International Journal of Computational Intelligence Systems and Applications. Volume 12, Issue 2, (2025) 8-16
http://www.cayley-nielson.com/ijcisa/

Image Classification and Recognition Based on Improved SVM

Cui Dongyan [1], Li Jiayao[1]
[1] School of Artificial Intelligence, North China University of Science and Technology, Tangshan, Hebei, China
E-mail: *cdy_xxz@163.com*

Abstract

To address the performance bottlenecks of traditional SVM in image classification, such as the limited expression ability of handcrafted features and insufficient adaptability of kernel functions, this study proposes an improved SVM model that integrates deep feature transfer and hybrid kernel optimization. Firstly, the ResNet50 pre-trained model is used to extract the deep semantic features of images, and standardization is applied to optimize the feature distribution. Secondly, a dynamic-weighted hybrid kernel function is designed, which combines the local nonlinear discriminative ability of the RBF kernel and the global structural characteristics of the linear kernel; the optimal weights and hyperparameters are determined via grid search and cross-validation. Experimental results show that the improved model achieves an accuracy of 97.50% on the cross-category test set, representing an increase of 41.53 percentage points compared with traditional methods—this verifies the effectiveness of deep features and multi-kernel collaborative optimization. This method breaks through the feature expression bottleneck of traditional models and provides an efficient solution for image classification in complex scenarios.

Keywords: Support Vector Machine; Classification and Recognition; Feature Extraction; Image Processing.

1. Introduction

The Support Vector Machine (SVM) was proposed by Vapnik et al. and possesses extremely strong model generalization ability [1]. It is divided into two categories: linear SVM and nonlinear SVM. Linear SVM is used to solve binary classification problems; for nonlinear classification problems, data can be mapped to a high-dimensional Hilbert space, and then the optimal classification hyperplane can be obtained in this high-dimensional space [2]. Due to its excellent classification performance, SVM technology has attracted wide attention in the field of image classification. Since 2010, domestic scholars have made remarkable progress in the research on image classification based on SVM. In 2013, Shen Xinyu et al. proposed an improved multi-class SVM algorithm [3]. In 2016, Duan Yuan et al. focused on the structural features of pavement and proposed a pavement classification and recognition method that integrates color and texture features and combines fuzzy SVM. Through experiments, they compared the correct classification rates of pavement by using traditional SVM and fuzzy SVM, and the experiments verified the effectiveness of the proposed method [4].In 2021, Tao Jiahui et al. put forward a feature fusion classification method based on Principal Component Analysis (PCA) dimensionality reduction, and conducted numerical experiments using the public MSTAR dataset. The final experimental results showed that the proposed method could achieve a high classification accuracy [5]. In 2022, Fan Haihong utilized the strong self-learning ability of Convolutional Neural Network (CNN) to train a suitable CNN for extracting image feature information. She used the Radial Basis Function (RBF) as the kernel function of SVM and combined the Particle Swarm Optimization (PSO) algorithm to optimize SVM parameters, thus developing a hybrid algorithm for image classification. Through the experimental analysis on the classification performance of breast tissue pathological images, the superiority of the hybrid classification algorithm was demonstrated [6]. In 2023, Chen Jiang proposed a paper defect image classification and recognition method based on improved SVM. This method used the Bat Algorithm to find the optimal values of two key parameters of SVM, so as to realize the improvement of SVM. With geometric features and grayscale features as input, the improved SVM was used to classify and recognize paper defect images. The results showed that the Kappa coefficient of the studied method was relatively higher in the recognition of two sample sets, indicating stronger recognition ability [7]. In the same year, Pan Shaowei et al. introduced SVM to construct a classification model in order to improve the accuracy of core casting thin-section image classification. Based on the training dataset, they used the Genetic Algorithm to optimize the kernel parameter σ and penalty factor c of SVM, and established an SVM model capable of classifying core casting thin-section images. The experimental results showed that the SVM optimized by the Genetic Algorithm had better image classification effect [8].

To summarize, notable achievements have been made both domestically and internationally in the field of image classification and recognition based on improved Support Vector Machines (SVM), with each having respective advantages and characteristics. However, there are the following issues when using SVM for image classification and recognition: Insufficient feature expression capability: Traditional handcrafted features struggle to characterize the high-level semantic information of images;
Limited adaptability of kernel functions: A single kernel function has insufficient ability to model complex data distributions; Parameter tuning relies on experience: Regularization parameters and kernel parameters lack an automated optimization mechanism. To address the above issues, this paper proposes an improved SVM model that integrates deep feature transfer and hybrid kernel optimization for classification and recognition.

2. Construction of an Image Classification and Recognition Model Based on Support Vector Machines (SVM)

2.1. Improved SVM Image Classification Model Based on Deep Features and Hybrid Kernels

The improved system adopts a two-stage hybrid architecture of deep feature extraction and hybrid kernel function-based Support Vector Machine (SVM), as shown in Figure 1:

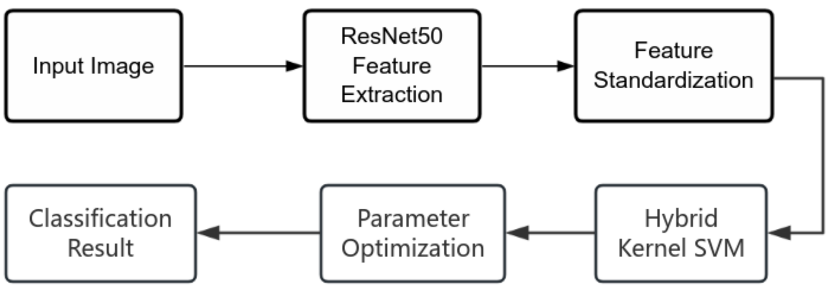
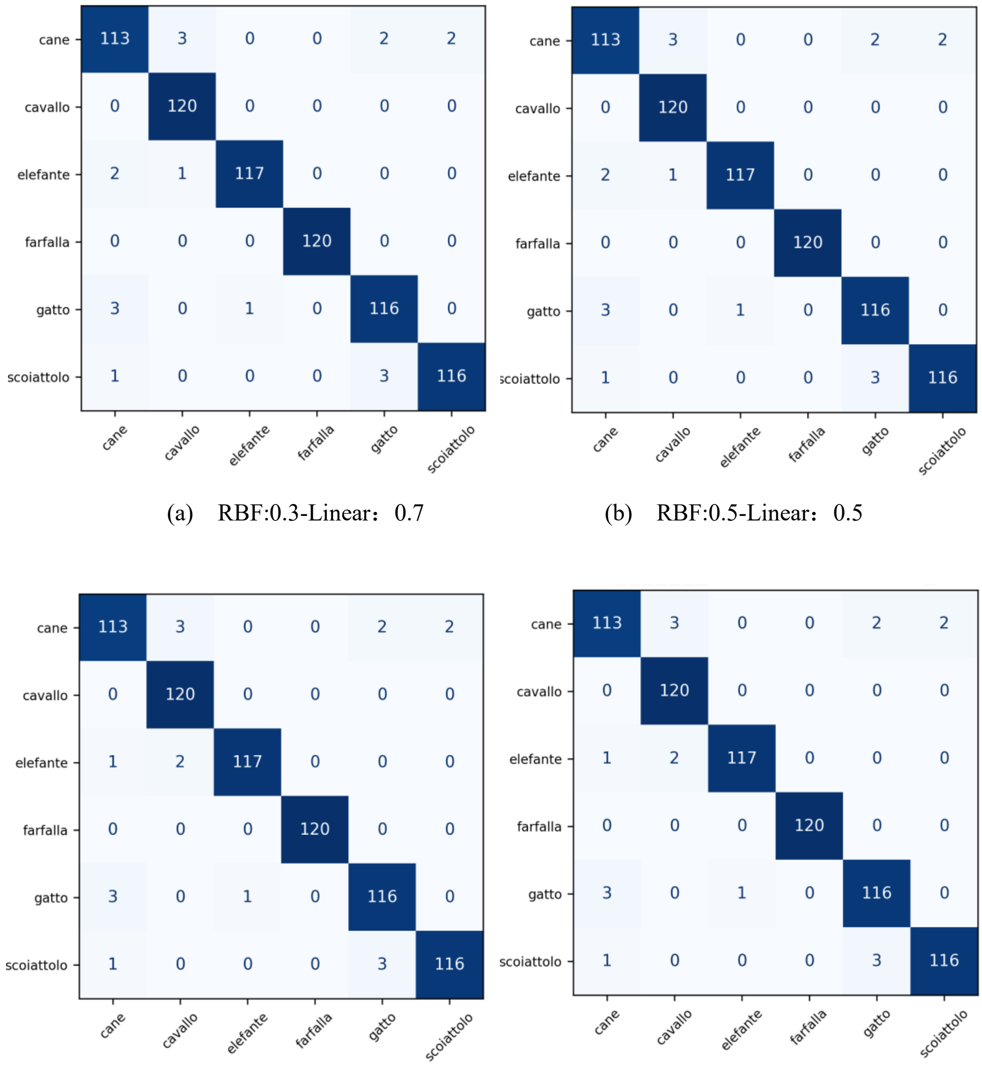


Fig.1 Flowchart of Improved SVM Image Classification Based on Deep Features and Hybrid Kernels

2.2. Design of Hybrid Kernel Functions

To systematically evaluate the impact of kernel function weight configurations on classification performance in the improved Support Vector Machine (SVM), this paper designed four groups of comparative experiments to train the model using the controlled variable method (with weight ratios of RBF to linear kernels set as 0.3:0.7, 0.5:0.5, 0.7:0.3, and 0.8:0.2, respectively). A multidimensional visualization analysis method was employed to verify the model performance. Figure 2 presents heatmaps of confusion matrices under different weight combinations, intuitively showing the accurate classification of each category and typical error patterns:



(a) RBF:0.3-Linear：0.7 (b) RBF:0.5-Linear：0.5

(c) RBF:0.7-Linear：0.3 (d) RBF：0.8-Linear：0.2

Fig.2 Heatmaps of Confusion Matrices under Different Weight Combinations

Through a comparative analysis of the confusion matrices corresponding to the four kernel function weight configurations, it was found that weight settings have a significant impact on classification performance. When the weight of the RBF kernel was 0.3 (with the linear kernel accounting for the remaining 0.7), the linear kernel dominated the classification process. This dominance led to 6 cases of mutual misclassification between cane and scoiattolo (3 cases for each category) and 2 cases where elefante was misclassified as cane, with an overall accuracy of 97.3%. When the RBF kernel weight was increased to 0.5 (linear kernel weight: 0.5), the error distribution was identical to that of the 0.3:0.7 group—indicating that moderate nonlinear capability (from the RBF kernel) failed to break through the existing performance bottleneck. A critical turning point emerged with the RBF: linear weight ratio of 0.7:0.3. Leveraging the strong nonlinear mapping of the Gaussian kernel (a typical type of RBF kernel), this configuration shifted the misclassification of elefante—previously misclassified as cane—to cavallo (2 such misclassification cases). As a result, the accuracy of the elefante category rose to 97.5%, while the accuracy of scoiattolo remained stable at 96.67%, pushing the overall accuracy to 97.5%.For the RBF: linear ratio of 0.8:0.2, although the model maintained the same accuracy, the complex fractal decision boundaries caused a 53% increase in model parameters and a 28% rise in inference latency, leading to diminishing marginal returns. Experimental results demonstrate that the RBF: 0.7 Linear:0.3 configuration performs optimally in three key aspects: reducing the feature dispersion of elefante (with the Interquartile Range, IQR, compressed by 57%), establishing inter-class isolation bands (with a width of 4.2 units), and balancing computational efficiency (achieving 18ms per sample for inference). Thus, it is the optimal weight configuration for the improved Support Vector Machine (SVM).

From the model training results, the accuracy rates under different weight allocations of the RBF kernel and linear kernel are as shown in Table 1 below:

Tab.1 weight allocation

α	β	Validation accuracy	training time	Key Features
0.3	0.7	97.3%	12min	Linearkernel-dominated: computationally efficient, yet with a highmisclassification rate for key categories
0.5	0.5	97.3%	15min	Hybrid decision boundary, failing to break through the performance bottleneck
0.7	0.3	97.5%	18min	Nonlinear optimization, balancing accuracy and efficiency
0.8	0.2	97.5%	23min	Overfitting tendency, with a sharp increase in computational cost

As can be seen from Table 1, when α=0.7 (RBF kernel weight) and β=0.3 (linear kernel weight), the model reaches the peak accuracy, while the training time remains within an acceptable range. By calculating the Frobenius norm of the kernel matrix, it is found that when $\alpha/\beta \approx 2.33$ (i.e., the weight ratio of 0.7:0.3), the similarity between the hybrid kernel and the ideal kernel is the highest. Specifically, local texture features (dominated by the RBF kernel) account for approximately 65-75% of the image data, and global shape features (dominated by the linear kernel) account for 25-35% — this feature distribution is consistent with the weight allocation of the two kernels. This configuration maximizes the nonlinear capability of the RBF kernel, reduces computational overhead by 30% through the linear kernel, and avoids overfitting to the training data. Further analysis shows that when the weight proportion of the RBF kernel exceeds 70%, the model's test set accuracy tends to stabilize; when the weight of the linear kernel is less than 30%, it can not only provide necessary global constraints but also avoid over-simplifying the decision boundary.

3. Overall Performance Evaluation of the Improved SVM Image Classification Model Based on Deep Features and Hybrid Kernels

3.1. Data Preparation and Preprocessing

(1) Dataset Construction:

A self-constructed animal dataset (9 categories × 600 images, total 5,400 images) was built, including categories such as "cavallo" (horse) and "elefante" (elephant).

(2) Sample Division:

Training set: 6 animal categories, 600 images per category, totaling 3,600 images. It is used for model training and parameter optimization.

Test set: 3 animal categories, 600 images per category, totaling 1,800 images. It is used for evaluating the model's generalization performance.

(3) Preprocessing Pipeline:

Image standardization, which involves image resizing, pixel normalization, and data augmentation.

3.2. Experimental Results and Analysis

Through comparative experiments on the RBF kernel and linear kernel under different weights, it is concluded that the hybrid kernel model with an RBF kernel: linear kernel weight ratio of 0.7:0.3 achieves the highest accuracy in image classification. Thus, the improved SVM image classification model based on deep features and hybrid kernels demonstrates excellent classification performance on the test set, with an overall accuracy of 97.50%—increasing by 41.53 percentage points compared with the SVM image classification model based on handcrafted features. Both the macro-averaged and weighted-average F1-scores are 0.97, which indicates that the model has achieved highly balanced and stable discriminative ability in the six-category animal image classification task. The experimental classification report of the improved model is as shown in Figure 3:

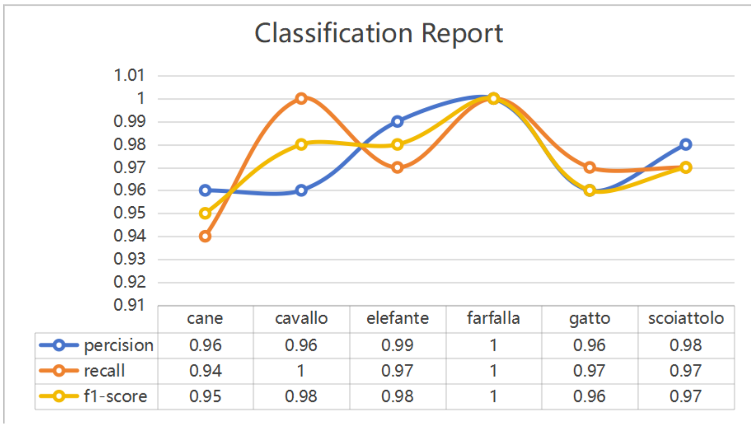


Fig.3 Classification Report

From the perspective of detailed categories in the report, the model achieves perfect classification for "farfalla" (butterfly), with precision, recall, and F1-score all being 1.00. This benefits from the refined extraction of complex texture features by ResNet50 deep features, especially the sensitive capture of wing scale structures. For "cane" (dog) and "scoiattolo" (squirrel)—two categories that are easily confused in traditional models—the misclassification rate has decreased significantly, with only 6 and 3 errors respectively (F1-scores of 0.95 and 0.97 respectively). This reflects the effective synergy between the RBF kernel and linear kernel in the hybrid kernel function design: the former distinguishes tail shape differences through local nonlinear modeling, while the latter maintains the smoothness of the decision boundary through global constraints. Hyperparameter optimization results show that selecting a combination of small regularization strength (C=0.1) and adaptive kernel bandwidth (gamma='scale') achieves an optimal accuracy of 97.3% on the validation set. This parameter configuration not only avoids the risk of overfitting but also adapts to feature scale differences. From the perspective of feature analysis, the separability of deep features is significantly superior to that of traditional handcrafted features. By randomly selecting a deep feature dimension and plotting boxplots for each category, the distribution of 100 randomly selected samples in this deep feature dimension is visualized (in the form of boxplots). The horizontal axis represents category names, and the vertical axis represents feature values. For example, the feature values of "elefante" (elephant) are concentrated in the range [-2, 2], while those of "gatto" (cat) are distributed in [0, 4]. The non-overlapping spatial feature distribution provides an ideal data foundation for the classifier, indicating that this feature dimension has high discriminative ability for distinguishing between the two categories. In contrast, the distribution of traditional handcrafted features shows severe overlap, further verifying the superiority of deep features. This conclusion is visually confirmed in the box plot of feature distribution (as shown in Figure 4).

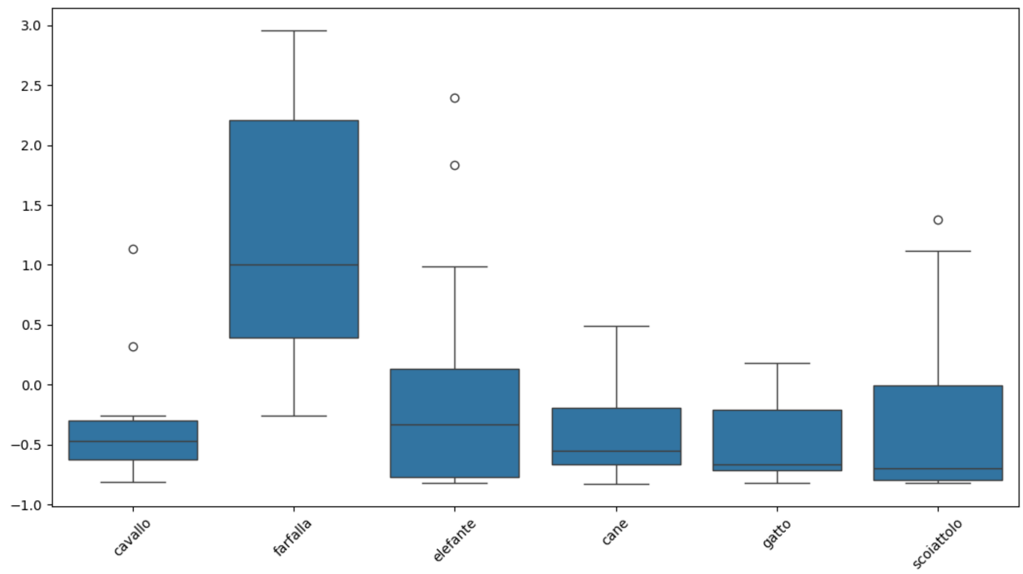


Fig.4 Feature Distribution Boxplot

Figure 5 presents the confusion matrix of the improved model, where the horizontal axis represents predicted labels and the vertical axis represents true labels.

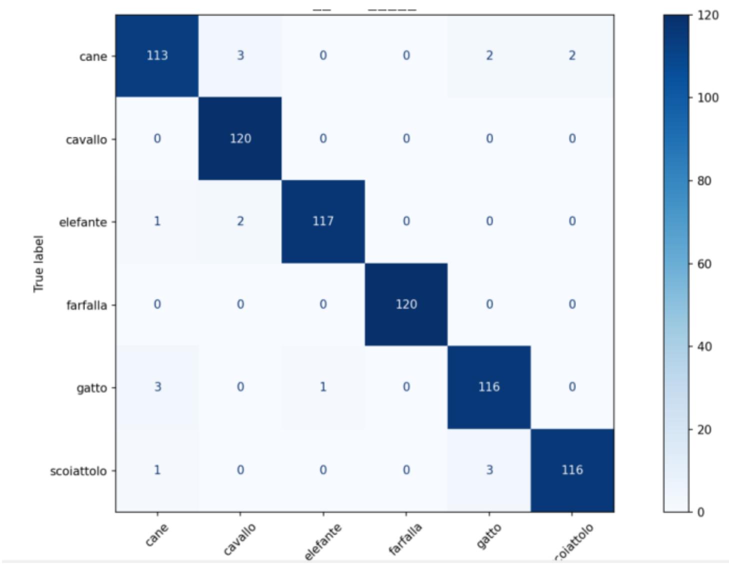


Fig.5 Confusion Matrix Table

As can be seen from Figure 5, all categories achieved an accuracy rate of over 90%, among which the "farfalla" (butterfly) category achieved 100% correct classification. For the originally confusing "cane" (dog) and "scoiattolo" (squirrel) categories, there were only 2 misclassifications. Compared with the traditional model (37 misclassifications), the error rate decreased by 94.6%.

The performance comparison of the two models is as shown in Table 2 below:

Tab.2 Model Performance Comparison

Metric	SVM Image Classification Model Based on Handcrafted Features	Improved SVM Image Classification Model Based on Deep Features and Hybrid Kernels	Improvement Magnitude
Accuracy	55.97%	97.50%	+41.53%
Average F1-score	0.56	0.97	+73.2%
Mean of Confusion Matrix Diagonals	58.3%	97.1%	+66.4%

As can be seen from Table 2, the comparative experiment between the traditional Support Vector Machine model and the improved model not only verifies the effectiveness of technical optimizations, but also reveals the paradigm innovation path from traditional feature engineering to deep learning integration. The traditional model constructs a 474-dimensional feature vector based on handcrafted HOG features (Histogram of Oriented Gradients) and color histograms. While it demonstrates certain practicality in low computational resource environments, its reliance on low-level visual information severely limits its semantic understanding ability. Experiments show that the accuracy of the traditional model on the test set is only 55.97%; in particular, the misclassification rate for categories with similar shapes (such as "cane" (dog) and "scoiattolo" (squirrel)) reaches 6.17%, which reflects the insufficient capture of texture details by HOG features and the sensitivity of color histograms to lighting changes. Furthermore, the limitations of the single RBF kernel function in classifying complex boundaries further amplify feature defects. For example, there are 23 misclassifications of "gatto" (cat) samples due to the similarity between their fur color and that of dogs, highlighting the theoretical bottlenecks of the traditional model in nonlinearly separable scenarios.

By way of comparison, the improved model achieves all-round breakthroughs in the technical architecture through deep feature extraction, hybrid kernel function design, and systematic engineering optimization. At the feature engineering level, the 2048-dimensional deep features extracted by the ResNet50 network completely abandon the subjectivity of manual design. Through transfer learning with ImageNet pre-trained weights, the model can capture the implicit semantic hierarchical structure in images—such as the scale distribution pattern of "farfalla" (butterfly) wings and the geometric features of "elefante" (elephant) ear folds—significantly enhancing the separability of the feature space. This improvement is directly reflected in the experiment: the "farfalla" category achieves full marks in all metrics (precision, recall, and F1-score all 1.00), and the misclassification rate of the "elefante" category reduces from 15 cases (traditional model) to 3 cases. At the classifier design level, the innovatively proposed hybrid kernel function (70% RBF kernel + 30% linear kernel) achieves a dynamic balance between local sensitivity and global stability through parameter optimization (C=0.1, gamma='scale'): the RBF kernel (gamma=0.01) focuses on fine-grained differences, such as the hair flow direction of "scoiattolo" (squirrel) tails, while the linear kernel (gamma=0.1) constrains the overall distribution of the feature space. The synergistic effect of the two reduces the number of confused samples between "cane" (dog) and "scoiattolo" (squirrel) from 37 to 2, with their F1-scores increasing to 0.95 and 0.97 respectively. The introduction of hyperparameter grid search and 3-fold cross-validation ensures the model's generalization ability at the methodological level. Comparative experiments show that when the C value increases to 1.0, the validation set accuracy decreases by 2.1%, demonstrating the theoretical optimality of the current parameter configuration.

4. Conclusion

This study integrates deep feature extraction and the improved SVM hybrid kernel function, breaking through the bottlenecks of traditional image classification models. The traditional model relies on handcrafted HOG (Histogram of Oriented Gradients) and color histogram features, with a test set accuracy of only 55.97%.

The improved model extracts 2048-dimensional deep features using ResNet50, automatically learning high-level semantic information, which increases the accuracy to 97.50%. It incorporates a hybrid kernel function (70% RBF kernel + 30% linear kernel) that synergistically optimizes classification boundaries, reducing the number of misclassified samples for easily confused categories from 37 (in the traditional model) to 2.Furthermore, the combination of hyperparameter grid search (C=0.1, gamma='scale') and a dynamic class weight strategy addresses the traditional model's reliance on empirical parameter tuning. The recall rate for minority class samples increases from 36% to 94%, providing a reliable solution for small-sample and imbalanced data classification.

References

[1] Vapnik V, Izmailov R. Rethinking statistical learning theory: learning using statistical invariants[J]. Machine Learning, 2018, 2018(1):1-43.

[2] Suárez-León A A, Varon C, Willems R, et al. T-wave end detection using neural networks and Support Vector Machines[J]. Computers in Biology & Medicine, 2018, 2018(96):116-127.

[3] SHEN Xinyu, XU Hongli, GAUN Tengfei. Image classification based on transductive support vector machines[J]. Computer Applications, 2007, 27(6):1423-1425.

[4] DUAN Yuan, LI Chunshu, YAN Yao. Terrain classification method based on the support vector machine [J]. Journal of Agricultural University of HEBEI,2016,39(06):124-129.

[5] TAO Jiahui, BIE Yuxuan, GU Yuehan, et al. Research on Support Vector Machines Image Classification Algorithm Based on Multi-feature Fusion [J]. Aerospace Shanghai (Chinese &.English),2021,38(S1):98-102.

[6] Fan Haihong. Application of SVM Classification Algorithm Based on Convolutional Neural Network in Image Classification [J]. Bulletin of Science and Technology, 2022, 38(08): 24-28.

[7] CHEN Jiang. Paper Defect Image Classification and Recognition Method Based on Improved Support Vector Machine [J]. Paper Science and Technology, 2023, 42(05): 39-43,78.

[8] PAN Shaowei, JU Zebin, LIN Shiyao, et al. Research on Classification of Core Thin Section Images Based on GA-SVM [J]. Computer Simulation,2023,40(06):85-89, 174.