文章编号：1001-9014(2025)06-0865-10

DOI：10. 11972/j. issn. 1001-9014. 2025. 06. 005

基于PSO-GA-SVM融合算法及荧光激光雷达遥测技术的生物气溶胶浓度预测

饶志敏[1,2]，李一成[1,2]，李一秀[1,2]，刘佳鑫[1,2]，巩 鑫[1,2]，赵 虎[1,2]，毛建东[1,2*]

（1. 北方民族大学 电气信息工程学院，宁夏 银川 750021；
2. 宁夏回族自治区大气环境遥感探测重点实验室，宁夏 银川 750021）

摘要：生物气溶胶粒子在空气中传播较广，高浓度的生物气溶胶对人体健康有着极大的危害。为实现大气生物气溶胶浓度的预警预测，本文以荧光激光雷达为探测工具，在获取生物气溶胶浓度廓线的基础上，结合大气环境相关参数，利用粒子群(PSO)和遗传算法(GA)优化支持向量机(SVM)，建立生物气溶胶浓度廓线预测模型，通过采用温度、湿度、PM2. 5、PM10、CO_2、SO_2、NO_2、O_3、风速等相关参数数据作为输入量，生物气溶胶浓度廓线数据作为输出量进行模型训练，确定预测模型参数配置，重新引入新的大气环境参数，利用训练好的模型预测生物气溶胶浓度廓线，并与荧光激光雷达探测的生物气溶胶浓度廓线进行比较，同时分析不同算法优化模型预测的生物气溶胶浓度及其相对误差。

关 键 词：荧光激光雷达；生物气溶胶；PSO-GA-SVM融合算法；浓度预测

中图分类号：P4　　**文献标识码：**A

Prediction of bioaerosol concentration based on PSO-GA-SVM fusion algorithm and fluorescence lidar

RAO Zhi-Min[1,2]，LI Yi-Cheng[1,2]，LI Yi-Xiu[1,2]，LIU Jia-Xin[1,2]，GONG Xin[1,2]，ZHAO Hu[1,2]，MAO Jian-Dong[1,2*]

（1. School of Electrical and Information Engineering，North Minzu University，Yinchuan 750021，China；
2. Key Laboratory of Atmospheric Environment Remote Sensing of Ningxia Province，Yinchuan 750021，China）

Abstract：Bioaerosol particles spread widely in the air，and high concentrations of bioaerosols pose a great threat to human health. To achieve early warning and prediction of atmospheric bioaerosol concentration，this paper uses fluorescence lidar as the detection tool. Based on the acquisition of bioaerosol concentration profiles，combined with relevant parameters of the atmospheric environment，particle swarm optimization（PSO）and genetic algorithm（GA）are used to optimize the support vector machine（SVM）to establish a bioaerosol concentration profile prediction model. Using temperature，humidity，PM2. 5，PM10，CO_2，SO_2，NO_2，O_3，wind speed and other related parameter data as inputs，and bioaerosol concentration profile data as outputs for model training，the prediction model parameter configuration is determined. New atmospheric environment parameters are reintroduced，and the trained model is used to predict the bioaerosol concentration profile，which is compared with the bioaerosol concentration profile detected by fluorescence lidar. At the same time，different algorithms are analyzed to optimize the model's predicted bioaerosol concentration and its relative error.

Key words：fluorescence lidar，bioaerosol，PSO-GA-SVM fusion algorithm，concentration prediction

收稿日期：2024-12-09，**修回日期：**2025-07-22　　**Received date：**2024-12-09，**Revised date：**2025-07-22

基金项目：国家自然科学基金(42465007、42105140、42265009)、宁夏自然科学优秀青年基金(2022AAC05032)、北方民族大学研究生创新项目(YCX24345)

Foundation items：Supported by the National Natural Science Foundation of China（42465007，42105140，42265009），the Natural Science Outstanding Youth Foundation of Ningxia Province(2022AAC05032)，the Graduate Innovation Project of North Minzu University(YCX24345)

作者简介(Biography)：饶志敏(1987—)，男，江西临川人，工学博士，主要研究领域为生物气溶胶荧光激光雷达、荧光光谱激光雷达. E-mail：1805695165@qq. com

*通讯作者(Corresponding author)：E-mail：mao_jiandong@163. com

引言

空气中的生物气溶胶来源多样，如人类活动、环境污染以及自然环境排放都会产生生物气溶胶。生物气溶胶含有多种生物性粒子，且体积小质量轻，随着粒径减小，其空气粘性作用越明显，在空气中沉降速度变慢，随着大气环境中的气流流动而四处扩散，造成病原性生物气溶胶的扩散和传播，给人体健康造成影响；如细菌、真菌、病毒和花粉等生物气溶胶被人体吸入后，会引发过敏性呼吸系统疾病等[1-3]。此外，空气中的生物气溶胶含量及分布容易受到大气环境污染类型和污染水平而变化，相比于清洁大气，污染天气中的生物气溶胶浓度和种类都会增加[4,5]。因此，对空气中的生物气溶胶进行实时监控和预警预测，以及深入研究其分布特征，对于预防和控制生物性疾病的扩散和预防至关重要。

随着人类活动不断地进步和发展，大气环境污染的程度也随之增加，如何实现污染状况有效预警预测是大气污染控制和治理的关键因素之一。神经网络模型作为一种高效的计算方法具有强大的非线性预测能力，通过对大气污染物、气象条件等因素进行分析与测算，广泛应用于大气环境污染程度的预警预测。Xiaonan Li[6]等开发了一种基于小波去噪的反向传播神经网络模型预测生物气溶胶浓度，并使用定点仪器采集的数据集对模型进行了实验验证，证明反向传播神经网络模型在预测大气生物气溶胶浓度方面的有效性。Gholamreza Goudarzi等[7]通过建立人工神经网络模型预测了城市市区的花粉总浓度，并评估整个城市花粉在空气中的分布状态以及花粉浓度与环境参数之间的相关性。因此，通过建立预测模型对影响生物气溶胶含量的众多因素进行综合分析，可以预测未来一段时间内的生物气溶胶含量分布状况，为大气环境污染控制和治理提供科学依据[8-14]。

鉴于生物气溶胶潜在的威胁及对实时远距离监测的迫切需求，将人工智能技术应用于生物气溶胶浓度预测已成为大气科学遥感探测领域新的研究方案。本文通过建立融合PSO（粒子群算法）、GA（遗传算法）以及SVM（支持向量机算法）的生物气溶胶浓度预测模型，以大气环境参数为输入量、生物气溶胶浓度廓线为输出量，采用PSO-GA优化持向量机，自动调整系统预测模型参数，达到全局搜索最优解，提高系统预测模型的稳定性，实现大气生物气溶胶浓度预测。

1 生物气溶胶浓度预测模型建立

1.1 基于荧光激光雷达的生物气溶胶探测原理

荧光激光雷达探测技术通过发射特定波长的激光脉冲，激发大气中的目标物质产生荧光信号，通过接收、分析这些信号来识别和量化目标物质，如真菌、细菌、病毒、孢子、花粉等。通过激光诱导目标荧光、信号接收与处理、数据分析过程，获得目标物质的浓度、分布状况和化学组成等信息，实现探测功能。大气生物气溶胶荧光激光雷达回波信号功率方程如下所示：

$$P(R)=\frac{1}{2}E_0\cdot c\frac{A_0}{R^2}\cdot e^{-[\alpha_1(\lambda_1,R)+\alpha_2(\lambda_2,R)]\cdot R}\cdot \xi(R)\cdot\Delta\lambda\cdot S\cdot N_{\mathrm{Bio}}(R)\cdot\eta_0 \quad , \quad (1)$$

式中，$P(R)$表示为荧光信号强度，E_0为激光器输出的激光脉冲，c为光速（$m\cdot s^{-1}$），A_0为雷达系统所使用的望远镜的接受面积（m^2），R为激光雷达探测距离（km），$\alpha_1(\lambda_1,R)$为激发波长在大气中总消光系数（km^{-1}）、$\alpha_2(\lambda_2,R)$为荧光波长在大气中总消光系数（km^{-1}），气溶胶和大气分子消光系数两部分组成激光脉冲在大气中总的消光系数，式中的$e^{-[\alpha_1(\lambda_1,R)+\alpha_2(\lambda_2,R)]\cdot R}$为激发波长和荧光波长在大气中的透过率，$\lambda_1$为激发波长（nm），$\lambda_2$为荧光波长（nm），$\xi(R)$为几何重叠因子，$\Delta\lambda$为滤光片带宽，$S$为荧光非弹性散射的有效截面积，$N_{\mathrm{bio}}(R)$为生物气溶胶粒子浓度，$\eta_0$为整个光学系统对荧光波长的接收效率。表1中所示为荧光激光雷达系统的部分主要参数。当荧光激光雷达探测到大气中确定的荧光回波信号范围内强度值，由式（1）可得到生物气溶胶浓度的反演表达式如下所示：

$$N_{\mathrm{Bio}}(R)=\frac{2P(R)R^2}{E_0\cdot c\cdot A_0\cdot e^{-[\alpha_1(\lambda_1,R)+\alpha_2(\lambda_2,R)]\cdot R}\cdot\xi(R)\cdot\Delta\lambda\cdot S\cdot\eta_0} \quad . \quad (2)$$

图1所示为荧光激光雷达基本结构及工作原理。

1.2 机器学习预测模型算法

1.2.1 支持向量机

支持向量机（SVM）是一种源于统计学习理论的机器学习方法，具备卓越的分类与预测性能。对于训练样本集(x_i,y_i)，其中$x_i\in R^n$为输入向量，$y_i\in R^n$为输出向量，通过引入非线性映射$\varphi(x_i)$，实现样本空间向高维的变化，对样本进行线性函数拟合，而对于非线性样本，估计函数$f(x)$转化为：

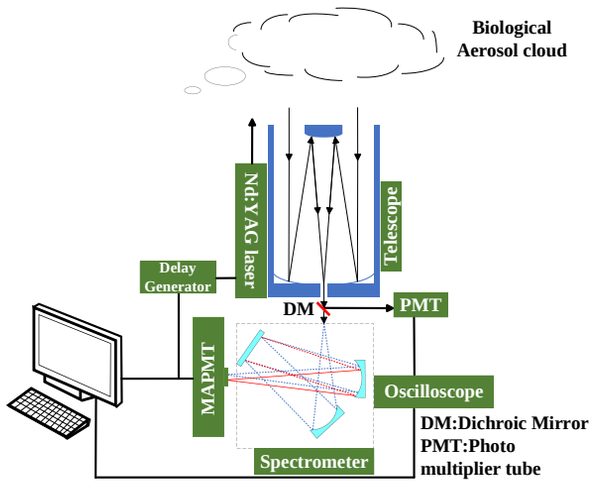


图1　荧光激光雷达系统工作原理图

Fig. 1　Working principle diagram of fluorescence lidar system

表1　荧光激光雷达系统部分主要参数

Table 1　Specifications of the fluorescence lidar

Definition	Reference value
Pulse energy	60 mJ
Field of view of telescope	0. 5 mrad
Quantum efficiency of the PMT	0. 2
Transmission of receiving optical train	0. 3
Filter bandwidth	10 nm
Diameter of telescope	25 cm
Laser wavelength	266 nm
Fluorescence wavelength	310~440 nm
Pulse repetition frequency	10 Hz
Detector frequency bandwidth	5 MHz

$$f(x) = W \times \phi(x) + b \quad , \quad (3)$$

式中W为权重系数,b为函数偏置项,$\varphi(x)$为非线性映射函数,作用是将样本映射到高维空间。

支持向量机引入估计函数解决非线性回归问题,假设训练样本在一定精度下进行线性拟合,向拟合中引入松弛因子后,则寻找偏置项与权重系数的问题转化为:

$$\min Q = \frac{1}{2}\| \omega \|^2 + C\sum\nolimits_{i=1}^{n}(\delta_i + \delta_i^*) \quad , \quad (4)$$

$$s.t.\begin{cases} y_i - \omega\phi(x_i) - b \leqslant \varepsilon + \delta_i^* \\ -y_i + \omega\phi(x_i) + b \leqslant \varepsilon + \delta_i^* \\ \delta_i \geqslant 0, \delta_i^* \geqslant 0 \end{cases} \quad , \quad (5)$$

其中,$\|\omega\|^2$为用于反映模型复杂程度的描述函数;c(cost)为惩罚因子,ε为拟合误差ε>0;δ_i、δ_i^*为松弛因子,y_i为输出向量。在满足约束条件的情况下引入Lagrange乘子,优化问题转为:

$$\max \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j K(x_i, x_j) \quad , \quad (6)$$

$$s.t.\begin{cases} \sum_{i=1}^{n} a_i y_i = 0 \\ a_i \geqslant 0 \end{cases} \quad , \quad (7)$$

式中,$K(x_i, x_j)$为核函数,a_i、a_i^*为拉格朗日乘子。

根据KKT(Karush-Kuhn-Tucker)条件求出b,SVM模型函数:

$$f(x) = \sum_{i,j=1}^{n}(a_i - a_i^*)K(x_i, x_j) + b \quad . \quad (8)$$

在本文所设计的模型中,我们选择了径向基核函数(RBF)为核函数,表达式为:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2g^2}\right) \quad , \quad (9)$$

其中,g(gamma)为核函数参数。

1. 2. 2　遗传算法

遗传算法(GA)是一种搜索算法,其灵感来源于自然选择和遗传学原理,用于解决优化和搜索问题。其基本流程如下[15-19]:

(1)定义适应度函数:依据问题的目标来设计适应度函数,经运算推理与迭代后,所得误差能够反映个体的优劣状况,适应度高的个体将被保留。

(2)编码与初始化种群:将性状转换为基因,相邻节点都有其对应的权值,该权值则可以成为个体的基因编码,其表达式为:

$$W = [w_{11}, w_{12}, \ldots w_{nn}] \quad , \quad (10)$$

每个个体由n^2个基因组成,w_{ii}=0(i=1,2,…n),个体中的w_{1i}组成一行,以此类推,所得的矩阵即为权值矩阵的标准形式,并随机生成一个初始种群。

(3)评估种群:使用适应度函数评估种群中每个个体的适应度。

(4)选择(Selection):个体依据适应度被挑选用于繁殖。适应度较高的个体能够将其遗传因子传递至下一代,由于适应度值的差异会直接对算法性能产生影响,故而其被选中的概率更大。

(5)交叉(Crossover):随机挑选两个个体,然后通过交换这两个父代个体的部分基因信息来生成两个子代个体。这一操作能够增加种群的多样性,对遗传算法更迅速地找到最优解有所助益。

(6)变异(Mutation):染色体的某些位会以一定概率发生改变,其变量可能以较小概率或者步长转变,这有一定概率提升种群的多样性。

(7)新种群形成:将交叉和变异后的个体加入到新种群中。

(8)迭代:重复评估、选择、交叉、变异的步骤,直到达到预设的停止条件,如迭代次数、适应度阈值或种群收敛。

(9)解码与结果:在最后一代中,找出适应度最高的个体,解码为问题的实际解,并作为最优解输出。

1.2.3 粒子群算法

粒子群算法(PSO)属于优化算法,其模拟自然界群体智能行为。通过粒子群算法自身的迭代过程,可求出问题的最优解,其中个体最优解记为*pBest*,全局最优解记为*gBest*。粒子速度的更新涉及粒子的惯性、认知分量(取决于个体极值)以及社会分量(取决于全局极值)[20-22]。PSO的速度更新公式和位置更新公式如下所示:

$$v_i(t+1)=\omega_g v_i(t)+c_1 r_1[pBest_i(t)-x_i(t)]+c_2 r_2[gBest_i(t)-x_i(t)] \quad , \quad (11)$$

$$x_i(t+1)=x_i(t)+v_i(t+1) \quad , \quad (12)$$

式中:ω_g为惯性权重,c_1、c_2为调整最大步长的学习因子,r_1、r_2为随机函数,$v_i(t)$、$v_i(t+1)$为t、$t+1$时刻的速度,$x_i(t)$、$x_i(t+1)$为t、$t+1$时刻的位置。

2 基于PSO-GA-SVM的生物气溶胶预测模型

2.1 数据组成

对于大气生物气溶胶浓度的预测,本文采用气象局发布的环境参数数据以及由荧光激光雷达反演得到的生物气溶胶浓度数据训练预测模型,选取温度、湿度、PM2.5、PM10、CO_2、SO_2、NO_2、O_3、风速等相关环境参数为输入数据,不同高度的生物气溶胶浓度为输出数据,实现生物气溶胶浓度廓线的预测,其中11个高度对应范围分别是0.3 km、0.35 km、0.4 km、0.45 km、0.5 km、0.55 km、0.6 km、0.65 km、0.7 km、0.75 km、0.8 km,分辨率为0.05 km;输入的不同环境参数变量以及输出变量如表2所示。

2.2 数据处理

数据预处理有利于提高预测模型的性能与运算速度,归一化可加快优化算法收敛速度,让所有输入特征处于同一尺度,便于优化算法找到最小值,减少模型在训练数据上的过拟合,提升模型泛化能力。文中采用标准归一化技术,分子得到的值有正有负,实现不同的参数朝着不同方向调整直至到达最优解,减少迭代次数,更新尽可能沿着最优路径进行,其公式如下所示。

表2 PSO-GA-SVM具体输入输出信息

Table 2 PSO-GA-SVM specific input and output information

输入/输出	输入参数编号	变量
输入	1	PM2.5
	2	PM10
	3	CO
	4	NO_2
	5	O_3
	6	SO_2
	7	温度
	8	湿度
	9	风级
	10	AQI
输出	1~11	11个不同高度的生物气溶胶浓度

$$x_i^s=\frac{x_i-E(X)}{\sqrt{Var(X)}} \quad , \quad (13)$$

其中:

$$E(X)=\frac{1}{n}\sum_{i=1}^{n}x_i \quad , \quad (14)$$

$$Var(X)=\frac{1}{n}\sum_{i=1}^{n}(x_i-E(X))^2 \quad , \quad (15)$$

式中,$X=\{x_i\}$,$i=1,\cdots n$,$E(X)$为样本均值,$Var(X)$为样本方差。

生物气溶胶浓度的预测是通过环境参数数据训练得到,预测值与真实值之间有误差,判断生物气溶胶浓度预测模型的性能是否满足预测要求要对这些差距进行分析,即可判断出算法的预测效率,本文中选用的误差分析标准为相对误差法,其公式为:

$$E_r=\left|\frac{E-E_a}{E}\right|\times 100\% \quad , \quad (16)$$

式中,E_r为相对误差,E为预测值,E_a为真实值。

2.3 PSO-GA-SVM预测模型

将GA和PSO与SVM进行融合,使GA与PSO干预SVM寻找最佳的惩罚系数(c)和核函数参数(g),提高数据预测的准确性[23-24]。图2显示为基于PSO-GA-SVM融合算法的预测模型流程图。

(1)数据预处理,数据不同特征可能具有不同的尺度和单位,这会影响机器学习算法的训练效果,归一化处理是建立模型的重要步骤之一,对数据集进行归一化处理,并分成训练集与测试集。

(2)函数获取训练数据集的标签和特征矩阵,这些数据将作为函数的基础输入,用于训练模型并评估其预测能力,同时需用于超参数调优,包括惩

罚参数c和核函数参数g的搜索范围、步长,以及交叉验证的折数等,从而使得模型在面对未知数据时具有更好的泛化能力和预测精度。

(3)使用PSO算法优化SVM的超参数c和g。PSO通过模拟群体智能搜索最优解,在粒子群的选项结构中,引入PSO定义粒子群的选项结构(如学习因子、最大代数、种群大小等),这些参数决定了算法的收敛性与全局搜索能力,使模型获得较优的c和g。

(4)引入GA对粒子群进行处理进一步优化c和g的值,GA与PSO相结合,形成一个混合优化框架。PSO首先提供一个较好的初始解,GA随后在此基础上进行进一步搜索和优化,确保能够避免局部最优解,并探索到全局最优解,从而获得更加精确的c和g参数。

(5)在(3)与(4)中的优化参数,GA与PSO算法能够实现自动优化参数,根据优化后的c和g值,定义SVM的参数,包括惩罚系数、核函数参数、SVM类型(e-SVR)、核函数类型(RBF核)、损失函数参数。

(6)采用数据集进行训练,构建PSO-GA-SVM预测模型。

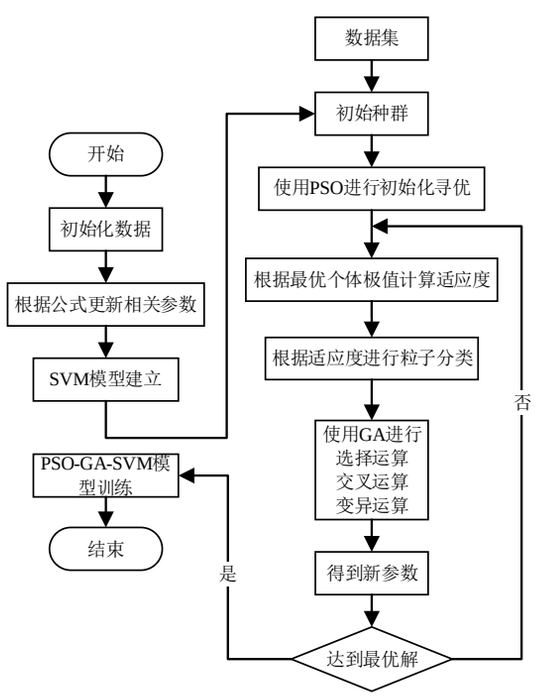


图2　PSO-GA-SVM预测模型流程图

Fig. 2　Flowchart of PSO-GA-SVM prediction model

3　大气生物气溶胶浓度廓线预测实验

3.1　SVM参数自动搜索

支持向量机(SVM)模型的参数调优过程中,网格搜索(Grid Search)是一项核心策略,用于精准定位最有效的超参数组合,过程涉及全面探索预设的超参数空间,对每个候选组合执行交叉验证(CV),以评估模型的泛化能力。通过比较不同组合的CV平均性能指标,能够甄选出表现最优的参数设定。图3和图4是在训练时对不同测试数据进行预测时所得的网格搜索的参数结果3D视图与等高线视图。在网格搜索过程中,首先设定一组可能的超参数值,其中SVM的c参数值为0.1至100,核函数g值为0.000 1至10,每一对参数组合使用交叉验证训练SVM模型,评估其性能(如准确率等)。通过计算所有参数组合的验证结果,网格搜索选择出最佳的参数组合,对于数据较少的数据集来说,网格搜索处理效率尚可,但是对于过大的数据量来说,网格搜索会带来巨大的计算压力,同样当参数粒度不够细时可能会错过最优解。

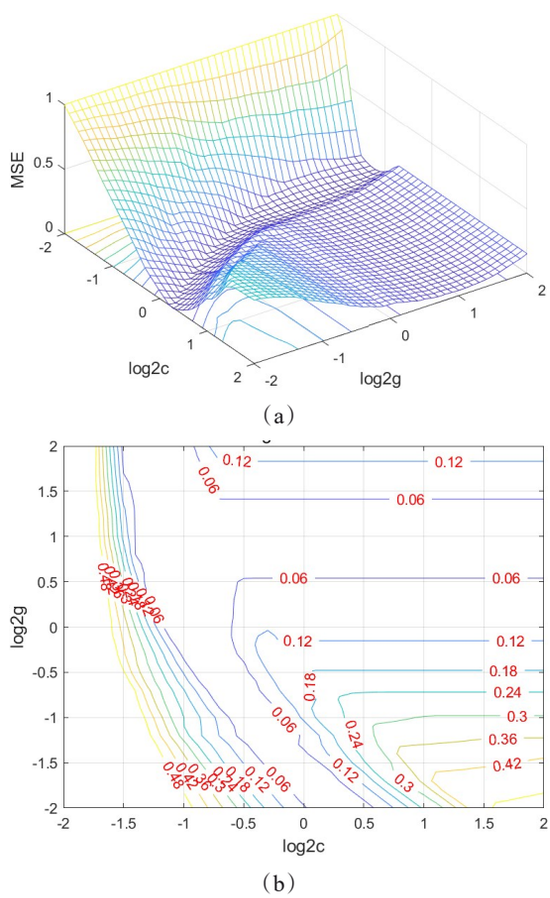
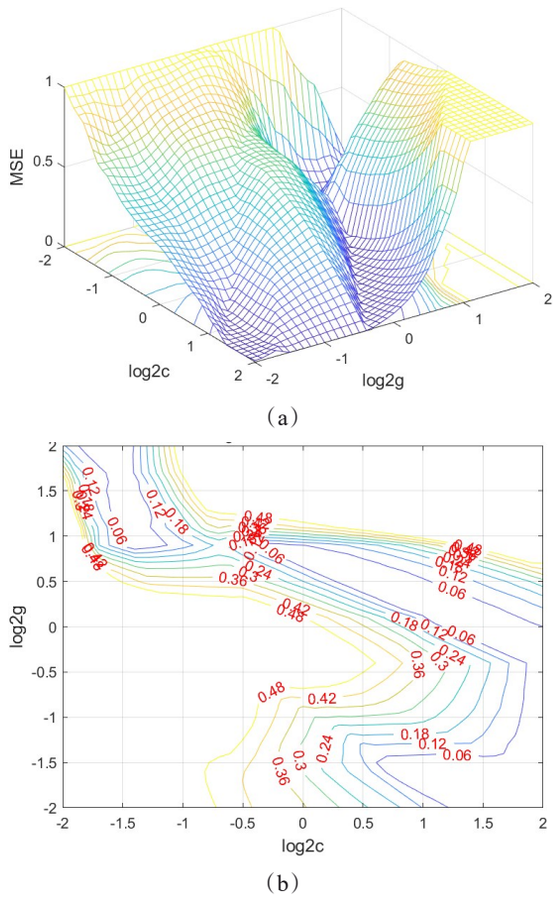


图3　基于测试数据集SVM运行结果的结果图:(a)相对误差3D视图;(b)等高线图

Fig. 3　Result plots based on the running results of SVM on the test dataset: (a) 3D view of relative error; (b) contour map

3.2　PSO-GA参数优化

为了避免网格搜索可能带来的错过最优解的



(a)

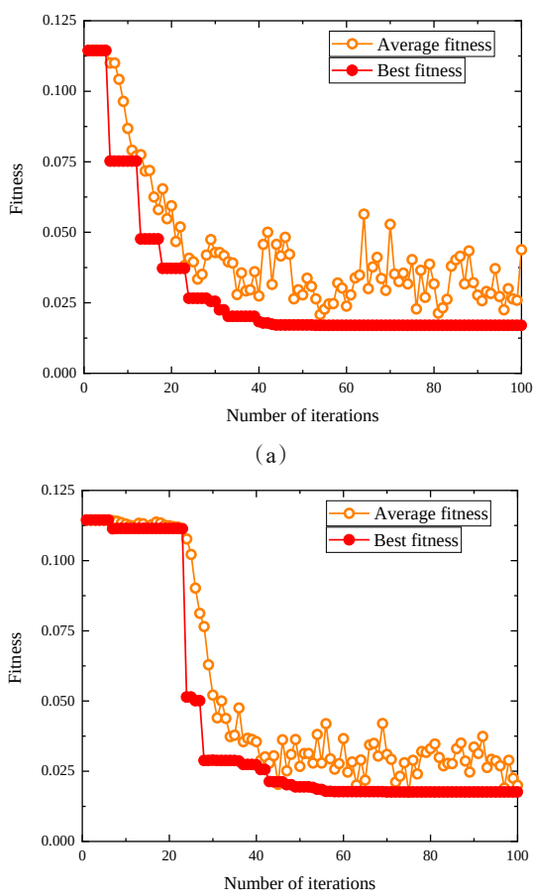
(b)

图4 基于测试数据集SVM运行结果的结果图:(a)相对误差3D视图;(b)等高线图

Fig. 4 Result plots based on the running results of SVM on the test dataset: (a) 3D view of relative error; (b) contour map

问题,本文引入PSO-GA算法对模型进行优化,通过对数据集分成的种群不断进行优化,形成新参数,达到最优数据集解,以下是模型训练时对测试数据进行预测时所得的粒子群种群与遗传种群的最佳适应度与平均适应度的曲线[25-28]。对于不同优化算法下该曲线会发生变化以达到最佳的效果,由图5(a)GA种群适应度曲线来看,种群初期波动较大,随着迭代的进行,最佳适应度曲线将趋于平稳,GA依赖于选择和交叉操作,因此适应度曲线通常呈现出渐进式的改善,GA中的变异操作可能导致适应度值出现较大的波动或突变,而从图5(b)可以看出,经过PSO优化后的GA种群适应度曲线,变化较为稳定且迅速,通过粒子群的协同搜索,PSO能够引导搜索向潜在的全局最优解逼近,从而避免GA中由于交叉和变异引发的搜索停滞现象。

由图6(a)PSO种群适应度曲线来看,由于粒子之间的协同作用,PSO的适应度曲线相对平滑,避免了像GA那样由于交叉和变异操作带来的剧烈波动,PSO具有较强的全局搜索能力,但它也可能在某些情况下陷入局部最优,特别是在搜索空间复杂、解的分布较为零散时。此时,适应度曲线可能会出现较长时间的停滞,虽然粒子会做出微小调整,但整体的适应度提升缓慢,图6(b)可以看出PSO和GA结合后,能够通过GA引入的多样性和全局搜索能力有效避免PSO陷入局部最优,并改善算法的收敛性和稳定性。GA帮助PSO在搜索过程中避免过早收敛,提高了群体的多样性,进而使得平均适应度曲线更加趋近于最佳适应度曲线。可知,两者相互结合可以显著提高模型寻找最优解时的稳定性与准确性[29-30]。



(a)

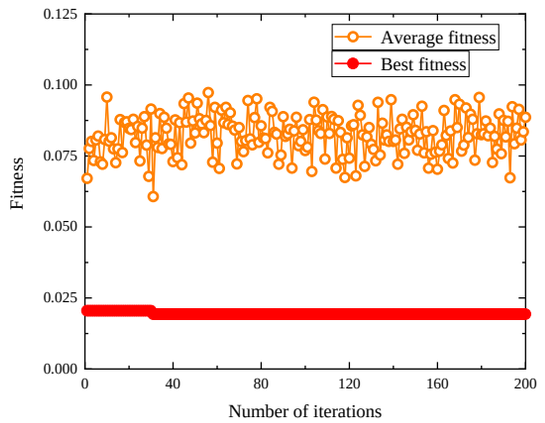
(b)

图5 不同算法的遗传种群适应度图:(a) GA-SVM算法遗传种群适应度;(b) PSO-GA-SVM融合算法的遗传种群适应度

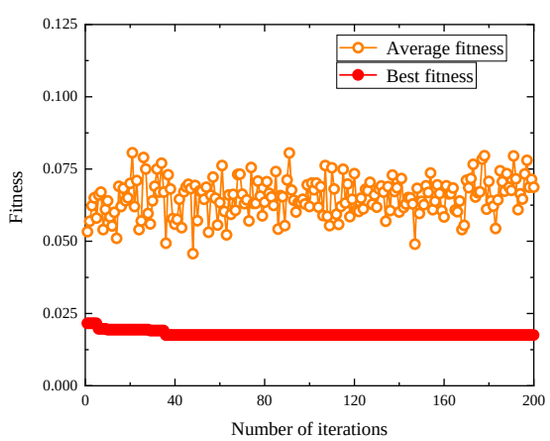
Fig. 5 Genetic population fitness maps for different algorithms: (a) genetic population fitness of GA-SVM algorithm; (b) genetic population fitness of PSO-GA -SVM algorithm

3.3 网络模型预测结果

采用确定好的模型进行训练并与不同算法训



(a)



(b)

图6　不同算法的粒子群种群适应度图:(a) PSO-SVM算法的粒子群种群适应度;(b) PSO-GA-SVM算法的粒子群种群适应度

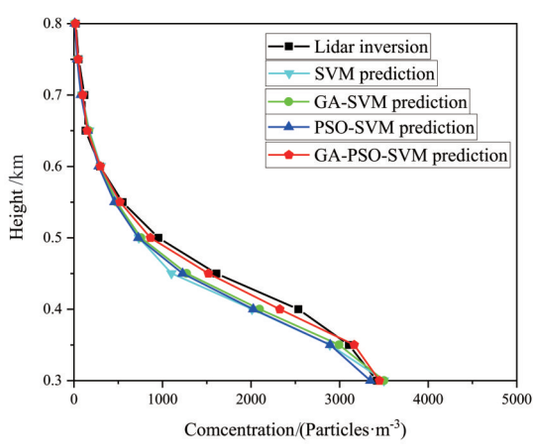
Fig. 6　Plot of particle swarm population fitness for different algorithms: (a) PSO-SVM algorithm for particle swarm population fitness; (b) PSO-GA-SVM algorithm for particle swarm population fitness

练好的模型进行对比,采用500组数据进行模型训练,前80%的数据组作为训练样本以达到良好的训练效果,其余20%数据组为测试样本。最后利用训练好的模型进行预测,并通过预测数据与激光雷达反演得到的数据进行对比,进一步验证模型的可靠性。由图7和图9中的预测结果图可以看出,在不同优化方式下,预测值与真实值的对比研究中,在生物气溶胶预测方面展现出更优的整体拟合性,其预测结果相较于其他模型更接近真实值。在图7和图9的相对误差能够清楚地看出PSO-GA-SVM预测模型整体误差更小。由图8和图10显示同一组数据分别使用SVM预测、GA-SVM预测、PSO-SVM预测与PSO-GA-SVM预测结果的相关系数图中可以看出PSO-GA-SVM预测模型的效果要优于其他三种算法的预测效果,拟合效果更佳。表3所示为PSO-GA-SVM模型相关参数。

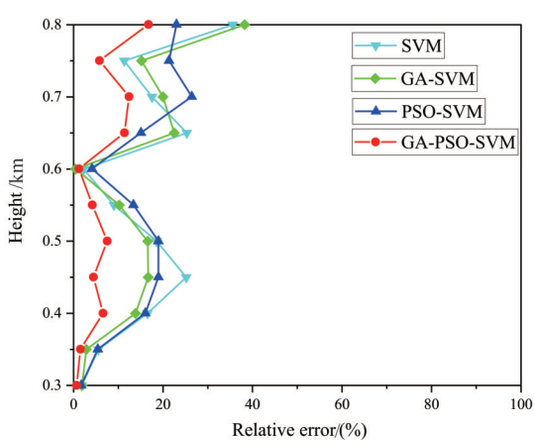
表3　PSO-GA-SVM模型相关参数

Table 3　PSO-GA-SVM model related parameters

参数	值
学习因子1	1.5
学习因子2	1.7
进化次数	200
种群规模	20
损失函数	0.1
交叉验证参数	3
速率更新速度弹性系数	1
种群更新速度弹性系数	2



(a)



(b)

图7　基于测试数据的不同算法模型预测图:(a)预测结果;(b)预测结果相对误差

Fig. 7　Prediction plots of different algorithmic models based on test data: (a) projected results; (b) relative error in projected results

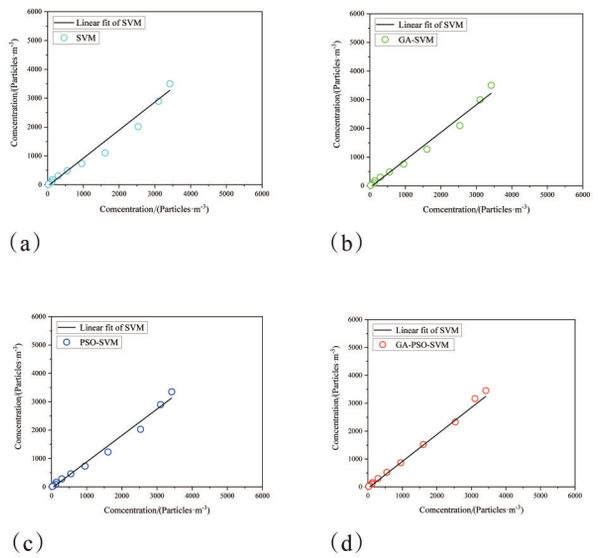


图8 不同算法模型预测结果的相关性图:(a) SVM算法预测相关性图;(b) GA-SVM算法预测相关性图;(c) PSO-SVM算法预测相关性图;(d) PSO-GA-SVM算法预测相关性图

Fig. 8 Correlation plot of the prediction results of different algorithmic models: (a) SVM algorithm predicted correlation plot; (b) GA-SVM algorithm predicted correlation plot; (c) PSO-SVM algorithm predicted correlation plot; (d) PSO-GA-SVM algorithm predicted correlation plot

在不同算法下平均误差信息如表4所示。

对于SVM模型而言,经不同算法优化后,主要体现为相关参数的改变,其中惩罚因子c和核函数参数g最为关键。c取值过大或过小都可能致使模型泛化能力下降,当选择径向基函数(RBF)作为核函数(kernel)时,g是该函数自带的参数,它隐含地决定了数据映射到新特征空间后的分布情况,g越大,支持向量越少;g越小,支持向量越多,而支持向量的个数对训练与预测速度存在影响。通过引入GA算法对支持向量机的核函数、惩罚系数以及径向基础核参数进行编码,从而获取最优值,再利用所得最优解的参数进行预测;另外,引入PSO算法,借助改变学习因子等参数的方式加快算法收敛速度,不易陷入局部最优。表4结果显示,GA算法和PSO算法单独优化后的预测效果均优于基础的SVM算法,而将GA与PSO算法相结合的预测结果较前三种算法均有所提升,这一结果证实了PSO-GA-SVM算法相较于SVM具有更好的预测性能。此外,生物气溶胶浓度廓线为垂直分布状态,可选取不同高度的环境参数进行模型训练,提高预测精准度;为了提升模型在其他环境特征参数变动情况下的预测准确度,可进一步提出更好的预测算法,使预测模型有更强的学习能力,并同时提升模型训练数据的多样性和复杂性。

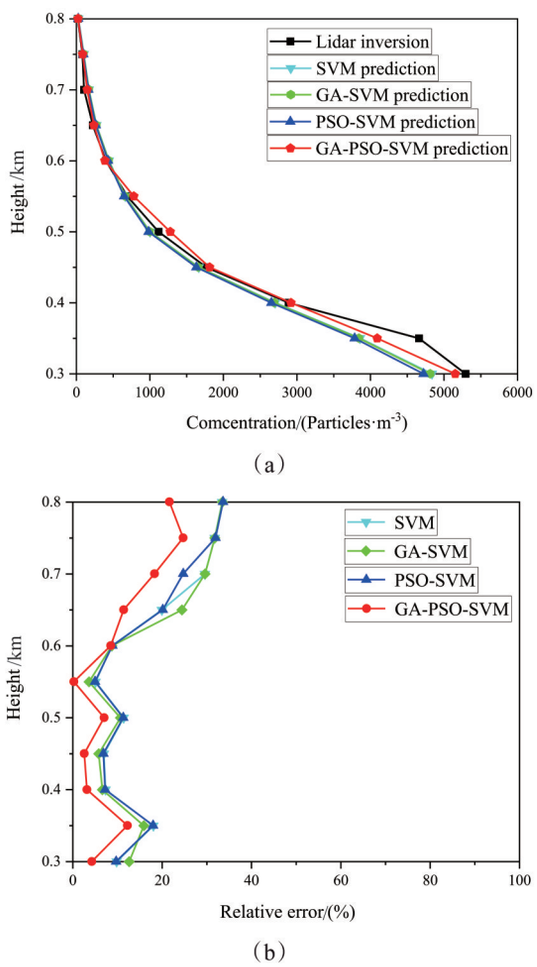


图9 基于测试数据的不同算法模型预测图:(a)预测结果;(b)预测结果相对误差

Fig. 9 Prediction plots of different algorithmic models based on test data: (a) forecast results; (b) relative error in forecast results

4 总结

本文研究了一种基于GA-PSO-SVM融合算法的机器学习预测模型,结合荧光激光雷达探测技术,将其应用于大气生物气溶胶浓度预测。基于生物气溶胶浓度受大气环境参数的显著影响,模型将温度、湿度、PM2.5、PM10、CO_2、SO_2、NO_2、O_3、风速等相关参数数据作为输入量,激光雷达测得的生物气溶胶浓度作为输出变量进行训练;通过结合粒子群算法(PSO)和遗传算法(GA)对支持向量机(SVM)模型的惩罚系数c和核函数参数g进行优化,避免了SVM在传统网格搜索中容易陷入局部最优解的问题,从而提高了模型的预测性能。实验结果显示,

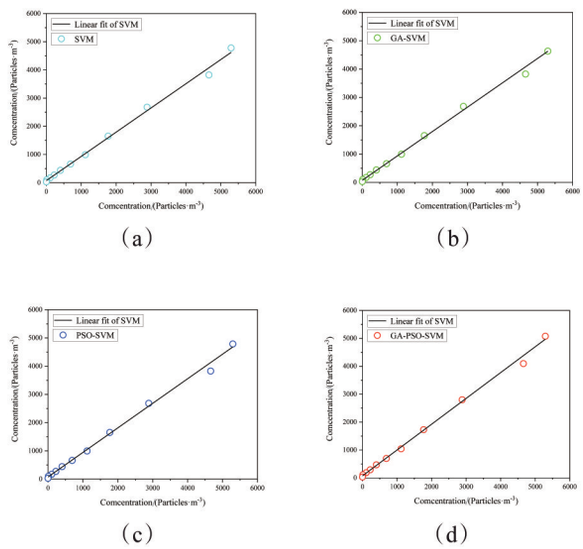


图10　不同算法模型预测结果的相关性图:(a) SVM算法预测相关性图;(b) GA-SVM算法预测相关性图;(c) PSO-SVM算法预测相关性图;(d) PSO-GA-SVM算法预测相关性图

Fig. 10　Correlation plot of the prediction results of different algorithmic models: (a) SVM algorithm predicted correlation plot; (b) GA-SVM algorithm predicted correlation plot; (c) PSO-SVM algorithm predicted correlation plot; (d) PSO-GA-SVM algorithm predicted correlation plot

表4　PSO-GA-SVM平均误差信息

Table 4　PSO-GA-SVM average error

优化算法	平均误差(%)
SVM	15.335
GA-SVM	14.411
PSO-SVM	14.958
PSO-GA-SVM	6.5931

PSO-GA-SVM预测模型的平均预测误差为6.5931%,明显低于PSO-SVM(14.958%)、GA-SVM(14.411%)及传统SVM(15.335%)模型的预测误差。结果表明,GA与PSO相结合的优化算法相较于单一GA或PSO优化算法具有更优的种群适应度,从而显著提升了PSO-GA-SVM模型的预测准确性。

References

[1] Kim K S, Lee I, Lee J. Synergetic chemo-mechano antimicrobial puncturable nanostructures for efficient bioaerosol removal[J]. BioChip Journal, 2024, 18(3): 439-452.

[2] Da Silva J C R, Clara D S L M, Prates K V M C, et al. Characterization of indoor airborne particulate matter and bioaerosols in wood-fired pizzeria kitchens [J]. Discover Environment, 2024, 2(1):107.

[3] Nasser N I, Al-Hadrawi M K, Oleiwi S A, et al. The diversity in dust fungal spores concentration at four districts of Al-Najaf environment and their potential correlation with asthma[J]. Journal of Pure & Applied Microbiology, 2019, 13(4): 2169-2176.

[4] Liu H, Zhang X, Zhang H, et al. Effect of air pollution on the total bacteria and pathogenic bacteria in different sizes of particulate matter [J]. Environmental Pollution, 2018, 233(Feb.): 483-493.

[5] Liggio J, Li S M. Organ sulfate formation during the uptake of pinon aldehyde on acidic sulfate aerosols [J]. Geophysical Research Letters, 2006, 33(13): 338-345.

[6] Li X N, Cheng X, Wu W J. An improved wavelet de-noising-based back propagation neural network model to forecast the bioaerosol concentration [J]. Aerosol Science and Technology, 2021, 55(1/3): 352-360.

[7] Goudarzi G, Birgani Y T, Assarehzadegan M A, et al. Prediction of airborne pollen concentrations by artificial neural network and their relationship with meteorological parameters and air pollutants [J]. Journal of Environmental Health Science and Engineering, 2022, 20(1): 251-264.

[8] Rao Z M, He T Y, Hua D X, et al. Remote sensing of particle mass concentration using multi-wavelength Lidar [J]. Spectroscopy and Spectral Analysis, 2018, 38(04): 1025-1030.
饶志敏,何廷尧,华灯鑫,等.多波段激光雷达颗粒物质量浓度探测方法[J]. 光谱学与光谱分析, 2018, 38(04): 1025-1030.

[9] Rao Z M, Hua D X, He T Y, et al. Research and analysis on lidar performance with intrinsic fluorescence biological aerosol measurements [J]. Acta Physica Sinica, 2016, 65(20): 36-43.
饶志敏,华灯鑫,何廷尧,等.基于本征荧光的生物气溶胶测量激光雷达性能[J]. 物理学报, 2016, 65(20): 36-43.

[10] Rao Z M, Hua D X, He T Y, et al. Performance analysis of double wavelength fluorescence Lidar in detecting atmospheric biological aerosols [J]. Spectroscopy & Spectral Analysis, 2017, 37(09): 2804-2808.
饶志敏,华灯鑫,何廷尧,等.双波长荧光雷达探测大气生物气溶胶的性能分析[J]. 光谱学与光谱分析, 2017, 37(09): 2804-2808.

[11] Shoshanim O, Baratz A. Daytime measurements of bioaerosol simulants using a hyperspectral laser-induced fluorescence LIDAR for biosphere research [J]. Journal of Environmental Chemical Engineering, 2020, 8(5): 104392.

[12] Yang T, Wang X, Jiang L, et al. Mobility, bacterial hosts, and risks of antibiotic resistome in submicron bioaerosols from a full-scale wastewater treatment plant [J]. Journal of Environmental Management, 2024(Feb. 1): 119771.

[13] Lin B L, Tokai A, Nakanishi J. Approaches for establishing predicted-no-effect concentrations for population-level ecological risk assessment in the context of chemical substances management [J]. Environmental Science & Technology, 2005, 39(13): 4833-4840.

[14] Kondo A. Developing an automatic asbestos detection method based on a convolutional neural network and support vector machine [J]. Applied Sciences, 2024, 14(20): 9408

[15] Ibrir A, Kerchich Y, Hadidi N, et al. Evaluation and pre-

diction of the effects of the dispersion of (VOCs) on the population in urban air using ANSYS CFX[J]. Algerian Journal of Environment Science and Technology, 2021,7(4):2135-2142.
[16] Karadurmus E, Berber R. Dynamic simulation and parameter estimation in river streams[J]. Environmental Technology, 2004, 25(4): 471-479.
[17] AlMohimeed A, Shehata M, Rashidy E N, et al. ViT-PSO-SVM: cervical cancer predication based on integrating vision transformer with particle swarm optimization and support vector machine[J]. Bioengineering, 2024, 11(7): 729-729.
[18] Eseye A T, Zhang J, Zheng D. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information[J]. Renewable Energy, 2017, 118(APR.): 357-367.
[19] Velikoslavinskii S D, Kotov A B, Sal'Nikova E B, et al. Age of the ilikan sequence from the Stanovoi complex of the Dzhugdzhur-Stanovoi superterrane, central-Asian foldbelt[J]. Doklady Earth Sciences, 2011, 438(Pt.1): 612-616.
[20] Fan M, Sharma A. Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0[J]. International Journal of Intelligent Computing and Cybernetics, 2021, 14(2): 145-157.
[21] Naveen M NA.V, R. R M V, et al. Multiobjective design of water distribution networks using modified NSGA-II algorithm[J]. Water Supply, 2023, 23(3): 1220-1233.
[22] Chang Y, Xue Y, Zhang Y, et al. PCB defect detection based on PSO-optimized threshold segmentation and SURF features[J]. Signal, Image and Video Processing, 2024, 18(5): 4327-4336.
[23] Suggala R K, Krishna M V, Swain S K. Discover the new factor for dengue fever outbreaks and predicted using Bayes Network-PSO (BN-PSO)[J]. Journal of Information Science and Engineering: JISE, 2023, 39(6): 1383-1401.
[24] Klempner S J, Chao J, Bailey M, et al. Genomic alterations (GA) predicted to confer lack of benefit from trastuzumab in advanced esophagogastric cancers (EGC): Analysis of 527 HER2 -amplified (HER2 amp) cases[J]. Journal of Clinical Oncology, 2018, 36(4_suppl): 44-44.
[25] Ahmad A, Yadav A K, Singh A, et al. A hybrid RSM-GA-PSO approach on optimization of process intensification of linseed biodiesel synthesis using an ultrasonic reactor: Enhancing biodiesel properties and engine characteristics with ternary fuel blends[J]. Energy, 2024(Feb. 1): 288.
[26] Shi G, Chan C S, Li W J, et al. Mobile human airbag system for fall protection using MEMS sensors and embedded SVM classifier[J]. IEEE Sensors Journal, 2009, 9(5): 495-503.
[27] Lee C P, Leu Y, Yang W N. Constructing gene regulatory networks from microarray data using GA/PSO with DTW[J]. Applied Soft Computing, 2012, 12(3): 1115-1124.
[28] Gulande P, Awale R N. SVM-ABC based cancer microarray (gene expression) hybrid method for data classification[J]. Computational Intelligence, 2023, 39(6): 1054-1072.
[29] Rebentrost P, Mohseni M, Lloyd S. Quantum support vector machine for big data classification[J]. Physical Review Letters, 2014, 113(13): 130503.
[30] Srivastava N, Vignesh D, Saxena N. Investigation of artificial neural network performance in the aerosol properties retrieval[J]. Journal of Water and Climate Change, 2021, 12(6): 2816-2834.
[31] Wei Y, Zhao J, Yang Z, et al. Fault detection method for flexible DC grid based on CEEMDAN multiscale entropy and GA-SVM[J]. Electrical Engineering, 2024, 106(5): 6171-6183.
[32] Zhang J, Yu M, He K. Research on high-efficiency optimization algorithm applied to near-field effect error correction[J]. International Journal of RF and Microwave Computer-aided Engineering, 2022, 32(12): 23530.