

引用格式:

杨睿, 宾俊, 苏家恩, 汪华国, 王文伦, 何承刚, 陈颐, 邹聪明. 基于近红外光谱与图像识别技术融合的烟叶成熟度的判别[J]. 湖南农业大学学报(自然科学版), 2021, 47(4): 406-411.

YANG R, BIN J, SU J E, WANG H G, WANG W L, HE C G, CHEN Y, ZOU C M. Identification of tobacco leaf maturity based on the fusion of near infrared spectroscopy and image recognition[J]. Journal of Hunan Agricultural University(Natural Sciences), 2021, 47(4): 406-411.

投稿网址: <http://xb.hunau.edu.cn>



## 基于近红外光谱与图像识别技术融合的烟叶成熟度的判别

杨睿<sup>1,2</sup>, 宾俊<sup>3</sup>, 苏家恩<sup>4</sup>, 汪华国<sup>5</sup>, 王文伦<sup>5</sup>, 何承刚<sup>2</sup>, 陈颐<sup>1\*</sup>, 邹聪明<sup>1</sup>

(1.云南省烟草农业科学研究院, 云南 昆明 650031; 2.云南农业大学烟草学院, 云南 昆明 650201; 3.贵州大学烟草学院, 贵州 贵阳 550025; 4.云南省烟草公司大理州公司, 云南 大理 671000; 5.云南省烟草公司楚雄州公司, 云南 楚雄 675005)

**摘要:**为提高鲜烟叶成熟度的识别精度, 提出基于近红外光谱和图像识别的多源信息融合技术的烟叶成熟度判别方法: 利用随机森林(RF)方法分别建立近红外光谱判别模型、图像判别模型和多源信息融合判别模型, 对烟叶成熟程度进行检测。近红外光谱模型对红花大金元、K326和云烟87等3个烤烟品种烟叶成熟度的识别正确率分别为91.27%、90.43%、89.44%, 图像模型的识别正确率分别为86.20%、86.96%、81.23%, 融合模型的识别正确率分别为94.08%、94.78%和92.96%。与近红外光谱模型相比, 融合模型的判别正确率平均提高了3.93%; 与图像模型相比, 融合模型的判别正确率平均提高了10.83%。

**关键词:** 烟叶成熟度; 近红外光谱; 图像识别; 数据融合; 判别

中图分类号: TP391.41

文献标志码: A

文章编号: 1007-1032(2021)04-0406-06

### Identification of tobacco leaf maturity based on the fusion of near infrared spectroscopy and image recognition

YANG Rui<sup>1,2</sup>, BIN Jun<sup>3</sup>, SU Jiaen<sup>4</sup>, WANG Huaguo<sup>5</sup>, WANG Wenlun<sup>5</sup>,  
HE Chenggang<sup>2</sup>, CHEN Yi<sup>1\*</sup>, ZOU Congming<sup>1</sup>

(1.Yunnan Academy of Tobacco Agricultural Sciences, Kunming, Yunnan 650031, China; 2.College of Tobacco, Yunnan Agricultural University, Kunming, Yunnan 650201, China; 3. College of Tobacco, Guizhou University, Guiyang, Guizhou 550025, China; 4.Dali Tobacco Company of Yunnan, Dali, Yunnan 671000, China; 5.Chuxiong tobacco company of Yunnan, Chuxiong, Yunnan 675005, China)

**Abstract:** In order to improve the identification accuracy of fresh tobacco leaf maturity, a tobacco leaf maturity identification method was proposed based on the multi-source information fusion technology of near infrared spectrum and image recognition. Using random forest method, the near infrared spectrum discriminant model, image discriminant model and multi-source information fusion discriminant model were established to detect and analyze the maturity of tobacco leaves. The accuracy rates of the near infrared spectrum model for the three flue-cured tobacco leaf maturity were 91.27%, 90.43% and 89.44%, and the accuracy rates of the image model for the three flue-cured tobacco leaf maturity were 86.20%, 86.96% and 81.23%, respectively. The accuracy of the fusion model for the identification of leaf maturity of the three flue-cured tobacco varieties was 94.08%, 94.78% and 92.96%, respectively. Compared with near infrared spectroscopy model and image model, the accuracy of fusion model is improved by 3.93% and 10.83% on average.

收稿日期: 2020-09-01

修回日期: 2021-05-02

基金项目: 中国烟草总公司重点项目(110202102007); 云南省烟草专卖局重点项目(2019530000241019)

作者简介: 杨睿(1995—), 女, 云南大理人, 硕士研究生, 主要从事烟草调制技术研究, 1215013156@qq.com; \*通信作者, 陈颐, 博士, 助理研究员, 主要从事烟草调制技术研究, 229781579@qq.com

**Keywords:** tobacco maturity; near infrared spectroscopy; image recognition; data fusion; discriminant analysis

烟叶成熟度是形成烟叶品质的重要因素<sup>[1-2]</sup>。准确把握烟叶成熟度是获得优质烟叶的前提和基础。采收成熟度好的烟叶,等级高,易烘烤,具有较好的香气与香气量,可增加烟农收益。将无损分析技术<sup>[3-5]</sup>应用于鲜烟叶成熟度的判别,可提高工作效率与准确率。李佛琳等<sup>[6]</sup>用高光谱技术判别烟叶成熟度,其训练样本的准确率为98%,验证样本的准确率为97%。梁寅等<sup>[7]</sup>利用高光谱遥感探测技术将云烟87初熟类、适熟类和过熟类烟叶分开,分类精度达90%。李向阳等<sup>[8]</sup>研究表明,利用室内光谱红边参数估测叶绿素含量效果较好,从而可以判断烟叶成熟度。王承伟等<sup>[9]</sup>应用近红外光谱技术结合随机森林算法,对烟株上、中、下3个部位的烟叶成熟度的判别正确率分别为92%、90%和91%,能快速实现烟叶成熟度的判别。史龙飞等<sup>[10]</sup>、周首峰<sup>[11]</sup>利用图像识别技术提取鲜烟叶的颜色特征和纹理特征,进行模式识别,能在一定程度上实现对烟叶成熟度的判别。王杰等<sup>[12]</sup>采用稀疏自编码器对烟叶图像特征进行提取和识别判断,从计算机视觉角度自动识别烟叶的成熟度,测试数据分类准确率

可达98.63%。刘剑君等<sup>[13]</sup>通过提取烟叶图像HSV颜色特征和RGB颜色特征,实现了对烤烟烟叶成熟度的判定。这些研究集中于对单一判别模型的探讨,且存在建模样本数少、模型识别正确率不高等问题。

笔者拟利用随机森林方法,构建近红外光谱判别模型、图像判别模型和多源信息融合判别模型,比较分析3种判别模型的准确度,建立烟叶成熟等级快速判别方法,旨在将光谱分析技术和图像识别技术应用于烟叶采收的质量控制,以及为开发基于无损分析技术的烟叶成熟度检测仪器提供支持。

## 1 材料与方法

### 1.1 材料

选用云南大理州弥渡烘烤工场和楚雄州牟定烘烤工场的红花大金元、K326和云烟87烟叶,按照GB/T 23219—2008《烤烟烘烤技术规程》,划分为欠熟(M1)、尚熟(M2)、成熟(M3)、完熟(M4)和过熟(M5)5个成熟等级(表1)。

表1 鲜烟叶样本的成熟度

Table 1 Maturity of fresh tobacco samples

品种	总样本数	样本数				
		M1	M2	M3	M4	M5
红花大金元	1184	132	324	102	290	336
K326	1148	274	198	165	201	310
云烟87	1135	164	298	256	286	131

### 1.2 方法

#### 1.2.1 采集烟叶的近红外光谱

采用NIRQuest便携式近红外光谱仪采集烟叶的近红外光谱。光谱仪配有InGaAs线阵检测器、标准探头和漫反射白板,仪器的平均分辨率小于10 nm,积分时间为5 ms,光谱采集范围为900~2500 nm。开机预热30 min后,对烟叶进行近红外光谱扫描。扫描时,避开主脉,在主脉两侧各取3个点扫描,取光谱平均值。

#### 1.2.2 采集烟叶图像

采用AI泛智能抓拍筒型网络摄像机(海康威视)采集烟叶图像。摄像机内置CMOS传感器、高效白

光阵列灯和电动变焦镜头。保持摄像机镜头与烟叶垂直高度60 cm,图像以JPG格式自动保存于计算机中,分辨率为3072像素×1728像素。

#### 1.2.3 数据融合

提取InGaAs采集器和CMOS传感器采集的烟叶特征数据信息,在数据融合之前采用主成分分析法(PCA)分别对近红外光谱数据和图像数据进行降维处理,再对融合数据进行归一化处理。

#### 1.2.4 判别分析方法

采用随机森林(RF)方法<sup>[14]</sup>构建判别模型。

1) 用自举法<sup>[15]</sup>从原始训练集中抽取分类树数目(ntree)自举样本,形成新的训练集。约2/3的原

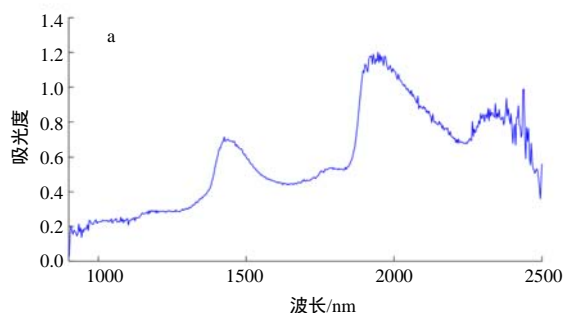
始训练样本会被抽到作为自举样本,用于决策树模型的建立,其余的 1/3 训练样本未能参与决策树的建立,使用这些样本估计模型的性能。

2) 对于每个自举样本,构建 1 个未修饰的分类树,并进行优化。在每个节点随机选择随机选择变量数(mtry)并从中得到最佳分离比,每棵树的构建中就具有了训练样本和随机选择的特性。

3) 将决策树的预测输出按照少数服从多数的原则,得到目标分类数据的最优分类结果。

### 1.3 模型评价

以分类正确率作为判别模型的评价指标,所有光谱、图像处理及RF源代码的编写和运行都在 Matlab R2012a软件平台上完成。

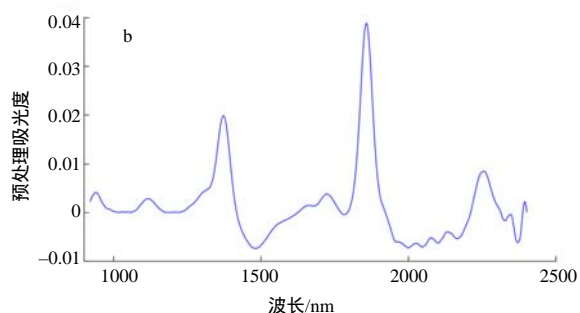


## 2 结果与分析

### 2.1 基于近红外光谱的烟叶成熟度的判别

#### 2.1.1 光谱预处理及样本划分

经过对比分析,选择一阶Norris导数(间隔为 5、分段为 5),结合Savitzky-Golay平滑(平滑窗口为 9)的预处理组合,对烟叶光谱进行预处理。由于光谱首尾波动较大,在应用中将首尾部分波长点去除,保留 920~2400 nm的光谱(468 个波长点),预处理前后的光谱见图 1。光谱数据按SPXY样本划分方法<sup>[16]</sup>划分为训练集和验证集,划分比例为 0.7 : 0.3,红花大金元、K326 和云烟 87 的训练集样本数分别为 829、803、794,验证集样本数分别为 355、345、341。



a 原始光谱; b 预处理后光谱。

图 1 烟叶的近红外光谱

Fig.1 Near infrared spectrum of tobacco leaves

#### 2.1.2 模型参数优化

在模型建立过程中,采用OOB错误率来评估分类树的数目(ntree)和随机选择变量数(mtry)。以红花大金元为例,经过优化,ntree选择 76 时,可得到最小OOB误差,ntree的变化趋势见图 2,各数据集

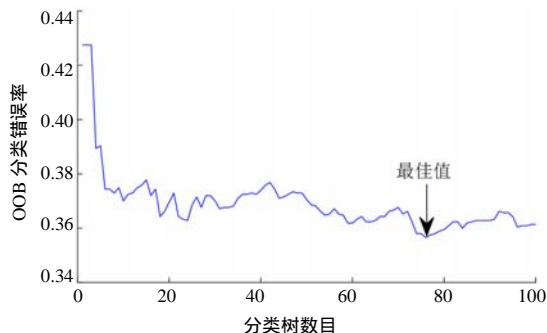


图 2 RF 分类器 ntree 参数的选择

Fig.2 Parameter selection of RF classifier ntree

选择的具体参数见表 2。

#### 2.1.3 近红外光谱判别模型的构建

近红外光谱判别模型的参数优化和预测结果列于表 2。3 个烟草品种模型的训练集的正确率都较高,验证集预测正确率也都达到或超过 90%,模型整体的判别性能较好。

表 2 近红外光谱模型参数优化和分类预测结果

Table 2 Parameter optimization and classification prediction results of near infrared spectrum model

品种	ntree	mtry	训练集正确率/%	验证集正确率/%
红花大金元	76	21	97.83(811/829)	91.27(324/355)
K326	68	21	96.51(775/803)	90.43(312/345)
云烟 87	64	21	96.22(764/794)	89.44(305/341)

括号内数据为分类正确样本数/样本总数。

## 2.2 图像识别技术对烟叶成熟度的判别

### 2.2.1 图像处理

采用小波变换(WT)<sup>[17]</sup>技术对烟叶图像进行去噪处理,选择coif2小波函数、3阶分解,通过滤波处理除去烟叶图像中的噪声,优化烟叶图像质量,图3-1为原始图像,图3-2为去噪后的烟叶图像。

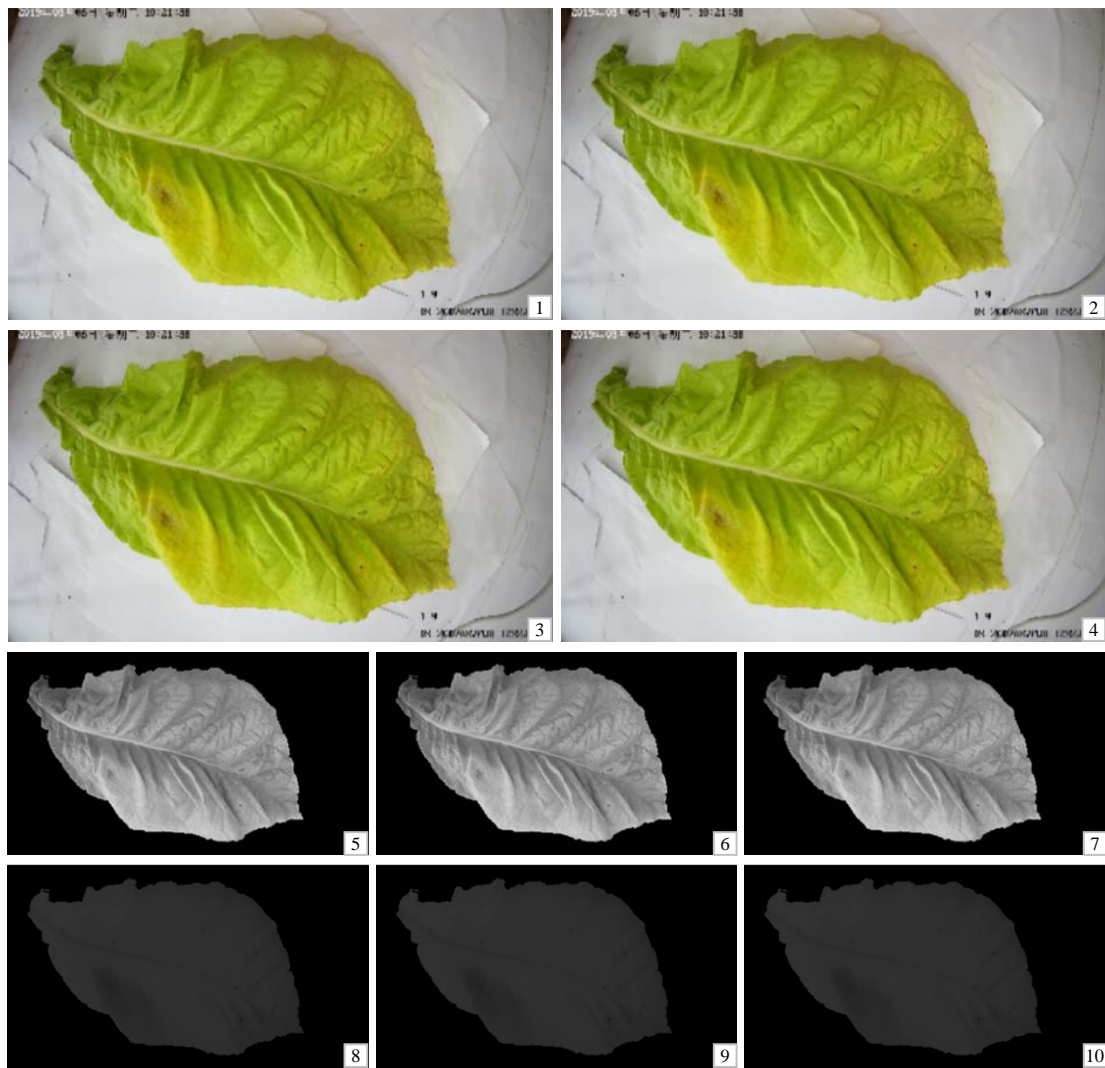
采用自适应模糊C均值聚类算法(AFCM)<sup>[18]</sup>进行烟叶图像的边缘提取,将原始图像处理为灰度直方图后,根据图像灰度来进行图像分割,图3-3为提取的烟叶轮廓。根据提取的烟叶轮廓,去除图像背景,保留烟叶本身,扣除背景后的图像见图3-4。

利用红(R)、绿(G)、蓝(B)3种颜色传感器提取烟叶颜色信息,得到彩色图像,将红、绿、蓝这三

种颜色按照不同的分量相加混合,可得到任意一种颜色,提取这3种颜色的平均值和它们的标准差作为图像的颜色特征之一<sup>[19]</sup>,图3-5、图3-6、图3-7分别为RGB通道的分量值。

将RGB颜色转换为HSV颜色系统,将HSV的均值以及标准差作为颜色的特征,图3-8、图3-9、图3-10分别为HSV3个颜色分量值。

利用灰度差分统计法<sup>[20]</sup>将RGB图像处理为灰度图像,再使用图像灰度共生矩阵方法从其中提取能量、熵、惯性矩、相关性4个纹理特征参数,将它们的均值及标准差作为纹理特征,根据颜色和纹理特征,从每个烟叶图像中提取20个特征变量。



1 烟叶原始图像; 2 去噪后烟叶图像; 3 提取的烟叶轮廓; 4 扣除背景后的烟叶轮廓; 5、6、7 分别为RGB通道的分量值; 8、9、10 分别为HSV颜色分量值。

图3 烟叶的图像处理

Fig.3 Image processing of tobacco leaves

### 2.2.2 图像判别模型的建立

为了方便烟草图像数据的处理分析,采用归一化方法对特征变量进行归一化处理,再利用RF法得出分类预测结果。图像判别模型的烟叶成熟度预测的正确率列于表3。3个品种的成熟度预测正确率都超过81%,红花大金元与K326的预测正确率较云烟87的高,这说明不同烟草品种的成熟度判别模型并不能通用。另外,模型的总体预测正确率都

表3 图像模型参数优化和分类预测结果

Table3 Parameter optimization and classification prediction results of image model

品种	ntree	mtry	训练集正确率/%	验证集正确率/%
红花大金元	37	4	89.02(738/829)	86.20(306/355)
K326	76	4	85.18(684/803)	86.96(300/345)
云烟87	32	4	85.01(675/794)	81.23(277/341)

括号内数据为分类正确样本数/样本总数。

表4 融合模型参数优化和分类预测结果

Table4 Parameter optimization and classification prediction results of fusion model

品种	光谱主成分数	图像主成分数	ntree	mtry	训练集正确率/%	验证集正确率/%
红花大金元	24	8	68	5	96.98(804/829)	94.08(334/355)
K326	22	8	56	5	97.01(779/803)	94.78(327/345)
云烟87	24	8	64	5	96.73(768/794)	92.96(317/341)

括号内数据为分类正确样本数/样本总数。

对表2、表3和表4进行分析比对,可以看出,融合模型分类预测的准确性要高于近红外光谱判别模型和图像判别模型。用融合模型对3个烤烟品种烟叶成熟度进行预测,其准确度分别比近红外光谱模型高3.1%、4.8%和3.9%,分别比图像特征模型高9.1%、9.0%和14.4%。与近红外光谱模型相比,融合模型的判别正确率平均提高了3.93%;与图像模型相比,融合模型的判别正确率平均提高了10.83%。

## 3 结论与讨论

许多研究表明,总氮和叶绿素含量可以用来判断烟叶的成熟度<sup>[21]</sup>。N—H键的伸缩振动一级倍频主要位于1500 nm附近,其伸缩振动与弯曲振动的组合频主要位于2000 nm附近<sup>[15]</sup>,叶绿素主要的特征波长都位于近红外光谱区间,所以进行烟叶的成熟度判别可以采用近红外光谱技术。在农艺上,烟叶主脉对烟叶成熟度判别有一定作用,所以在对烟叶成熟度判别时,不可忽视叶脉这一重要因素,但

偏低,说明图像采集的烟叶特征信息区分度较小,图像识别技术无法较好地识别此类特征。

### 2.3 基于近红外光谱技术与图像识别技术融合的烟叶成熟度的判别

采用特征级融合方式进行数据融合。为了降低计算的复杂程度,提高预测判别的精度,在数据融合之前,采用PCA分别将2种技术所得到的高维数据转化为低维数据,将总得分超过99%作为保留特征主成分的标准,采用数据融合技术将两类数据的特征主成分进行融合,再对融合数据进行统一,消除不同数据间量纲的影响。从融合模型分类正确率(表4)来看,光谱与图像融合后仅保留了约30个变量,比光谱数据有极大的减少,与图像特征数据相当,3个品种的融合模型预测正确率都超过了92%,模型的预测能力有较明显的提高。

近红外光谱在进行数据采集时,避开主脉采集两侧信息,忽略了叶脉对成熟度的影响。

烟叶的叶片颜色是判断烟叶成熟的重要依据<sup>[22]</sup>。烟叶的生长过程中伴随着许多生理生化的改变,而叶片由绿变黄是因为内部的叶绿素和类胡萝卜素含量的降低,叶黄素含量的升高,不同烟叶成熟度有着不同的叶绿素、类胡萝卜素和叶黄素含量,也就表示有着不同的颜色表现。综上,可以利用烟叶的颜色和纹理特征,对其进行信息提取,使用图像特征模型进行烟叶成熟度类别分析,但此种方法不能判定烟叶这几种成分的具体含量,因此,其判别分类的准确度并不高。

将近红外光谱模型和图像特征模型分类结果进行比较,发现近红外光谱模型的性能要优于图像特征模型,可能是图像特征模型需对烟叶颜色进行信息提取,针对一些极为相近的颜色在判别上会有所模糊,特别是两个相邻等级的烟叶成熟度之间,易造成误差,降低了图像特征模型的准确度,而近红外光谱模型所提取的是烟叶某些成分的含

量信息, 没有其他不利因素的干扰, 所以近红外光谱模型分类预测结果会比图像特征模型更好。

采用的数据融合技术是将近红外光谱技术和图像识别技术的信息进行融合处理, 实现更加精确、完整的决策<sup>[23]</sup>。但在进行数据融合之前需制定合理的融合等级方法, 不同的融合等级方法有着不同的融合效果。一般地, 数据融合有三个层次, 分别为数据级融合、特征级融合以及决策级融合<sup>[24]</sup>。本试验采用近红外仪器的 InGaAs 采集器与摄像机的 CMOS 采集器, 二者的性质不同, 所以不能将这两种采集器所采集到的信息进行简单直接地融合, 基于这两类信息数据得不到相应的分析判决, 决策级融合的方式不予适用, 所以所采用的是特征级融合方式进行数据融合。

将近红外光谱技术与图像识别技术进行融合, 利用随机森林算法构建了鲜烟叶 5 个成熟等级的融合判别模型, 3 个烟草品种模型预测正确识别率均超过 93%, 在一定程度上实现了烟叶成熟程度的准确判别, 且建模样本数多, 样本来自不同的烟叶产区, 模型的适应性较广。近红外光谱技术采集的是烟叶内在成分含量信息, 图像识别技术获取的是烟叶的颜色和纹理特征信息, 融合数据将二者融合在一起, 能够更全面, 更完善地描述烟叶的成熟特征。融合模型显示出较光谱或图像模型明显的优势, 避免了单一传感器采集信息片面的缺点, 而且融合数据的容错率高, 覆盖率广, 为烟叶成熟度的分类判别提供一种辅助手段。

#### 参考文献:

- [1] 许自成, 赵瑞蕊, 王龙宪, 等. 烟叶成熟度的研究进展[J]. 东北农业大学学报, 2014, 45(1): 123-128.  
XU Z C, ZHAO R R, WANG L X, et al. Research advance of maturity of flue-cured tobacco leaves[J]. Journal of Northeast Agricultural University, 2014, 45(1): 123-128.
- [2] 杨树勋. 准确判断烟叶采收成熟度初探[J]. 中国烟草科学, 2003, 24(4): 34-36.  
YANG S X. Preliminary study on accurate determination of leaf tobacco harvest maturity[J]. Chinese Tobacco Science, 2003, 24(4): 34-36.
- [3] BI Y M, LI S T, ZHANG L L, et al. Quality evaluation of flue-cured tobacco by near infrared spectroscopy and spectral similarity method[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 215: 398-404.
- [4] ZHANG F, ZHANG X H. Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation[J]. Sensors, 2011, 11(3): 2369-2384.
- [5] JI J T, DENG M L, HE Z T, et al. Extraction of tobacco leaves color features based on HSI color space[J]. Applied Mechanics and Materials, 2014, 651/652/653: 2424-2429.
- [6] 李佛琳, 赵春江, 王纪华, 等. 不同成熟度烤烟鲜叶的高光谱响应及其判别分析[J]. 福建农林大学学报(自然科学版), 2008, 37(6): 565-569.  
LI F L, ZHAO C J, WANG J H, et al. Spectral characteristic of flue-cured virginia tobacco leaves of different maturity grades and spectra discriminant classification[J]. Journal of Fujian Agriculture and Forestry University(Natural Science Edition), 2008, 37(6): 565-569.
- [7] 梁寅, 张云伟, 李军营. 基于支持向量机的云烟 87 烟叶成熟度高光谱遥感识别[J]. 西南农业学报, 2013, 26(3): 957-962.  
LIANG Y, ZHANG Y W, LI J Y. Spectral feature extraction for flue-cured Virginia tobacco leaves of different maturity grades[J]. Southwest China Journal of Agricultural Sciences, 2013, 26(3): 957-962.
- [8] 李向阳, 刘国顺, 史舟, 等. 利用室内光谱红边参数估测烤烟叶片成熟度[J]. 遥感学报, 2007, 11(2): 269-275.  
LI X Y, LIU G S, SHI Z, et al. Predicting leaf maturity of flue-cured tobacco using red edge characteristics of laboratory spectrometry[J]. Journal of Remote Sensing, 2007, 11(2): 269-275.
- [9] 王承伟, 宾俊, 范伟, 等. 基于近红外光谱技术结合随机森林的烟叶成熟度快速判别[J]. 西南农业学报, 2017, 30(4): 931-936.  
WANG C W, BIN J, FAN W, et al. Rapid discrimination of maturity of tobacco leaf based on near-infrared spectroscopy and random forest[J]. Southwest China Journal of Agricultural Sciences, 2017, 30(4): 931-936.
- [10] 史龙飞, 宋朝鹏, 贺帆, 等. 基于机器视觉技术的烤烟鲜烟叶成熟度检测[J]. 湖南农业大学学报(自然科学版), 2012, 38(4): 446-450.  
SHI L F, SONG C P, HE F, et al. Determination of the maturity grades of fresh leaves for flue-cured tobacco[J]. Journal of Hunan Agricultural University(Natural Sciences), 2012, 38(4): 446-450.
- [11] 周首峰. 基于图像处理的烟叶成熟度检测技术研究[D]. 杨凌: 西北农林科技大学, 2013.  
ZHOU S F. Tobacco maturity detection research based on image processing[D]. Yangling, China: Northwest A & F University, 2013.