寒旱农业科学 2023,2(10):952-961
Journal of Cold-Arid Agricultural Sciences

机器视觉的烟叶物理特性表征方法

梁耀星[1]，古政坤[1]，刘晓涵[2]，曹燕琼[2]，李俊鑫[2]，刘程炜[4]，张　建[3]，罗海燕[2]
（1. 广东中烟工业有限责任公司技术中心，广东　广州　510385；2. 广东韶关烟叶复烤有限公司，广东　韶关　512000；3. 上海创和亿电子科技发展有限公司，上海　200082；
4. Mc Master University, Hamilton Ontario Canada L8S4L8）

摘要： 为了研究梅州地区各等级烟叶的外观特征与物理特性间的关系，找到一种通过烟叶外观特征表征其物理特性的方法，选取了梅州6个产地、12个等级的初烤烟叶共977片。使用机器视觉设备和质构仪分别检测了烟叶样本的外观特征和物理特性。选取其中781片烟叶样本作为训练集，使用了弹性网络、极端随机树、支持向量机等回归模型以及模型融合技术分别构建了基于烟叶外观特征的最大拉力、剪切力和撕裂度的表征模型。选取196片烟叶样本作为测试集，以平均绝对误差为模型评价指标，评估了3种表征模型的泛化性能。结果表明，对于最大拉力的表征模型而言，模型在测试集上的预测值与真实值的相关系数超过0.73，拟合优度为0.54；对于剪切力的表征模型而言，模型在测试集上的预测值与真实值的相关系数超过0.78，拟合优度为0.60；对于撕裂度的表征模型而言，模型在测试集上的预测值与真实值的相关系数超过0.75，拟合优度为0.56。烟叶的外观特征对于烟叶的最大拉力、剪切力和撕裂度具有一定的表征能力。

关键词： 物理特性；外观特征；机器视觉；表征模型
中图分类号： S-3　　**文献标志码：** A　　**文章编号：** 2097-2172(2023)10-0952-010
doi:10.3969/j.issn.2097-2172.2023.10.013

Characterization Method of Tobacco Leaves Physical Properties Based on Machine Vision

LIANG Yaoxing[1], GU Zhengkun[1], LIU Xiaohan[2], CAO Yanqiong[2], LI Junxin[2], LIU Chengwei[4], ZHANG Jian[3], LUO Haiyan[2]
(1. China Tobacco Guangdong Industrial Co., Ltd., Guangzhou Guangdong 510385, China; 2. Guangdong Shaoguan Tobacco Recuring Co., Ltd., Shaoguan Guangdong 512000, China; 3. Shanghai Micro Vision Technology Ltd., Shanghai 200082, China; 4. Mc Master University, Hamilton Ontario L8S4L8, Canada)

Abstract: In order to study the relationship between the appearance characteristics and the physical properties of various grades of tobacco in Meizhou, an attempt was made to find a method about characterizing the physical properties of tobacco through its appearance characteristics. A total of 977 pieces of first-roasted tobacco leaves of 12 grades from six origins in Meizhou were selected. The appearance characteristics and physical properties of the tobacco samples were examined using machine vision equipment and texture analyzer, respectively. A total of 781 tobacco samples were selected as the training set. Regression models such as Elastic Net, Extremely Randomized Trees and Support Vector Machine were used along with the Ensemble technique to construct the characterization models of maximum tensile force, shear force and tearing degree based on the appearance characteristics of tobacco samples. A total of 196 tobacco samples were selected as the test set, and the generalization performance of the three characterization models were evaluated using the mean absolute error. The results indicated that the maximum tensile force model exhibited a correlation coefficient over 0.73 between the predicted and true values of the samples in the test set, with a goodness of fit of 0.54. Similarly, the shear force model demonstrated a correlation coefficient exceeding 0.78 and a goodness of fit of 0.60. Additionally, the tearing degree model displayed a correlation coefficient surpassing 0.75 and a goodness of fit of 0.56 for the predicted and true values of the samples in the test set. The appearance characteristics of tobacco leaves have certain ability to

收稿日期： 2023-05-25；**修订日期：** 2023-09-07
基金项目： 广东中烟工业有限责任公司项目(Q/GDZY207011-02)。
作者简介： 梁耀星(1984—)，男，广东阳江人，农艺师，硕士，主要从事烟叶质量检验及研究工作。Email: liangyaoxing@gdzygy.com。
通信作者： 罗海燕(1965—)，女，广东梅州人，高级农艺师，主要从事打叶复烤工艺技术研究工作。Email: sgxcl@126.com。

characterize the maximum tensile force, shear force and tearing degree of tobacco leaves.

Key words: Physical characteristic; Appearance characteristic; Machine vision; Characterization model

烟叶的物理特性是烟叶加工的重要指标之一，直接影响加工过程的造碎程度，进而影响烟叶原料的损耗及其加工质量[1]。烟叶的物理特性与其部位等级密切相关，此外，烟叶的部位等级又与其外观特征有着强烈的相关性[2]，因此，烟叶的物理特性在一定程度上可由烟叶的外观特征进行表征。长期以来，人们都在对烟叶加工过程中叶梗分离段如何提高烟叶质量并减少造碎进行研究，提高加工中的烟叶质量并减少造碎对各工业公司具有非常重要的作用和影响[3]。马雨佳等[4]研究了烟叶抗破碎指数与物理特性的关联性，发现烟叶抗破碎指数与抗张强度、延伸率、回复性、厚度呈显著正相关。另外，在基于理化特性的烟叶耐加工性研究中发现烟叶耐加工性与主要物理特性间存在不同程度的相关，但目前少有学者对烟叶外观特征与物理特性之间的关系进行研究。为此，我们选取了梅州 6 个产地、12 个等级的初烤烟叶共 977 片。使用机器视觉设备和质构仪分别检测了烟叶样本的外观特征和物理特性。选取其中 781 片烟叶样本作为训练集，使用了弹性网络（Elastic Net）、极端随机树（Extremely Randomized Trees）、支持向量机（Support Vector Machine）等回归模型以及模型融合（Ensemble）技术分别构建了基于烟叶外观特征的最大拉力、剪切力和撕裂度的表征模型。选取 196 片烟叶样本作为测试集，以平均绝对误差为模型评价指标，评估了 3 种表征模型的泛化性能，探讨了烟叶外观特征与物理特性间的关系，并尝试通过构建机器学习回归模型以基于烟叶外观特征表征物理特性。与生产线上工人们根据经验及眼观手摸的方式来评估烟叶物理特性相比，基于机器学习算法的物理特性表征模型更具客观性和准确性，这种表征模型在烟叶加工中能够更好地评估烟叶物理特性，进而对后续润叶、打叶过程中工艺参数的调节提供可靠的数据支持。

1　材料与方法

1.1　材料

1.1.1　供试样品　于 2020 年和 2021 年采集梅州 6 个地区、12 个等级的初烤烟叶样本共 977 片。由烟叶分级专家按照烤烟国家标准（GB2635—1992）进行等级分选[5]，所选初烤烟叶产地为广东省梅州市下的大埔县、丰顺县、蕉岭县、梅县区、平远县和五华县，等级分别为上部橘黄一级烟（B1F）、上部橘黄二级烟（B2F）、上部橘黄三级烟（B3F）、上部橘黄四级烟（B4F）、中部橘黄一级烟（C1F）、中部橘黄二级烟（C2F）、中部橘黄三级烟（C3F）、中部橘黄四级烟（C4F）、下部橘黄一级烟（X1F）、下部橘黄二级烟（X2F）、下部橘黄三级烟（X3F）、下部橘黄四级烟（X4F）。

1.1.2　实验设备　烟叶综合测试台 GTM 600（上海创和亿，中国）；质构仪 CTX（AMETEK Brookfield，美国）；恒温恒湿箱 KBF115-E6（Binder，德国）。

1.2　*方法*

1.2.1　外观特征测定　将采集的烟叶样本展平后置于综合测试台内采集烟叶样本图像，从中提取外观特征，包括重量、长度、宽度、周长、面积、颜色深浅、颜色均匀度、油分、厚度和结构；并从图像中提取更细致的颜色特征[6]，包括 RGB 颜色空间中的 B 均值、G 均值、R 均值，HSV 颜色空间中的 V 均值、S 标准偏差以及 Lab 颜色空间中的 L 均值、a 均值和 b 均值。烟叶的颜色深浅和均匀度分别以烟叶颜色的 H 均值和 H 标准偏差表征，油分以 S 均值来表征，厚度以烟叶透光强度的倒数来表征，结构则由烟叶重量与烟叶面积之比来表征。

1.2.2　物理特性测定　将每个产地、每个部位等级的烟叶样本置于恒温恒湿箱内，对恒温恒湿箱设置不同的温度和湿度，平衡 48 h 后将烟叶样本取出。将每片烟叶样本按叶尖、叶腰、叶基裁剪出检测叶片，每张检测叶片要求宽度为 15 mm、长度不小于 40 mm。使用质构仪对裁剪好的检测叶片测量最大拉力和剪切力，进而通过计算得到撕裂度、拉伸长度和撕裂距离。平衡烟叶样本时恒温恒湿箱设定的温度分别为20.0、25.0、30.0 ℃，湿度分别为 60.0%、65.0%、70.0%（表1）。

表1 恒温恒湿箱设定的温湿度

梯度	温度/℃	湿度/%
1	20.0	60.0
2	20.0	65.0
3	20.0	70.0
4	25.0	60.0
5	25.0	65.0
6	25.0	70.0
7	30.0	60.0
8	30.0	65.0
9	30.0	70.0

1.3 数据挖掘

1.3.1 数据划分 按照烟叶的部位等级对样本数据集进行分层抽样，训练集与测试集比例设定为80%：20%。为进一步判断数据划分合理性，使用主成分分析（Principal Component Analysis, PCA）方法将训练集和测试集样本投影至二维平面并观察它们的分布情况。

主成分分析是一种数据降维技术[7]，其本质是一个线性变换，通过该线性变换，数据集被变换至一个新的坐标系中，使数据投影的第一大方差对应的方向在第一个坐标轴上（称为第一主成分），第二大方差对应的方向在第二个坐标轴上（第二主成分），依此类推。通过主成分分析，通常能够使原始数据集的维数降低，同时保留数据集中尽可能多的信息[8]。

1.3.2 数据探索 对烟叶样本的各物理特性指标以及外观特征指标绘制频数直方图，检查数据分布情况。对烟叶样本的各外观特征及物理特性指标绘制相关系数热力图，检查各指标之间的线性相关程度。特征间的线性相关程度很高说明不同特征所包含的信息存在一定重复，即数据集中存在冗余信息，若不对这些冗余信息加以处理，则可能导致后续机器学习模型的训练和推断速度变慢。为此以0.8作为相关系数的阈值，对外观特征进行筛选，以消除数据集中的冗余信息。最后绘制经过筛选后的外观特征与物理特性之间的相关系数热力图，据此确定要构建回归模型的物理特性。

1.3.3 特征工程 在现有外观特征的基础上，通过构建如下组合特征可以扩充特征空间，以得到更为丰富且全面的烟叶外观特征，组合特征的定义方式如表2所示。绘制扩充后的外观特征与物理特性之间的相关系数热力图，并计算所有烟叶外观特征与各物理特性的皮尔逊相关系数 r 的95%置信区间和对应的 p 值，以观察各外观特征与目标物理特性之间的相关性。

表2 组合特征

组合特征	组合方式
重长比	重量/长度
重宽比	重量/宽度

1.3.4 模型训练与融合 常用的机器学习回归模型包括：弹性网络（Elastic Net, EN），极端随机树（Extremely Randomized Trees, ERT），支持向量机（Support Vector Machine, SVM）和多层感知机（Multi-layer Perceptron, MLP）等。弹性网络是一种同时具有L1和L2正则化的线性模型[9]，它具备自动选择特征的能力。极端随机树与随机森林类似，都是由许多决策树集成的模型，但其完全随机的分裂方式使其在模型训练上比随机森林更快。另外在各种实践中，极端随机树与随机森林在泛化性能上往往难分伯仲。支持向量机原本是基于最大化分类间隔的一种分类算法，对其方法稍做修改，将数据尽可能多地纳入“间隔”中，即可得到一种强有力的回归模型，该算法尤其适用于中小型复杂数据集。多层感知机是一种前向结构的人工神经网络，包含输入层、输出层及若干隐藏层[10]，理论上它能够拟合任何函数，并且已被证明是一种通用的近似算法[11]。基于扩充后的外观特征分别构建上述4种基础机器学习回归模型：弹性网络、极端随机树、支持向量机和多层感知机，并在训练集上进行五折交叉验证，通过网格搜索获得各模型的最佳超参数组合。在最佳超参数组合下计算各回归模型在训练集上交叉验证的平均绝对误差，以判断各物理特性下的每个回归模型的泛化性能的优劣。

为进一步提升模型泛化能力，针对每个目标物理特性，选择对应平均绝对误差的均值最小的3个模型进行等权重的模型融合。一般而言，不同的基础模型基于不同的假设，对同一数据集的适应程度也各不相同，因此在预测新样本时，不同

模型“所犯的错误”也有所不同(即不同基础模型给出的预测值与真实值之差也不相同)，因此，将基础模型进行融合得到的融合模型通常比单个基础模型有着更强的泛化能力[12]。

1.3.5　模型测试及评估　本文选用平均绝对误差作为评价融合回归模型泛化性能的指标，这种评价指标在数据集中可能存在异常值时表现较好，因此适合于使用质构仪检测得到的烟叶物理特性数据[13]。平均绝对误差的计算公式如下：

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|$$

式中，y_i 为第 i 个样本的标签值，$\hat{y}_i$ 为第 i 个样本的预测值，n 为样本容量。

2　结果与分析

2.1　数据划分结果

将完整数据集(c)划分为训练集(a)与测试集(b)，并分别绘制三个数据集中所有部位等级的分布图。从图 1 可见，采用分层抽样法可以基本保持训练集与测试集中每个部位等级的样本所占的比例。

全部梅州烟叶样本经分层抽样后，781 个样本用于构建预测烟叶物理特性的回归模型，196 个样本用于评估所建回归模型的泛化性能（在未知数据上给出准确预测值的能力)。

训练集与测试集经过主成分分析法投影至二维平面后，所得散点图如图 2 所示。从图 2 可以看出，训练集与测试集的样本点在二维平面上的分布基本一致，说明测试集的选取对全体样本数据集有较好的代表性，因而使用该测试集对模型进行泛化性能评估所得的结果具有较高的可靠性。

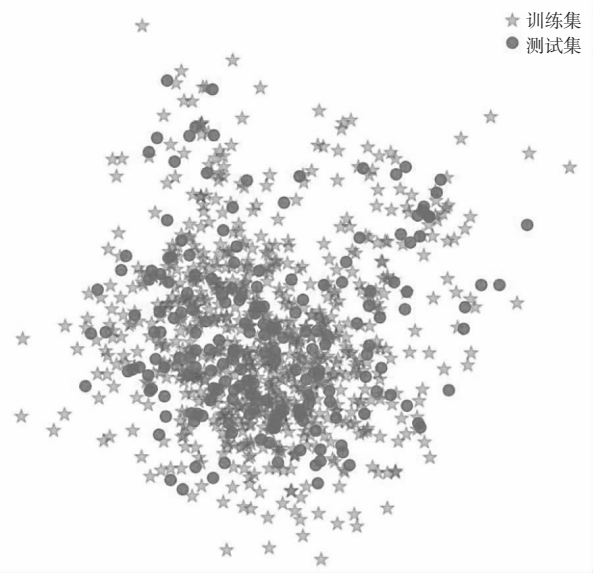

图 2　训练集与测试集的投影

2.2　数据探索结果

各烟叶外观特征的分布情况如图 3 所示，各烟叶物理特性的分布情况如图 4 所示。

观察图 3 与图 4，发现多数烟叶外观特征基本呈现正态分布，长度、油分和厚度则呈现出一定的偏态分布。烟叶物理特性则基本呈现正态分布，没有明显异常。对训练集中所有 18 个外观特征绘制相关系数热力图，如图 5 所示。

从图 5 可以发现，重量和宽度均与面积存在较高的相关性，深浅和部分颜色特征(如 B 均值、G 均值、R 均值、L 均值和 b 均值等）也存在线性相关程度较高的现象。除此之外，B 均值、G 均值、R 均值、L 均值和 b 均值之间也存在强烈的线

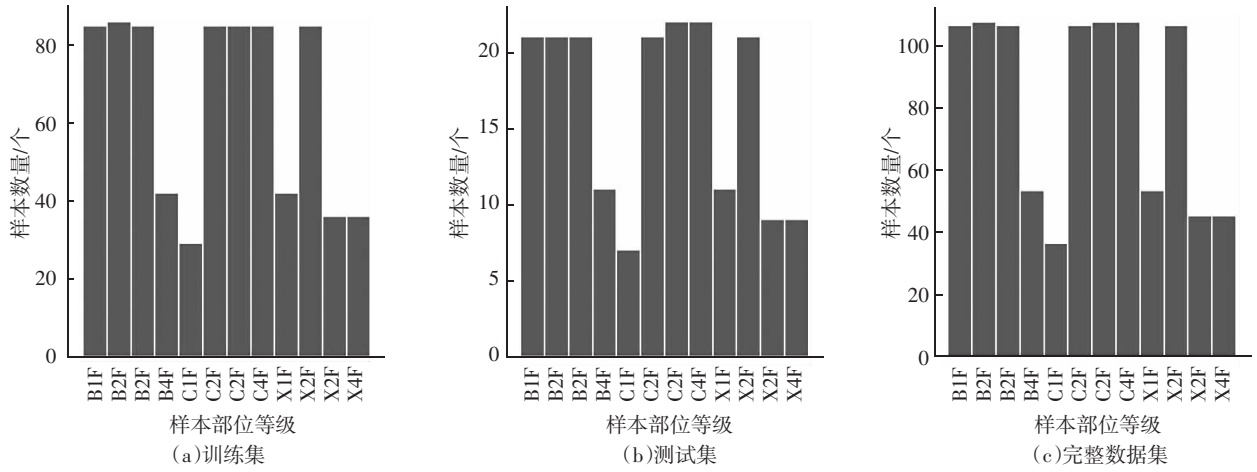

图 1　训练集、测试集和完整数据集中的部位等级分布

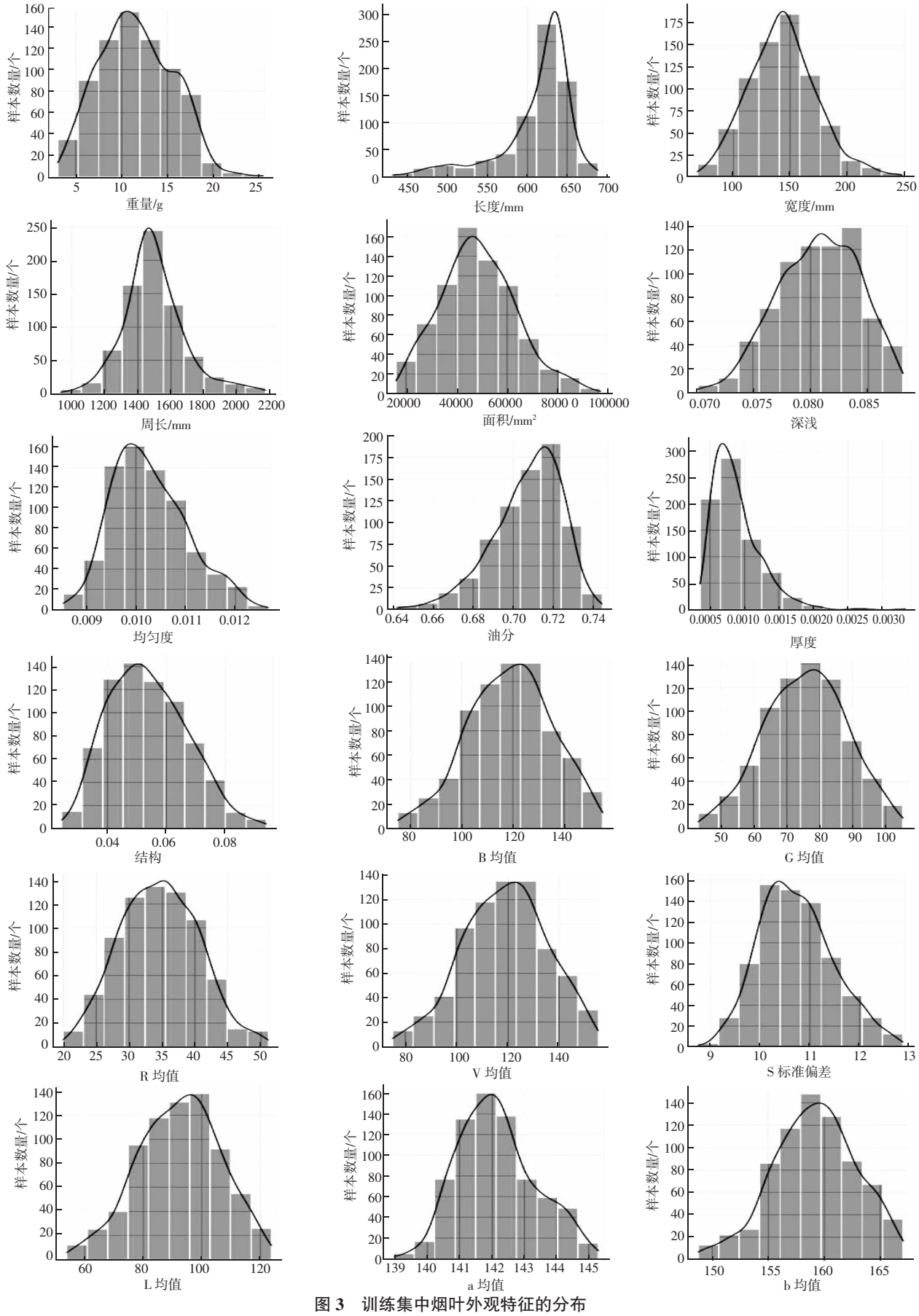

图 3　训练集中烟叶外观特征的分布

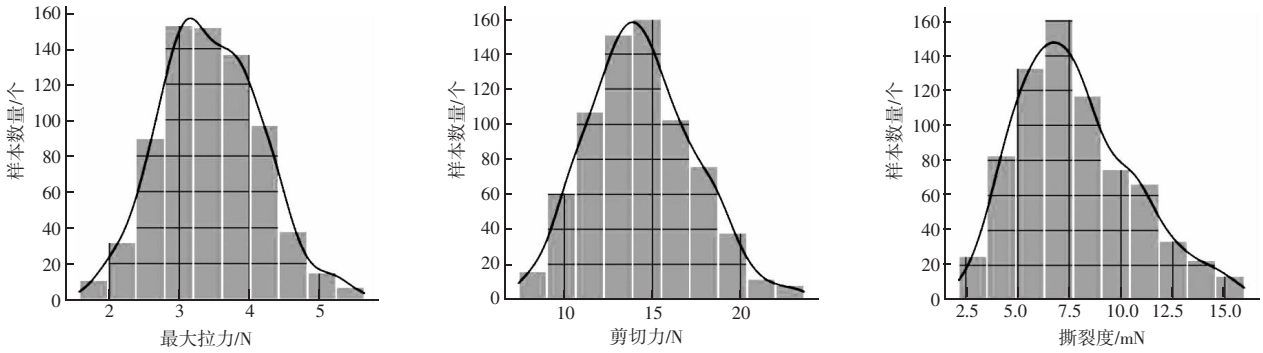

图 4　训练集烟叶物理特性的分布

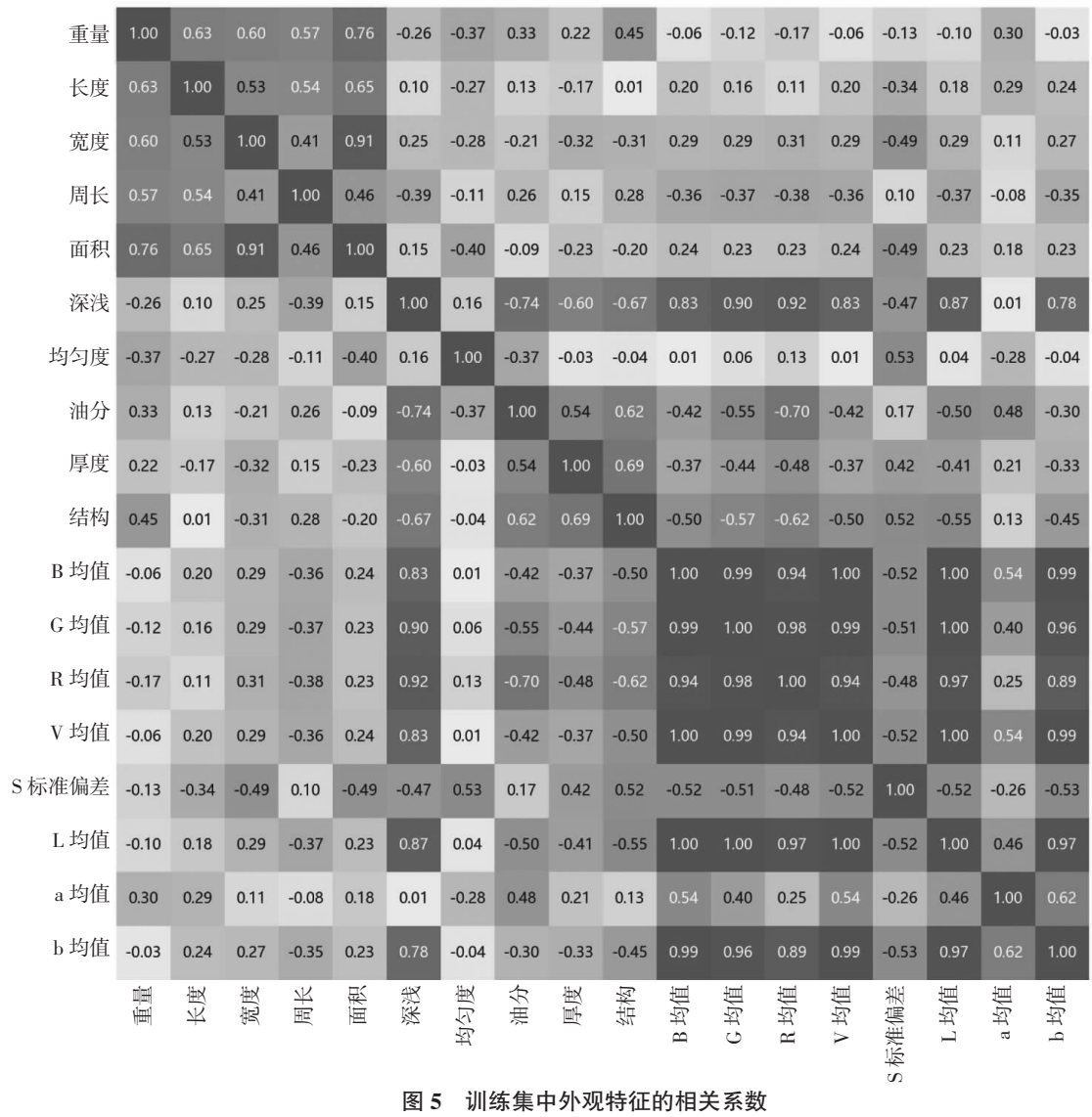

图 5　训练集中外观特征的相关系数

性相关性，相关系数均在 0.85 以上。

设定相关系数阈值并经过特征筛选后，剩余的外观特征为 12 个，对这些外观特征重新绘制相关系数热力图可以看到，高相关的外观特征已经不存在了，具体如图 6 所示。

绘制筛选后的外观特征与物理特性之间的相关系数热力图，结果如图 7 所示。发现拉伸长度和撕裂距离与所有外观特征的相关性都较弱，因此只对最大拉力、剪切力和撕裂度这 3 个物理特性构建回归模型。

2.3 特征工程结果

绘制扩充后的外观特征与物理特性之间的相关系数热力图，发现所创建的组合特征中重长比和重宽比与最大拉力和剪切力之间有较明显的正相关性，如图8所示。

重长比、重宽比和其他烟叶外观特征与各物理特性的皮尔逊相关系数 r 的95%置信区间和对应的 p 值如表3所示。

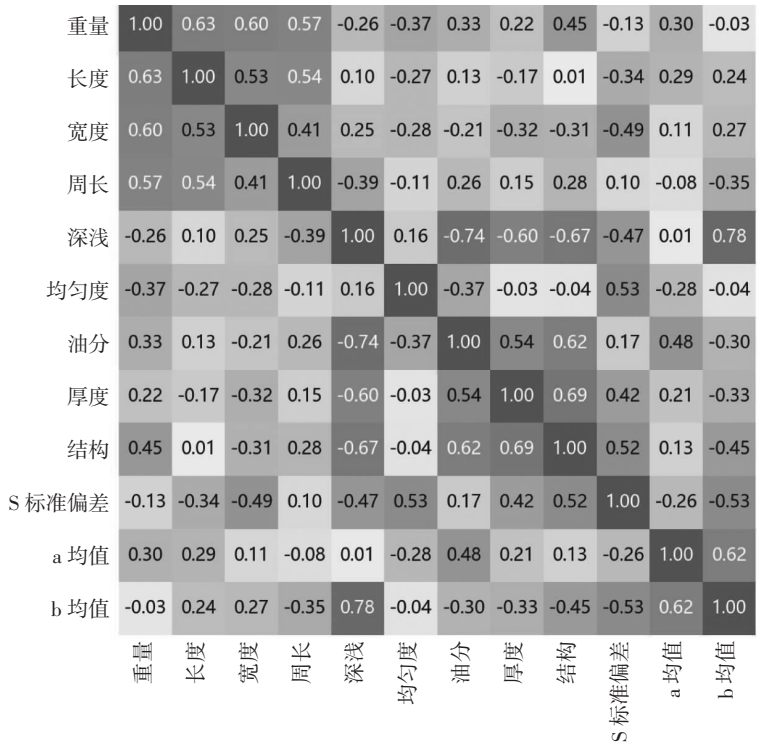

图6 筛选后的外观特征的相关系数

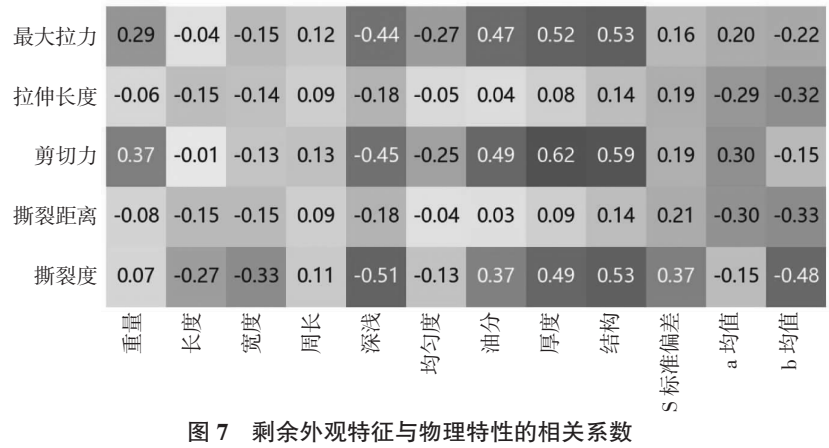

图7 剩余外观特征与物理特性的相关系数

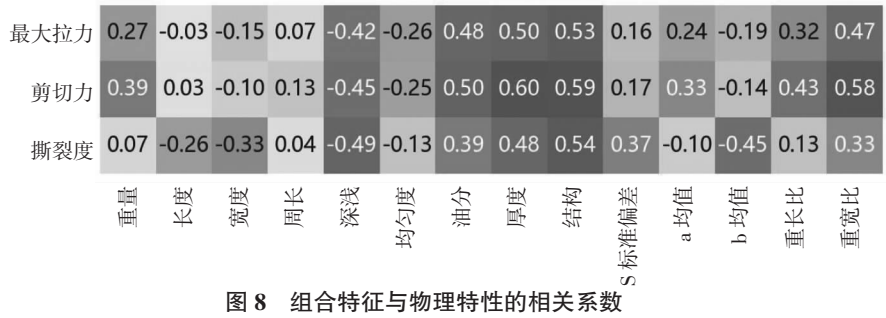

图8 组合特征与物理特性的相关系数

2.4　模型训练与融合结果

结合五折交叉验证与网格搜索，对各物理特性下的每个模型确定了其最佳超参数组合，结果如表 4 所示。在最佳超参数组合下计算 4 种基础回归模型在训练集上交叉验证的平均绝对误差(Mean Absolute Error, MAE)的均值，如表 5 所示。

对每个目标物理特性，选择对应平均绝对误差的均值最小的 3 个模型进行融合，根据表 5，预测最大拉力的融合回归模型由 SVR、EN 和 ERT 组成，预测剪切力的融合回归模型由 SVR、EN 和 MLP 组成，预测撕裂度的融合回归模型由 SVR、

表 5　基础模型的交叉验证 MAE 均值

物理特性	EN	ERT	SVR	MLP
最大拉力	0.450 2	0.451 6	0.439 6	0.451 6
剪切力	1.679 6	1.711 2	1.656 6	1.709 7
撕裂度	1.683 7	1.711 4	1.649 5	1.689 0

表 3　烟叶外观特征与各物理特性相关系数的 95%置信区间和 *p* 值

	最大拉力			剪切力			撕裂度		
	r	95%置信区间	*p*值	*r*	95%置信区间	*p*值	*r*	95%置信区间	*p*值
重量	0.27	[0.20, 0.34]	6.23×10^{-14}	0.39	[0.32, 0.45]	1.36×10^{-27}	0.07	[−0.01, 0.14]	6.95×10^{-2}
长度	−0.03	[−0.10, 0.04]	3.80×10^{-1}	0.03	[−0.04, 0.11]	3.60×10^{-1}	−0.26	[−0.33, −0.19]	5.50×10^{-13}
宽度	−0.15	[−0.22, −0.08]	6.40×10^{-5}	−0.10	[−0.17, −0.02]	9.17×10^{-3}	−0.33	[−0.39, −0.26]	6.03×10^{-20}
周长	0.07	[−0.00, 0.14]	6.00×10^{-2}	0.13	[0.05, 0.20]	6.92×10^{-4}	0.04	[−0.03, 0.12]	2.32×10^{-1}
深浅	−0.42	[−0.48, −0.36]	2.15×10^{-33}	−0.45	[−0.50, −0.39]	3.42×10^{-37}	−0.49	[−0.54, −0.43]	1.91×10^{-45}
均匀度	−0.26	[−0.33, −0.19]	6.36×10^{-13}	−0.25	[−0.32, −0.18]	6.38×10^{-12}	−0.13	[−0.20, −0.06]	3.96×10^{-4}
油分	0.48	[0.42, 0.53]	9.12×10^{-43}	0.50	[0.45, 0.56]	2.46×10^{-48}	0.39	[0.32, 0.45]	2.51×10^{-27}
厚度	0.50	[0.44, 0.55]	1.93×10^{-47}	0.60	[0.55, 0.64]	3.84×10^{-72}	0.48	[0.42, 0.53]	8.28×10^{-43}
结构	0.53	[0.47, 0.58]	8.30×10^{-54}	0.59	[0.54, 0.63]	1.72×10^{-69}	0.54	[0.49, 0.59]	6.87×10^{-58}
S标准偏差	0.16	[0.09, 0.23]	1.70×10^{-5}	0.17	[0.10, 0.24]	3.00×10^{-6}	0.37	[0.31, 0.43]	4.00×10^{-25}
a均值	0.24	[0.17, 0.31]	3.36×10^{-11}	0.33	[0.27, 0.40]	1.56×10^{-20}	−0.10	[−0.17, −0.02]	9.92×10^{-3}
b均值	−0.19	[−0.26, −0.12]	3.37×10^{-7}	−0.14	[−0.21, −0.07]	1.40×10^{-14}	−0.45	[−0.5, −0.39]	4.29×10^{-37}
重长比	0.32	[0.25, 0.38]	9.60×10^{-19}	0.43	[0.37, 0.49]	1.18×10^{-34}	0.13	[0.06, 0.20]	2.85×10^{-4}
重宽比	0.47	[0.41, 0.52]	2.57×10^{-41}	0.58	[0.53, 0.62]	4.24×10^{-66}	0.34	[0.27, 0.40]	5.22×10^{-21}

表 4　基础模型的最佳超参数组合

物理特性	EN		ERT		SVR(kernel='rbf')			MLP	
	alpha	l1_ratio	max_depth	min_samples_leaf	k	C	gamma	alpha	hidden_layer_sizes
最大拉力	0	0	12	4	12	26.37	0	26.37	(30, 50, 20)
剪切力	0.05	1.00	7	3	8	112.88	0.04	233.57	(20, 30, 10)
撕裂度	0.03	0	11	3	12	112.88	0.01	2.98	(20, 30, 10)

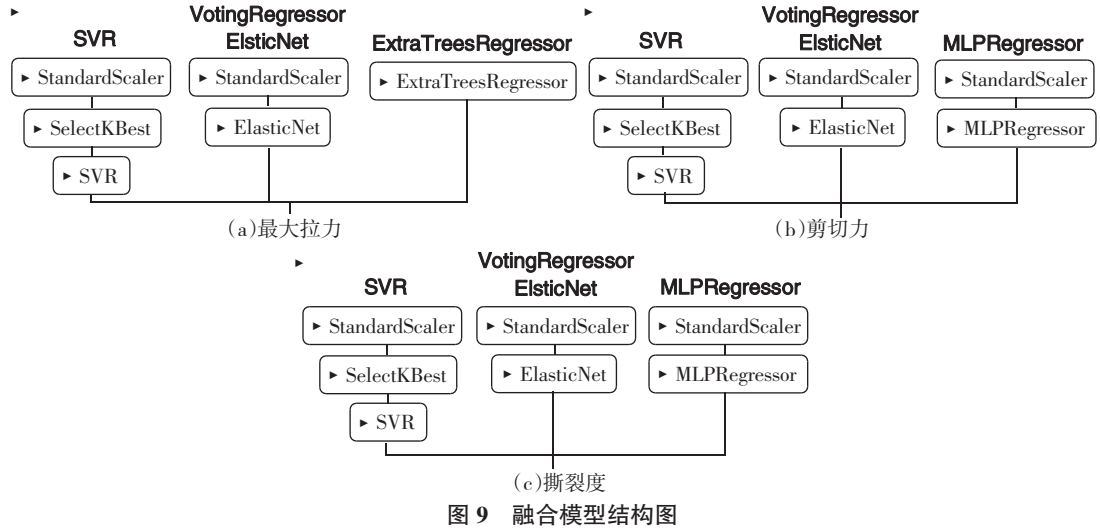

图 9　融合模型结构图

EN 和 MLP 组成。图 9 展示了这 3 个融合回归模型的结构。

为有效地评估 3 个融合回归模型的泛化性能，计算了它们在测试集上的预测值与真实值的皮尔逊相关系数，并绘制了相应的散点图，如图 10 所示。其中每个点的横坐标和纵坐标分别代表单个测试集样本的物理特性的真实值和融合回归模型在该样本点上的预测值，而黑色虚线表示预测值与真实值完全相等的情况，也就是说，当某个测试集样本对应的点落在黑色虚线上时，模型对该样本的预测值等于其真实值。

由图 10 可知，3 个融合回归模型给出的预测结果较理想，对于预测剪切力的融合回归模型而言，其在测试集上的预测值与真实值间的相关系数接近 0.80；而对于泛化性能稍弱的预测最大拉力的融合回归模型而言，其在测试集上的预测值与真实值间的相关系数也接近 0.75。

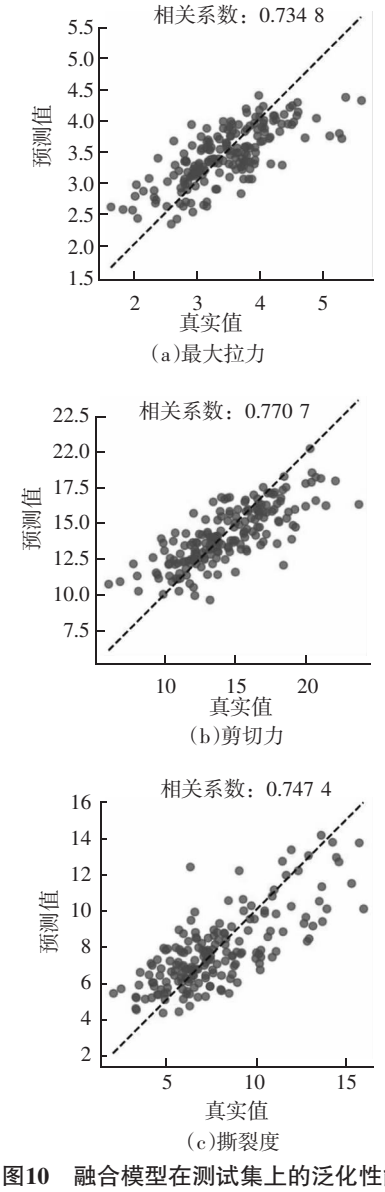

图10 融合模型在测试集上的泛化性能

另外，为了更全面地评估模型的泛化性能，还计算了 3 个融合回归模型在测试集上的拟合优度，如表 6 所示。从表 6 中可以看到，预测每个物理特性的融合回归模型在测试集上的拟合优度均在 0.50～0.60，其中剪切力模型的拟合优度最高，比较接近 0.60；而最大拉力模型的拟合优度稍低，略高于 0.50，这里的评估结果与图 10 中展示的结果一致。

表 6 3 个融合模型在测试集上的拟合优度

融合模型	拟合优度
最大拉力	0.53
剪切力	0.58
撕裂度	0.56

3 讨论与结论

数据集的合理划分对于模型的构建与验证非常重要，通过绘制训练集与测试集的二维投影散点图，可以快速观察到训练集与测试集的分布是否有明显差异。本研究对梅州 6 个产地的烟叶，利用机器学习回归模型针对烟叶外观特征与物理特性的关系进行了研究。基于烟叶外观特征构建机器学习回归模型，并将多种模型进行融合以表征烟叶物理特性的这种研究方法与其他学者常用的统计、回归分析方法相比，在模型选择上的限制较少，允许复杂度更高的模型，因而更适合于表征烟叶外观特征与物理特性之间的复杂关系[13-14]。因此本研究通过绘制二维投影散点图，确定了按分层抽样方法划分数据集的合理性，进而保证了模型验证结果的可靠性。

探索特征的分布情况对于模型构建也有指导性意义，具体而言，通过观察特征分布情况，可以发现样本中可能存在的异常值，有些模型（例如支持向量机）容易受到数据集中异常值的影响[15]，当模型对异常值进行拟合时，其泛化性能也必然会下降。本研究通过绘制变量的分布直方图，确认了数据集中没有明显的异常值，这为后续的建模打下了良好的数据基础。探索特征之间的相关系数能够帮助及早发现数据集中的冗余信息，使用主成分分析[16]，或其他特征选择算法

对这些冗余信息进行剔除，通常可以在不显著降低模型泛化性能的情况下缩短模型的训练和推断时间，同时也能够得到更具解释性的机器学习模型。由于本研究的数据集中包含了不同颜色空间的颜色特征，这些特征之间必然存在较高的相关性，也即特征中的信息存在冗余，为此采用了 Drop CorrelatedFeatures 特征选择算法，将原始的 18 个外观特征筛选为 12 个，同时确保了筛选后的外观特征间的相关系数均不超过 0.8。在此基础上探索外观特征与物理特性间的相关系数，进而确定最大拉力、剪切力和撕裂度作为烟叶外观特征来表征其物理特性，而其他物理特性因其与外观特征相关性较弱，可以预见对它们构建表征模型的效果不会很好，因此本研究中放弃对它们建模。

本研究表明，基于烟叶的外观特征，表征梅州烟叶的物理特性，在一定程度上具有可行性的，从融合模型在测试集上的泛化性能展示的结果看，当前的外观特征对剪切力有较好的表征能力，而对最大拉力的表征能力稍弱，表征效果仍有待提高。一个值得尝试的思路是寻找一些与最大拉力的相关性更强的外观特征，例如烟叶的纹理等，这需要将烟叶纹理进行数字化表征[17]。本研究选取了弹性网络、极端随机树、支持向量机和多层感知机算法对所选烟叶物理特性进行建模，结果表明，对于最大拉力的表征模型而言，模型在测试集上的预测值与真实值的相关系数超过 0.73，拟合优度为 0.54；对于剪切力的表征模型而言，模型在测试集上的预测值与真实值的相关系数超过 0.78，拟合优度为 0.60；对于撕裂度的表征模型而言，模型在测试集上的预测值与真实值的相关系数超过 0.75，拟合优度为 0.56。烟叶的外观特征对于烟叶的最大拉力、剪切力和撕裂度具有一定的表征能力。SVR 对 3 种物理特性的泛化性能最优，为打破模型假设的限制，进一步提升模型泛化性能，本研究中还采取了模型融合技术，将在每个物理特性上泛化性能最佳的前 3 个基础模型进行融合，分别得到了对物理特性有着更强泛化性能的 3 个融合回归模型。在这 3 个模型中，剪切力模型具有最优的泛化能力，这得益于剪切力与外观特征之间有着更高的相关性。与此同时，最大拉力与外观特征之间的相关性相对较弱，这也在一定程度上导致了最大拉力模型的泛化能力稍差。

参考文献：

[1] 马雨佳. 基于理化特性的烟叶耐加工性模型建立及应用[D]. 郑州：郑州轻工业大学，2022.

[2] 李　波，张仲文，章　程，等. 浅谈不同颜色模型在烟叶颜色数字化中的运用[J]. 天津农业科学，2021，27（7）：48–51.

[3] 邓　凯. 烟叶初加工过程中叶梗分离线三级打叶风分提高烟叶品质及减少造碎的相关研究[J]. 中国标准化，2017(18)：50–51.

[4] 马雨佳，纪晓楠，刘志洋，等. 烟叶抗破碎指数与物理特性的关联性分析[J]. 轻工学报，2022，37(3)：101–107.

[5] 国家技术监督局. 中华人民共和国国家标准：烤烟 GB2635–1992[S]. 北京：中国标准出版社，1992.

[6] 唐　岚. 密集化烟叶烘烤中图像特征提取及应用研究[D]. 重庆：重庆大学，2015.

[7] 郭奕通. 主成分分析在球坐标系下的分析与研究[D]. 广州：广东工业大学，2020.

[8] 池陈帆. 福建省耕地面积变化及其驱动因子研究[D]. 福州：福建农林大学，2011.

[9] 冯明皓. 自适应弹性网神经网络模型及算法研究[D]. 大连：大连海事大学，2020.

[10] 王春柳，杨永辉，赖辉源，等. 基于开放域对话系统的自动化评测方法研究[J]. 计算机应用研究，2020，37(5)：1456–1459.

[11] 顾成露. 基于神经网络的直扩信号捕获算法研究[D]. 成都：电子科技大学，2020.

[12] AURÉLIEN G. Hands–on machine learning with scikit–learn, keras, and tensorflow[M]. Sebastopol: O′Reilly Media, 2019.

[13] 李　晓，陈科冰，韩　明，等. 质构仪在烟叶力学特性检测中的应用进展[J]. 轻工学报，2021，36(3)：63–69.

[14] 秦　琅. 基于模型融合的烟叶烘烤过程状态预测方法研究[D]. 武汉：华中科技大学，2022.

[15] 陈思昂，王　敏，杜　薇，等. 基于原烟外观图像和近红外光谱的烟叶感官质量模型研究[J]. 寒旱农业科学，2023，2(3)：260–269.

[16] 姜有虎，李玉梅，李旭林，等. 基于主成分分析的嘉峪关产区马瑟兰葡萄最佳采收期确定[J]. 甘肃农业科技，2022，53(1)：94–98.

[17] 李嘉康，陶智麟，徐　波，等. 基于随机森林的烟叶纹理定量分析[J]. 湖北农业科学，2022，61(14)：155–159.