particularly under higher noise rates where robustness is most consequential. In small-sample, high-dimensional problems such as *Leukemia* and *ALLAML*, the proposed formulation shows clear advantages under moderate to severe label noise. This behavior suggests that the signed decomposition mechanism is effective in managing substantial labeling uncertainty. Taken together, these findings indicate that the proposed model retains the robustness characteristics of existing approaches while relying on a substantially simpler optimization structure. Overall, the results indicate that the proposed SD-DRESVM preserves the classification capability of existing robust SVM formulations while substantially simplifying the underlying optimization model. Across most datasets and noise levels, SD-DRESVM achieves prediction accuracy statistically comparable to CDRESVM, indicating that the elimination of Big-$M$ constraints does not compromise the robustness of the classifier. This observation suggests that the proposed reformulation successfully preserves the optimization landscape of the original model while providing a more compact mathematical representation.

To provide an aggregate comparison across datasets and noise levels, the average rank of each method is reported in Table 3.

The ranking analysis indicates that SD-DRESVM attains the best overall average rank across all noise levels, followed by CDRESVM and DRESVM.

The ranking analysis indicates that SD-DRESVM achieves the best average rank across all considered noise levels, followed closely by CDRESVM. However, as confirmed by the subsequent Nemenyi post-hoc analysis, the difference between these two methods is not statistically significant. Consequently, the ranking results should be interpreted as indicating that SD-DRESVM preserves the predictive performance of the strongest existing formulation rather than establishing statistically superior classification accuracy. The associated critical difference diagram is reported in Figure 1.

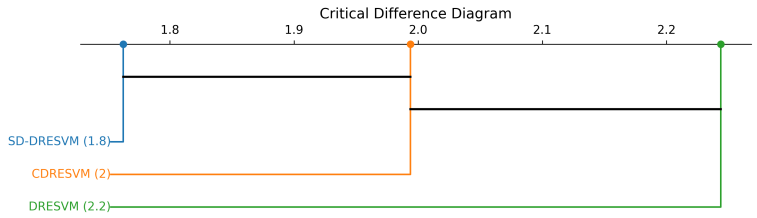


FIGURE 1. Critical difference diagram based on average ranks of classification accuracy across all datasets and noise levels. Methods connected by a horizontal bar are not significantly different according to the post-hoc Nemenyi test. Lower ranks indicate better performance.

The diagram indicates that SD-DRESVM achieves the best average rank, followed closely by CDRESVM. Both methods fall within the same statistical group under the Nemenyi post-hoc procedure, suggesting that their predictive performance cannot be distinguished at the selected significance level. By contrast, DRESVM receives a weaker rank and is separated from SD-DRESVM, reflecting inferior overall performance. Therefore, the principal advantage of the proposed formulation lies in maintaining the predictive robustness of the strongest existing robust SVM formulation while providing a simpler optimization structure and improved computational efficiency.

To complement the critical difference diagram, Table 4 reports the pairwise p-values obtained from the Nemenyi post-hoc procedure. The results indicate that SD-DRESVM and CDRESVM are statistically indistinguishable ($p = 0.3091$), confirming that the proposed reformulation preserves the predictive performance of the strongest existing robust SVM formulation. In contrast, SD-DRESVM achieves a statistically significant improvement over DRESVM ($p = 0.0066$), while no statistically significant difference is observed between CDRESVM and DRESVM ($p = 0.2538$). These findings are consistent with the critical difference diagram and support the conclusion that the primary contribution of SD-DRESVM lies in maintaining competitive predictive performance while improving the underlying optimization formulation.

The distribution of classification accuracy across datasets is illustrated in Figure 2.

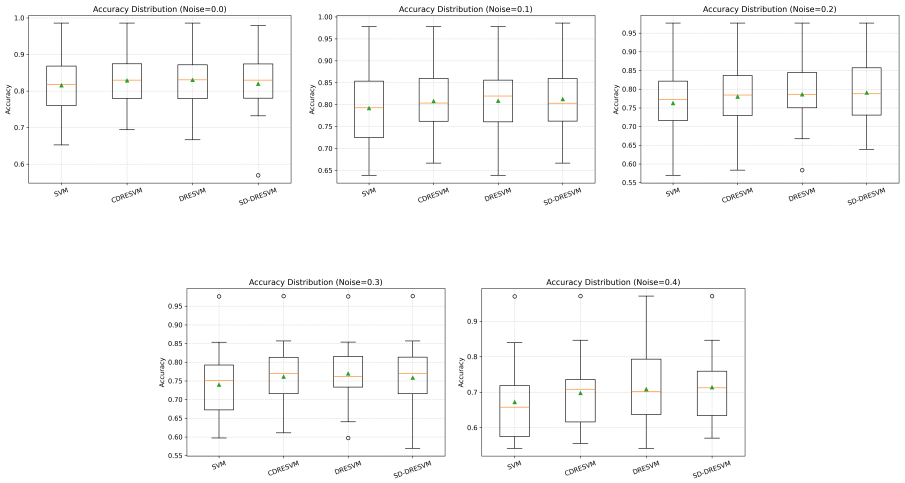


FIGURE 2. Distribution of classification accuracy across datasets for different label noise levels.

At low contamination levels, all methods exhibit broadly similar predictive performance with relatively limited variability. As the proportion of noisy labels increases, the accuracy of the standard SVM declines sharply and its dispersion widens substantially. In contrast, the robust formulations retain noticeably tighter performance distributions. Among them, SD-DRESVM shows a stable interquartile range together with competitive median accuracy across all noise levels, indicating both robustness and consistency.

The aggregate relationship between average classification accuracy and noise level is illustrated in Figure 3. While all methods experience a monotonic decline in performance, SD-DRESVM exhibits a more gradual degradation compared to SVM and remains competitive with the best-performing robust methods, particularly in moderate to high noise regimes.

Table 6 reports the training runtime of the robust formulations over all datasets and noise levels, offering a detailed comparison of computational efficiency. A clear and recurring pattern is observed: DRESVM is the most computationally demanding method on the majority of instances, frequently by a considerable margin. This effect is especially evident on medium- and large-scale datasets such as *mammoOrdinal*, *sonar*, and *twonorm*, where the runtime of DRESVM exceeds that of the competing formulations by several multiples. The observed overhead is consistent with the additional computational burden induced by auxiliary variables and Big-$M$ coupling constraints.

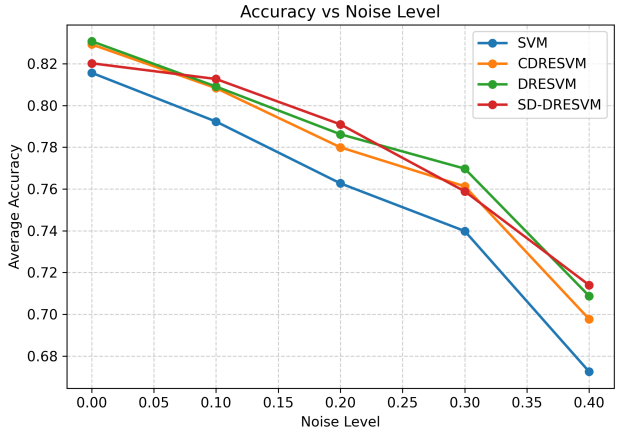


FIGURE 3. Average classification accuracy as a function of label noise level.

By contrast, the proposed SD-DRESVM formulation exhibits substantially stronger computational efficiency in most settings. On many datasets, including *parkinsons*, *promoters*, *wpbc*, and *vertebra-column*, SD-DRESVM consistently records the lowest, or nearly the lowest, runtime across all noise levels, while maintaining relatively stable performance as contamination increases. In larger-scale problems such as *twonorm*, SD-DRESVM clearly outperforms both CDRESVM and DRESVM, illustrating a tangible advantage in scalability. CDRESVM generally remains competitive and often appears as the second fastest formulation, yet it is slower than SD-DRESVM in most experimental scenarios. Some isolated variability is observed in specific instances, such as the *Leukemia* dataset at zero noise, where SD-DRESVM shows an atypically large runtime. This effect is not persistent, however, and disappears at higher noise levels. The computational improvements become more pronounced on datasets with larger optimization problems, where the branch-and-bound algorithm benefits from the reduced model complexity of the proposed formulation. Since SD-DRESVM eliminates the Big-$M$ linking constraints and requires fewer auxiliary variables, the solver generally explores a simpler feasible region, leading to shorter solution times while maintaining essentially the same classification performance.

Figure 4 presents the distribution of logarithmic runtimes. DRESVM exhibits the highest median runtime and the largest variability, indicating substantial computational burden. In contrast, SD-DRESVM achieves the lowest median runtime with a relatively compact distribution, demonstrating improved efficiency and stability. CDRESVM performs competitively but remains slightly slower on average.

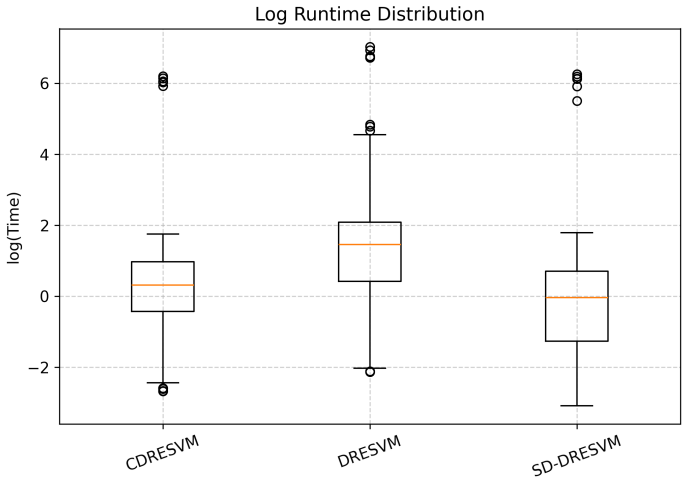


FIGURE 4. Boxplot of logarithmic computational runtime for the robust models across all datasets and noise levels. Lower values indicate faster solution times.

The relationship between runtime and noise level is shown in Figure 5. Across all noise levels, DRESVM consistently incurs the highest computational cost. SD-DRESVM achieves the most favorable performance in several regimes, particularly at moderate noise levels, and remains competitive even as noise increases.

Finally, the trade-off between predictive accuracy and computational cost is illustrated in Figure 6. The results indicate that while DRESVM may achieve marginally higher accuracy in some cases, this comes at a substantially higher computational cost. CDRESVM offers faster runtimes but at the expense of slightly reduced accuracy. The proposed SD-DRESVM provides a balanced compromise, achieving accuracy comparable to the best-performing methods while maintaining significantly lower computational cost.

The experimental results demonstrate that the proposed SD-DRESVM formulation successfully combines statistically competitive predictive performance with consistently improved computational efficiency. Across the considered benchmark datasets, it preserves the robustness of existing mixed integer robust SVM formulations while reducing computational cost through a simpler optimization model that avoids Big-$M$ constraints and auxiliary variables.

Overall, the experimental results indicate that the principal advantage of SD-DRESVM lies in its optimization formulation rather than in improved predictive accuracy. Across all noise levels, the proposed model consistently achieves classification performance comparable to CDRESVM while requiring

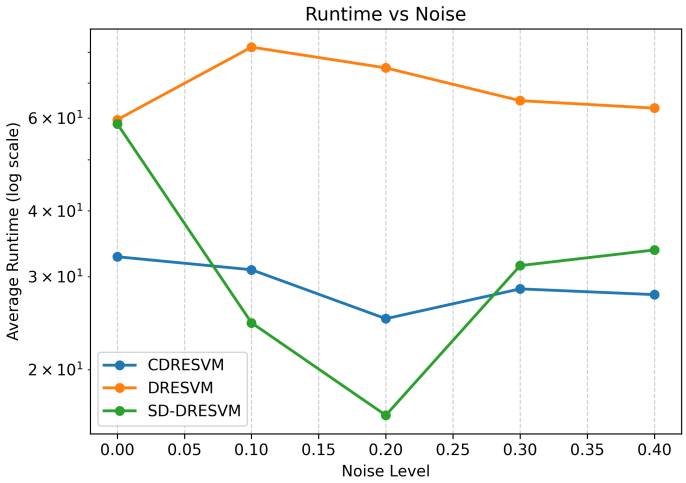


FIGURE 5. Average computational runtime versus label noise level for the robust models. The vertical axis is displayed on a logarithmic scale.

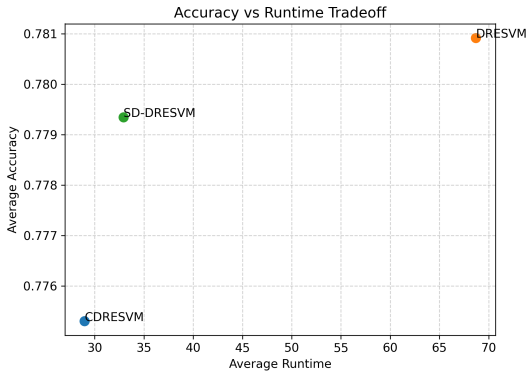


FIGURE 6. Accuracy versus average runtime tradeoff across all methods.

substantially less computational time for most benchmark datasets. These improvements are most pronounced on datasets where the elimination of Big-$M$ constraints leads to a simpler mixed-integer optimization model. Conversely, on a small number of datasets the computational advantage becomes less evident, suggesting that the benefit of the proposed reformulation depends on the complexity of the underlying optimization instance. This observation highlights

the trade-off between predictive performance and computational efficiency and provides further insight into the practical behavior of the proposed formulation.

## 5. Conclusion

A new mixed integer reformulation for robust support vector machines, termed SD-DRESVM, has been presented to address structural and computational limitations of existing approaches under label noise. The model replaces conventional Big-$M$-based linearization with a signed variable decomposition, in which each dual variable is expressed as the difference of two nonnegative components controlled by a binary decision. This construction embeds label correction directly into the variable domain and avoids auxiliary coupling constraints. The resulting formulation is structurally simpler owing to the elimination of Big-$M$ constraints and auxiliary coupling variables. By avoiding the need for an artificially selected Big-$M$ parameter, the proposed model provides a more compact mixed-integer optimization formulation while preserving the label correction mechanism of the underlying robust SVM model. In addition, the decomposition provides a clear interpretation of how individual observations contribute to the decision boundary under potential mislabeling. Computational experiments conducted on sixteen benchmark datasets demonstrate that SD-DRESVM achieves predictive performance that is statistically comparable to the strongest existing robust SVM formulation, CDRESVM, while consistently exhibiting lower computational cost than Big-$M$-based alternatives. Rather than improving predictive accuracy, the principal contribution of the proposed formulation is the redesign of the optimization model itself. By eliminating Big-$M$ constraints and auxiliary variables through a signed decomposition of the dual variables, the proposed formulation preserves robustness under label noise while providing a structurally simpler and computationally more efficient mixed-integer optimization formulation. Future work may consider extensions to kernelized settings and large-scale instances, as well as the incorporation of explicit regularization schemes for controlling label adjustments. Further analysis of the relaxation properties of the proposed formulation may also provide additional insight into its computational behavior. A further direction for future work is to investigate the practical competitiveness of the proposed formulation against representative noisy-label learning algorithms developed for deep learning.

## References

[1] Akhtar, M., Quadir, A., Tanveer, M., & Arshad, M. (2024). Flexi-Fuzz least squares SVM for Alzheimer's diagnosis: Tackling noise, outliers, and class imbalance. arXiv preprint arXiv:2410.14207. https://doi.org/10.48550/arXiv.2410.14207

[2] Akhtar, M., Tanveer, M., & Arshad, M. (2024). RoBoSS: A robust, bounded, sparse, and smooth loss function for supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(1), 149-160. https://doi.org/10.1109/TPAMI.2024.3465535

[3] Bertsimas, D., Dunn, J., Pawlowski, C., & Zhuo, Y. D. (2019). Robust classification. INFORMS Journal on Optimization, 1(1), 2-34. https://doi.org/10.1287/ijoo.2018.0001
[4] Blanco, V., Japón, A., & Puerto, J. (2022). A mathematical programming approach to SVM-based classification with label noise. Computers & Industrial Engineering, 172, 108611. https://doi.org/10.1016/j.cie.2022.108611
[5] Du, K.-L., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024). Exploring kernel machines and support vector machines: Principles, techniques, and future directions. Mathematics, 12(24), 3935. https://doi.org/10.3390/math12243935
[6] Duan, Y., & Wu, O. (2016). Learning with auxiliary less-noisy labels. IEEE Transactions on Neural Networks and Learning Systems, 28(7), 1716-1721. https://doi.org/10.1109/TNNLS.2016.2546956
[7] Ekambaram, R., Fefilatyev, S., Shreve, M., Kramer, K., Hall, L. O., Goldgof, D. B., & Kasturi, R. (2016). Active cleaning of label noise. Pattern Recognition, 51, 463-480. https://doi.org/10.1016/j.patcog.2015.09.020
[8] Fazekas, A., & Szeghalmy, S. (2024). Effect of label-noise filtering on classification of imbalanced data sets with SVM. Proceedings of the Future Technologies Conference, 194-204. Springer. https://doi.org/10.1007/978-3-031-73110-5-14
[9] Gao, R., Qi, K., & Yang, H. (2024). Fused robust geometric nonparallel hyperplane support vector machine for pattern classification. Expert Systems with Applications, 236, 121331. https://doi.org/10.1016/j.eswa.2023.121331
[10] Khatibi Bardsiri, A. (2025). A new model for lung cancer prediction based on differential evolution algorithm and effective feature selection. Journal of Mahani Mathematical Research, 14(1), 345-367. https://doi.org/10.22103/jmmr.2024.23134.1597
[11] Khyathi, G., Indumathi, K. P., Jumana Hasin, A., Lisa Flavin Jency, M., Krishnaprakash, G., & others. (2025). Support vector machines: a literature review on their application in analyzing mass data for public health. Cureus, 17(1), e77169. https://doi.org/10.7759/cureus.77169
[12] Kumar, R., & Swarnkar, M. (2025). QuIDS: A quantum support vector machine-based intrusion detection system for IoT networks. Journal of Network and Computer Applications, 234, 104072. https://doi.org/10.1016/j.jnca.2024.104072
[13] Li, J., Li, Y., Song, J., Zhang, J., & Zhang, S. (2024). Quantum support vector machine for classifying noisy data. IEEE Transactions on Computers, 73(9), 2233-2247. https://doi.org/10.1109/TC.2024.3416619
[14] Maggioni, F., & Spinelli, A. (2025). A novel robust optimization model for nonlinear support vector machine. European Journal of Operational Research, 322(1), 237-253. https://doi.org/10.1016/j.ejor.2024.12.014
[15] Quadir, A., Sajid, M., & Tanveer, M. (2024). Granular ball twin support vector machine. IEEE Transactions on Neural Networks and Learning Systems, 36(7), 12444-12453. https://doi.org/10.1109/TNNLS.2024.3476391
[16] Roy, A., & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. Reliability Engineering & System Safety, 233, 109126. https://doi.org/10.1016/j.ress.2023.109126
[17] Sahleh, A., Salahi, M., & Eskandari, S. (2025). A dual SVM approach to noisy labels relabeling. Journal of the Operations Research Society of China, 1-18. https://doi.org/10.1007/s40305-024-00575-8
[18] Shiri, M. A., Omidi, M. R., & Mansouri, N. (2024). A new hybrid filter-wrapper feature selection using equilibrium optimizer and simulated annealing. Journal of Mahani Mathematical Research Center, 13(1). https://doi.org/10.22103/jmmr.2023.21150.1411
[19] Shrivastava, S., Shukla, S., & Khare, N. (2024). Support vector machine with eagle loss function. Expert Systems with Applications, 238, 122168. https://doi.org/10.1016/j.eswa.2023.122168

[20] Song, B., Zhao, S., Dang, L., Wang, H., & Xu, L. (2025). A survey on learning from data with label noise via deep neural networks. Systems Science & Control Engineering, 13(1), 2488120. https://doi.org/10.1080/21642583.2025.2488120
[21] Song, H., Kim, M., Park, D., Shin, Y., & Lee, J.-G. (2022). Learning from noisy labels with deep neural networks: A survey. IEEE Transactions on Neural Networks and Learning Systems, 34(11), 8135-8153. https://doi.org/10.1109/TNNLS.2022.3152527
[22] Thulasidasan, S., Bhattacharya, T., Bilmes, J., Chennupati, G., & Mohd-Yusof, J. (2019). Combating label noise in deep learning using abstention. arXiv preprint arXiv:1905.10964. https://doi.org/10.48550/arXiv.1905.10964
[23] Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE Transactions on Neural Networks, 10(5), 988-999. https://doi.org/10.1109/72.788640
[24] Xu, Y., Cao, P., Kong, Y., & Wang, Y. (2019). L_dmi: A novel information-theoretic loss function for training deep nets robust to label noise. Advances in Neural Information Processing Systems, 32. https://doi.org/10.48550/arXiv.1909.03388
[25] Zhang, H., Zhang, Y., Li, J., Liu, J., & Ji, L. (2025). A survey on learning with noisy labels in Natural Language Processing: How to train models with label noise. Engineering Applications of Artificial Intelligence, 146, 110157. https://doi.org/10.1016/j.engappai.2025.110157
[26] Zhang, J., Song, B., Wang, H., Han, B., Liu, T., Liu, L., & Sugiyama, M. (2024). Badlabel: A robust perspective on evaluating and enhancing label-noise learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(6), 4398-4409. https://doi.org/10.1109/TPAMI.2024.3355425
[27] Zhang, W., Wang, G., & Du, J. (2025). Robust optimization models for nonparallel support vector machine. Digital Signal Processing, 105616. https://doi.org/10.1016/j.dsp.2025.105616
[28] Zhang, X.-Y., Zhang, X.-P., Yu, H.-G., & Liu, Q.-S. (2025). A confident learning-based support vector machine for robust ground classification in noisy label environments. Tunnelling and Underground Space Technology, 155, 106128. https://doi.org/10.1016/j.tust.2024.106128

Ehsan Ahmadi
Orcid number: 0009-0005-1904-1202
Department of Computer Engineering
Salman Farsi University of Kazerun
Kazerun, Iran
*Email address*: e.ahmadi@kazerunsfu.ac.ir

Saeed Fallahi
Orcid number: 0000-0003-4661-5345
Department of Mathematics and Computer Science
Salman Farsi University of Kazerun
Kazerun, Iran
*Email address*: s.fallahi@kazerunsfu.ac.ir

TABLE 5. Classification Accuracy (Mean ± Std) under Different Label Noise Levels

| Dataset | Flip | SVM | CDRESVM | DRESVM | SD-DRESVM |
|---|---|---|---|---|---|
| fertility | 0.0 | 0.878±0.006 | 0.880±0.000 | 0.880±0.000 | 0.880±0.000 |
| | 0.1 | 0.877±0.009 | 0.880±0.000 | 0.881±0.003 | 0.880±0.000 |
| | 0.2 | 0.853±0.021 | 0.873±0.015 | 0.867±0.018 | 0.873±0.015 |
| | 0.3 | 0.807±0.025 | 0.857±0.021 | 0.834±0.021 | 0.857±0.021 |
| | 0.4 | 0.634±0.048 | 0.709±0.034 | 0.694±0.037 | 0.709±0.034 |
| parkinsons | 0.0 | 0.868±0.010 | 0.886±0.006 | 0.880±0.004 | 0.886±0.006 |
| | 0.1 | 0.856±0.014 | 0.874±0.012 | 0.861±0.010 | 0.873±0.012 |
| | 0.2 | 0.835±0.019 | 0.861±0.014 | 0.846±0.016 | 0.862±0.014 |
| | 0.3 | 0.788±0.020 | 0.821±0.011 | 0.809±0.012 | 0.821±0.011 |
| | 0.4 | 0.682±0.036 | 0.721±0.031 | 0.709±0.031 | 0.721±0.030 |
| pima | 0.0 | 0.769±0.003 | 0.774±0.002 | 0.771±0.003 | 0.774±0.002 |
| | 0.1 | 0.767±0.006 | 0.772±0.005 | 0.769±0.005 | 0.772±0.005 |
| | 0.2 | 0.761±0.009 | 0.765±0.008 | 0.763±0.008 | 0.764±0.009 |
| | 0.3 | 0.737±0.011 | 0.740±0.011 | 0.744±0.007 | 0.740±0.011 |
| | 0.4 | 0.703±0.015 | 0.708±0.013 | 0.718±0.013 | 0.708±0.013 |
| heart | 0.0 | 0.834±0.005 | 0.844±0.004 | 0.850±0.004 | 0.844±0.004 |
| | 0.1 | 0.831±0.012 | 0.841±0.010 | 0.848±0.006 | 0.841±0.010 |
| | 0.2 | 0.817±0.012 | 0.830±0.009 | 0.844±0.009 | 0.830±0.009 |
| | 0.3 | 0.765±0.021 | 0.784±0.017 | 0.834±0.011 | 0.784±0.017 |
| | 0.4 | 0.695±0.025 | 0.730±0.023 | 0.807±0.009 | 0.730±0.023 |
| haberman | 0.0 | 0.726±0.004 | 0.732±0.004 | 0.736±0.009 | 0.732±0.004 |
| | 0.1 | 0.730±0.006 | 0.736±0.005 | 0.732±0.004 | 0.736±0.004 |
| | 0.2 | 0.733±0.006 | 0.736±0.005 | 0.736±0.007 | 0.737±0.006 |
| | 0.3 | 0.733±0.008 | 0.738±0.008 | 0.735±0.008 | 0.738±0.008 |
| | 0.4 | 0.709±0.035 | 0.716±0.029 | 0.712±0.031 | 0.716±0.029 |
| mammoOrdinal | 0.0 | 0.827±0.002 | 0.827±0.002 | 0.825±0.003 | 0.827±0.002 |
| | 0.1 | 0.801±0.008 | 0.796±0.008 | 0.815±0.003 | 0.797±0.008 |
| | 0.2 | 0.778±0.009 | 0.777±0.007 | 0.813±0.003 | 0.778±0.007 |
| | 0.3 | 0.776±0.007 | 0.777±0.007 | 0.806±0.005 | 0.777±0.007 |
| | 0.4 | 0.750±0.011 | 0.753±0.010 | 0.793±0.007 | 0.753±0.010 |
| promoters | 0.0 | 0.784±0.026 | 0.804±0.024 | 0.831±0.018 | 0.805±0.025 |
| | 0.1 | 0.703±0.030 | 0.768±0.027 | 0.824±0.025 | 0.768±0.024 |
| | 0.2 | 0.647±0.033 | 0.709±0.024 | 0.790±0.034 | 0.712±0.025 |
| | 0.3 | 0.600±0.039 | 0.651±0.033 | 0.744±0.032 | 0.652±0.033 |
| | 0.4 | 0.547±0.052 | 0.622±0.051 | 0.668±0.047 | 0.632±0.046 |
| Spect | 0.0 | 0.810±0.011 | 0.832±0.009 | 0.831±0.011 | 0.832±0.009 |
| | 0.1 | 0.785±0.018 | 0.811±0.007 | 0.799±0.010 | 0.810±0.007 |
| | 0.2 | 0.767±0.021 | 0.797±0.007 | 0.782±0.018 | 0.798±0.007 |
| | 0.3 | 0.715±0.020 | 0.763±0.020 | 0.747±0.019 | 0.764±0.020 |
| | 0.4 | 0.603±0.043 | 0.635±0.035 | 0.622±0.039 | 0.635±0.035 |
| wpbc | 0.0 | 0.799±0.012 | 0.813±0.010 | 0.801±0.011 | 0.813±0.010 |
| | 0.1 | 0.766±0.014 | 0.786±0.008 | 0.779±0.009 | 0.787±0.007 |
| | 0.2 | 0.730±0.016 | 0.762±0.015 | 0.755±0.016 | 0.762±0.014 |
| | 0.3 | 0.693±0.032 | 0.737±0.023 | 0.728±0.024 | 0.737±0.023 |
| | 0.4 | 0.609±0.030 | 0.652±0.032 | 0.642±0.035 | 0.653±0.033 |
| sonar | 0.0 | 0.732±0.012 | 0.782±0.012 | 0.782±0.012 | 0.782±0.012 |
| | 0.1 | 0.708±0.014 | 0.746±0.018 | 0.736±0.016 | 0.746±0.018 |
| | 0.2 | 0.676±0.024 | 0.706±0.027 | 0.706±0.023 | 0.707±0.028 |
| | 0.3 | 0.612±0.019 | 0.653±0.018 | 0.682±0.016 | 0.653±0.019 |
| | 0.4 | 0.577±0.039 | 0.601±0.032 | 0.648±0.027 | 0.601±0.033 |
| vertebra-column | 0.0 | 0.870±0.005 | 0.873±0.003 | 0.869±0.004 | 0.872±0.003 |
| | 0.1 | 0.830±0.018 | 0.834±0.015 | 0.830±0.017 | 0.834±0.015 |
| | 0.2 | 0.813±0.010 | 0.817±0.009 | 0.815±0.010 | 0.817±0.009 |
| | 0.3 | 0.805±0.006 | 0.811±0.005 | 0.806±0.006 | 0.811±0.005 |
| | 0.4 | 0.785±0.026 | 0.803±0.017 | 0.794±0.023 | 0.803±0.017 |
| hayes-roth | 0.0 | 0.684±0.019 | 0.706±0.020 | 0.750±0.018 | 0.744±0.031 |
| | 0.1 | 0.651±0.026 | 0.672±0.022 | 0.682±0.016 | 0.673±0.022 |
| | 0.2 | 0.614±0.033 | 0.637±0.037 | 0.667±0.024 | 0.641±0.040 |
| | 0.3 | 0.601±0.025 | 0.616±0.027 | 0.641±0.022 | 0.616±0.027 |
| | 0.4 | 0.546±0.048 | 0.571±0.050 | 0.606±0.025 | 0.571±0.049 |
| australian | 0.0 | 0.851±0.002 | 0.855±0.000 | 0.854±0.002 | 0.855±0.000 |
| | 0.1 | 0.853±0.003 | 0.855±0.000 | 0.854±0.004 | 0.855±0.000 |
| | 0.2 | 0.855±0.001 | 0.856±0.001 | 0.856±0.002 | 0.856±0.001 |
| | 0.3 | 0.853±0.005 | 0.854±0.005 | 0.854±0.004 | 0.854±0.005 |
| | 0.4 | 0.840±0.016 | 0.847±0.011 | 0.844±0.013 | 0.847±0.011 |
| twonorm | 0.0 | 0.978±0.000 | 0.979±0.000 | 0.979±0.000 | 0.979±0.000 |
| | 0.1 | 0.978±0.000 | 0.978±0.000 | 0.978±0.000 | 0.978±0.000 |
| | 0.2 | 0.977±0.000 | 0.977±0.000 | 0.977±0.000 | 0.977±0.000 |
| | 0.3 | 0.976±0.000 | 0.977±0.000 | 0.976±0.000 | 0.977±0.000 |
| | 0.4 | 0.971±0.000 | 0.971±0.000 | 0.971±0.000 | 0.971±0.000 |
| Leukemia | 0.0 | 0.986±0.000 | 0.986±0.000 | 0.986±0.000 | 0.927±0.000 |
| | 0.1 | 0.903±0.000 | 0.917±0.000 | 0.917±0.000 | 0.986±0.000 |
| | 0.2 | 0.778±0.000 | 0.792±0.000 | 0.778±0.000 | 0.903±0.000 |
| | 0.3 | 0.778±0.000 | 0.792±0.000 | 0.778±0.000 | 0.792±0.000 |
| | 0.4 | 0.569±0.000 | 0.569±0.000 | 0.569±0.000 | 0.778±0.000 |
| ALLAML | 0.0 | 0.653±0.000 | 0.694±0.000 | 0.667±0.000 | 0.569±0.000 |
| | 0.1 | 0.639±0.000 | 0.667±0.000 | 0.639±0.000 | 0.667±0.000 |
| | 0.2 | 0.569±0.000 | 0.583±0.000 | 0.583±0.000 | 0.639±0.000 |
| | 0.3 | 0.597±0.000 | 0.611±0.000 | 0.597±0.000 | 0.569±0.000 |
| | 0.4 | 0.542±0.000 | 0.556±0.000 | 0.542±0.000 | 0.597±0.000 |

TABLE 6. Training Runtime (Mean ± Std) under Different Label Noise Levels

| Dataset | Flip | CDRESVM | DRESVM | SD-DRESVM |
|---|---|---|---|---|
| fertility | 0.0 | 0.070±0.002 | 0.121±0.015 | 0.046±0.004 |
| | 0.1 | 0.069±0.003 | 0.118±0.018 | 0.047±0.003 |
| | 0.2 | 0.075±0.010 | 0.137±0.042 | 0.057±0.020 |
| | 0.3 | 0.112±0.029 | 0.132±0.023 | 0.129±0.058 |
| | 0.4 | 0.108±0.034 | 0.162±0.041 | 0.103±0.053 |
| parkinsons | 0.0 | 1.073±0.044 | 5.148±0.393 | 0.896±0.329 |
| | 0.1 | 1.074±0.157 | 1.350±0.137 | 0.977±0.423 |
| | 0.2 | 1.025±0.124 | 2.138±1.240 | 0.875±0.307 |
| | 0.3 | 1.042±0.174 | 2.123±0.709 | 0.909±0.321 |
| | 0.4 | 0.949±0.208 | 1.596±0.298 | 0.681±0.458 |
| pima | 0.0 | 2.429±0.471 | 4.862±1.122 | 3.191±1.268 |
| | 0.1 | 1.961±0.335 | 3.498±0.360 | 1.821±1.399 |
| | 0.2 | 1.970±0.325 | 3.175±0.142 | 1.373±0.786 |
| | 0.3 | 1.697±0.180 | 15.300±13.680 | 1.038±0.423 |
| | 0.4 | 1.929±0.310 | 22.834±9.383 | 2.128±3.105 |
| heart | 0.0 | 0.975±0.147 | 9.546±3.324 | 1.099±0.340 |
| | 0.1 | 0.786±0.179 | 8.474±0.741 | 0.818±0.410 |
| | 0.2 | 0.726±0.226 | 8.167±1.503 | 0.635±0.584 |
| | 0.3 | 0.860±0.137 | 8.348±0.637 | 0.918±0.273 |
| | 0.4 | 0.830±0.166 | 8.090±0.564 | 0.760±0.341 |
| haberman | 0.0 | 0.274±0.144 | 3.052±2.855 | 0.466±0.428 |
| | 0.1 | 0.163±0.051 | 0.904±1.807 | 0.269±0.333 |
| | 0.2 | 0.162±0.035 | 0.485±0.383 | 0.202±0.180 |
| | 0.3 | 0.164±0.037 | 0.434±0.279 | 0.162±0.064 |
| | 0.4 | 0.159±0.040 | 0.295±0.113 | 0.140±0.060 |
| mammoOrdinal | 0.0 | 5.734±0.714 | 61.285±17.393 | 3.444±0.755 |
| | 0.1 | 4.727±0.808 | 126.983±39.340 | 2.626±0.795 |
| | 0.2 | 4.979±0.469 | 119.649±17.290 | 3.129±1.797 |
| | 0.3 | 4.301±0.368 | 106.503±15.380 | 2.795±1.059 |
| | 0.4 | 4.382±0.624 | 95.431±26.684 | 3.117±1.672 |
| promoters | 0.0 | 1.866±0.308 | 4.606±0.371 | 1.092±0.509 |
| | 0.1 | 2.128±0.614 | 4.322±0.360 | 1.218±0.591 |
| | 0.2 | 2.102±0.527 | 4.470±0.372 | 1.533±0.545 |
| | 0.3 | 2.001±0.237 | 4.388±0.421 | 1.530±0.433 |
| | 0.4 | 2.151±0.632 | 4.504±0.363 | 1.819±0.515 |
| Spect | 0.0 | 1.169±0.244 | 4.196±3.943 | 0.605±0.346 |
| | 0.1 | 1.559±0.314 | 2.643±0.732 | 1.335±0.697 |
| | 0.2 | 1.360±0.160 | 1.886±0.195 | 1.209±0.509 |
| | 0.3 | 1.635±0.393 | 2.767±1.372 | 1.798±0.912 |
| | 0.4 | 1.495±0.247 | 3.126±1.386 | 1.199±0.596 |
| wpbc | 0.0 | 2.147±0.238 | 3.552±0.436 | 1.202±0.315 |
| | 0.1 | 1.851±0.258 | 2.367±0.363 | 1.151±0.552 |
| | 0.2 | 1.827±0.375 | 2.074±0.318 | 0.960±0.531 |
| | 0.3 | 1.826±0.289 | 2.454±0.769 | 0.849±0.406 |
| | 0.4 | 1.962±0.336 | 2.367±0.856 | 1.024±0.668 |
| sonar | 0.0 | 4.018±0.108 | 4.896±0.062 | 1.466±1.026 |
| | 0.1 | 5.008±1.198 | 8.317±3.184 | 1.996±0.751 |
| | 0.2 | 5.220±1.741 | 11.893±2.014 | 2.553±1.353 |
| | 0.3 | 4.973±1.813 | 10.650±3.164 | 2.120±1.448 |
| | 0.4 | 5.763±1.750 | 11.268±2.234 | 2.742±1.303 |
| vertebra-column | 0.0 | 0.475±0.093 | 1.141±0.255 | 0.398±0.097 |
| | 0.1 | 0.336±0.071 | 0.763±0.300 | 0.303±0.156 |
| | 0.2 | 0.276±0.048 | 0.525±0.236 | 0.253±0.118 |
| | 0.3 | 0.258±0.038 | 0.633±0.335 | 0.199±0.064 |
| | 0.4 | 0.284±0.058 | 0.584±0.250 | 0.309±0.228 |
| hayes-roth | 0.0 | 0.074±0.017 | 0.586±0.243 | 0.049±0.001 |
| | 0.1 | 0.097±0.028 | 0.964±0.518 | 0.090±0.050 |
| | 0.2 | 0.088±0.024 | 0.875±0.557 | 0.089±0.062 |
| | 0.3 | 0.088±0.019 | 1.091±0.574 | 0.075±0.031 |
| | 0.4 | 0.075±0.019 | 0.870±0.630 | 0.065±0.049 |
| australian | 0.0 | 4.283±0.227 | 6.836±0.523 | 2.587±0.226 |
| | 0.1 | 4.458±0.333 | 7.026±1.237 | 6.025±2.463 |
| | 0.2 | 3.801±0.277 | 6.081±0.344 | 2.435±0.774 |
| | 0.3 | 3.449±0.202 | 5.843±0.424 | 1.643±0.841 |
| | 0.4 | 3.668±0.312 | 6.057±0.907 | 2.509±1.257 |
| twonorm | 0.0 | 496.941±0.000 | 834.592±0.000 | 462.808±0.000 |
| | 0.1 | 467.931±0.000 | 1137.272±0.000 | 373.004±0.000 |