Dual-Strategy Optimization of SVM Using Improved Whale Optimization Algorithm for Multi-Class Data Mining

Yang Wang[1*], Zhuliang Cheng[2], Yongyong Tian[2]
E-mail: onezntz@163.com; chengzhuliang@yjglj.beijing.gov.cn; tianyongyong@yjglj.beijing.gov.cn
*Corresponding author
[1]Cybersecurity Management Center of Beijing Municipal Bureau of Economy and Information Technology
Beijing 100101, China
[2]Beijing Emergency Command and Support Center
Beijing 101101, China

Keywords: SVM, WOA, multi-classification, data processing, data mining

Received: February 11, 2025

To address the challenges of high computational complexity and limited generalization ability in traditional support vector machines (SVM) for large-scale and multiclass datasets, this study proposes an SVM optimization model integrating an Improved Whale Optimization Algorithm (IWOA) and dual strategies. Specifically, a nonlinearly decreasing convergence factor and adaptive inertia weight are introduced to enhance the global and local search capabilities of IWOA. A redundant sample removal strategy based on K-means clustering and Fisher projection is designed to filter low-value training data. Furthermore, a distributed parallel optimization strategy with time feedback is introduced to balance node load and improve optimization efficiency. The experimental results on several public datasets (Iris, Wine, CIFAR-10, and Fashion-MNIST) demonstrated that the proposed model outperformed other benchmark algorithms, achieving the highest classification accuracy of 97.8%. In addition, the model achieved a minimum classification error of only 0.15 on the test set, significantly lower than the other three comparison models. Therefore, by incorporating the IWOA and dual strategies, the proposed model effectively enhances the classification accuracy and computational efficiency of SVM in large-scale multiclass data mining tasks.

Povzetek: Za področje večrazrednega podatkovnega rudarjenja je predstavljen DS-IWOA-SVM, ki optimira SCM z izboljšanim kitovim algoritmom: nelinearno padajoči konvergenčni faktor in adaptivna inercijska utež uravnotežita globalno iskanje in lokalno izpopolnjevanje. Redundanco odstrani K-means + Fisherjeva projekcija, porazdeljena paralelna optimizacija z časovnim povratkom pa uravnoteži obremenitve.

1 Introduction

The continuous development of data mining technology has made Support Vector Machine (SVM) one of the most widely used algorithms in multi-class data processing due to its powerful classification performance and efficient processing ability for nonlinear data [1-2]. SVM can effectively distinguish different categories of data by finding the optimal hyperplane and has good generalization ability. However, its performance is highly dependent on the setting of hyperparameters [3]. In recent years, the performance of swarm intelligence optimization algorithms in solving complex optimization problems has been widely recognized. Among them, the Whale Optimization Algorithm (WOA) has gradually become the main algorithm for SVM parameter optimization due to its strong global search ability and fast convergence speed [4]. However, traditional WOA has shortcomings in balancing Global Search and Local Search (GS-LS), which can easily lead to low optimization accuracy or falling into local optima. In addition, redundant samples and unevenly distributed samples in large-scale multi-class datasets can significantly increase the training cost of the model and reduce classification efficiency. Accordingly, an IWOA-SVM model combining Improved WOA and SVM is proposed. The WOA is enhanced by incorporating a nonlinear decreasing convergence factor and an Adaptive Inertia Weight (AIW), effectively improving its GS-LS capabilities. Meanwhile, to address the issue of redundant samples in large-scale datasets, this study innovatively combines K-Means Clustering (K-Means) and Fisher projection to propose a Redundant Data Sample Removal Strategy Based on K-means Clustering and Fisher Projection (KCFJ). To cope with the computational complexity of multi-class datasets, an innovative Distributed Parallel Optimization Strategy Based on Time Feedback (DPOS-TF) has been introduced to improve the efficiency of parameter optimization and model training. This study is expected to provide a new optimization solution for the fields of data mining and multi-class data processing.

This study aims to address the following core issues, including improving the parameter optimization

efficiency and classification accuracy of SVM in multi-class tasks, reducing the negative impact of redundant training samples on model performance, and achieving efficient and scalable model optimization in distributed environments.

2 Related works

SVM is a supervised learning model, mainly used for tasks such as classification, regression, and anomaly detection. Currently, many scholars have utilized this model to efficiently handle tasks such as classification and prediction. Wang H et al. proposed a sparse and robust SVM model aimed at reducing the computational cost of the model by truncating and smoothing the absolute deviation loss function. In addition, the minimum support vector was defined based on the near stationary point, and an alternating direction multiplier method based on the working set was proposed. The final proposed algorithm had higher classification accuracy, fewer support vectors, and faster computation speed on large-scale datasets [5]. Yan X et al. proposed an SVM algorithm based on an analytical exploratory grey wolf optimizer for electrocardiogram emotion recognition tasks. This method achieved average accuracies of 93.37% and 95.93% on the iRealcare and benchmark WESAD datasets, indicating that the algorithm exhibits higher reliability compared to existing supervised learning methods [6]. Samantaray et al. proposed an optimization model called phase space reconstruction SVM firefly algorithm and applied it to monthly traffic prediction tasks. The accuracy of the hybrid SVM firefly algorithm model was improved by extracting information and features from the flow time series through phase space reconstruction. To evaluate the performance of the model, the Nash Sutcliffe coefficient, root mean square error, and Wilmot index were calculated. This model performed better than other application methods in monthly traffic forecasting, with the Wilmot index being the best [7].

Brand L et al. proposed a novel primal-dual multi-instance SVM algorithm aimed at efficiently processing large-scale data. This method was based on a multi-block variant of the Alternating Direction Method of Multipliers (ADMM), which decomposed the original optimization problem into several sub-problems that can be solved in parallel. Each sub-problem only dealt with local variables, thereby avoiding repeated solving of global Quadratic Programming (QP) during the iteration process and significantly improving computational efficiency. Additionally, to enhance performance when handling large feature sets and data batches, extra optimization steps were introduced to circumvent solving least squares problems. The experimental results demonstrated that the method exhibited good scalability and classification performance [8]. Singgalen Y A et al. put forth an emotion classification method built on SVM and synthetic minority oversampling techniques for analyzing audience comments on YouTube travel channels. This method adopted a cross-industry standard data mining process, using SVM combined with synthetic minority oversampling technique to classify sentiment in comment data. This model performed well in sentiment classification, with an accuracy of 84.26% and a precision of 100%. In contrast, SVM models that did not use synthetic minority oversampling techniques performed weaker in distinguishing positive and negative emotions [9]. H. Huang proposed a two-stage feature selection algorithm based on the fusion of information gain and the maximum correlation minimum redundancy algorithm for the redundancy and uneven distribution of text data. They proposed an improved SVM algorithm based on the Fourier mixed kernel function, thereby improving the accuracy and efficiency of text data classification. The experimental results showed that the performance of this improved algorithm on the IMDB dataset was superior to other algorithms. The F1 value has increased by 1% to 3%, and the number of correctly classified texts has increased by 20 to 45 [10]. The summary of the existing literature is shown in Table 1.

To sum up, although the existing SVM optimization methods have achieved certain results in specific tasks, there are still common problems such as low parameter optimization efficiency, difficulty in adapting to large-scale parallel training, and insufficient processing of redundant samples. In contrast, the DS-IWOA-SVM model combining Dual Strategies and IWOA-SVM improves the parameter optimization efficiency by introducing nonlinear decreasing convergence factors and AIWs. This model combines the KCFJ redundant sample elimination strategy for data reduction and adopts the DPOS-TF to enhance computational scalability. This significantly enhances the robustness and adaptability of the model and effectively breaks through the bottlenecks of traditional SVM in terms of efficiency and generalization ability.

Table 1: Comparative summary of existing SVM optimization methods.

Author & Ref.	Method	Main Contribution	Limitation
Wang H et al. [5]	Sparse and robust SVM (truncated smooth loss + ADMM)	Improved classification accuracy and reduced support vectors for large-scale datasets	Complex structure and still reliant on manual parameter tuning
Yan X et al. [6]	EGWO-SVM (Exploratory Grey Wolf Optimizer)	Achieved high accuracy in emotion recognition tasks (95.93% on WESAD)	Prone to local optima; lacks handling of redundant samples
Samantaray S et al. [7]	PSR-FA-SVM (Phase Space Reconstruction + Firefly Algorithm)	Enhanced time series modeling for flow prediction	Limited to time-series applications; lacks generalization
Brand L et al. [8]	Primal-dual multi-instance SVM (blockwise ADMM)	Avoided repeated QP solving; improved efficiency on big data	Inadequate feature selection for high-dimensional redundant data
Singgalen Y A et al. [9]	SMOTE + SVM (oversampling minority classes)	Significantly improved sentiment classification on imbalanced datasets	Prone to overfitting; limited generalizability
H. Huang. [10]	Two-stage feature selection + Fourier hybrid kernel SVM	Improved classification accuracy on text data; reduced feature subsets	Struggles with high-dimensional image data

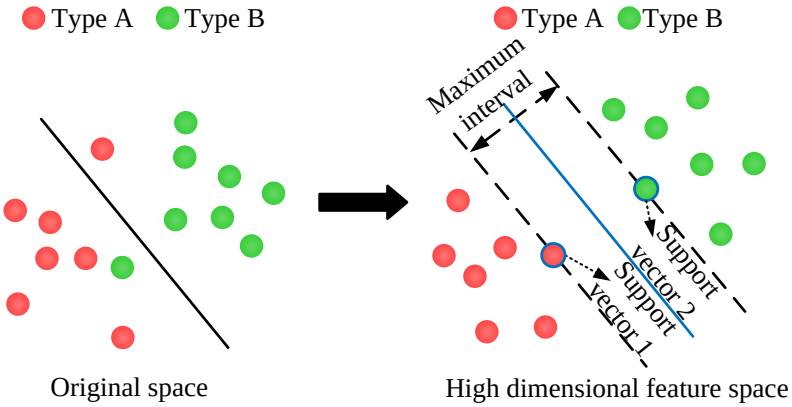


Figure 1: Schematic diagram of SVM optimal hyperplane.

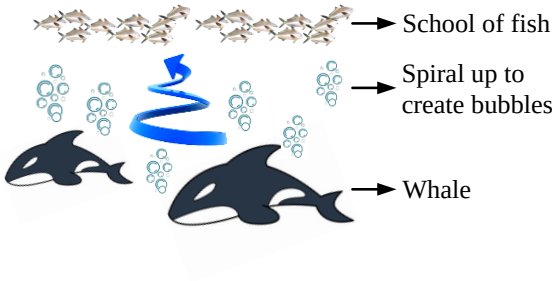


Figure 2: Diagram of whale bubble feeding process.

3 Multi-Class data mining and processing based on improved SVM

To improve the classification performance of SVM in multi-class datasets, this study proposes three different strategies to optimize SVM. Firstly, improvements are made to WOA by proposing the IWOA-SVM algorithm. Secondly, by combining KCFJ and DPOS-TF, the IWOA-SVM algorithm is further improved, and the final multi-class data processing model is constructed.

3.1 Design of improved SVM algorithm based on WOA

In practical multi-class classification scenarios, SVM has shown strong adaptability to complex data characteristics such as high feature dimensionality, sparse distributions, and limited labeled samples [11]. Unlike traditional linear classifiers, SVM leverages kernel functions to map data into higher-dimensional spaces, enabling accurate decision boundaries even under nonlinearly separable conditions [12]. The basic principle of SVM is to effectively distinguish different categories of data in the feature space by finding an optimal hyperplane. The optimal hyperplane classification process of SVM is shown in Figure 1.

In Figure 1, SVM will determine an optimal hyperplane through training data. The core goal of this hyperplane is to maximize the classification interval, even if the distance between the hyperplane and the nearest data point is maximized, thereby enhancing the robustness and generalization ability. However, the performance of SVM is highly dependent on the setting of hyperparameters, and how to effectively optimize these parameters is a key issue in improving SVM performance. Intelligent optimization algorithms have been widely used to address parameter optimization issues due to their efficient global search capabilities. Among them, WOA, as a new type of swarm intelligence algorithm, has received widespread attention for its search efficiency and global convergence [13]. The operation process of traditional WOA mainly simulates the bubble hunting behavior of whales, as shown in Figure 2.

In Figure 2, the whale forms a bubble network by spiraling upwards and rotating around its prey to limit its range of activity. Subsequently, by controlling the shape and density of the bubbles, it gradually approaches the optimal solution. Finally, the whale narrows down the bubble range, focuses on the local area for an in-depth search, and finds a more accurate optimal solution. The entire predation process of WOA is divided into three strategies: trapping prey, spiral updating position, and random search. The calculation for surrounding prey is given by equation (1) [14-15].

$$\begin{cases} D = \left| C \cdot X^* - X \right| \\ X\left(t+1\right) = X^* - A \cdot D \end{cases} \tag{1}$$

In equation (1), X^* and X denote the position vectors of the prey and the current position vector of the whale. D is the current distance from the whale to its prey. A is the linear convergence factor that controls the

search range, with a range of $[-a,a]$. t is time. C is a random weight vector used to enhance the randomness of the search direction, as shown in equation (2).

$$C = 2 \cdot r_1 \tag{2}$$

In equation (2), r_1 is a random number ranging from [0,1]. The formula for spiral updating position is given by equation (3).

$$X(t+1) = D \cdot e^{b \cdot l} \cdot \cos(2\pi l) + X^* \tag{3}$$

In equation (3), $X(t+1)$ represents the updated value of the whale's current position in generation $t+1$. b is the helix shape constant used to control the helix shape. l means a random number of [-1,1] used to generate a random spiral path. $e^{b \cdot l}$ is the exponential decay factor that controls the trend of spiral contraction. When the whale cannot obtain effective information about the current optimal solution, WOA will conduct random search to simulate the whale's exploration behavior in unknown areas, as shown in equation (4).

$$X(t+1) = X_{rand} - A \cdot D \tag{4}$$

In equation (4), X_{rand} is a randomly selected whale position. Although WOA can dynamically switch between GS-LSes, its linear convergence factor A may lead to insufficient global search capability and lower accuracy in later local searches. In addition, traditional WOA has poor ability to balance GS-LS and may fall into local optima in the later stage. Therefore, this study introduces a nonlinear decreasing convergence factor and AIW value for improvement, that is IWOA. In IWOA, A adopts a nonlinear decreasing method to enhance the dynamic adjustment ability of the algorithm's GS-LS, as shown in equation (5).

$$a(t') = a_{\max} \cdot \left(1 - \frac{t'}{T}\right)^n \tag{5}$$

In equation (5), t' and T are the current and maximum iteration counts. $a(t')$ and $a_{\max}$ are the convergence factor and maximum convergence factor at t' iterations. n is a nonlinear decreasing control parameter, and when $n>1$, the convergence factor decreases exponentially. After introducing the convergence factor of nonlinear descent, the trapping formula at this time is given by equation (6).

$$\begin{cases} X(t+1) = X^* - A' \cdot D \\ A' = 2a(t') \cdot r_1 - a(t') \end{cases} \tag{6}$$

In equation (6), A' represents the scalar value of the nonlinearly decreasing convergence factor, which is used to adjust the step size between the current position and the optimal solution. To further balance GS-LS, IWOA also introduces the AIW value ω, whose weight update formula is given by equation (7).

$$\omega(t') = \omega_{\max} - \frac{t'}{T} \cdot (\omega_{\max} - \omega_{\min}) \tag{7}$$

In equation (7), $\omega_{\min}$ and $\omega_{\max}$ are the minimum and maximum of inertia weights, and the search strategy of IWOA is expressed as equation (8).

$$X(t+1) = \omega(t') \cdot (X^* - A' \cdot D) + (1 - \omega(t')) \cdot X_{rand} \tag{8}$$

In equation (8), the introduction of inertia weight can dynamically adjust the degree of dependence on the current optimal solution and random solution, thereby providing better adaptability in different search stages. IWOA is used to optimize SVM penalty parameters and kernel function parameters. The running process of IWOA-SVM algorithm is shown in Figure 3.

In Figure 3, IWOA-SVM first initializes a population consisting of penalty parameters and kernel function parameters and sets other input parameters. Secondly, cross-validation is conducted on the training set to calculate fitness values. The next step is to find the individual with the lowest fitness value in the current population as the prey position and execute three strategies: hunting prey, spiral updating position, and random search to update the population position. Then, the adaptive weight dynamic adjustment algorithm is introduced to enhance the GS-LS capabilities, and the fitness values of each individual are recalculated. Finally, the above steps are repeated until the termination condition for the maximum iterations is met, and then the optimal penalty parameters and kernel function parameters are input to train the SVM model.

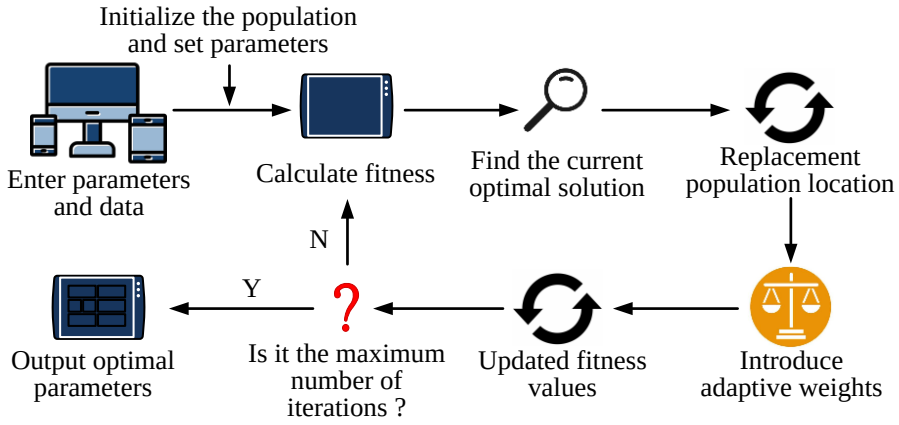


Figure 3: IWOA-SVM flow chart.

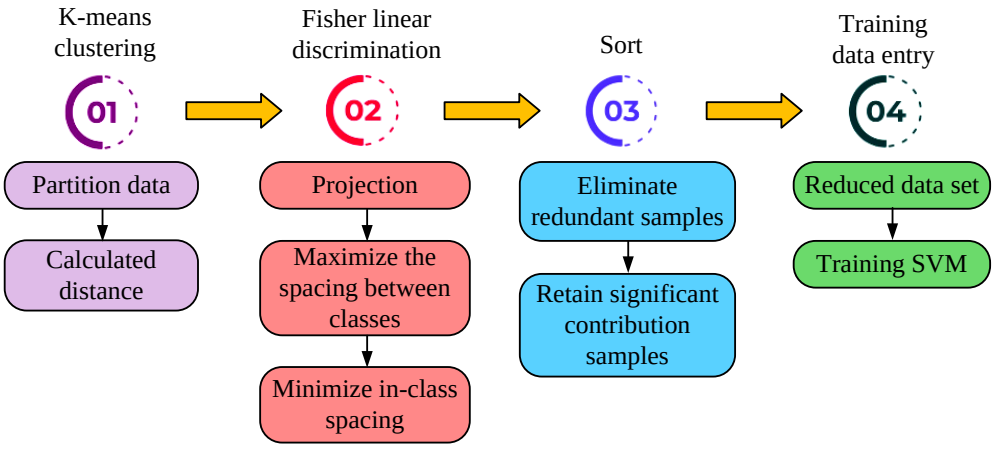


Figure 4: KCFJ policy steps.

3.2 Construction of multi-class data processing model based on improved IWOA-SVM

In practical data mining and multi-classification tasks, large-scale training datasets typically contain a large number of redundant samples. Although these samples contribute relatively little to the construction of classification boundaries, they significantly increase the computational complexity and training time of the model [16-17]. Therefore, to improve the training efficiency of IWOA-SVM, this study further proposes the KCFJ strategy and its implementation steps are shown in Figure 4.

In Figure 4, firstly, K-Means is performed on the training dataset. The data are divided into several category centers to preliminarily partition the distribution structure of the data, and the importance of each data point is measured by calculating the distance between each data point and its category center. Secondly, the data are projected using Fisher's linear discriminant analysis. While maximizing the inter-class interval, the intra-class interval is minimized to reduce data dimensionality and enhance sample discriminability, thereby more clearly identifying key samples that are close to the classification boundary in the projection space. Then, based on the projection results and the distance from the sample to the center of the category, the samples are sorted. Firstly, samples that are far from the category center and have high discriminability in the projection space are retained, while redundant samples that are closer to the category center and have low discriminability in the projection space are removed. This process helps optimize the training dataset and retains only the samples that contribute significantly to the classification boundaries. Finally, the processed simplified dataset is used as the training data input for SVM to reduce computational complexity, improve training efficiency, and ensure that classification performance is not significantly affected. In KCFJ, K-means is first used to cluster the data. The data are divided into K clusters, and the objective function is the minimum sum of the squared distances from each sample to the cluster center, as shown in equation (9) [18-20].

$$J = \sum_{k=1}^{K} \sum_{x_i \in C'_k} \left\| x_i - u_k \right\|^2 \tag{9}$$

In equation (9), x_i is the i-th sample point, C'_k is the sample set of the k-th cluster, and u_k is the centroid of the k-th cluster, i.e., the cluster center. $x_i \in C'_k$ indicates that sample point x_i belongs to C'_k. By iteratively updating the cluster allocation relationship between cluster centers and sample points, it is possible to gradually converge to the local optimal partition. Based on the Euclidean distance from the sample to the center of its cluster, its impact on the classification boundary can be determined. The larger the distance, the higher the importance of the sample. Next, the data are mapped from high to low dimensional space, maximizing inter-class differences while minimizing intra-class differences. The expression for the projection target is shown in equation (10).

$$w = \arg\max_{w} \frac{w^T S_B w}{w^T S_W w} \tag{10}$$

In equation (10), w is the Fisher projection direction vector. S_B and S_W are the inter-class divergence matrix and intra-class divergence matrix, and their specific expressions are shown in equation (11).

$$\begin{cases} S_B = \sum_{k=1}^{K} |C'_k| (u_k - u)(u_k - u)^T \\ S_W = \sum_{k=1}^{K} \sum_{x_i \in C'_k} (x_i - u_k)(x_i - u_k)^T \end{cases} \tag{11}$$

By using equations (10) and (11), the value of w can be calculated, and then the sample can be projected into 1D space to obtain the projected sample coordinates, as shown in equation (12).

$$z_i = w^T x_i \tag{12}$$

In equation (12), z_i is the projection value of sample x_i in the Fisher projection direction. After KCFJ, the comprehensive importance score of the samples is defined as shown in equation (13).

$$S(x_i)=\alpha\cdot\frac{\|x_i-u_k\|}{\max_j\|x_j-u_k\|}+\beta\cdot\frac{|z_i-z_{mean}|}{\max_j|z_j-z_{mean}|} \quad (13)$$

In equation (13), $S(x_i)$ is the importance score value. α and β are weight factors used to balance the impact of K-means and Fisher projections on importance assessment. z_{mean} is the mean of all samples after Fisher projection. j is the index set of all sample points. After calculating the importance scores of all samples, samples with lower scores are more likely to be redundant data. Finally, samples are removed according to the set threshold, as shown in equation (14).

$$D_{reduced}=\{x_i\in D'|S(x_i)>\tau\} \quad (14)$$

In equation (14), $D_{reduced}$ and D' are the simplified dataset and initial training dataset obtained through redundant sample removal strategy processing. τ is the threshold for removing importance scores. To better meet the parallel processing and efficient scheduling requirements of multi-classification tasks, and reduce the load imbalance during training, this study also introduces the DPOS-TF strategy to perfect the computational efficiency and resource utilization of the algorithm. The main operational steps of the DPOS-TF strategy are shown in Figure 5.

In Figure 5, the first step is to divide the multi-class dataset into multiple subsets. Each subset is allocated to different Map nodes for processing based on data volume and category ratio. Secondly, each Map node independently processes the assigned subset of data. Then it enters the Reduce stage, which requires weighting the local fitness based on the sample ratio and calculating the global fitness value to generate the global optimal individual as the basis for population update. Next, the Reduce stage monitors the task processing time of each node, calculates the average task time, and adjusts the task allocation weights based on the time feedback mechanism to balance the load on the Reduce nodes. Finally, the IWOA population is updated based on the global fitness results, and the updated population is reassigned to the Map node for the next round of optimization. The structure of the DS-IWOA-SVM model is depicted in Figure 6.

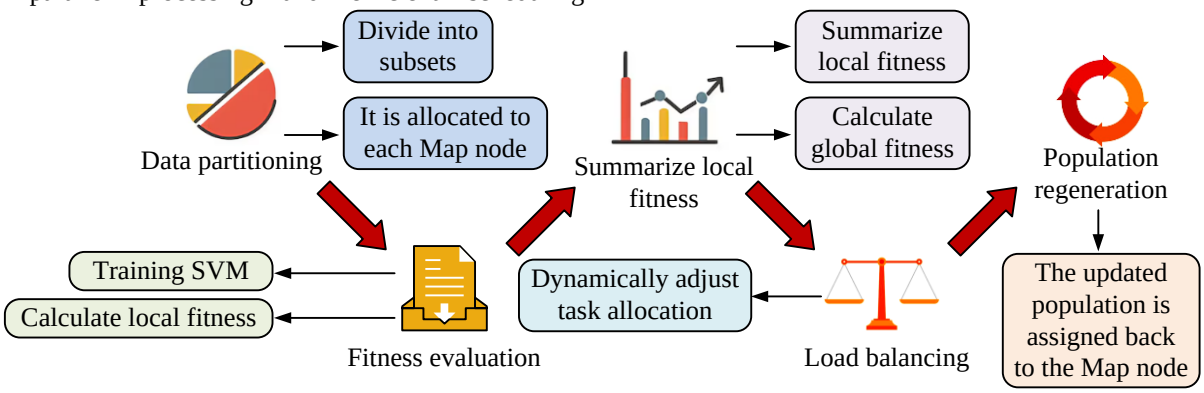


Figure 5: DPOS-TF policy steps.

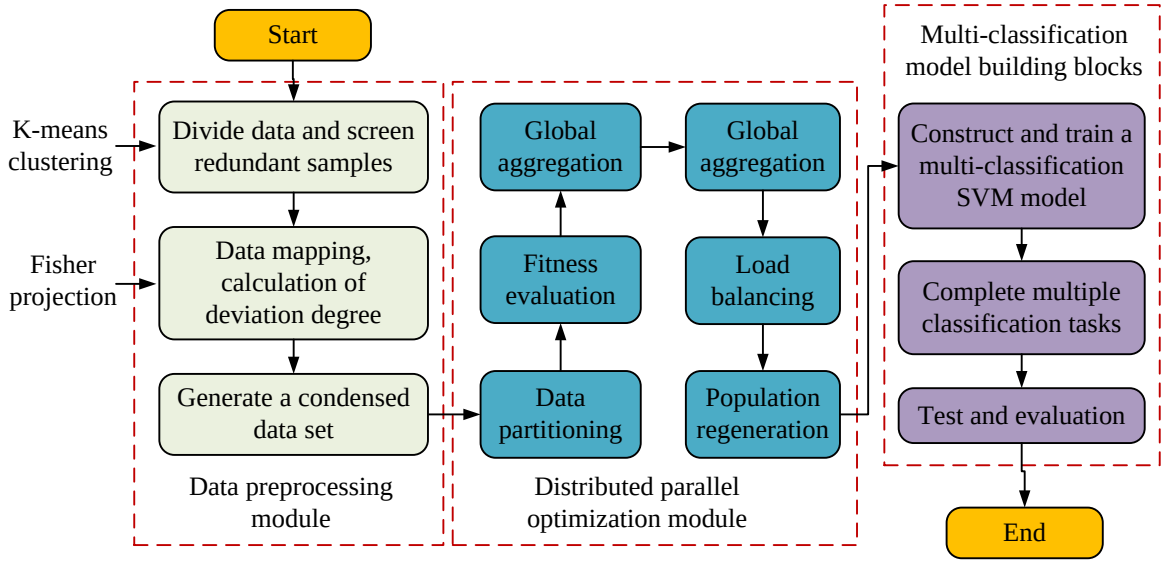


Figure 6: DS-IWOA-SVM structure diagram.

In Figure 6, the DS-IWOA-SVM model has three main modules: data preprocessing (M1), distributed parallel optimization (M2), and multi-classification model construction (M3). In M1, the KCFJ strategy filters and simplifies the original dataset to remove redundant samples and irrelevant features. This strategy combines KCFJ discriminant analysis. By maximizing inter-class spacing and minimizing intra-class spacing, the dimensionality of the dataset can be effectively reduced while retaining samples that significantly contribute to classification performance. In M2, the DPOS-TF strategy will complete the global optimization of the IWOA population in a distributed framework. At M3, the globally optimal parameter combinations outputted after M2 will be used to construct a multi-classification SVM model through the one-to-many strategy. Eventually, DS-IWOA-SVM will complete the efficient training of multi-classification data in the training set and verify the model performance through the test set.

4 Results

To verify the performance of DS-IWOA-SVM in multi-class data mining tasks, multiple experimental evaluations were designed. Firstly, ablation tests were conducted on each module of the algorithm. Secondly, benchmark performance comparisons were conducted between DS-IWOA-SVM and various mainstream-optimized SVM models. Finally, the algorithm was applied to actual multi-class datasets to verify its effectiveness in large-scale data mining tasks.

4.1 Ablation testing and algorithm performance testing

This study establishes a high-performance experimental platform for benchmark testing of the DS-IWOA-SVM algorithm. The operating system is Ubuntu 20.04, equipped with Nvidia GeForce RTX 3090 GPU, and runs on Python 3.9 and PyTorch 1.11 frameworks. The publicly available multi-class dataset used is the Cora dataset. All datasets have been standardized and segmented into training and testing sets in an 8:2 ratio. To ensure the fairness and reliability of the results, all experiments are performed under the same environment and parameter settings. Table 2 shows the specific equipment parameters.

In Table 2, the population size of IWOA is 50, the maximum iterations are 500, the initial learning rate is 0.01, and the weight decay coefficient is 0.001. The study first conducts a hyperparameter sensitivity analysis. Different population sizes (30, 50, 70, 100) and convergence factors (0.3, 0.5, 0.7) are set for the experiments. The results are shown in Table 3.

Table 3 shows that when the population size is 50 and the convergence factor is 0.5, the convergence accuracy is the highest, reaching 96.2%. This configuration performs better than other population sizes. A smaller population size (such as 30) leads to poor model performance due to the smaller search space. However, for larger population sizes (such as 100), although the convergence accuracy is improved, the computational complexity increases, and the improvement in convergence accuracy is not significant. Therefore, the population size is set at 50 and the convergence factor is 0.5.

Table 2: Experimental setup and model parameters.

Category	Parameter/Equipment	Description
Experimental setup	Operating System	Ubuntu 20.04
	GPU	NVIDIA GeForce RTX 3090
	CPU	Intel Core i9-10900K
	Memory	64GB DDR4
	Deep Learning Framework	PyTorch 1.11
	Programming Language	Python 3.9
Model parameters	IWOA Population Size	50
	Maximum Iterations	500
	Initial Learning Rate	0.01
	Weight Decay	0.001
	SVM Kernel Type	Radial Basis Function (Gaussian Kernel)

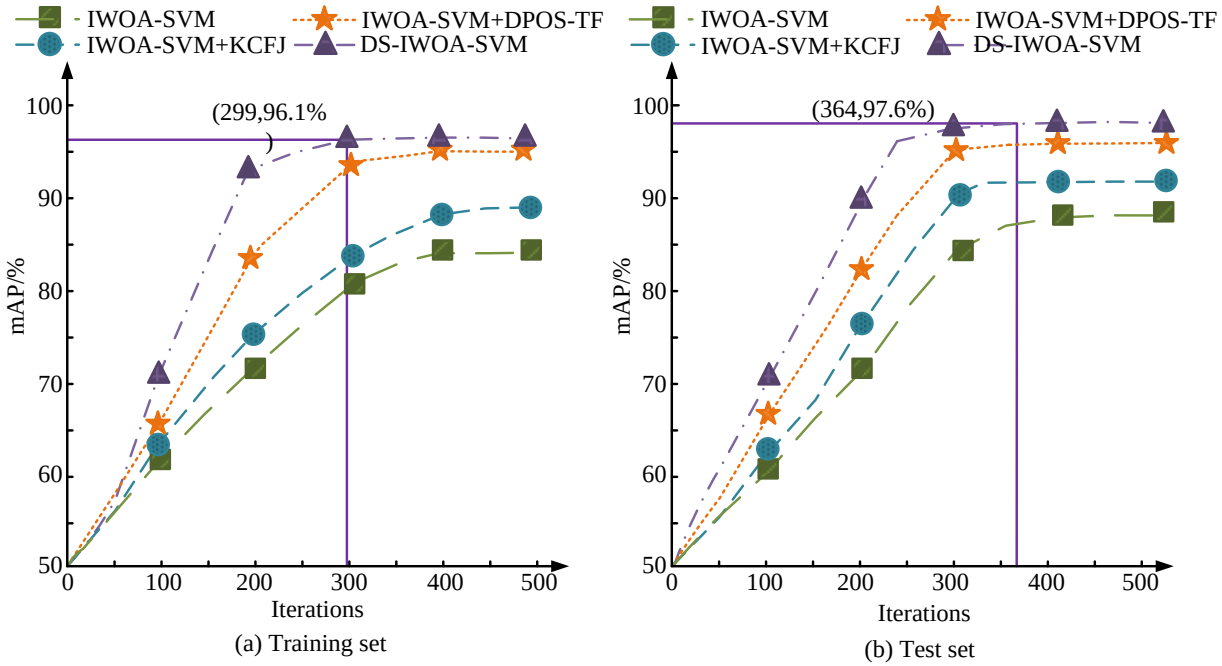
Table 3: Results of hyperparameter sensitivity analysis.

Population Size	Convergence Factor	Convergence Accuracy (%)
30	0.3	92.4
	0.5	93.1
	0.7	91.8
50	0.3	94.5
	0.5	96.2
	0.7	94.1
70	0.3	95.0
	0.5	95.9
	0.7	95.2
100	0.3	95.6
	0.5	96.0
	0.7	95.5

Due to the fact that the DS-IWOA-SVM model is composed of multiple modules, this study tests the classification accuracy of different combinations in DS-IWOA-SVM using mean Average Precision (mAP) as the metric, as shown in Figure 7.

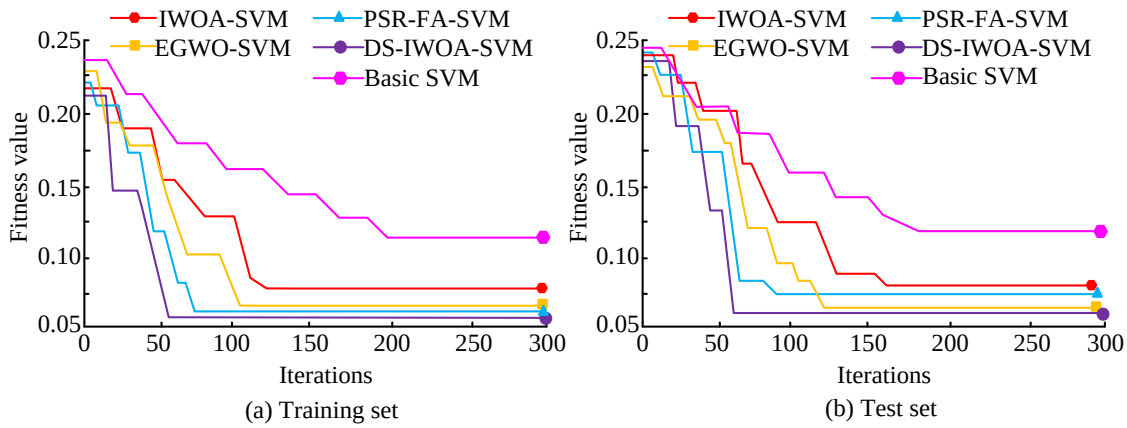
Figure 7 shows the mPA values for different ablation combinations in the training and testing sets. The ablation combinations include the IWOA-SVM model, IWOA-SVM+KCFJ, IWOA-SVM+DPOS-TF, and DS-IWOA-SVM. In Figure 7 (a), IWOA-SVM, as the base model, has the lowest mAP value in the training set, only 83.5%. When combining KCFJ and DPOS-TF strategies simultaneously, the mAP value of DS-IWOA-SVM on the training set reaches as high as 96.1%. In Figure 7 (b), when IWOA-SVM, IWOA-SVM+KCFJ, IWOA-SVM+DPOS-TF, and DS-IWOA-SVM are iterated to a stable state, the mPA values of the four combinations in the test set are 87.4%, 92.5%, 95.8%, and 97.6%. In the ablation test, KCFJ and DPOS-TF can significantly lift the classification accuracy and convergence velocity of the DS-IWOA-SVM, and demonstrate collaborative optimization ability in DS-IWOA-SVM.

This study further selects IWOA-SVM, SVM based on Exploratory Grey Wolf Optimizer (EGWO-SVM), and SVM based on Phase Space Reconstruction and Firefly Algorithm (PSR-FA-SVM) as comparative algorithms. To ensure the fairness of the experimental comparison, the three benchmark models are experimented under the same data division, hardware platform, and evaluation metrics, and fixed hyperparameter values are set. The specific hyperparameter settings are as follows: The population size is 50, the maximum number of iterations is 400, the SVM penalty factor C is 1, and the width of the RBF kernel function γ is 0.01. Among the specific parameters of other algorithms, the exploration factor α of EGWO is set to 0.5, the attraction degree β of PSR-FA is 0.3, and the randomness factor α is 0.5. In addition, the standard SVM is introduced as a baseline for comparison. The iterative curves of the five algorithms on the training and test sets are shown in Figure 8.



(a) Training set (b) Test set

Figure 7: The mPA values for the different ablation combinations.



(a) Training set (b) Test set

Figure 8: Iterative curves of the four algorithms.

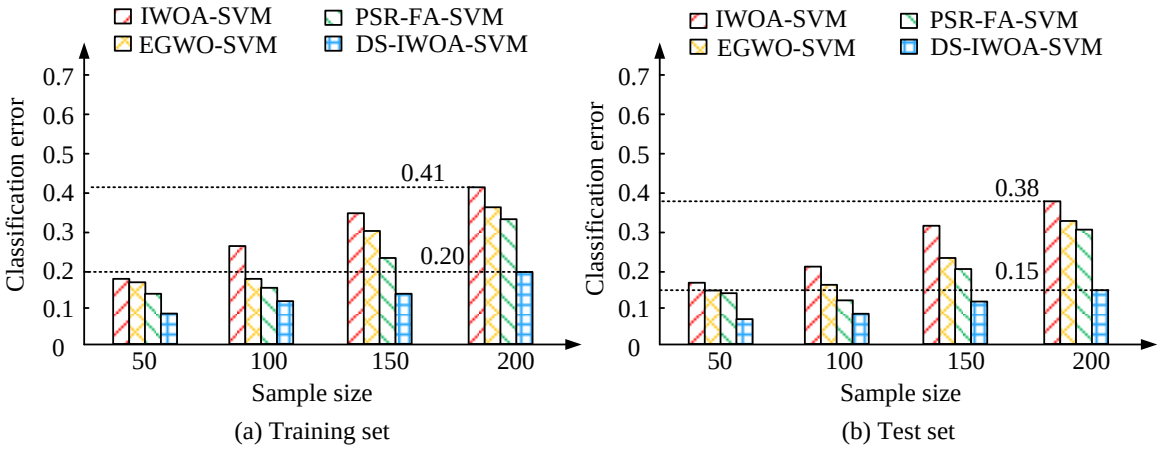


Figure 9: Classification error of the four algorithms.

Table 4: Classification accuracy of the different models.

Data set	Dataset Size (samples)	IWOA-SVM	EGWO-SVM	PSR-FA-SVM	DS-IWOA-SVM
Iris	150	92.3%	94.1%	95.4%	97.8%
Wine	178	85.6%	89.7%	91.2%	94.5%
Digits	1797	88.9%	91.3%	92.7%	96.2%
Fashion-MNIST	60000	78.5%	81.9%	84.3%	88.6%
CIFAR-10	60000	70.2%	74.6%	76.9%	82.1%
20 Newsgroups	18000	83.7%	86.4%	89.1%	92.3%
HAR	6000	88.5%	90.7%	92.5%	95.8%
Glass Identification	214	71.3%	74.9%	77.2%	80.6%
Yeast	1484	64.5%	69.8%	72.4%	78.3%
Emotion	2000	86.2%	89.5%	91.3%	94.7%

As shown in Figure 8(a), DS-IWOA-SVM demonstrates faster convergence compared to the other four algorithms. After 51 iterations, it reaches a stable fitness value of 0.052. In contrast, IWOA-SVM exhibits a slower decline in fitness and only stabilizes after 124 iterations, with a final fitness value of 0.079. The fitness curves of EGWO-SVM and PSR-FA-SVM fall between those of IWOA-SVM and DS-IWOA-SVM. Their convergence speed and final optimization performance are better than IWOA-SVM but inferior to DS-IWOA-SVM. The standard SVM performs the worst, showing the slowest convergence and the highest fitness value. Similarly, in Figure 8(b), DS-IWOA-SVM again reaches a stable state the fastest. Ultimately, the best fitness values achieved by IWOA-SVM, EGWO-SVM, PSR-FA-SVM, and DS-IWOA-SVM in the stable state are 0.081, 0.069, 0.075, and 0.058, respectively. The standard SVM reaches a fitness value of 0.126. Figure 9 compares the classification errors of four algorithms in two datasets.

Figures 9 (a) and (b) show the classification errors of four algorithms in two datasets, with DS-IWOA-SVM having the lowest classification error across all sample sizes. In Figure 9 (a), DS-IWOA-SVM exhibits strong learning and generalization abilities. When the sample size reaches 200, the classification error of DS-IWOA-SVM increases to 0.20, but it is still significantly better than the classification error of the IWOA-SVM model at 0.41. In Figure 9 (b), as the sample size rises to 200, the classification error of DS-IWOA-SVM increases to 0.15, while the classification errors of IWOA-SVM, EGWO-SVM, and PSR-FA-SVM are as high as 0.38, 0.35, and 0.32.

4.2 Application effect analysis

To validate the classification performance of the DS-IWOA-SVM in practical multi-classification tasks, 10 different types of publicly available multi-classification datasets are selected as the research objects. These datasets cover a variety of data types such as images, text, sensor signals, etc., with categories ranging from 3 to 10 and data scales ranging from thousands to hundreds of thousands, making them widely representative. Table 4 compares the classification accuracy of IWOA-SVM, EGWO-SVM, PSR-FA-SVM, and DS-IWOA-SVM on 10 datasets.

In Table 4, DS-IWOA-SVM achieves the highest classification accuracy of 97.8% on all datasets. The classification performance of PSR-FA-SVM is inferior to DS-IWOA-SVM, with a classification accuracy of up to 95.4%. The accuracy of this model is similar on most datasets, but there is a certain gap on high-complexity datasets such as CIFAR-10 and Yeast. The classification ability of EGWS-SVM and IWOA-SVM is poor, with the highest classification accuracies of 94.1% and 92.3%. It is worth noting that CIFAR-10, as a typical high-dimensional image dataset, has a complex spatial structure and strong feature redundancy. SVM models have difficulty giving full play to their advantages on such data. However, datasets such as Wine and Fashion-MNIST have lower feature dimensions and clear category boundaries, which can better reflect the performance improvement of DS-IWOA-SVM in sample compression and parameter optimization. This indicates that the characteristics of different datasets affect the

improvement extent of the model to a certain extent. To further verify whether the performance improvement of DS-IWOA-SVM on multiple datasets is statistically significant, paired t-test and Wilcoxon signed-rank test are conducted on its classification accuracy results with IWOA-SVM, EGWO-SVM and PCR-FA-SVM. The results show that on 10 datasets, the accuracy improvement of DS-IWOA-SVM compared with other models is significant at the 95% confidence level (p<0.05), indicating that the performance improvement of this model is not accidental.

The classification performance of four models on ten datasets is demonstrated using clustering in the dimensional space, as shown in Figure 10.

In Figure 10 (a), the classification results of IWOA-SVM are quite chaotic, with many overlapping areas between sample points of different categories. This indicates that IWOA-SVM has insufficient boundary learning ability in complex multi-classification tasks. In Figures 10 (b) and (c), compared to IWOA-SVM, EGWS-SVM, and PSR-FA-SVM have improved classification results, with some boundaries between categories becoming clearer. However, there is still a certain degree of confusion in the classification of a few samples, indicating that the two models still have shortcomings in dealing with highly similar categories. In Figure 10 (d), the classification results of DS-IWOA-SVM are the most ideal, with a clear distribution and good separation of sample points in the dimensional space for each category. This indicates that DS-IWOA-SVM has the best boundary capture ability and robustness in complex multi-classification tasks. To further verify the computational efficiency and scalability of the DS-IWOA-SVM model in practical applications, a training time and complexity comparison experiment is conducted across multiple datasets. Four representative datasets (Cora, Wine, Fashion-MNIST, and CIFAR-10) are selected for training time evaluation. Each experiment is independently run five times, and the total training time is recorded and averaged as the final result. The results are shown in Table 5.

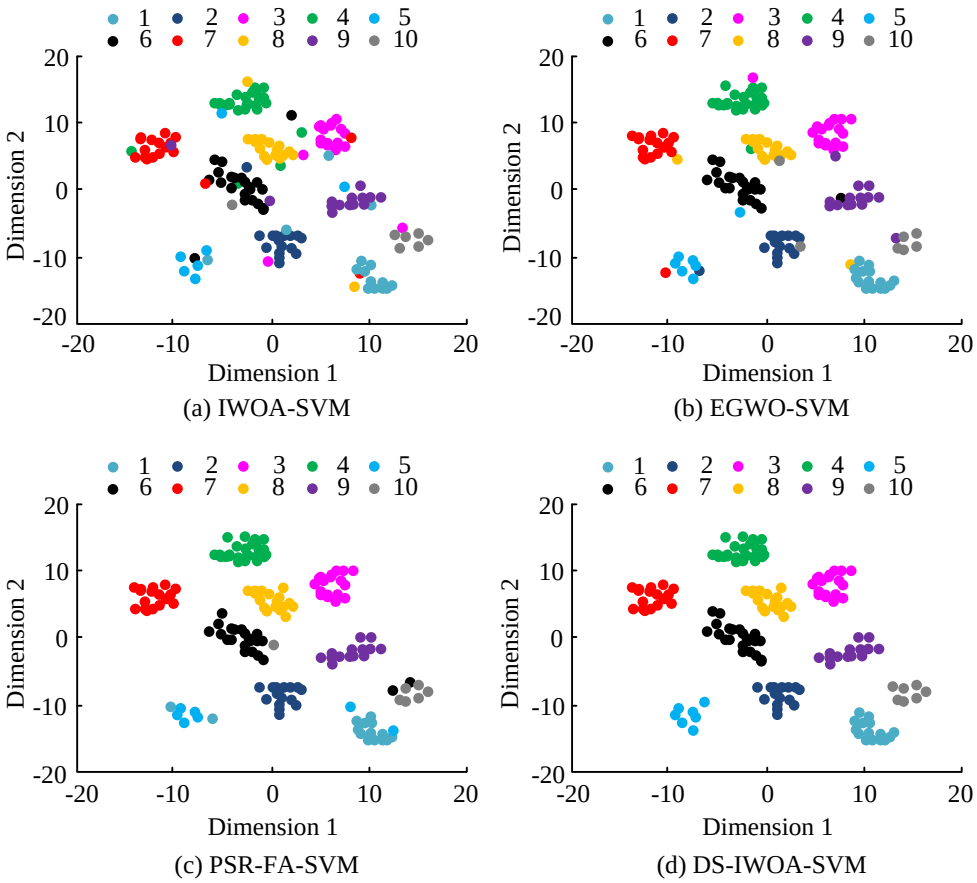


Figure 10: Classification results of the four models in the dimensional space.

Table 5: Average training time of each model on different datasets.

Model	Cora	Wine	Fashion-MNIST	CIFAR-10	Average Time
EGWO-SVM	12.5 s	7.8 s	21.3 s	34.6 s	19.1 s
PSR-FA-SVM	13.1 s	8.8 s	23.0 s	35.4 s	19.9 s
IWOA-SVM	14.4 s	9.0 s	24.5 s	37.2 s	21.3 s
DS-IWOA-SVM	11.6 s	6.9 s	19.2 s	31.8 s	17.4 s

As shown in Table 5, DS-IWOA-SVM achieves lower average training time across all four datasets compared to the other three baseline models. In low dimensional small sample tasks such as Wine, the model only requires 6.9 seconds of training time, which is 23.3% less than IWOA-SVM. For mid-dimensional to high-dimensional datasets such as Fashion-MNIST and CIFAR-10, the training times are 19.2 seconds and 31.8 seconds, respectively, both shorter than those of EGWO-SVM and PSR-FA-SVM. These results demonstrate the practical effectiveness of the KCFJ strategy in reducing the training dataset size, thereby alleviating the computational burden introduced by the SVM component.

5 Discussion

The proposed DS-IWOA-SVM model demonstrates superior classification performance over existing optimization methods across multiple public multiclass datasets. The experimental results show that compared with benchmark models such as EGWO SVM and PSR-FA-SVM, DS-IWOA-SVM has higher accuracy on complex datasets such as CIFAR-10, Wine, Fashion MNIST, etc. This indicates that DS-IWOA-SVM has stronger boundary learning and generalization abilities in multi-class distribution and complex feature scenes. The improvement in classification performance is primarily attributed to the introduction of a nonlinearly decreasing convergence factor and AIW in the IWOA. In the early stages of optimization, these mechanisms enhance the population's global exploration ability and help avoid local optima. In the later stages, the AIW facilitates fine-grained local search, thereby improving the stability and precision of parameter optimization. In terms of computational cost, although the improved IWOA introduces additional computational overhead, the KCFJ strategy effectively removes redundant samples, significantly reducing the training set size for the SVM. Overall, the model achieves better performance in both convergence iterations and training time compared to IWOA-SVM without KCFJ, confirming the overall efficiency gain brought by sample compression.

Nevertheless, the DS-IWOA-SVM model still has certain limitations. In extremely high-dimensional feature spaces (such as text or genomic data), the model may face challenges such as insufficient feature selection mechanisms and slower convergence in parameter optimization. In addition, the current DPOS-TF strategy relies on a predefined node partitioning scheme, which may limit its adaptability to imbalanced data distributions. Moreover, the impact of the KCFJ strategy on data distribution during redundant sample removal has not yet been visualized, restricting the intuitive understanding of structural adjustments to the training data. Future research can further enhance the model performance in the following ways: On the one hand, integration with other feature selection algorithms or deep learning models can be explored to improve adaptability on high-dimensional data; On the other hand, the DPOS-TF strategy can be optimized. Through adaptive node partitioning and dynamic load balancing techniques, the performance of the model on imbalanced data can be enhanced. In addition, clustering algorithms or incremental learning methods can be combined to further improve the algorithm's efficiency on ultra-high dimensional datasets. Meanwhile, for the KCFJ strategy, future research can introduce a visual analysis of data distribution to evaluate the effect of eliminating redundant samples and optimize the sample selection strategy to ensure more efficient data processing and classification performance.

6 Conclusion

To improve the processing efficiency and classification performance of multi-class datasets, this study combined IWOA, KCFJ, and DPOS-TF to construct a novel dual strategy improved SVM model, namely DS-IWOA-SVM. In the ablation test, DS-IWOA-SVM achieved the highest mAP score of 97.6%. Compared with IWOA-SVM, EGWO-SVM, and PSR-FA-SVM, this model could iterate to a stable state faster and maintain optimal fitness values of 0.052 and 0.058 in both the training and testing sets. Additionally, the classification error of DS-IWOA-SVM was obviously lower than that of the comparison model, with a minimum of only 0.15. In practical applications, DS-IWOA-SVM achieved the highest classification accuracy of 97.8% in 10 multi-class datasets, significantly higher than IWOA-SVM's 92.3%, EGWO SVM's 94.1%, and PSR-FA-SVM's 95.4%. In addition, through visual analysis of dimensional spatial clustering, DS-IWOA-SVM could better capture boundary features between categories and achieve a clear separation of samples from different categories. Overall, DS-IWOA-SVM effectively addresses the efficiency and accuracy issues of traditional SVM in large-scale datasets and multi-classification tasks by combining multiple optimization strategies. It has demonstrated excellent classification performance and computational efficiency in practical tasks. However, this model still has certain computational complexity limitations when dealing with ultra-high dimensional data. Future research can optimize the structure and distributed framework of algorithms, reduce computational costs, and explore their adaptability and generalization ability in more practical scenarios.

References

[1] Hassan Tanveer, Muhammad Ali Adam, Muzammil Ahmad Khan, Muhammad Awais Ali, and Abdul Shakoor. Analyzing the performance and efficiency of machine learning algorithms, such as deep learning, decision trees, or support vector machines, on various datasets and applications. The Asian Bulletin of Big Data Management, 3(2):126-136, 2023. https://doi.org/10.62019/abbdm.v3i2.83

[2] Maryam Cheraghy, Meysam Soltanpour, Hemn Barzan Abdalla, and Amir Hosein Oveis. SVM-based factor graph design for max-SR problem of SCMA networks. IEEE Communications Letters, 28(4):877-881, 2024. https://doi.org10.1109/LCOMM.2024.3366426

[3] Salsabila Rabbani, Dea Safitri, Nadila Rahmadhani, and Al Amin Fadillah Sani. Perbandingan evaluasi kernel SVM untuk klasifikasi sentimen dalam analisis kenaikan harga BBM: Comparative evaluation of SVM kernels for sentiment classification in fuel price increase analysis. Indonesian Journal of Machine Learning and Computer Science, 3(2):153-160, 2023. https://doi.org/10.57152/malcom.v3i2.897

[4] Zhi Quan, and Luoxi Pu. An improved accurate classification method for online education resources based on support vector machine (SVM): Algorithm and experiment. Education and Information Technologies, 28(7):8097-8111, 2023. https://doi.org/10.1007/s10639-022-11514-6

[5] Huajun Wang, and Yuanhai Shao. Sparse and robust SVM classifier for large scale classification. Applied Intelligence, 53(16):19647-19671, 2023. https://doi.org/10.1007/s10489-023-04511-w

[6] Xucun Yan, Zihuai Lin, Zhiyun Lin, and Branka Vucetic. A novel exploitative and explorative GWO-SVM algorithm for smart emotion recognition. IEEE Internet of Things Journal, 10(11):9999-10011, 2023. https://doi.org/10.1109/JIOT.2023.3235356

[7] Sandeep Samantaray, and Abinash Sahoo. Prediction of flow discharge in Mahanadi River Basin, India, based on novel hybrid SVM approaches. Environment, Development and Sustainability, 26(7):18699-18723, 2024. https://doi.org/10.1007/s10668-023-03412-9

[8] Lodewijk Brand, Hoon Seo, Lauren Zoe Baker, Carla Ellefsen, Jackson Sargent, and Hua Wang. A linear primal-dual multi-instance SVM for big data classifications. Knowledge and Information Systems, 66(1):307-338, 2024. https://doi.org/10.1007/s10115-023-01961-z

[9] Yerik Afrianto Singgalen. Performance evaluation of SVM algorithm in sentiment classification: A visual journey of wonderful indonesia content. KLIK: Kajian Ilmiah Informatika dan Komputer, 4(4):2078-2087, 2024. https://doi.org/10.30865/klik.v4i4.1709

[10] Hua Huang. Feature extraction and classification of text data by combining two-stage feature selection algorithm and improved machine learning algorithm. Informatica, 48(8):137-150, 2024. https://doi.org/10.31449/inf.v48i8.5763

[11] Issa Mohammed Saeed Ali, and D. Hariprasad. Hyper-heuristic salp swarm optimization of multi-kernel support vector machines for big data classification. International Journal of Information Technology, 15(2):651-663, 2023. https://doi.org/10.1007/s41870-022-01141-2

[12] Xu Wang, Xiaobo Long, Guangwei Li, Jing Li, and Yuweijia Zhao. Application method and least squares support vector machine optimization for leak fault diagnosis in heat pipe networks. Informatica, 49(14):199-212, 2025. https://doi.org/10.31449/inf.v49i16.6990

[13] Leyla Zeynallı-Hüseynzade. Evaluation of machine learning algorithms' performance in digital transformations: a comparative analysis. Scientific Collection, 41(185):510-518, 2024. https://doi.org/10.51582/interconf.19-20.01.2024.062

[14] Muhammad Junaid, Sajid Ali, Isma Farah Siddiqui, Choonsung Nam, Nawab Muhammad Faseeh Qureshi, Jaehyoun Kim, and Dong Ryeol Shin. Performance evaluation of data-driven intelligent algorithms for big data ecosystem. Wireless Personal Communications, 126(3):2403-2423, 2022. https://doi.org/10.1007/s11277-021-09362-7

[15] Akshit Kurani, Pavan Doshi, Aarya Vakharia, and Manan Shah. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 10(1):183-208, 2023. https://doi.org/10.1007/s40745-021-00344-x

[16] Vivine Nurcahyawati, and Zuriani Mustaffa. Improving sentiment reviews classification performance using support vector machine-fuzzy matching algorithm. Bulletin of Electrical Engineering and Informatics, 12(3):1817-1824, 2023. https://doi.org/10.11591/eei.v12i3.4830

[17] D. Oktavia, Y. R. Ramadahan, and M. Minarto. Analisis sentimen terhadap penerapan sistem e-tilang pada media sosial twitter menggunakan algoritma Support Vector Machine (SVM). KLIK: Kajian Ilmiah Informatika dan Komputer, 4(1):407-417, 2023. https://doi.org/10.30865/klik.v4i1.1040

[18] Ashish Namdeo, and Dileep Singh. WITHDRAWN: Challenges in evolutionary algorithm to find optimal parameters of SVM: A review. Materials Today: Proceedings, 16(4):298-326, 2024. https://doi.org/10.1016/j.matpr.2021.03.288

[19] Wenqin Zhao, Yaqiong Lv, Jialun Liu, Carman K. M. Lee, and Lei Tu. Early fault diagnosis based on reinforcement learning optimized-SVM model with vibration-monitored signals. Quality Engineering, 35(4):696-711, 2023. https://doi.org/10.1080/08982112.2023.2193255

[20] Mehdi Gheisari, Mehdi Gheisari, Yang Liu, Peyman Saedi, Arif Raza, Ahmad Jalili, Hamidreza Rokhsati, and Rashid Amin. Data mining techniques for web mining: A survey. Artificial Intelligence and Applications, 1(1):3-10, 2022. https://doi.org/10.47852/bonviewAIA2202290