

杨 浩,张 通,阳苇丽,等. 基于图像处理的雪茄烟叶晾制期间含水率预测模型比较[J]. 江苏农业学报,2023,39(9):1891-1900.  
doi:10.3969/j.issn.1000-4440.2023.09.011

## 基于图像处理的雪茄烟叶晾制期间含水率预测模型比较

杨 浩<sup>1</sup>, 张 通<sup>1</sup>, 阳苇丽<sup>2</sup>, 向 欢<sup>3</sup>, 郭仕平<sup>4</sup>, 刘晓丽<sup>1</sup>, 张洪淇<sup>1</sup>, 刘 雷<sup>1</sup>,  
刘雅洁<sup>1</sup>, 杨兴有<sup>4</sup>, 曾淑华<sup>1</sup>

(1.四川农业大学农学院,四川 成都 611130; 2.四川省烟草公司达州市公司,四川 达州 635000; 3.四川省烟草公司德阳市公司,四川 德阳 618400; 4.四川省烟草公司,四川 成都 610017)

**摘要:** 为探索雪茄烟叶晾制期间的水分变化规律并实现含水率的快速准确预测,在雪茄烟叶各个晾制阶段拍摄数字图像,同时以杀青烘干法测定烟叶含水率。提取图像中的颜色和纹理特征作为初始特征,经正交偏最小二乘(OPLS-DA)确定优选特征。分别以初始特征和优选特征为输入,含水率为输出,建立前馈神经网络(BPNN)、遗传算法优化前馈神经网络(GA-BPNN)、支持向量机(SVM)、遗传算法优化支持向量机(GA-SVM)、极限学习机(ELM)、粒子群算法优化极限学习机(PSO-ELM)模型。结果表明,①随着晾制时间推移,烟叶含水率逐渐降低,各阶段含水率差异显著;②优选特征建立的 GA-SVM 对全晾制阶段的含水率整体预测能力相较于其他模型表现最佳,决定系数( $R^2$ )和均方根误差(RMSE)分别为0.969 3、0.044 7;③优选特征建立的 GA-SVM 对各晾制阶段的含水率预测准确度较高,其中干筋期含水率预测准确度最低,但也高于 87.0%。说明,采用 OPLS-DA 优选的颜色特征、纹理特征建立的 GA-SVM 可准确预测雪茄烟叶晾制期间的含水率。

**关键词:** 雪茄烟叶; 含水率预测; 数字图像; BP 神经网络; 支持向量机; 极限学习机

**中图分类号:** TS453 **文献标识码:** A **文章编号:** 1000-4440(2023)09-1891-10

## Comparison of prediction models for moisture content of cigar tobacco leaves during drying period based on image processing

YANG Hao<sup>1</sup>, ZHANG Tong<sup>1</sup>, YANG Wei-li<sup>2</sup>, XIANG Huan<sup>3</sup>, GUO Shi-ping<sup>4</sup>, LIU Xiao-li<sup>1</sup>,  
ZHANG Hong-qi<sup>1</sup>, LIU Lei<sup>1</sup>, LIU Ya-jie<sup>1</sup>, YANG Xing-you<sup>4</sup>, ZENG Shu-hua<sup>1</sup>

(1. *Agriculture College of Sichuan Agricultural University, Chengdu 611130, China*; 2. *Sichuan Provincial Tobacco Company Dazhou Branch, Dazhou 635000, China*; 3. *Sichuan Provincial Tobacco Company Deyang Branch, Deyang 618400, China*; 4. *Sichuan Provincial Tobacco Company, Chengdu 610017, China*)

**Abstract:** To explore the law of moisture change during the drying period of cigar tobacco leaves and realize fast and accurate prediction of moisture content, digital images were taken at each drying stage of cigar tobacco leaves, and the moisture contents of tobacco leaves was determined by method of curing and drying. The color and texture features in the images were extracted as the initial features, and the optimal features were determined by orthogonal partial least squares-discriminant analysis (OPLS-DA). Models of back

收稿日期:2022-11-24

基金项目:四川省烟草公司科技项目(SCYC202121)

作者简介:杨 浩(1997-),云南昭通人,硕士研究生,研究方向为烟草栽培。(E-mail)2551958014@qq.com

通讯作者:刘雅洁,(E-mail)31917272@qq.com;杨兴有,(E-mail)to-baccoboy@163.com;曾淑华,(E-mail)zshgsp@163.com

propagation neural network (BPNN), genetic algorithm optimized BPNN (GA-BPNN), support vector machine (SVM), genetic algorithm optimized SVM (GA-SVM), extreme learning machine (ELM), particle swarm optimized (PSO)-ELM were established with initial features

and optimal features as input and moisture content as output, respectively. The results showed that, with the drying time increased, the moisture content of tobacco leaves gradually decreased, and the moisture contents between different stages showed significant differences. Compared with other models, the overall prediction ability of moisture content in the whole drying stage by GA-SVM established by optimized characteristics was the best, coefficient of determination ( $R^2$ ) and root mean squared error (RMSE) were 0.969 3 and 0.044 7, respectively. The prediction accuracy of GA-SVM established by the optimized features for each drying stage was high, and the prediction accuracy of moisture content in dry gluten stage was the lowest, but was also higher than 87.0%. The results indicated that, GA-SVM constructed by the color and texture features optimized by OPLS-DA can predict the moisture content of cigar tobacco leaves during drying stage accurately.

**Key words:** cigar tobacco leaves; moisture content prediction; digital images; BP neural network; support vector machine; extreme learning machine

近年来雪茄烟发展空间良好,已成为烟草市场和科研领域的研究重点<sup>[1]</sup>。含水率是雪茄烟叶在晾制期间的重要物理参数,对烟叶质量影响重大,烟叶在晾制期间的温湿度调控、晾制密度等是基于烟叶含水率进行设置的,晾制过程中的凋萎和变黄实际上也是烟叶含水率变化的体现,只有控制好含水率才能使烟叶含有的物质朝着人们期望的方向转化。目前生产上雪茄烟叶含水率的判断往往凭借眼观手摸,主观性强且误差大,定量测定方法如烘箱法<sup>[2]</sup>、快速水分测定法<sup>[3]</sup>、卡尔费休法<sup>[4]</sup>等耗时,且对烟叶有损,难以满足实际需求。

图像处理技术可以量化图像的颜色、纹理等特征,而颜色、纹理特征反映烟叶的变黄、失水、皱缩过程,是含水率预测的主要依据<sup>[5-6]</sup>,图像处理技术结合机器学习的方法因具有快速、实时、准确等优点,近年来在作物含水率预测、病虫害识别、成熟度判别等方面得到广泛应用。前人对烤烟烘烤过程中烟叶含水率的预测进行了初步探索,陈飞程等<sup>[7]</sup>提取烟叶图像的颜色、纹理特征建立前馈神经网络(BPNN)对含水率进行预测,模型 $R^2$ 为0.998 7。杜海娜等<sup>[8]</sup>对烘烤阶段的烟叶图像进行拍摄,采用称质量法计算烟叶失水率,以颜色、纹理特征建立萤火虫算法优化支持向量机(GS-SVM)、遗传算法优化前馈神经网络(GA-BPNN)、极限学习机(ELM),结果所建模型 $R^2$ 均大于0.996 1。李增盛等<sup>[9]</sup>提取烟叶颜色、纹理特征建立遗传算法优化支持向量机(GA-SVM)、粒子群算法优化前馈神经网络(PSO-BPNN)、极限学习机,模型准确率分别为93.27%、89.35%、85.05%。此外,Ghosh等<sup>[10]</sup>提取 Flavia 和 Swedish 2 个植物叶片的颜色、纹理特征,并建立人工神经网络(ANN)对其进行分类识别,2 个数据集的 ANN 准确率均超过 93%。宋佳<sup>[11]</sup>对生菜叶片图

像的颜色、纹理、形状特征进行提取,建立偏最小二乘回归模型(PLSR)含水率预测模型,结果 $R^2$ 为0.902,均方根误差仅为0.302。但基于图像处理的机器学习方法对晾制期间雪茄烟叶含水率的预测目前鲜见报道,前人的研究主要集中于烤烟,而雪茄烟叶的调制方法相比烤烟差异大,调制阶段不同,其含水率也不同,所建模型在雪茄烟叶含水率预测上难以达到预期效果。

目前常用于植物分类和预测研究的机器学习算法有 BPNN、支持向量机(SVM)、ELM 等。其中 BPNN 是一种多隐含层网络,通过梯度下降方式不断修正训练权重,其特点在于逼近数据间的相关性。ELM 则是单隐含层网络,与梯度下降方法不同的是,ELM 采用随机的输入层权重,输出层权重则通过隐含层输出的广义逆矩阵确定,其特点在于学习速度快、泛化能力强。在回归问题上,SVM 通过将不同维度空间特征最大程度地映射到同一平面上而达到逼近效果,在处理中小型样本、高维特征时其效果理想,且 SVM 的核函数类型众多,从而可以灵活解决各种非线性的回归问题。本研究采用 BPNN、GA-BPNN、SVM、GA-SVM、ELM、粒子群算法优化极限学习机(PSO-ELM) 6 种模型对烟叶含水率的预测效果进行比较,得出优选模型,以期对雪茄烟叶的智能晾制发展提供参考。

## 1 材料与方法

### 1.1 试验材料

试验于 2022 年在四川省达州市峰城镇蜜甜香雪茄烟种植基地进行,供试品种为川雪 1 号,株距 0.4 m,行距 1.1 m,施氮量为 202.5 kg/hm<sup>2</sup>,N 含量:P<sub>2</sub>O<sub>5</sub> 含量:K<sub>2</sub>O 含量为 1.0:1.1:2.0,其他相关栽培措施按照达州当地生产技术规范进行。采收

时间为移栽后 88 d,采收完后烟叶串杆挂入晾房晾制,每杆烟叶约 50 片,晾制房为新型密集式晾房,长、宽、高分别为 15 m、6 m、8 m,共 4 层 2 路。

1.2 试验方法

1.2.1 烟叶图像采集 参考五段式晾制技术,分别在凋萎期、变黄期、变褐期、干湿交替匀色期、干筋期采集烟叶图像。图像采集装置如图 1 所示,上下左右均为黑色背景,左右两侧、顶部装有补光设备,光源类型为模拟晾制房的自然光,下地板为烟叶拍摄区,摄像头至烟叶距离固定为 0.8 m。摄像头为海康威视 $8\times 10^6$  像素镜头工业相机,镜头型号为 MVL-MF0828M-8MP。晾制阶段的判定和图像采集数量如表 1 所示。

1.2.2 含水率的测定 各晾制时期烟叶的含水率使用杀青烘干法<sup>[12]</sup>测定。

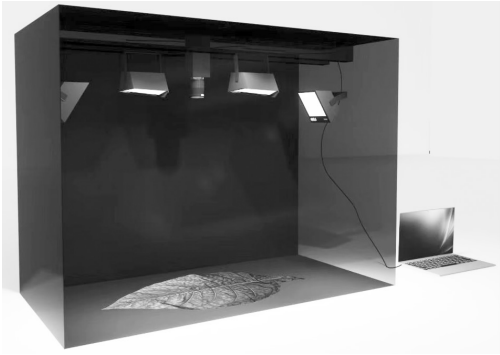


图 1 图像采集装置示意  
Fig.1 Diagram of the image acquisition device

表 1 雪茄烟晾晒时期特征和图像采集数量  
Table 1 Characteristics of cigars in the drying period and number of images collected

晾制时期	外观特征	图像数量 (张)
凋萎期	叶面轻度凋萎,叶尖变黄 3~5 cm,叶缘开始变黄	275
变黄期	叶面变黄 80%~90%,主脉两侧含绿,支脉变软,烟叶凋萎至塌架	269
变褐期	叶面 70%~90%转为褐色,呈大卷筒形,主脉部分含绿	273
干湿交替 匀色期	叶片颜色均匀一致,主脉完全变褐但未干筋	271
干筋期	叶面、主脉颜色一致,主脉完全干筋	366

1.2.3 图像预处理 本试验拍摄设备对外界环境适应性较强,但对拍摄背景颗粒物和相机温度升高

导致的电子动能增加及光电荷转移不确定性增强所产生的成像噪点适应性较差。需对图像进行预处理,增强图像信息,剔除冗余信息。本研究使用 Matlab 自带的图像分割器对图像背景和烟叶区域进行分割,再对分割的图像进行  $5\times 5$  的中值滤波,达到降噪目的。原始图像、分割后图像和中值滤波后图像如图 2、图 3、图 4 所示。

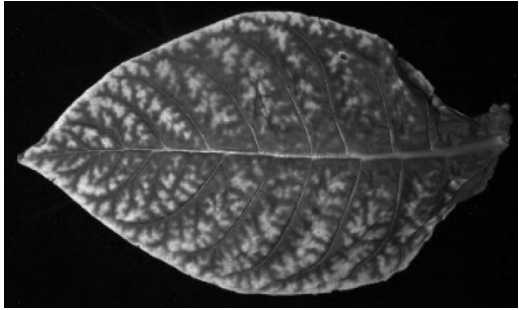


图 2 烟叶原始图像  
Fig.2 Original image of tobacco leaf

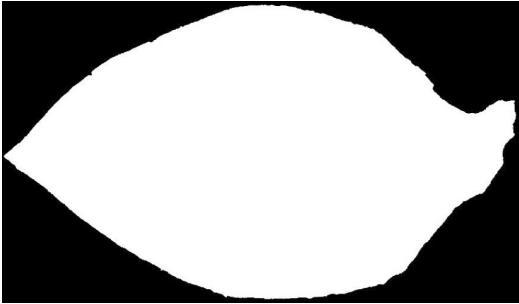


图 3 烟叶和背景分割后图像  
Fig.3 Image of tobacco leaf after segmented from the background

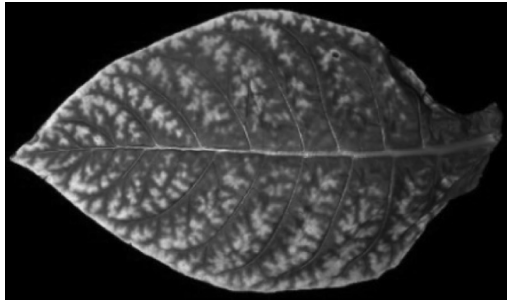


图 4 烟叶中值滤波后图像  
Fig.4 Median filtered image of tobacco leaf

1.2.4 图像特征提取 本研究通过图像灰度共生矩阵计算反映纹理变化的能量( $E$ )、同一性( $H_g$ )、

对比度 ( $Ct$ )、相关性 ( $Cl$ )、均值 ( $Mean$ )、方差 ( $Var$ ) 6 个特征作为烟叶的纹理特征。提取 RGB、HSV、Lab、YCrCb 4 个颜色空间的 12 个单通道颜色特征,按表 2 中的计算公式,衍生出 13 种颜色相关的特征,最终共获得 25 个特征作为颜色特征。

**1.2.5 数据处理和预测模型的建立** 采用隔三选一法,将样本图像划分为训练集和测试集,其中训练集 1 086 个样本,测试集 368 个样本。以最初提取的 31 个特征作为初始特征,使用 SIMCA 13.0 对 31 个特征进行正交偏最小二乘 (OPLS-DA) 分析,得出优选特征。基于 Matlab 2021a,以初始特征和优选特征作为模型输入,含水率作为输出,分别建立 BPNN、GA-BPNN、SVM、GA-SVM、ELM 和 PSO-ELM。比较所建模型的准确度,筛选出预测准确度最优的模型。

表 2 图像特征指标计算公式

Table 2 Calculation formulas of image feature indices

特征	计算公式
归一化红色指数 ( $r$ )	$R/(R+G+B)$
归一化绿色指数 ( $g$ )	$G/(R+G+B)$
归一化蓝色指数 ( $b$ )	$B/(R+G+B)$
超红指数 ( $ExR$ )	$1.4r-g^{[13]}$
超绿指数 ( $ExG$ )	$2g-r-b^{[14]}$
超蓝指数 ( $ExB$ )	$1.4b-g^{[15]}$
归一化红绿差值指数 ( $NGRDI$ )	$(g-r)/(g+r)^{[14]}$
绿叶指数 ( $GLI$ )	$(2g-b-r)/(2g+b+r)^{[16]}$
修正的红绿植被指数 ( $MGRVI$ )	$g^2-r^2/g^2+r^2^{[17]}$
深绿颜色指数 ( $DGCV$ )	$[(H-60/60)+(1-S)+(1-V)]/3$
绿-红差值指数 ( $GMR$ )	$g-r^{[18]}$
黄色指数 ( $YI$ )	$G+R^{[19]}$
蓝绿指数 ( $CI$ )	$G+B^{[19]}$

$R$ 、 $G$ 、 $B$ 、 $H$ 、 $S$ 、 $V$  分别表示红色通道值、绿色通道值、蓝色通道值、色调、饱和度、明度。

**1.2.6 模型预测准确度评价方法** 各模型对测试集的整体预测效果的评价由决定系数 ( $R^2$ ) 和均方根误差 ( $RMSE$ ) 衡量。对各阶段的预测效果以准确度 ( $Accuracy$ ) 衡量,准确度计算公式如下。

$$Accuracy = 1 - \left( \frac{|T_{\text{真实}} - P_{\text{预测}}|}{T_{\text{真实}}} \right)$$

式中,  $T_{\text{真实}}$  为雪茄烟叶实际含水率,  $P_{\text{预测}}$  为模型预测含水率。

## 2 结果与分析

### 2.1 各晾晒阶段烟叶含水率测定结果

各晾晒阶段烟叶的含水率测定结果如表 3 所示,由表 3 可知,随着晾晒进行,烟叶含水率逐渐下降。各阶段含水率间均有显著差异,其中凋萎期含水率最高,平均 84.6%,干筋期最低,平均 21.4%。

### 2.2 初始特性的提取结果

1 454 个样本的颜色特征和纹理特征提取结果见表 4,由方差分析结果可知,色调 ( $H$ )、绿色通道值 ( $G$ )、蓝色通道值 ( $B$ )、灰阶值 ( $Y$ )、归一化红色指数 ( $r$ )、归一化绿色指数 ( $g$ )、归一化蓝色指数 ( $bI$ )、超红指数 ( $ExR$ )、超蓝指数 ( $ExB$ )、黄色指数 ( $YI$ )、蓝绿指数 ( $CI$ )、方差 ( $Var$ )、亮度 ( $L^*$ ) 在干湿交替匀色期和干筋期间没有显著差异,在凋萎期、变黄期、变褐期间差异均达到显著水平;红色差异值 ( $Cr$ )、超绿指数 ( $ExG$ )、归一化红绿差值指数 ( $NGRDI$ )、绿叶指数 ( $GLI$ )、修正的红绿植被指数 ( $MGRVI$ )、对比度 ( $Ct$ )、能量 ( $E$ )、绿-红差值指数 ( $GMR$ )、深绿颜色指数 ( $DGCV$ )、明度 ( $V$ )、均值 ( $Mean$ )、黄蓝色度值 ( $b^*$ ) 在 5 个晾晒时期间有显著差异;同一性 ( $Hg$ )、红绿色度值 ( $a^*$ ) 在变褐期、干湿交替匀色期和干筋期间差异不显著,其余 2 个时期间差异显著;饱和度 ( $S$ )、红色通道值 ( $R$ )、蓝色差异值 ( $Cb$ ) 分别在变褐期和干筋期间、凋萎期和干筋期间、变褐期和干湿交替匀色期间差异不显著,其余 3 个时期间均差异显著;相关性 ( $Cl$ ) 在 5 个时期间均没有显著差异。

表 3 烟叶含水率测定结果

Table 3 Determination results of moisture content of tobacco leaves

晾晒时期	均值 (%)	最大值 (%)	最小值 (%)	变异系数
凋萎期	84.6a	89.2	70.6	0.081
变黄期	77.9b	80.4	68.1	0.022
变褐期	68.3c	74.5	64.1	0.030
干湿交替匀色期	50.5d	59.4	30.7	0.122
干筋期	21.4e	34.2	12.1	0.193

同一列数据后不同小写字母表示在 0.05 水平上差异显著。

### 2.3 特征优选结果

在方差分析的基础上,进一步对初始特征进行 OPLS-DA 判别分析,优选出 10 个映射重要性值



(Variable importance of projection, *VIP*) 排名前 10 的特征,结果如图 5 所示。10 个特征的 *R2X*(表示特征对总特征矩阵的百分比解释能力)如表 5 所示,*R2X* 累计为 0.995 4,说明所提取的 10 个特征对总特征的解释能力达 99.54%。最终确立颜色特征饱和度(*S*)、*Cr*、*NGRDI*、*MGRVI*、*GLI*、*GMR*、

*DGCV*、*H*、红色通道值(*R*)和纹理特征 *E* 为优选特征。

## 2.4 预测模型的建立

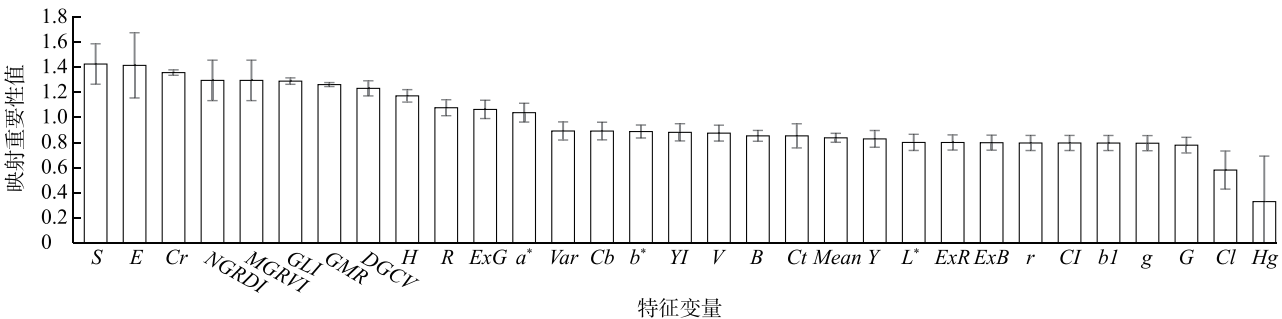
本研究对初始特征和优选特征分别建立预测模型,最终比较平均准确度,确定预测效果最佳的模型。

表 4 图像 31 个初始特征提取结果

Table 4 Extraction results of 31 initial features of images

特征	凋萎期	变黄期	变褐期	干湿交替匀色期	干筋期
色调( <i>H</i> )	0.255±0.015a	0.156±0.017b	0.089±0.008c	0.078±0.004d	0.079±0.004d
饱和度( <i>S</i> )	0.507±0.041d	0.544±0.022c	0.632±0.021b	0.655±0.016a	0.631±0.085b
明度( <i>V</i> )	0.347±0.068b	0.549±0.074a	0.321±0.060c	0.273±0.026d	0.256±0.027e
红色通道值( <i>R</i> )	67.700±16.000cd	135.900±19.500a	81.800±15.400b	69.600±6.700c	65.400±6.800d
绿色通道值( <i>G</i> )	88.400±17.200b	129.700±18.400a	58.700±13.300c	44.800±4.400d	43.500±4.300d
蓝色通道值( <i>B</i> )	42.700±5.700b	63.700±9.100a	30.700±6.700c	24.200±2.500d	24.500±2.500d
灰阶值( <i>Y</i> )	87.500±14.200b	130.300±16.000a	73.100±12.100c	61.500±4.400d	59.700±4.300d
红色差异值( <i>Cr</i> )	110.900±5.000d	98.100±4.500e	112.300±3.300c	115.300±1.200b	116.400±1.300a
蓝色差异值( <i>Cb</i> )	122.200±1.600d	135.400±3.000c	140.100±1.700a	140.300±1.400a	138.900±1.400b
亮度( <i>L</i> <sup>*</sup> )	34.300±6.700b	52.700±7.100a	26.200±5.700c	20.700±2.200d	19.800±2.100d
红绿色度值( <i>a</i> <sup>*</sup> )	-16.500±2.500b	-10.100±2.100c	-2.900±2.100a	-0.700±0.600a	-0.400±0.100a
黄蓝色度值( <i>b</i> <sup>*</sup> )	23.200±5.600b	35.700±4.400a	20.700±3.600c	17.500±1.500d	16.100±1.600e
归一化红色指数( <i>r</i> )	132.100±22.600b	194.500±27.200a	90.400±20.000c	70.100±6.800d	68.900±6.700d
归一化绿色指数( <i>g</i> )	132.400±22.500b	194.400±27.200a	90.100±20.100c	69.700±6.800d	68.600±6.700d
归一化蓝色指数( <i>bI</i> )	131.700±22.500b	193.900±27.200a	89.700±20.000c	69.400±6.800d	68.300±6.700d
超红指数( <i>ExR</i> )	52.500±9.100b	77.800±10.900a	36.400±7.900c	28.300±2.700d	27.900±2.600d
超绿指数( <i>ExG</i> )	0.989±0.115a	0.443±0.086b	0.051±0.064c	-0.058±0.034e	-0.044±0.031d
超蓝指数( <i>ExB</i> )	52.000±9.000b	77.100±10.900a	35.500±7.900c	27.400±2.700d	27.100±2.600d
绿-红差值指数( <i>GMR</i> )	20.600±4.000a	-6.100±6.900b	-23.100±3.400d	-24.700±3.000e	-21.900±3.000c
黄色指数( <i>YI</i> )	156.100±32.900b	265.700±37.200a	140.400±28.500c	114.400±10.900d	108.900±10.900d
蓝绿指数( <i>CI</i> )	131.100±22.600b	193.500±27.200a	89.300±20.000c	69.100±6.800d	67.900±6.700d
归一化红绿差值指数( <i>NGRDI</i> )	0.013±0.004a	-0.001±0.001b	-0.017±0.005c	-0.026±0.003e	-0.024±0.003d
绿叶指数( <i>GLI</i> )	0.019±0.004a	0.058±0.0013b	0.012±0.001c	-0.021±0.0013e	-0.016±0.001d
修正的红绿植被指数( <i>MGRVI</i> )	0.025±0.003a	-0.002±0.005b	-0.034±0.007c	-0.051±0.003e	-0.049±0.002d
深绿颜色指数( <i>DGCV</i> )	-0.035±0.072a	-0.238±0.004b	-0.316±0.026d	-0.334±0.022e	-0.303±0.027c
方差( <i>Var</i> )	0.018±0.008b	0.044±0.013a	0.010±0.005c	0.005±0.001d	0.004±0.002d
对比度( <i>Ct</i> )	0.023±0.004e	0.026±0.005d	0.009±0.003c	0.006±0.001b	0.006±0.001a
相关性( <i>Cl</i> )	0.984±0.058a	0.991±0.058a	0.987±0.005a	0.981±0.003a	0.978±0.005a
能量( <i>E</i> )	0.572±0.055e	0.601±0.058d	0.735±0.054c	0.768±0.042b	0.802±0.034a
同一性( <i>Hg</i> )	0.989±0.002b	0.984±0.036c	0.993±0.026a	0.996±0.001a	0.997±0.001a
颜色均值( <i>Mean</i> )	17.900±3.600b	25.900±5.300a	9.600±2.900c	6.600±1.300d	5.300±1.100e

同一行数据后不通小写字母表示在 0.05 水平上差异显著。



S:饱和度;E:能量;Cr:红色差异值;NGRDI:归一化红绿差值指数;MGRVI:修正的红绿植被指数;GLI:绿叶指数;GMR:绿-红差值指数;DGCv:深绿颜色指数;H:色调;R:红色通道值;ExG:超绿指数;a\*:红绿色度值;Var:方差;Cb:蓝色差异值;b\*:黄蓝色度值;YI:黄色指数;V:明度;B:蓝色通道值;Ct:对比度;Mean:均值;Y:灰阶值;L\*:亮度;ExR:超红指数;ExB:超蓝指数;r:归一化红色指数;CI:蓝绿指数;bI:归一化蓝色指数;g:归一化绿色指数;G:绿色通道值;Cl:相关性;Hg:同一性。

图 5 颜色、纹理特征变量映射重要性值  
Fig.5 Variable importance of projection value of color and texture feature variables

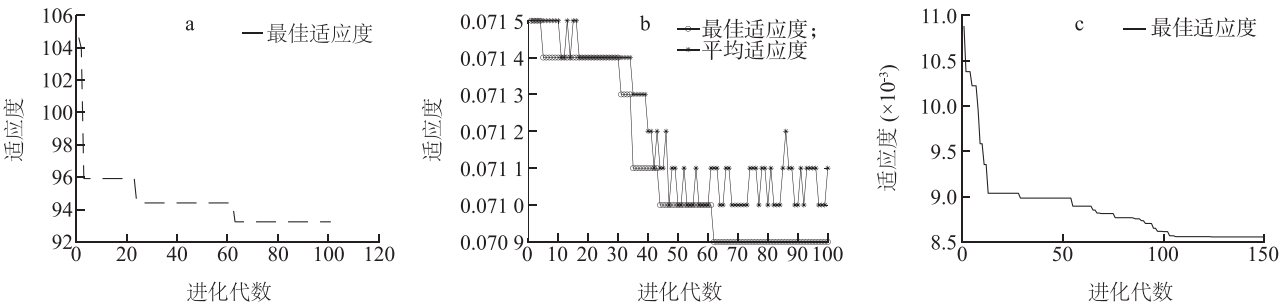
表 5 优选特征的 R2X  
Table 5 Preferred features of R2X

主成分	R2X
色度(H)	0.691 0
饱和度(S)	0.187 0
红色分量值(R)	0.007 0
红色偏移度(Cr)	0.002 2
绿-红差值指数(GMR)	0.027 0
归一化红绿差值指数(NGRDI)	0.031 2
绿叶指数(GLI)	0.023 7
修正的红绿植被指数(MGRVI)	0.014 0
深绿颜色指数(DGCv)	0.004 8
能量(E)	0.007 4
总和	0.995 4

R2X 表示特征对总特征矩阵的百分比解释能力。

2.4.1 初始特征的含水率预测模型建立结果 利用初始特征的训练集样本作为输入,含水率作为输出,建立 BPNN、GA-BPNN、SVM、GA-SVM、ELM、PSO-ELM 预测模型。其中 GA-BPNN 的初始权值和阈值、GA-SVM 的误差惩罚因子( $c$ )和核函数参数( $g$ )由遗传算法寻优获得,PSO-ELM 的输入层权值、输出层权值、偏置由粒子群优化算法寻优确定,寻优过程如图 6 所示。GA-BPNN、GA-SVM 在进化代数分别为 63、62 时,达到最佳适应度值,PSO-ELM 在进化代数为 103 时,达到最佳适应度值。

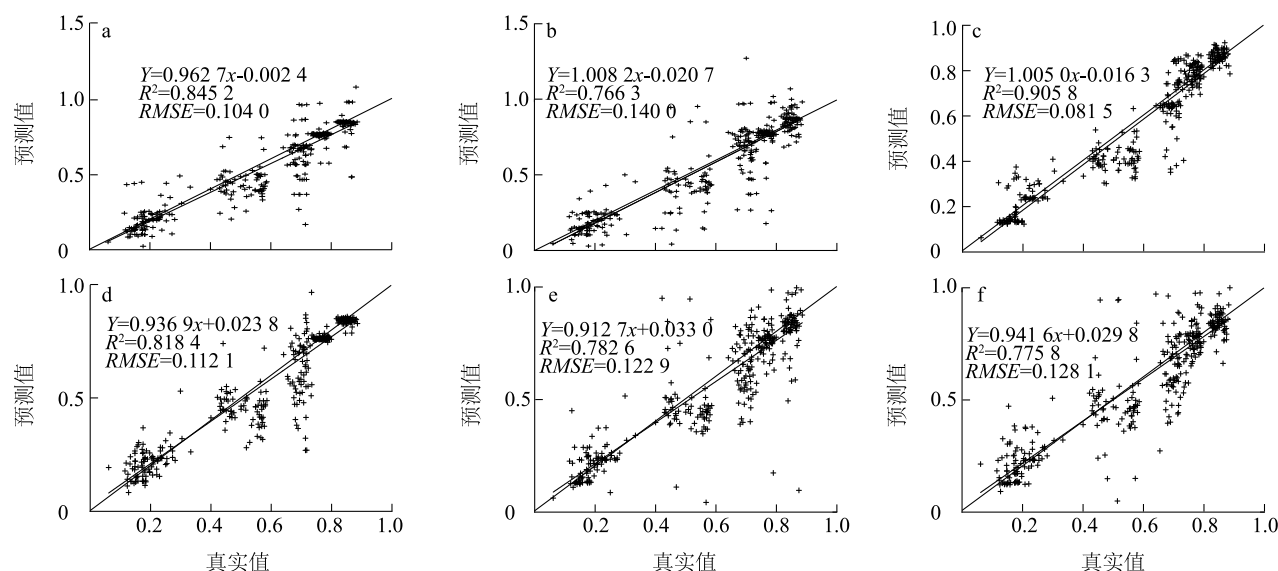
模型建立后,利用测试集样本对每个模型进行测试,得到各个模型的预测值、测定值散点图(图 7)。从图 7 可以看出,GA-SVM 的预测值和真实值拟合度最优, $R^2$ 、 $RMSE$  分别为 0.905 8、0.081 5,其次是 GA-BPNN,BPNN 的  $R^2$ (0.766 3)最低。



GA-BPNN:遗传算法优化前馈神经网络;GA-SVM:遗传算法优化支持向量机;PSO-ELM:粒子群算法优化极限学习机。

图 6 初始特征的 GA-BPNN (a)、GA-SVM (b)、PSO-ELM (c) 的参数寻优过程

Fig.6 Parameter optimization process of genetic algorithm optimized back propagation neural network (GA-BPNN) (a), genetic algorithm optimized support vector machine (GA-SVM) (b) and particle swarm optimized extreme learning machine (PSO-ELM) (c) of initial characteristics



GA-BPNN:遗传算法优化前馈神经网络;BPNN:前馈神经网络;GA-SVM:遗传算法优化支持向量机;SVM:支持向量机;PSO-ELM:粒子群算法优化极限学习机;ELM:极限学习机。 $R^2$ :决定系数; $RMSE$ :均方根误差。

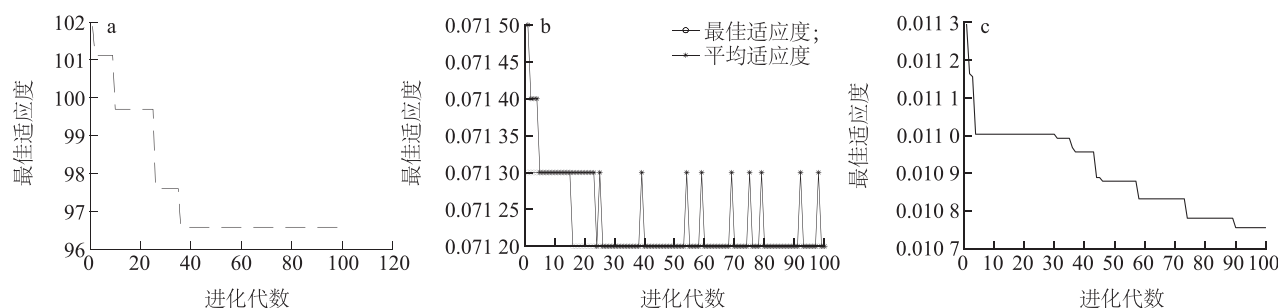
图7 初始特征的GA-BPNN(a)、BPNN(b)、GA-SVM(c)、SVM(d)、PSO-ELM(d)、ELM(e)测试集的预测值和真实值散点图

Fig.7 Scatter plot of predicted and true values of genetic algorithm optimized back propagation neural network (GA-BPNN) (a), back propagation neural network (BPNN) (b), genetic algorithm optimized support vector machine (GA-SVM) (c), support vector machine (SVM) (d), particle swarm optimized extreme learning machine (PSO-ELM) (d), extreme learning machine (ELM) (e) test sets for initial features

2.4.2 优选特征的含水率预测模型建立结果 利用优选特征的训练集样本作为输入,含水率作为输出,建立BPNN、GA-BPNN、SVM、GA-SVM、ELM、PSO-ELM预测模型。其中GA-BPNN的初始权值和阈值、GA-SVM的误差惩罚因子和核函数参数以及PSO-ELM的输入层权值、输出层权值、偏置的寻优过程如图8所示。GA-BPNN、GA-SVM在进化代数

分别为38、17时,达到最佳适应度值,PSO-ELM在进化次数为90时,达到最佳适应度值。

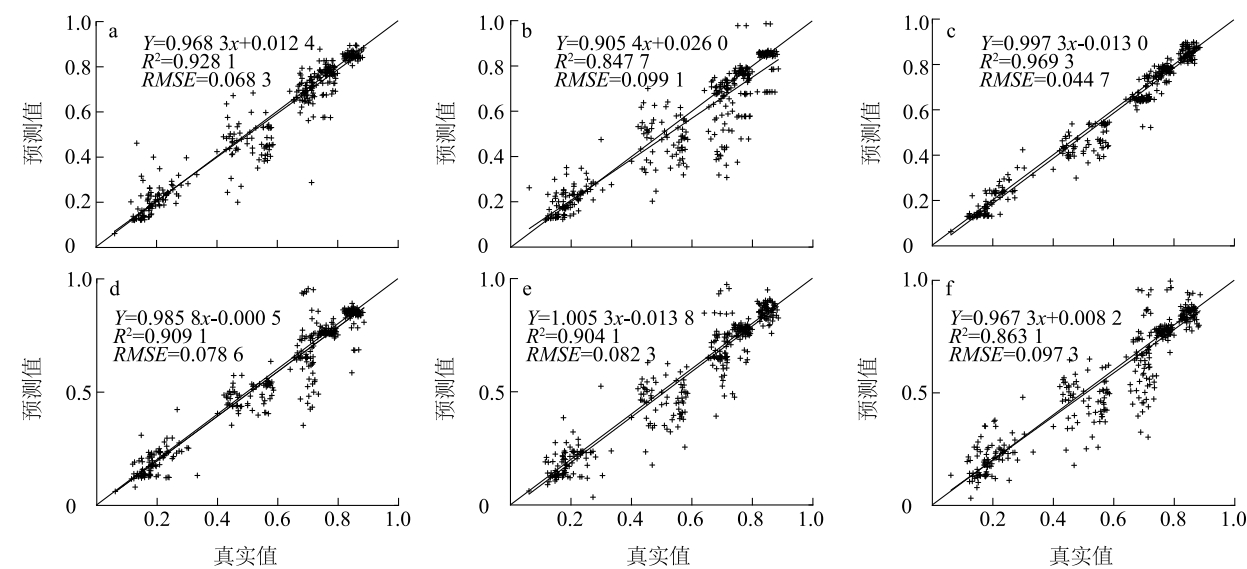
模型建立后,利用测试集样本对每个模型进行测试,得到各个模型的预测值、真实值散点图(图9)。从图9c可以看出,GA-SVM的拟合度最优, $R^2$ 、 $RMSE$ 分别为0.9693、0.0447,其次是GA-BPNN, BPNN的 $R^2$ (0.8477)最低。



GA-BPNN:遗传算法优化前馈神经网络;GA-SVM:遗传算法优化支持向量机;PSO-ELM:粒子群算法优化极限学习机。

图8 优选特征的GA-BPNN(a)、GA-SVM(b)、PSO-ELM(c)的参数寻优过程

Fig.8 Parameter optimization process of genetic algorithm optimized back propagation neural network (GA-BPNN) (a), genetic algorithm optimized support vector machine (GA-SVM) (b) and particle swarm optimized extreme learning machine (PSO-ELM) (c) of preferred features



GA-BPNN:遗传算法优化前馈神经网络;BPNN:前馈神经网络;GA-SVM:遗传算法优化支持向量机;SVM:支持向量机;PSO-ELM:粒子群算法优化极限学习机;ELM:极限学习机。 $R^2$ :决定系数; $RMSE$ :均方根误差。

图 9 优选特征的 GA-BPNN(a)、BPNN(b)、GA-SVM(c)、SVM(d)、PSO-ELM(e)、ELM(f) 测试集的预测值和真实值散点图  
Fig.9 Scatter plot of predicted and true values of genetic algorithm optimized back propagation neural network (GA-BPNN) (a), back propagation neural network (BPNN) (b), genetic algorithm optimized support vector machine (GA-SVM) (c), support vector machine (SVM) (d), particle swarm optimized extreme learning machine (PSO-ELM) (e), extreme learning machine (ELM) (f) test sets for preferred features

2.5 对 5 个晾晒阶段含水率预测准确度结果分析

根据方法 1.2.6 计算初始特征和优选特征对各晾晒阶段的预测准确度,结果见表 6。由表 6 可知,各模型对凋萎期和变黄期的预测效果理想,准确度均高于 90.0%,优选特征建立的 GA-SVM 对变黄期的预测准确度最高(98.9%),但对于干湿交替匀色期

和干筋期的预测准确度较低,初始特征的 ELM 对干筋期的预测准确度最低(68.5%)。综合考虑 5 个晾晒阶段的预测准确度,以优选特征的 GA-SVM 准确度最高,对干筋期和干湿交替匀色期的预测准确度也高于 87.0%。

表 6 各模型对各晾晒阶段的预测准确度

Table 6 Prediction accuracy of each model for each drying stage

特征	晾晒阶段	模型准确度(%)					
		BPNN	GA-BPNN	PSO-ELM	ELM	GA-SVM	SVM
优选特征	凋萎期	95.4	98.1	97.4	97.0	97.9	96.3
	变黄期	91.4	94.7	95.8	95.9	98.9	97.3
	变褐期	85.2	94.8	88.2	82.4	93.9	84.3
	干湿交替匀色期	80.3	83.4	80.7	80.0	88.3	90.8
	干筋期	77.8	80.2	77.2	72.7	87.1	81.8
	平均值	86.0	90.2	87.9	85.6	93.2	90.1
初始特征	凋萎期	94.0	94.7	93.7	94.4	97.5	97.0
	变黄期	92.3	96.4	93.0	90.3	96.9	98.0
	变褐期	80.0	83.0	85.9	81.9	88.8	78.0
	干湿交替匀色期	71.8	80.2	76.8	76.0	81.0	80.5
	干筋期	71.7	72.5	74.7	68.5	78.9	72.6
	平均值	82.0	85.4	84.8	82.2	88.6	85.2



### 3 讨 论

雪茄烟叶晾制过程中的变黄、失水状态目前主要依靠人为判断,存在较强的主观性和盲目性,而烘箱法、快速水分测定法等测量速度缓慢,难以实现对含水率快速实时的判断。研究表明,简单的线性关系分析难以确定图像信息和含水率的关系<sup>[20-21]</sup>,而通过颜色特征和纹理特征结合机器学习方法可以对含水率进行准确估算。在其他作物上,程晓婷<sup>[22]</sup>基于番茄图像的纹理、颜色特征建立神经网络对番茄含水率进行预测,模型均方根误差为1.222 7;徐腾飞等<sup>[23]</sup>通过提取玉米叶片的颜色特征对含水率进行预测,模型标准误差仅0.021;王甜等<sup>[24]</sup>提取沉香叶片的颜色、纹理特征建立 RFE-RF 模型预测含水率,其精度达到了91.62%。在烤烟方面,段史江等<sup>[25]</sup>对烘烤过程中烟叶的颜色特征和纹理特征进行提取,建立 BPNN、GA-SVM 含水率预测模型,其标准误差分别为0.048 5和0.020 0,预测准确度理想;詹攀等<sup>[26]</sup>提取了烤烟鲜烟叶的颜色、纹理特征建立极限学习机模型对其含水率进行预测,模型平均绝对误差仅0.007 9。本研究首次对雪茄烟叶晾制期间的图像进行采集,并测定其含水率,提取出颜色特征和纹理特征,分别选用初始特征和 OPLS-DA 筛选的优选特征作为输入,建立含水率预测模型。结果表明,用优选特征建立的 GA-SVM 的决定系数( $R^2$ )达到了0.969 3,均方根误差仅0.044 7,预测效果理想,这与陈飞等<sup>[7]</sup>、张娟利等<sup>[27]</sup>所建模型准确度接近。此外,在本研究中,随着晾制进行,烟叶含水率逐渐降低,与赵晨等<sup>[28]</sup>、范宁波<sup>[29]</sup>的研究结果一致。

本研究中,无论是使用初始特征还是优选特征,GA-SVM 预测效果均好于 GA-BPNN 和 PSO-ELM,这与谢滨瑶等<sup>[30]</sup>、李增盛等<sup>[9]</sup>的研究结论一致。原因是 BPNN 和 ELM 2 类算法是模拟神经元传递信息,采用非线性函数的形式通过连接权值和偏置对输入特征进行映射,以达到预测效果,而晾制阶段中的颜色特征较为接近,尤其是干湿交替匀色期和干筋期,因此难以达到十分精准的预测。SVM 是基于结构最小风险原理将不同维度样本映射到高维空间,在空间中确定一个平面,使得样本尽可能映射在该平面上,从而达到较好的预测效果。此外,本研究采用遗传算法对 SVM 的  $c$ 、 $g$  参数进行寻优,解决了常规 SVM 的  $c$ 、 $g$  参数敏感、难以确定的问题。OPLS-DA 筛选的特征建立的模型准确度

均高于初始特征所建模型,表明 OPLS-DA 可以作为优选方法,筛选与预测响应相关性高的特征指标。

本研究的烟叶拍摄环境是在补光均匀的暗箱内,最大程度地降低了外界光照、风力、人为走动等的影响,将烟叶在晾制期间的变黄、失水与皱缩状态充分体现在所拍摄的图像上,使得所建模型的平均预测准确度均高于80.0%。本研究发现,优选特征和初始特征所建立的模型对干湿交替匀色期、干筋期样本的预测准确度较低,ELM 的干筋期初始特征预测准确度仅68.5%。原因是在雪茄烟晾制过程中,变褐期以后叶面颜色基本确定,更多的是烟筋的干燥过程,而这2个时期叶面颜色比较接近,故导致预测准确度较低,但干湿交替匀色期和干筋期是烟筋最后的干燥时期,重要性比前3个时期低,在实际预测中造成的偏差对最终烟叶质量影响小。在今后的研究中,可对这2个时期进一步细化晾制的前、中、后阶段,提取和含水率相关性更高的特征进行预测。

### 4 结 论

通过以上研究,可得出以下结论:①各晾制阶段烟叶含水率差异显著,随着晾制进行,含水率逐渐下降,凋萎期含水率平均84.6%,干筋期含水率平均21.4%。②优选特征建立的 GA-SVM 对雪茄烟叶全晾制阶段的含水率整体预测能力相较于本研究其他模型表现最佳, $R^2$ 和 RMSE 分别为0.969 3、0.044 7。③优选特征建立的 GA-SVM 对各晾制阶段含水率的预测准确度较高,其中干筋期预测准确度最低,但也高于87.0%。

### 参考文献:

- [1] 杨月先,李宗平,付庆灵,等.采收和晾制方式对雪茄烟中上部烟叶品质的影响[J].烟草科技,2022,55(8):28-34.
- [2] 国家烟草质量监督检验中心.Y烟草及烟草制品试样的制备和水分的测定-烘箱法;C/T 31-1996[S].北京:中国标准出版社,1996.
- [3] 中国烟草标准化研究中心.烟草水分的测定-卡尔费休法;YC/T 165-2003[S].北京:中国标准出版社,2003.
- [4] 李小玲.复烤烟叶快速水分测定法和烘箱水分测定法的比较[J].工业计量,2006,16(4):28-29.
- [5] 郭朵朵.基于机器视觉的烟草烘烤品质检测及控制系统的设计[D].杨凌:西北农林科技大学,2017.
- [6] HARALICK R M, SHANMUGAM K, DINSTEN I. Textural features for image classification[J]. IEEE Transactions on Systems

- Man & Cybernetics, 1973, 3(6): 610-621.
- [7] 陈飞程, 杨懿德, 李常军, 等. 基于图像信息的烤烟烘烤过程中烟叶含水率预测[J]. 西南农业学报, 2021, 34(11): 2378-2384.
- [8] 杜海娜, 孟令峰, 王松峰, 等. 基于机器学习的密集烘烤过程烟叶失水率预测模型对比[J]. 烟草科技, 2022, 55(9): 81-88.
- [9] 李增盛, 孟令峰, 王松峰, 等. 基于图像处理的烟叶烘烤阶段判别模型优选[J]. 中国烟草学报, 2022, 28(2): 65-76.
- [10] GHOSH A, ROY P. An automated model for leaf image-based plant recognition: an optimal feature-based machine learning approach[J]. Innovations in Systems and Software Engineering, 2022, 3: 1-14.
- [11] 宋佳. 基于图像处理的生菜叶片含水率检测研究及 Android 平台的实现[D]. 镇江: 江苏大学, 2017.
- [12] 张志良. 植物生理学实验指导[M]. 北京: 高等教育出版社, 1990: 1-10.
- [13] MEYER G E, HINDMAN T W, LAKSMI K. Machine vision detection parameters for plant species identification[J]. Proceedings of Spie the International Society for Optical Engineering, 1999, 1. DOI: 10.1117/12.336896.
- [14] WOEBBECKE D M, MEYER G E, VON BARGEN K, et al. Color indices for weed identification under various soil, residue, and lighting conditions[J]. Transactions of the ASAE, 1995, 38(1): 259-269.
- [15] GUIJARRO M, PAJARES G, RIOMOROS I, et al. Automatic segmentation of relevant textures in agricultural images[J]. Computers and Electronics in Agriculture, 2010, 75(1): 75-83.
- [16] LOUHAICHI M, BORMAN M M, JOHNSON D E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat[J]. Geocarto International, 2001, 16(1). DOI: 10.1080/10106040108542184.
- [17] LIANGSHENG Y Q Y H. Estimation of leaf area index of sugarcane using crop surface model based on UAV image[J]. Journal of Agricultural Engineering, 2017, 33(8): 104-111.
- [18] WANG Y, WANG D, ZHANG G, et al. Estimating nitrogen status of rice using the image segmentation of G-R thresholding method[J]. Field Crops Research, 2013, 149: 33-39.
- [19] SMITS B. An RGB-to-spectrum conversion for reflectances[J]. Journal of Graphics Tools, 2000, 1: 1-10.
- [20] 白竣文, 田潇瑜, 马海乐. 基于 BP 神经网络的葡萄气体射流冲击干燥含水率预测[J]. 现代食品科技, 2016, 32(12): 198-203.
- [21] VAN HENTEN E J. Non-destructive crop measurement by image processing for crop growth control[J]. Journal of Agricultural Engineering Reasearch, 1995, 61: 97-105.
- [22] 程晓婷. 基于番茄叶片图像的土壤含水率研究[D]. 银川: 宁夏大学, 2021.
- [23] 徐腾飞, 韩文霆, 孙瑜. 基于图像处理的玉米叶片含水率诊断方法研究[J]. 干旱地区农业研究, 2013, 31(1): 95-100.
- [24] 王甜, 王雪峰, 刘嘉政. 基于 RFE\_RF 算法的幼龄沉香叶片含水率预估模型[J]. 南京林业大学学报(自然科学版), 2022, 46(4): 177-184.
- [25] 段史江, 宋朝鹏, 马力, 等. 基于图像处理的烘烤过程中烟叶含水量检测[J]. 西北农林科技大学学报(自然科学版), 2012, 40(5): 74-80.
- [26] 詹攀, 谢守勇, 刘军, 等. 基于支持向量机回归的鲜烟叶含水量预测模型[J]. 西南大学学报(自然科学版), 2016, 38(4): 165-170.
- [27] 张娟利, 韩文霆, 杨六生, 等. 基于图像处理烤烟鲜烟叶含水率诊断研究[J]. 现代农业科技, 2017(9): 3-5.
- [28] 赵晨, 王爱霞, 范宁波, 等. 基于 CIE-Lab 的雪茄烟叶晾晒过程水分与颜色相关性研究[J]. 湖南文理学院学报(自然科学版), 2022, 34(2): 60-66.
- [29] 范宁波. 雪茄烟叶晾晒过程中颜色及相关生理变化特性分析[D]. 郑州: 河南农业大学, 2021.
- [30] 谢滨瑶, 祝诗平, 黄华. 基于 BPNN 和 SVM 的烟叶成熟度鉴别模型[J]. 中国烟草学报, 2019, 25(1): 45-50.

(责任编辑: 陈海霞)