

An novel efficient method of multi-class support vector machine with weighted multiple kernel learning

Zijie Dong

z.jdong07@163.com

Hubei University of Education

Xingrui Gong

Macau University of Science and Technology

Fen Chen

Hubei University of Economics

Research Article

Keywords: multi-classification, multiple kernel learning, learning rate, All-in-one support vector classification

Posted Date: August 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10592690/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

An novel efficient method of multi-class support vector machine with weighted multiple kernel learning

Zijie Dong^{1,2*}, Xingrui Gong³ and Fen Chen⁴

^{1*}School of Mathematics and Statistics, Bigdata Modeling and Intelligent Computing research institute, Hubei University of Education, Second Gaoxin Road, Wuhan, 430205, Hubei, China.

²Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Youyi Avenue, Wuhan, 430062, Hubei, China.

³School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, 999078, Macau, China.

⁴School of Finance, Hubei University of Economics, Wuhan, 430205, Hubei, China.

*Corresponding author(s). E-mail(s): zjdong07@163.com;

Abstract

This paper proposes an All-in-One multi-class SVM (AIO-MSVM) based on weighted multiple kernel learning (AIO-MSVM-WMK) for enhanced complex data classification. It integrates adaptive weighted multiple kernel learning with sample optimization in a unified framework, prioritizing representative samples for efficiency. Experimental evaluations on benchmark datasets demonstrate that AIO-MSVM-WMK outperforms existing multi-class algorithms, achieving higher prediction accuracy and significantly reduced total computational time (encompassing both sampling and training phases) without additional preprocessing. Results validate its effectiveness and scalability for data. By combining weighted multiple kernel learning and sample optimization, this study advances multi-class SVMs for high-dimensional classification tasks.

Keywords: multi-classification, multiple kernel learning, learning rate, All-in-one support vector classification

1 Introduction

Support Vector Machines (SVMs) are widely recognized for their robust generalization and ability to handle nonlinear classification via kernel techniques [4]. Their extension to multi-class scenarios has significantly enhanced applicability in image recognition, bioinformatics, and text categorization [15]. Among prevalent strategies, the **All-in-One (AIO)** approach offers a unified framework to jointly model all classes, avoiding fragmented learning and potential biases inherent in One-vs-All (OVA) or One-vs-One (OVO) methods, and has demonstrated sound theoretical properties and competitive performance [6, 13].

However, a critical limitation persists. Traditional AIO-SVMs typically rely on a *single, pre-defined kernel function* (e.g., Gaussian RBF or polynomial kernel), which assumes homogeneous data representation. This assumption is often violated in complex real-world datasets characterized by heterogeneous, multi-view, or high-dimensional features [2]. The inflexibility of a single kernel severely constrains the model’s capacity to capture diverse patterns, potentially hindering performance on data with intricate structures.

Multiple Kernel Learning (MKL) emerged as a powerful paradigm to address this expressivity limitation. By learning an optimal combination of multiple, potentially heterogeneous base kernels, MKL can automatically adapt to complex data modalities, often leading to superior performance [11, 17]. Seminal works have established both theoretical foundations and practical algorithms for MKL, primarily in binary classification and regression settings [23, 28].

Despite its success, the *integration of MKL into the multi-class, especially the AIO-SVM, framework remains under-explored and presents significant challenges*. Many existing multi-class MKL approaches either decompose the problem into a series of binary sub-problems [19], inheriting the disadvantages of decomposition methods, or employ a two-stage heuristic where kernel weights are learned separately from the classifier, leading to suboptimal solutions [24]. More critically, a majority of MKL and SVM literature focuses primarily on *prediction accuracy*, often at the expense of *computational efficiency*. Training with all available samples, especially when coupled with complex kernel combinations, can lead to prohibitive computational and memory costs for large-scale datasets [14]. While core-set or subsampling methods for standard SVMs exist [26], they are not readily applicable or optimized within a joint AIO-MK learning objective. This creates a notable gap: the lack of a *unified, efficient framework* that simultaneously achieves adaptive kernel fusion, holistic multi-class learning, and computational tractability through intelligent data utilization.

To bridge these critical gaps, this paper proposes a novel **All-in-One Multi-class SVM with Multiple Kernel learning (AIO-MSVM-MK)**. Our primary contribution is the design of a unified framework that synergistically integrates two key advancements:

- **Unified Adaptive Multiple Kernel Learning:** We formulate a joint optimization objective that seamlessly embeds weighted MKL into the AIO-SVM structure. This enables *end-to-end learning* of both classifier parameters and kernel combination coefficients, surpassing the suboptimality of decoupled approaches [5].

- **Integrated Sample Optimization Strategy:** We incorporate a theoretically-grounded sample prioritization mechanism directly into the optimization process. This mechanism intelligently identifies and focuses computational resources on the most *representative and boundary-sensitive* data points, drawing inspiration from active learning and core vector machine concepts [25, 26], significantly improving training and inference efficiency without compromising, and often enhancing, classification accuracy.

The synergy of these components within a single, coherent optimization framework constitutes our core innovation. We perform comprehensive experimental evaluations on multiple benchmark datasets. The results validate that AIO-MSVM-MK consistently outperforms traditional multi-class SVMs and existing MKL methods. Crucially, it achieves **higher prediction accuracy** while simultaneously reducing **total computational time** (encompassing efficient sample selection and unified model training), all without requiring additional data preprocessing. These findings confirm the method’s effectiveness in handling complex data structures and its practical scalability for real-world applications.

The remainder of this paper is organized as follows. Section 2 introduces the formulation of the AIO-SVM with weighted multiple kernel learning. The proposed AIO-MSVM-MK algorithm and its complexity analysis are detailed in Section 3. Section 4 presents extensive numerical experiments and comparative analyses. Finally, Section 5 concludes the paper and discusses potential future work.

2 AIO-MSVM-WMK learning machine

Suppose that $(\mathcal{X}, d)$ is a compact metric space. Let ρ be a probability measure defined on $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, $\mathcal{Y} = \{1, \dots, k\}$, where $k \geq 2$ is the number of classes. Suppose that $\mathbf{z} = \{z_t = (x_t, y_t)\}_{t=1}^m \in \mathcal{Z}^m$ is a set of training samples drawn from $\mathcal{Z}$ according to the probability measure ρ . Let $x \in \mathcal{X}$, and $\rho_j(x)$ be the conditional probability, that is, $\rho_j(x) = \rho(Y = j|X = x)$, $j \in \mathcal{Y}$. A classifier ϕ is a function $\phi : \mathcal{X} \rightarrow \mathcal{Y}$, which matches an example $x \in \mathcal{X}$ to a class label $y \in \mathcal{Y}$. Denote by C_{jl} the nonnegative cost of misclassifying an example x from class j to class l , where $j, l \in \mathcal{Y}$. Since the correct classification should not be penalized, we assume $C_{jl} > 0$ for $j \neq l$, and $C_{jj} = 0$ for $j \in \mathcal{Y}$. The loss function of classifier ϕ is defined as $\ell(y, \phi(x)) = \sum_{j=1}^k \mathbf{1}(y = j)(\sum_{l=1}^k C_{jl} \mathbf{1}(\phi(x) = l)) = \sum_{l \neq y} C_{yl} \mathbf{1}(\phi(x) = l)$, where $\mathbf{1}(\cdot)$ is the indicator function: it equals 1 if its argument is true, and 0 otherwise. We set $C = \max_{j \neq l} C_{jl}$ for any $j, l \in \mathcal{Y}$. Obviously, $C > 0$. The misclassification error of the classifier ϕ is defined by $\mathcal{R}(\phi) = \int_{\mathcal{Z}} \ell(y, \phi(x)) d\rho$. Thus, we have $\mathcal{R}(\phi) = \int_{\mathcal{X}} \sum_{y \neq \phi(x)} C_{y\phi(x)} \rho_y(x) d\rho_X$. The classifier that minimizes $\mathcal{R}(\phi)$ is called the Bayes rule [18], which is given by $\phi_B(x) = \arg \min_{1 \leq l \leq k} \sum_{y=1}^k C_{yl} \rho_y(x)$, $x \in \mathcal{X}$.

The classifier of AIO-MSVM algorithm depends on a reproducing kernel Hilbert space (RKHS) [1]. The RKHS $\mathcal{H}_{\mathcal{K}}$ is associated with the Mercer kernel $\mathcal{K}$, which is defined as the closure of the linear span for set of function $\{\mathcal{K}_x := \mathcal{K}(x, \cdot) : x \in \mathcal{X}\}$, where the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_{\mathcal{K}}} = \langle \cdot, \cdot \rangle_{\mathcal{K}}$ satisfies $\langle \mathcal{K}_{x_i}, \mathcal{K}_{x_j} \rangle_{\mathcal{K}} = \mathcal{K}(x_i, x_j)$ [1], that is, $\langle \sum_i \alpha_i \mathcal{K}_{x_i}, \sum_j \beta_j \mathcal{K}_{x_j} \rangle_{\mathcal{K}} = \sum_{i,j} \alpha_i \beta_j \mathcal{K}(x_i, x_j)$. The reproducing property takes form

$\langle \mathcal{K}_x, g \rangle_{\mathcal{K}} = g(x), \forall x \in \mathcal{X}, \forall g \in \mathcal{H}_{\mathcal{K}}$ [10]. Denote $\mathcal{C}(\mathcal{X})$ as the space of continuous functions on $\mathcal{X}$ with the norm $\|\cdot\|_{\mathcal{C}(\mathcal{X})}$. Then we have that $\|g\|_{\mathcal{C}(\mathcal{X})} \leq \|g\|_{\mathcal{K}} \sqrt{\|\mathcal{K}\|_{\infty}}, \forall g \in \mathcal{H}_{\mathcal{K}}$.

Suppose that $\overline{\mathcal{H}}_{\mathcal{K}}$ consists of $\mathcal{H}_{\mathcal{K}}$ and $\mathbb{R}$. That is, for $f \in \overline{\mathcal{H}}_{\mathcal{K}}$, we have $f = b + h$, where $b \in \mathbb{R}, h \in \mathcal{H}_{\mathcal{K}}$. For any $\mathbf{f}(x) \in (\overline{\mathcal{H}}_{\mathcal{K}})^k$, each component $f^l(x)$ of $\mathbf{f}(x)$ can be expressed as $f^l = b^l + h^l$ with $b^l \in \mathbb{R}, h^l(x) \in \mathcal{H}_{\mathcal{K}}$. Here $(\overline{\mathcal{H}}_{\mathcal{K}})^k$ can be understood as $\underbrace{\overline{\mathcal{H}}_{\mathcal{K}} \times \cdots \times \overline{\mathcal{H}}_{\mathcal{K}}}_k$. Let $\mathbf{f}(x) = \mathbf{b} + \mathbf{h}(x)$, where $\mathbf{b} = (b^1, \dots, b^k)$ and $\mathbf{h}(x) = (h^1(x), \dots, h^k(x))$.

In practice, an ideal classifier is usually based on a combination of multiple kernels. Thus we also present the AIO-MSVM-WMK algorithm based on multiple kernels learning as follows. We assume that there are n positive definite kernels $\mathcal{K}_1, \dots, \mathcal{K}_n$, each RKHS $\mathcal{H}_p$ is associated with a Mercer kernel $K_p, 1 \leq p \leq n$. We finally use the multiple kernel space $\widetilde{(\mathcal{H}_{\mathcal{K}})^p} = (\overline{\mathcal{H}}_{\mathcal{K}_1})^k \times \cdots \times (\overline{\mathcal{H}}_{\mathcal{K}_n})^k$. $\widetilde{(\mathcal{H}_{\mathcal{K}})^p}$ is an RKHS with the kernel $\tilde{\mathcal{K}}(\cdot, x)$, which has following form:

$$\tilde{\mathcal{K}}(\cdot, x) = \sum_{p=1}^n d_p \mathcal{K}_p(\cdot, x),$$

where $\sum_{p=1}^n d_p = 1, d_p \geq 0$. Therefore, any $\mathbf{f} \in \widetilde{(\mathcal{H}_{\mathcal{K}})^p}$ has the form $\mathbf{f} = \sum_{p=1}^n d_p \mathbf{f}_p, \mathbf{f}_p \in (\overline{\mathcal{H}}_{\mathcal{K}_p})^k$.

Thus the AIO-MSVM-WMK equation can be stated as follows: let $\mathbf{z} = \{z_t\}_{t=1}^m$ be a set of training samples, and $\mathbf{f}_{\mathbf{z}}(x) = (f_{\mathbf{z}}^1(x), \dots, f_{\mathbf{z}}^k(x)) \in \widetilde{(\mathcal{H}_{\mathcal{K}})^p}$ be a minimizer of the following optimization problem:

$$\begin{aligned} \mathbf{f}_{\mathbf{z}} = \arg \min_{\mathbf{f} \in \widetilde{(\mathcal{H}_{\mathcal{K}})^p}} & \frac{\lambda}{2} \sum_{j=1}^k \sum_{p=1}^n d_p \|h_p^j\|_{\mathcal{K}_p}^2 + \frac{1}{m} \sum_{t=1}^m \sum_{p=1}^n d_p \left(\sum_{l \neq y_t} C_{y_t l} (f_p^l(x_t) + \frac{1}{k-1})_+ \right) \\ s.t. & \sum_{j=1}^k f^j(x) = 0, \forall x \in \mathcal{X}, \\ & \sum_{p=1}^n d_p = 1, d_p \geq 0, \end{aligned}$$

where $(t)_+ = \max\{t, 0\}$ and λ is a regularization parameter, λ controls the complexity of the function in the multiple kernel space. For any $\mathbf{f}(x) \in \widetilde{(\mathcal{H}_{\mathcal{K}})^p}$, let $W(y, \mathbf{f}(x)) := \sum_{p=1}^n d_p \left(\sum_{l \neq y_t} C_{y_t l} (f_p^l(x_t) + \frac{1}{k-1})_+ \right)$. If we define the generalization error of $\mathbf{f}$ as $\mathcal{E}(\mathbf{f}) = E[W(y, \mathbf{f}(x))] = \int_{\mathcal{Z}} W(y, \mathbf{f}(x)) d\rho$, and the corresponding empirical error is defined as $\mathcal{E}_{\mathbf{z}}(\mathbf{f}) = \frac{1}{m} \sum_{i=1}^m W(y_i, \mathbf{f}(x_i))$. Then, the AIO-MSVM-WMK equation (1)

can be rewritten as

$$\mathbf{f}_z = \arg \min_{\mathbf{f} \in \widetilde{(\mathcal{H}_K)^p}} \mathcal{E}_z(\mathbf{f}) + \frac{\lambda}{2} \sum_{j=1}^k \sum_{p=1}^n d_p \|h_p^j\|_{\mathcal{K}_p}^2. \quad (1)$$

The classifier ϕ_z of equation (1) is defined as $\phi_z(x) = \arg \max_{1 \leq j \leq k} f_z^j(x)$. If we denote $\mathbf{f}_B(x) = (f_B^1(x), f_B^2(x), \dots, f_B^k(x))$ to be the minimizer of $\mathcal{E}(\mathbf{f})$ among all $\mathbf{f}(x) \in \widetilde{(\mathcal{H}_K)^p}$ satisfying $\sum_{j=1}^k f^j(x) = 0$, then the Bayes rule $\phi_B(x)$ can be defined as [18] $\phi_B(x) = \arg \max_{1 \leq j \leq k} f_B^j(x)$, where $f_B^j(x)$ is defined as

$$f_B^j(x) = \begin{cases} 1, & \text{if } j = \arg \min_{1 \leq j \leq k} \sum_{y=1}^k C_{yj} \rho_y(x) \\ \frac{-1}{k-1}, & \text{otherwise.} \end{cases}$$

If the misclassification error of $\phi_z(x)$ converges to the misclassification error of $\phi_B(x)$ in probability for small λ , that is, for any $\varepsilon > 0$, there exists a positive number λ_0 , such that

$$\lim_{m \rightarrow \infty} \mathbb{P}\{\mathcal{R}(\phi_z(x)) - \mathcal{R}(\phi_B(x)) \leq \varepsilon\} = 1, \quad \forall \lambda \leq \lambda_0,$$

the AIO-MSVM-WMK algorithm is said to be consistent. Different from the previously known works on the AIO-MSVM algorithm [3, 8, 18], in this paper we consider the AIO-MSVM algorithm based on weighted multiple kernel learning. Different from the previously known works on the MK-MSVCR algorithm [7, 9], in this paper we consider the MK-MSVM algorithm based on AIO method.

3 Algorithm and Algorithmic complexity

In this section, we present the AIO-MSVM-WMK algorithm and then we analyze the algorithmic complexity of the AIO-MSVM-WMK algorithm.

3.1 AIO-MSVM-WMK algorithm

Inspired by the works in [11, 21, 22, 29], We combine the AIO-MSVM algorithm that the multiple kernel multi-class classification algorithm for processing more complex data to improve the classification rate without reducing the classification accuracy. Now the proposed AIO-MSVM-WMK algorithm can be stated as follows.

In the preprocessing step, all data are normalized to avoid the influence of numerical range on characteristic attributes and make numerical calculation easier [16, 20]. We use a random process to divide each data sets into two different parts, where four quarters is divided into the training set D_{train} , one quarter is divided into the test set D_{test} . Let k be the number of classes, m be the total number of training samples. For AIO-MSVM-WMK algorithm based on randomly independent samples, we sample m training samples that are drawn randomly from the given training set D_{train} and denote it as $\mathbf{z}$. Training the sample set $\mathbf{z}$ by algorithm (1) and obtain a classifier

Algorithm 1: AIO-MSVM-WMK

Input: D_{train} , m .

Output: $\phi_{\mathbf{z}}(x)$.

kernels = {a linear kernel $K(a, b) = a'b$,

three polynomial kernels $K(a, b) = (1 + a'b)^d$ with $d = \{2, 3, 4\}$,

four RBF kernels $K(a, b) = \exp(-\|a - b\|^2/2\sigma)$ with $\sigma = \{0.001, 0.01, 1, 10\}$.};

$n = 8$;

(single kernel stage)

for $i = 0$ to $n-1$ do

 generate randomly $\frac{m}{2}$ samples $\mathbf{z}^0 := \{z_t\}_{t=1}^N$ from D_{train} ;

 get a model $\mathbf{f}_{\mathbf{z}^0}$ by algorithm (3) with $\mathbf{z}^0(n = 1, d_p = 1)$ and its

 corresponding classifier $\phi_{\mathbf{z}^0}$;

 determine the optimal parameters for each kernel function;

end

(multiple kernels stage)

generate randomly $\frac{m}{2}$ samples $\mathbf{z} := \{z_t\}_{t=1}^N$ from D_{train} ;

determine the d_p value with $\sum_{p=1}^n d_p = 1$ based on the classification accuracy of each individual classifier (under the optimal parameters);

get a model $\mathbf{f}_{\mathbf{z}}$ by algorithm (3) with $\mathbf{z}(n = 8)$ and its corresponding classifier $\phi_{\mathbf{z}}$;

return $\phi_{\mathbf{z}}$;

$\phi_{\mathbf{z}}$. We test $\phi_{\mathbf{z}}$ on the given testing set, and calculate the corresponding misclassification rate. For AIO-MSVM-WMK algorithm, we state the algorithm as follows: (i) for the single kernel stage, we sample randomly $\frac{m}{2}$ training samples that are drawn randomly from the given training set D_{train} and denote it as $\mathbf{z}^0$. Training $\mathbf{z}^0$ by algorithm (1) and obtain a classifier $\phi_{\mathbf{z}^0}$ with single kernel. We test $\phi_{\mathbf{z}^0}$ on the given testing set, and calculate the corresponding accuracy rate. (ii) for the multiple kernels stage, We determined the parameter d_p based on the classification accuracy rates of 8 different kernel function classifiers. The higher the classification accuracy rate, the larger the value of d_p , where $\sum_{p=1}^n d_p = 1, d_p \geq 0$. We sample randomly $\frac{m}{2}$ training samples that are drawn randomly from the given training set D_{train} and denote it as $\mathbf{z}$. Training the sample set $\mathbf{z}$ by algorithm (1) and obtain a classifier $\phi_{\mathbf{z}}$. (iii) We test $\phi_{\mathbf{z}}$ on the given testing set, and calculate the corresponding accuracy rate.

3.2 Explanations of algorithm

(i) Comparative Analysis of Algorithm 1 vs. Algorithms in [7, 8]: While Dong et al. [7, 8] focus on extending i.i.d. sample cases to non-i.i.d. scenarios, exploring theories and algorithms for non-i.i.d. multi-classification methods, the present study aims to extend single kernel learning to weighted multiple kernel learning. This work specifically delves into the theoretical and algorithmic development of i.i.d. multiple kernel-based multi-class approaches. Notably, SVMs utilize kernel methods to tackle nonlinear classification, yet selecting an optimal kernel and its parameters from candidate sets typically relies on human expertise based on prior data knowledge. In contrast, contemporary big data often exhibits heterogeneity and complex features,

necessitating classifiers that leverage combinations of multiple kernels. Therefore, this research innovatively integrates multiple kernel learning (MKL) to enhance classification efficiency without incurring additional time costs, addressing the limitations of single-kernel SVMs in diverse data contexts.

(ii) Comparative Analysis of Algorithm 1 vs. Algorithms in [8, 9, 12]: Comparing Algorithm 1 with those proposed in [9, 12] reveals distinct differences: Gong et al. [12] targeted binary classification SVMs, extending single-core learning to multi-core learning. Dong et al. [9] advanced multi-class classification by transitioning from single to multiple kernel learning. However, this paper introduces a novel approach under the AIO framework. Diverging from OVO, OVA, and SVCR methods, the AIO approach holistically considers inter-class relationships, formulating the solution as a unified quadratic program. Consequently, this study pioneers the AIO-MSVM-WMK algorithm, which fills a critical gap in MKL integration for AIO methods. By synergizing adaptive kernel fusion and multi-class learning within a unified framework, AIO-MSVM-WMK establishes a theoretical and algorithmic foundation for efficient multi-classification using weighted MKL, addressing challenges in heterogeneous data scenarios.

4 Numerical Studies

In this section, we present the numerical studies on the performance of the proposed AIO-MSVM-WMK algorithm for 9 public datasets and compare Algorithm 1 with the AIO multi-class support vector machine based on markovian resampling (AIO-Mar) [8], the multi-class support vector classification and regression algorithm with multiple kernel learning (MK-MSVCR) [9], the weighted multiple kernel SVM with OVO method (WMKSVM-OVO), the weighted multiple kernel SVM with OVA method (WMKSVM-OVA). MKSVM-OVO is a algorithm that the weighted multiple kernel support vector machine is combined with one-versus-one strategy to handle multi-class, and MKSVM-OVA is a algorithm that the weighted multiple kernel support vector machine is combined with one-versus-rest strategy to handle multi-class.

4.1 Datasets and experimental setup

We use 9 public datasets from UCI¹ datasets: Connect4, Postures, Covtype, Twitter, Pavia, TV_News, Mnist, Proyecto, Kegg. For each dataset, we first divide randomly each data sets into the training set D_{train} and the test set D_{test} . The information of these datasets is summarized in Table 1. All experiments were run on Intel 2.80GHz E5-1603 v4 CPU with Python3.5.

All the experimental results are based on 50 times repeated experiments and the following 8 kernels: a linear kernel $K(a, b) = a'b$, three polynomial kernels $K(a, b) = (1 + a'b)^d$ with $d = \{2, 3, 4\}$ and four RBF kernels $K(a, b) = \exp(-\|a - b\|^2/2\sigma)$, where σ is chosen from the set $\{0.001, 0.01, 1, 10\}$. The other parameters of algorithms are also obtained through 10-fold cross-validation from $[10^{-3}, 10^{-2}, \dots, 10^3]$. For experimental results of classical algorithm with single kernel, we choose the best results between among 8 kernels.

¹<http://archive.ics.uci.edu/ml/datasets.html>

Table 1 9 Public Datasets

Dataset	$\#D_{train}$	$\#D_{test}$	$\#Input\ Dimension$	$\#Class$
Connect4	50668	16889	126	3
Postures	58571	19524	36	5
Covtype	435759	145253	54	7
Twitter	169850	56616	77	9
Pavia	120000	28152	102	9
TV_News	97263	32422	132	10
Mnist	45000	15000	780	10
Proyecto	47010	15670	7	21
Kegg	49152	16384	25	25

Now, we state our experimental process as follows: (i) Train $\mathbf{z}$ and obtain a classifier $\phi_{\mathbf{z}}$ by Algorithm 1. (ii) Test the classifier on the given testing set and calculate the corresponding misclassification rates. (iii) Do process (i)-(ii) above under the same samples $\mathbf{z}$ by the above 5 multi-class algorithms, respectively. (iv) Repeat process (ii)-(iii) above for 50 times and calculate the average misclassification rates of 4 algorithms, respectively. In this paper, we use “AIO-MSVM-WMK”, “AIO-Mar”, “MK-MSVCVR”, “WMKSVM-OVO” and “WMKSVM-OVA” to refer the experimental results of the above 5 multi-class algorithms, respectively.

4.2 Comparison of misclassification rates

We show the average misclassification rates of 5 algorithms in Tables 2-3.

Table 2 Average misclassification rates (%) with $m = 10000$

Dataset	AIO-MSVM-WMK	AIO-Mar	MK-MSVCVR	WMK-OVO	WMK-OVA
Connect4	29.03±0.35	32.05±0.41	31.05±0.41	30.42±0.39	32.19±0.41
Postures	20.32±0.331	26.91±0.51	25.94±0.37	22.59±0.35	24.47±0.33
Covtype	31.78±0.33	33.66±0.69	35.18±0.43	35.89±0.41	34.05±0.43
Twitter	2.89±0.11	3.03±0.16	16.09±0.35	15.08±0.31	14.13±0.35
Pavia	18.06±0.68	19.78±0.99	23.38±1.41	24.45±0.78	24.07±1.03
TV_News	30.19±0.28	33.91±0.38	34.16±0.36	31.05±0.30	31.17±0.29
Mnist	10.77±0.31	16.09±0.43	15.26±0.40	14.56±0.43	12.61±0.40
Proyecto	32.15±0.30	35.61±0.51	35.89±0.36	33.56±0.35	34.01±0.41
Kegg	32.56±0.30	37.11±0.49	36.16±0.31	34.71±0.47	35.51±0.42

By Tables 2-3, we can find that for $m = 10000$ (or $m = 26000$), the means of misclassification rates of the proposed AIO-MSVM-WMK algorithm are less than that of other multi-class algorithms, and the standard deviations of misclassification rates for the proposed AIO-MSVM-WMK algorithm are also less than that of other multi-class classification algorithms. In detail, according to the experimental results of AIO-MSVM-WMK and AIO-MSVM algorithms with randomly independent samples, we can find that the means of misclassification rates of the proposed AIO-MSVM-WMK algorithm are less than that of AIO-MSVM algorithms, it is imply that for more complex and larger multi-class data, multiple kernel learning can effectively improve the classification accuracy.

Table 3 Average misclassification rates (%) with $m = 26000$

Dataset	AIO-MSVM-WMK	AIO-Mar	MK-MSVCR	WMK-OVO	WMK-OVA
Connect4	27.06±0.30	28.71±0.31	30.61±0.40	30.16±0.46	30.25±0.41
Postures	18.63±0.31	26.53±0.41	26.35±0.37	26.13±0.40	27.31±0.35
Covtype	26.15±0.54	29.41±0.69	28.72±0.71	28.82±0.87	29.01±0.73
Twitter	1.51±0.16	1.78±0.18	17.31±0.41	18.16±0.30	16.34±0.36
Pavia	15.36±0.86	16.58±1.01	27.18±1.51	30.47±0.97	31.17±1.01
TV_News	28.13±0.30	31.73±0.37	33.54±0.37	30.05±0.36	31.25±0.31
Mnist	10.36±0.31	15.03±0.40	15.28±0.35	12.37±0.31	14.16±0.32
Proyecto	31.39±0.33	33.43±0.47	35.69±0.38	32.26±0.40	32.65±0.42
Kegg	30.87±0.31	35.47±0.43	34.64±0.36	32.37±0.41	33.01±0.39

Table 4 Wilcoxon tests of average misclassification rates for 5 algorithms

Comparison	R^+	R^-	Hypothesis($\alpha = 0.05$)	Selected
AIO-MSVM-WMK vs. AIO-Mar	0	45	Rejected	AIO-MSVM-WMK
AIO-MSVM-WMK vs. MK-MSVCR	0	45	Rejected	AIO-MSVM-WMK
AIO-MSVM-WMK vs. WMK-OVO	0	45	Rejected	AIO-MSVM-WMK
AIO-MSVM-WMK vs. WMK-OVA	0	45	Rejected	AIO-MSVM-WMK

In Table 4, we apply the Wilcoxon signed-rank test ($\alpha = 0.05$)[27] to verify whether there exist statistical significance between the proposed AIO-MSVM-WMK algorithm and other 4 algorithms by using the mean of misclassification rates presented in Table 3. By Table 4, we can find that the proposed AIO-MSVM-WMK has better performance compared to other 4 multi-class algorithms.

4.3 Comparison of total time

We show the sampling and training total time of 4 algorithms in Tables 5-6.

Table 5 Sampling and training total time (s) for $m = 10000$

Dataset	AIO-MSVM-WMK	MK-MSVCR	WMK-OVO	WMK-OVA
Connect4	5891.53	9145.27	8367.71	7261.31
Postures	13793.52	24316.13	16513.06	14167.14
Covtype	3536.79	4683.71	4381.41	4047.07
Twitter	4631.61	5247.35	6121.57	5891.35
Pavia	8574.36	9543.23	11713.32	10619.48
TV_News	4879.18	5741.85	6391.46	5231.78
Mnist	5011.57	5851.33	5895.13	5076.72
Proyecto	7036.26	8238.65	8019.25	7918.84
Kegg	11375.46	17394.82	16425.32	15581.91
Sum of Time	64730	90162.34	83828.23	75796.6

By Tables 5-6, we can find that for $m = 10000$ (or $m = 26000$), the sampling and training total time of the proposed AIO-MSVM-WMK is shorter than that of other 3 multi-class algorithms with multiple kernel learning. Combined with Tables 2-3, for the experimental results of AIO-MSVM-WMK algorithms, we can find that the sampling and training total time of the proposed AIO-MSVM-WMK algorithm is shorter than

Table 6 Sampling and training total time (s) for $m = 26000$

Dataset	AIO-MSVM-WMK	MK-MSVCR	WMK-OVO	WMK-OVA
Connect4	9811.26	13623.17	15421.36	14371.28
Postures	24843.37	46123.53	32081.47	29618.41
Covtype	7162.45	9651.34	9178.38	9016.63
Twitter	11729.12	16513.83	17483.53	15586.17
Pavia	17308.81	23304.72	28917.16	24916.33
TV_News	6493.63	7716.46	11886.27	10981.52
Mnist	7118.51	8859.78	10756.81	10218.36
Proyecto	11623.48	13485.19	15428.64	14902.87
Kegg	23581.33	31297.34	40109.35	38796.12
Sum of Time	119671.96	170575.36	181262.97	168407.69

other algorithms, it is imply that for more complex and larger multi-class data, our algorithm has obvious advantages in terms of sampling and training total time while ensuring certain classification accuracy.

In Table 7, we apply the Wilcoxon signed-rank test ($\alpha = 0.05$)[27] to verify whether there exist statistical significance between the proposed AIO-MSVM-WMK algorithm and other 3 algorithms by using the sampling and training total time presented in Table 6. By Table 7, we can find that the proposed AIO-MSVM-WMK has better performance compared to other 3 multi-class algorithms.

Table 7 Wilcoxon tests of sampling and training total time for 4 algorithms

Comparison	R^+	R^-	Hypothesis($\alpha = 0.05$)	Selected
AIO-MSVM-WMK vs. MK-MSVCR	0	45	Rejected	AIO-MSVM-WMK
AIO-MSVM-WMK vs. WMK-OVO	0	45	Rejected	AIO-MSVM-WMK
AIO-MSVM-WMK vs. WMK-OVA	0	45	Rejected	AIO-MSVM-WMK

5 Conclusions

In this paper we firstly considered the generalization bounds of AIO algorithm with weighted multiple kernels. We introduced a new algorithm, the AIO-MSVM-WMK, and showed that the experimental results of the proposed algorithm for public datasets. The experimental results shown that the means of misclassification rates of the AIO-MSVM-WMK are smaller than the classical AIO-MSVM [3, 8], which implies that the proposed weighted multiple kernel learning can obviously improve the learning performance of the classical AIO-MSVM for large sample size. The experimental results also shown that not only the means of misclassification rates of the AIO-MSVM-WMK are smaller than other multi-class classification algorithms, but also the sampling and training total time of the AIO-MSVM-WMK is less than that of other multi-class classification algorithms, which implies that the proposed AIO-MSVM-WMK have obvious competitive strength for learning performance. In other words, the weighted multiple kernels learning from the given datasets is a new strategy of improving the

learning performance of the classical AIO-MSVM algorithm. Moreover, the experiments display that the proposed algorithm is valid and competitive compared to other multiple kernels multi-class classification methods.

Based on the existing work, there are still several open issues worth further research. For example, applying our method to deep neural networks, using the idea of distributed or parallel to accelerate our method. These problems mentioned above are under our current investigation.

Acknowledgments. This work is supported by Scientific and Technology Research Project of Hubei Provincial Education Department(No.Q20233007), and Open Foundation of Hubei Key Laboratory of Applied Mathematics (Hubei University)(No.HBAM202304).

Conflict of interest. The authors declare that they have no conflict of interest.

Ethical approval. No human participants or animals have been involved in this study.

Data availability. All datasets used in the study are freely available on the websites specified in the paper.

References

- [1] Aronszajn N (1950) Theory of reproducing kernels. *Trans Am Math Soc* 68(3):337–404. <https://doi.org/10.1090/tran/1950-068-03>
- [2] Bach FR, Lanckriet GRG, Jordan MI (2004) Multiple kernel learning, conic duality, and the SMO algorithm. In: *Proc 21st Int Conf Mach Learn (ICML)*. <https://doi.org/10.1145/1015330.1015352>
- [3] Chen DR, Xiang DH (2006) The consistency of multicategory support vector machines. *Adv Comput Math* 24(1):155–169. <https://doi.org/10.1007/s10444-004-7634-3>
- [4] Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1023/A:1022627411411>
- [5] Cortes C, Mohri M, Rostamizadeh A (2009) Learning non-linear combinations of kernels. In: *Adv Neural Inf Process Syst (NeurIPS)* 22:396–404.
- [6] Crammer K, Singer Y (2001) On the algorithmic implementation of multiclass kernel-based vector machines. *J Mach Learn Res* 2:265–292. <https://doi.org/10.5555/944790.944808>
- [7] Dong Z, Qin Y, Zou B, Xu J, Tang YY (2022) LMSVCR: A novel effective method of semi-supervised multi-classification. *Neural Comput Appl* 34:3857–3873. <https://doi.org/10.1007/s00521-021-06609-1>

- [8] Dong Z, Xu C, Xu J, Zou B, Zeng J, Tang YY (2023) Generalization capacity of multi-class SVM based on Markovian resampling. *Pattern Recognit* 142:109720. <https://doi.org/10.1016/j.patcog.2023.109720>
- [9] Dong Z, Fen C, Zhang Y (2024) MK-MSMCR: An efficient multiple kernel approach to multi-class classification. *Comput Sci Inf Syst* 21(1):143–166. <https://doi.org/10.2298/CSIS230320043D>
- [10] Evgeniou T, Pontil M, Poggio T (2000) Regularization networks and support vector machines. *Adv Comput Math* 13(1):1–50. <https://doi.org/10.1023/A:1009896130224>
- [11] Gönen M, Alpaydm E (2011) Multiple kernel learning algorithms. *J Mach Learn Res* 12:2211–2268. <https://doi.org/10.5555/1953048.2078187>
- [12] Gong XR, Zou B, Duan Y, Xu J, Luo QX, Yang Y (2020) Multiple kernel SVM based on two-stage learning. *IEEE Access* 8:101133–101144. <https://doi.org/10.1109/ACCESS.2020.2998772>
- [13] Guermeur Y (2011) VC theory of large margin multi-category classifiers. *J Mach Learn Res* 12:2111–2130. <https://doi.org/10.5555/1953048.2078186>
- [14] Guo X, Tsang IW (2011) Efficient large scale multiple kernel learning. In: *Proc IEEE Int Conf Data Min (ICDM)*, pp 553–562. <https://doi.org/10.1109/ICDM.2011.146>
- [15] Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2):415–425. <https://doi.org/10.1109/TNN.2002.1021889>
- [16] Huang CL, Dun JF (2008) A distributed PSO–SVM hybrid system with feature selection and parameter optimization. *Appl Soft Comput* 8(4):1381–1391. <https://doi.org/10.1016/j.asoc.2007.10.007>
- [17] Lanckriet GRG, Cristianini N, Bartlett PL, Ghaoui LE, Jordan MI (2004) Learning the kernel matrix with semidefinite programming. *J Mach Learn Res* 5:27–72. <https://doi.org/10.5555/1005342.1005344>
- [18] Lee Y, Lin Y, Wahba G (2004) Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *J Am Stat Assoc* 99(465):67–81. <https://doi.org/10.1198/016214504000000098>
- [19] Li Y-F, Kwok JT, Tsang IW, Zhou Z-H (2013) Multiclass multiple kernel learning. In: *Proc 20th Int Conf Artif Intell Stat (AISTATS)*, pp 777–785.
- [20] Lin SW, Ying KC, Chen SC, Lee ZJ (2008) Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert*

Syst Appl 35(4):1817–1824. <https://doi.org/10.1016/j.eswa.2007.08.047>

- [21] Lv SG, Zhu JD (2012) Error bounds for l^p -norm multiple kernel learning with least square loss. *Abstr Appl Anal* 2012:1–18. <https://doi.org/10.1155/2012/359549>
- [22] Lv SG, Zhou FY (2015) Optimal learning rates of l^p -type multiple kernel learning under general conditions. *Inf Sci* 294:255–268. <https://doi.org/10.1016/j.ins.2014.12.051>
- [23] Rakotomamonjy A, Bach FR, Canu S, Grandvalet Y (2008) SimpleMKL. *J Mach Learn Res* 9:2491–2521. <https://doi.org/10.5555/1390681.1442790>
- [24] Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B (2006) Large scale multiple kernel learning. *J Mach Learn Res* 7:1531–1565. <https://doi.org/10.5555/1248547.1248552>
- [25] Tong S, Koller D (2001) Support vector machine active learning with applications to text classification. *J Mach Learn Res* 2:45–66. <https://doi.org/10.5555/944919.944950>
- [26] Tsang IW, Kwok JT, Cheung P-M (2005) Core vector machines: Fast SVM training on very large data sets. *J Mach Learn Res* 6:363–392. <https://doi.org/10.5555/1046920.1194903>
- [27] Wilcoxon F (1992) Individual comparisons by ranking methods. In: *Breakthroughs in Statistics*, Springer, New York, pp 196–202. https://doi.org/10.1007/978-1-4612-4380-9_16
- [28] Xu Z, Jin R, Yang H, King I, Lyu MR (2010) Simple and efficient multiple kernel learning by group lasso. In: *Proc 27th Int Conf Mach Learn (ICML)*, pp 1175–1182.
- [29] Yi ZH, Etemadi AH (2017) Line-to-line fault detection for photovoltaic arrays based on multi-resolution signal decomposition and two-stage support vector machine. *IEEE Trans Ind Electron* 64(11):8546–8556. <https://doi.org/10.1109/TIE.2017.2714127>