



基于标志点法的烟草叶形提取与判别

钟培阁¹, 周也莹¹, 张彦², 石屹², 郭焱¹, 李保国¹, 马毓韬^{1*}

(1. 中国农业大学土地科学与技术学院, 北京 100193; 2. 中国农业科学院烟草研究所, 农业农村部烟草生物学与加工重点实验室, 山东 青岛 266101)

摘要 本研究采用标志点法并基于39个烟草品种叶片图像提取叶形信息, 比较分析了不同生长时期、不同品种和不同叶位间的叶形差异。采用主成分分析法实现数据降维并对叶形差异进行可视化表达, 采用决策树、随机森林和支持向量机法对不同叶形进行判别分析。主成分分析结果表明, 烟草叶片间差异的主要来源分别为叶片宽和最大宽位置、叶片扭转程度以及叶柄部3方面, 分别占总差异的42.7%、21.3%、10.7%。叶形判别结果表明, 采用标志点数据的判别精度为52%~62%, 而采用常用叶形指标的判别精度为51%~54%。上部叶、中部叶的判别精度高于下部叶10%左右, 具有更明显的品种特征。团棵期叶片由于处于营养生长期, 判别精度低于开花期叶片近10%。去除12类非典型品种后, 标志点法判别精度上升至77%。标志点数据在叶形信息提取上的效果优于常用叶形指标, 为叶形信息的自动化提取提供了新的思路。

关键词 几何形态学; 标志点法; 烟草; 叶形; 机器学习

中图分类号 S 24 **文献标志码** A

引用格式 钟培阁, 周也莹, 张彦, 等. 基于标志点法的烟草叶形提取与判别[J]. 浙江大学学报(农业与生命科学版), 2022, 48(4): 533-542. DOI: 10.3785/j.issn.1008-9209.2021.07.091

ZHONG Peige, ZHOU Yeying, ZHANG Yan, et al. Extraction and discrimination of tobacco leaf shape based on landmark method[J]. *Journal of Zhejiang University (Agriculture & Life Sciences)*, 2022, 48(4): 533-542.

Extraction and discrimination of tobacco leaf shape based on landmark method

ZHONG Peige¹, ZHOU Yeying¹, ZHANG Yan², SHI Yi², GUO Yan¹, LI Baoguo¹, MA Yuntao^{1*} (1. College of Land Science and Technology, China Agricultural University, Beijing 100193, China; 2. Key Laboratory of Tobacco Biology and Processing, Ministry of Agriculture and Rural Affairs, Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, Shandong, China)

Abstract The shape information of leaves from 39 tobacco varieties was extracted by using landmark method. The differences in leaf shapes were compared and analyzed among different varieties and different leaf positions at different growth stages. Principal component analysis was used to reduce the dimensionality of the data. The sources of differences were visualized among different leaf shapes. Decision tree, random forest and support vector machine were used to perform discriminant analysis on tobacco leaf shapes. The results of the principal component analysis showed that the first three principal components accounted for 42.7%, 21.3% and 10.7% of the total differences in tobacco leaves at the flowering stage, which were characterized by leaf width and the

基金项目: 中国烟草总公司山东省公司重点项目“山东‘中棵烟’特征及形成机制研究”(201803)。

*通信作者(Corresponding author): 马毓韬(<https://orcid.org/0000-0002-6583-0342>), E-mail: yuntao.ma@cau.edu.cn

第一作者(First author): 钟培阁(<https://orcid.org/0000-0002-1743-5698>), E-mail: zhongpg1998@163.com

收稿日期(Received): 2021-07-09; 接受日期(Accepted): 2022-01-28

maximum width position, leaf torsion, and petiole size, respectively. The discriminant results of tobacco leaf shape based on machine learning showed that the discriminant accuracy based on landmark data was 52%–62%, while the value was 51%–54% for common leaf shape indicators. The discriminant accuracy on superior or medial leaves was about 10% higher than that of inferior leaves, representing more obvious characteristics of variety. Due to the growth of the leaves, the discriminant accuracy of the leaves at rosette stage was nearly 10% lower than flowering stage. The discriminant accuracy of landmark method increased to 77% after removing 12 atypical varieties. The effect of the landmark method on leaf shape information extraction is better than the common leaf shape indicators, which provides a new idea for the automated extraction of leaf shape information.

Key words geometric morphometrics; landmark method; tobacco; leaf shape; machine learning

叶片是植物的光合器官,不同种类植物的叶片具有不同的特征。基于叶片的植物表型特征提取在种质资源的保护和利用、作物品种间的分类识别以及优质表型特征的筛选等方面具有重要的现实意义^[1]。随着计算机图像处理、模式识别以及机器学习技术的发展,对植物叶片特征的提取也从传统的手动测量与人工鉴定发展到采用计算机自动提取^[2-4],不仅减少了工作量,也去除了主观因素的影响,大大增加了特征提取的精度与效率。

植物叶片特征可以从叶形、颜色、叶脉、纹理等多个方面获取。LARESE等^[5]采用机器学习方法根据叶脉特征实现了豆科植物的自动分类。MUNISAMI等^[6]开发了一个能够利用叶片形状和颜色直方图等信息实现植物特征识别的系统。郑一力等^[7]提取叶片的形状和纹理作为叶片多特征,实现了对植物叶片的快速分类识别。在以上各种叶片表型研究中,对叶形特征的研究主要集中在一些基本几何形状的提取上,如叶片长、叶片宽、叶面积、叶周长、质心、叶片长宽比等^[3]。若要对叶形的细微差异进行捕捉,需要采用更为精确的测量方法。

几何形态学是一门对形态差异进行定量分析和比较的学科^[8]。采用几何形态的测量方法可以量化物体的轮廓信息,便于后续的数据处理和分析。标志点法是实现几何形态测量的最主要的手段。其主要原理为在目标物体上选择具有代表性的标志点并剔除轮廓和大小信息,将得到的二维点阵投射到相同空间进行样本间的比较分析^[9]。随着标志点测量技术的发展,很多学者将其应用到植物研究中。CHITWOOD等^[10]采用标志点和椭圆傅里叶描述符对不同种类和不同叶片节点的西番莲属植物叶片进行比较分析。FELDMANN等^[11]将包括标志点法在内的多种几何形态测量方法与机器学习算

法结合进行草莓形状分类。在国内,标志点法在植物叶片表型上的应用较少,仅有少量学者采用该方法对树叶进行研究^[12],其在农作物叶片形状特征提取上几乎没有应用。

烟草(*Nicotiana tabacum* L.)为管状花目、茄科植物,是一种重要的叶用经济作物。不同品种的烟草叶片形状不同,同一株烟草不同生长时期、不同叶位的叶片形状也存在差异。目前对烟草叶形的定义主要基于人工经验判别,按照叶片最宽位置、叶片长宽比来描述,缺少精确的定量化评价方法。本研究以烟草为研究对象,基于几何形态测量方法自动提取烟草叶形信息,并采用机器学习方法对不同的烟草叶形进行判别分析来对比标志点数据和常用叶形指标之间的差异,旨在为烟草叶形信息的精确提取提供新方法。

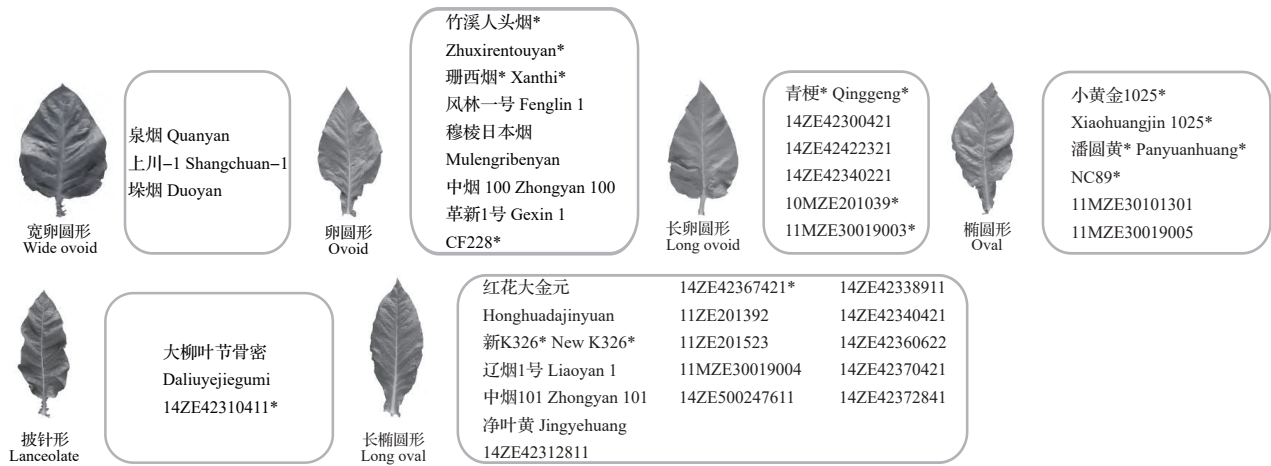
1 材料与方法

1.1 试验材料

田间试验于2019年5月5日—8月30日在山东省诸城市贾悦镇琅埠农场(35°35'24" N, 119°14'24" E)进行,试验共选择39个烟草品种,每个品种3个重复。品种具体信息见图1。目前对于烟草叶形的分类主要基于专家目测和对叶片长宽比的测量,根据39个烟草品种的叶形描述信息,可以将其划分为卵圆形、宽卵圆形、长卵圆形、椭圆形、长椭圆形、披针形6个类别。由于烟草叶形有典型类和非典型类之分,一些具有非典型叶形的烟草品种在人为分类时难以界定叶形,存在模糊性。非典型类烟草品种共12类,在图1中用“*”号标出。

1.2 叶片图像获取与处理

待烟草生长至团棵期及开花期(第一朵中心花开



*表示非典型类烟草品种。

Single asterisk (*) indicates atypical class of tobacco variety.

图 1 烟草品种与对应叶形分类

Fig. 1 Classification of tobacco varieties and the corresponding leaf shapes

放的时期),按照由下至上的顺序采集叶片,平铺,并采用佳能 EOS M3 相机拍照(图 2A),相机镜头距离叶片约 1.5 m,图像分辨率为 2 880 像素×1 920 像素。采用 Python 3.8.5 软件(<https://www.python.org/>)及 OpenCV 4.4.0 数据库(<https://opencv.org/>)进行叶片图像前景背景分离、轮廓提取与叶位划分等数据处理工作,结果如图 2B~D 所示。

采用超绿算法^[13],根据 RGB 图像中红(R)、绿(G)、蓝(B)3 个颜色分量构造出超绿指标 2G-R-B,对图像进行灰度化处理(图 2B),进一步突出植物叶片图像。采用最大类间方差法(OTSU 算法)^[14]确定灰度图像的最佳分割阈值,划分出属于烟草叶片的像素点。由得到的二值化图像提取出烟草叶片轮廓点云(图 2C)。根据背景上放置的 5 cm×5 cm 的

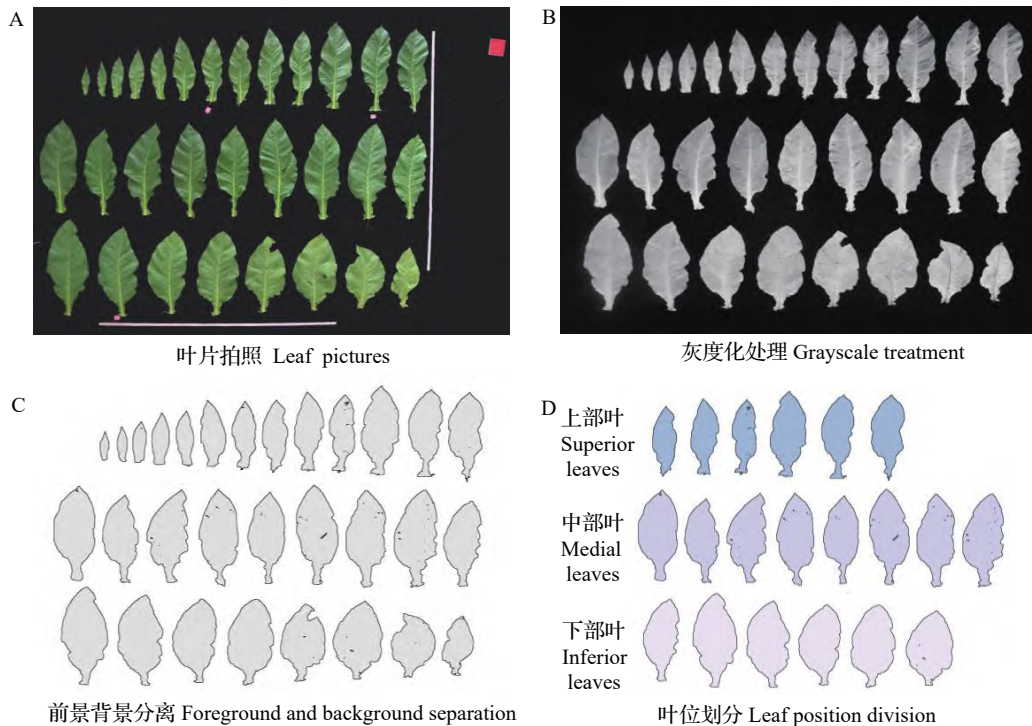


图 2 烟草叶片图像获取与处理流程图

Fig. 2 Flowchart of tobacco leaf pictures acquisition and processing

红色正方形卡片进行叶片大小校正。烟株上的烟叶自下而上通常划分为5个部位:脚叶、下二棚叶、腰叶、上二棚叶和顶叶。不同部位的烟叶间也存在差异。分离出烟草叶片图像后,去除研究意义较小的花叶、脚叶和一些缺损叶片,根据生产经验按照6-8-6的分布顺序将烟叶分为上部叶、中部叶、下部叶3个部分(图2D)。共获得烟草叶片2 865片。

1.3 各类叶形指标的提取

1.3.1 标志点数据提取

烟草叶片以椭圆形和卵圆形为主,为了尽量详尽地描述出叶片形状的变化,同时减少叶片缺损及褶皱等其他因素的干扰,本研究从叶片的最高点到最低点,每隔1/8分位在轮廓的最左端和最右端各选取一个像素点,共18个标志点^[15](图3A)。这样

的标志点排布不仅可以精确地刻画出叶柄、叶身、叶尖等位置的信息,还可以检测出叶片的扭转和不对称性。基于Python 3.8.5编程软件自动在叶片轮廓点云上提取标志点,获取各个点的位置坐标。

由于各个叶片的大小、方向、所在位置不同,在进行不同叶片对比分析前,需要先去除了干扰信息。本研究采用普氏叠印分析法(generalized Procrustes analysis, GPA)对位于不同叶片上的标志点进行归一化处理。其原理为采用最小二乘法最小化样本间存在的平移和旋转差异,实现叶片构型的标准化(图3B)^[16-17]。在将不同叶片的相应标志点坐标移动到对应位置后,即可进行每个叶片对应部位之间的差异比较。采用R 4.0.0编程软件(<https://www.r-project.org/>)中的shapes包^[18]对提取的叶片标志点坐标进行普氏叠印分析法处理。

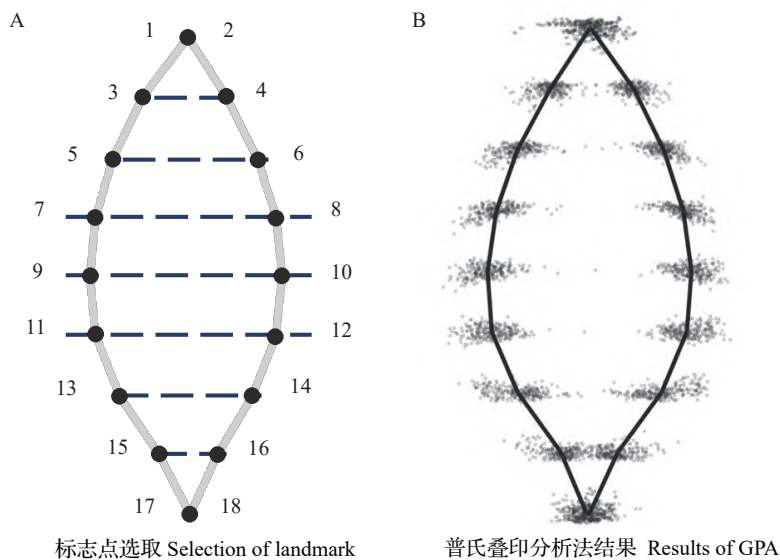


图3 烟草叶片标志点的选取及普氏叠印分析法结果示意图

Fig. 3 Schematic diagrams of the selection of tobacco leaf landmarks and the results of GPA

1.3.2 常用叶形指标的提取

基于Python 3.8.5软件及其OpenCV 4.4.0数据库提取叶片长、叶片宽、叶片长宽比、叶面积、叶面积指数5个常见的烟草叶片形状指标。其中,叶面积指数的计算方法为:

$$\text{叶面积指数} = \frac{\text{叶片实际面积}}{\text{叶片长} \times \text{叶片宽}}$$

1.4 统计分析 with 烟草叶形的判别

由于标志点数量众多,因此对其进行多元方差分析(multivariate analysis of variance, MANOVA),验证采用标志点数据是否可以检测出不同生长时

期、不同品种和不同叶位的烟草叶形差异。同时采用主成分分析方法对标志点坐标结果进行降维,提取出叶形差异的主要来源。采用R 4.0.0编程软件对以上数据进行分析。

采用标志点数据和其他常用叶形指标,将传统的烟草人工叶形分类结果作为真值,采用决策树(decision tree, DT)、随机森林(random forest, RF)和支持向量机(support vector machine, SVM)3种机器学习算法对各类烟草叶片进行分类建模和判别^[9],以此比较标志点数据与常用叶形指标的可分性差异。以上3类机器学习算法的原理如下。

决策树算法是一种应用较广泛的分类算法,其使用树的结构作为表现形式,内部的每一个节点代表对一个特征的测试,以递归的方式自上而下地对数据属性值进行比较,得到的分支代表特征属性的输出值。建立决策树的常用算法有很多,本研究采用分类和回归数(classification and regression tree, CART)算法^[20]。随机森林算法以决策树算法为基础,将多个随机决策树通过集成学习组合到一起,并根据多个分类器的投票决定最优的分类结果,分类精度相对更高^[21]。支持向量机是一种二分类模型,其基本思想为使用核函数定义的非线性特征映射将待分类数据映射到高维特征空间使其线性可分^[22],目前被广泛应用于模式分类、回归分析等多个领域。

采用十折交叉验证对以上 3 种模型进行分类器性能测试。将数据集分为 10 份,其中 9 份作为训练集来训练分类器,1 份作为测试集对分类器进行测试。上述过程循环 10 次并取 10 次验证结果的均值。采用准确率作为评判模型结果的指标,即预测正确的样本占总样本的比例。以上机器学习模型的训练及验证均采用 R 4.0.0 编程软件进行。

2 结果与分析

2.1 叶片标志点数据方差分析结果

标志点的多元方差分析结果表明,对于团棵期及开花期的烟草叶片,不同叶形类别、不同叶位以及两者间的交互效应均整体存在极显著差异。进

一步对各标志点单独进行双因素方差分析并将得到的 F 值结果可视化,以开花期的叶片数据为例。从图 4 中可知,黑色实心轮廓线为所有叶片的平均叶形,周围的灰色轮廓线为单独考虑不同叶类、不同叶位以及交互效应分组时各类别叶片的平均叶形。位于标志点的灰色圆形为该点处 F 值的可视化表达。其半径越大, F 值越大,即该标志点处不同类别叶片间的叶形差异越大。若标志点为深色实心圆,则代表该标志点处不同类别叶片间的叶形不存在显著差异。

图 4A 为 7 种叶形类别间由各标志点捕捉到的叶形差异,主要集中在叶片的中下部。图 4B 为 3 种叶位间由各标志点捕捉到的叶形差异,与图 4A 相比叶位间的差异上移,叶片上、中部的差异都较大,表明烟草不同叶位间的差异主要体现在叶片宽的变化上。图 4C 是叶形与叶位间交互效应的可视化表达。尽管这种交互效应十分微小,但显著性分析结果表明,对于不同的烟草品种,其上、中、下部叶的叶片形状变化也不同,且此交互效应在叶片上存在不对称性。团棵期烟草叶片数据的双因素方差分析结果(数据未列出)与开花期近似。团棵期烟草由于生长时期较短,叶片特征不明显,存在于品种间和叶位间的叶形差异不大,相比开花期烟草叶片的 F 值较小。

2.2 叶片标志点主成分分析结果

对获取的标志点数据进行主成分分析,降维得到不同烟草叶形的差异类型。对于开花期的烟草

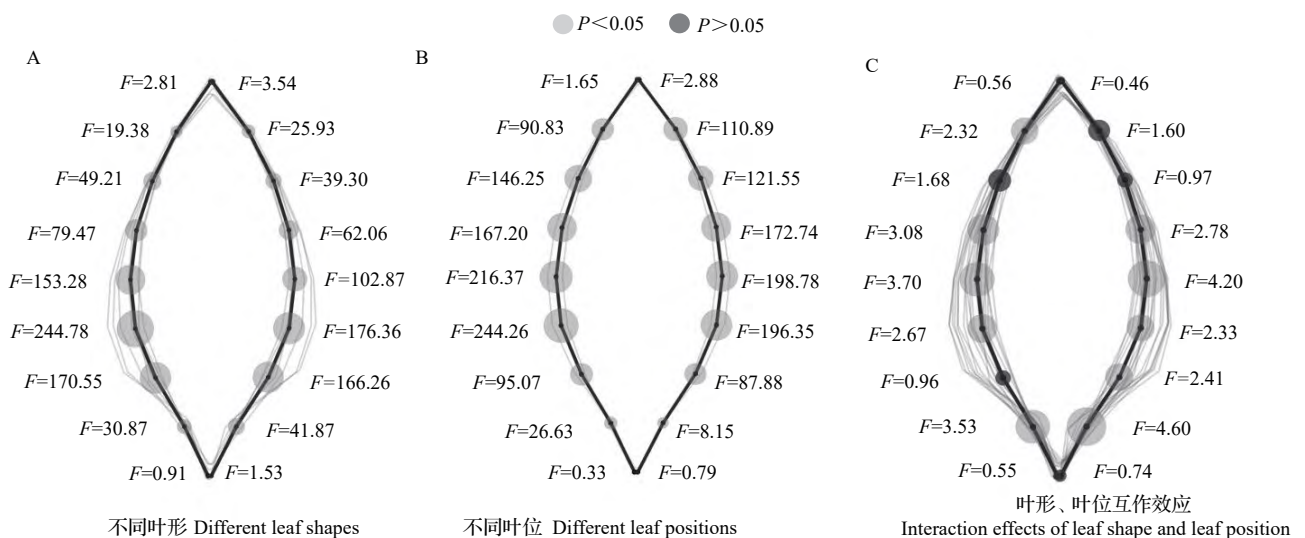


图 4 开花期不同叶形、叶位以及两者交互效应在各标志点的 F 值

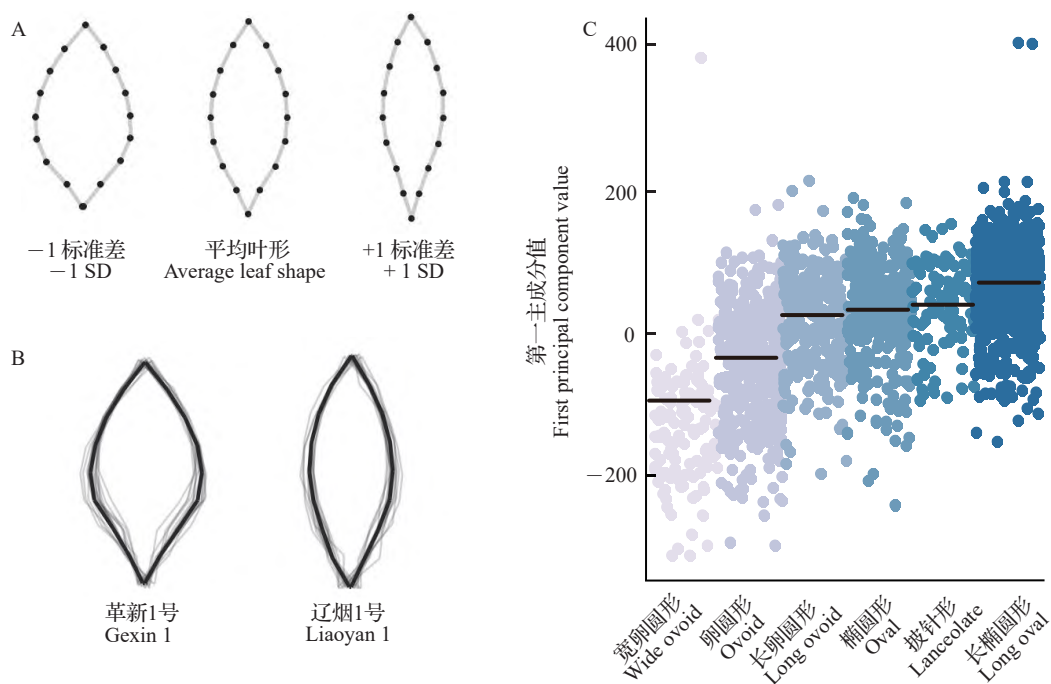
Fig. 4 F values of different leaf shapes, leaf positions and their interaction effects at each landmark at the flowering stage

叶片,第一主成分占总差异的42.7%,第二主成分占总差异的21.3%,第三主成分占总差异的10.7%。前三者累计占总差异的约75%,能够解释叶片间存在的主要差异。团棵期烟草叶片标志点数据的降维结果与开花期的结果相似,但3个主成分占总差异的比例不同,第一主成分占总差异的45.2%,第二主成分占总差异的26.3%,第三主成分占总差异的8.5%。

以开花期的烟草叶片为例,图5~7分别为第一、第二、第三主成分分析结果。3个主成分可以捕捉到来自不同方向的叶片差异。图5A、6A、7A均为可视化的主成分分析结果,平均叶形左右两边分别为将此主成分效应在平均标志点数据结果中放大1倍标准差与缩小50%标准差得到的叶片示意图。图5B、6B、7B均为表现出该主成分特征的典型烟草品种的叶形,其中黑色实心轮廓线为该品种所有叶

片的平均形状,周围的灰色轮廓线为该品种全部叶片形状,以此与可视化的主成分分析结果进行对比。

图5A显示:第一主成分值越大,叶片越偏向长椭圆形;第一主成分值越小,叶片越偏向宽卵圆形。该结果表明叶片的宽度和叶片最宽处的位置在一定程度上表现出同步变化的趋势。从第一主成分中可以捕捉到长叶片与宽叶片、椭圆形叶片与卵圆形叶片的差异。以具体品种为例,通过对比‘革新1号’与‘辽烟1号’的叶形可以看出第一主成分对烟草叶片形态的影响(图5B)。该主成分也是所有烟草叶形变化的最主要来源。图5C为各叶形类别烟叶的第一主成分值对比,中间的实线为该叶形的第一主成分均值。按照第一主成分的差异可以将几种叶形从宽卵圆形到长椭圆形进行清晰的划分。



A. 可视化第一主成分结果;B. 典型品种叶片平均形状;C. 各类别叶片第一主成分值。

A. Visualization of the first principal component results; B. Average leaf shapes of typical varieties; C. First principal component values of various types of leaves. SD: Standard deviation (the same as below).

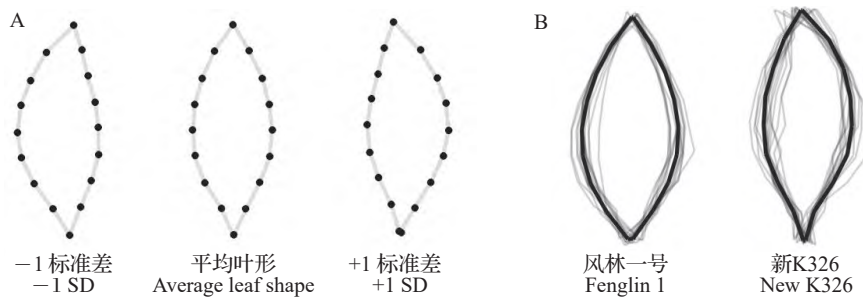
图5 开花期第一主成分结果对比

Fig. 5 Comparison of the results of the first principal component at the flowering stage

第二主成分值较大时叶片有向右扭转的趋势,反之则向左扭转(图6A)。部分品种的叶片呈现出较大的扭转趋势,并且左偏、右偏均有出现,因此其第二主成分绝对值较大;而其他品种叶片则很少出现扭转,因此第二主成分绝对值较小。图6B中‘风林一号’的第二主成分特征值品种内标准差较

小,烟草叶形较为对称;而‘新K326’的第二主成分特征值品种内变异很大,从形状上可以观察到该品种叶片出现了明显的扭转。

由图7A可知,第三主成分主要体现叶柄部分的差异性。第三主成分的值越大,叶柄越突出;其值越小,叶柄越不明显。由图7B可知,对比‘珊西烟’和

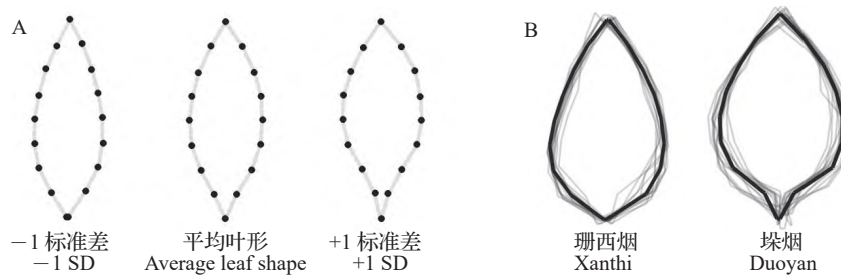


A. 可视化第二主成分结果;B. 典型品种叶片平均形状。

A. Visualization of the second principal component results; B. Average leaf shapes of typical varieties.

图6 开花期第二主成分结果对比

Fig. 6 Comparison of the results of the second principal component at the flowering stage



A. 可视化第三主成分结果;B. 典型品种叶片平均形状。

A. Visualization of the third principal component results; B. Average leaf shapes of typical varieties.

图7 开花期第三主成分结果对比

Fig. 7 Comparison of the results of the third principal component at the flowering stage

‘垛烟’2个烟草品种,尽管两者叶形相似,但叶柄的差异导致两者叶片形态并不相同。相较于开花期的叶片,团棵期的烟草叶片第一、第二主成分占比很大,而第三主成分占比有所下降。表明此时不同品种间的差异主要体现在叶片宽和叶形扭转上,叶柄处差异的权重有所降低。

2.3 基于机器学习的烟草叶类判别结果

为了验证由标志点法提取得到的叶形数据和常用叶形指标相比是否能对烟草叶形进行更加精确的描述,分别采用决策树、随机森林、支持向量机对不同生长时期的全部烟草品种进行叶类判别。以开花期的判别结果(表1)为例。根据全部叶片形状数据的判别准确度可知,采用标志点数据进行判别的精度范围为52%~62%,高于常用叶形指标判别的精度范围(51%~54%)。两者的差异主要在于常用叶形指标无法对叶片的最大宽位置进行区分,而标志点法可以捕捉到这种差异。

在3种机器学习算法中,支持向量机的判别精度约为65%,略优于随机森林,两者均优于决策树方法。各方法对上部叶和中部叶的判别精度比下

部叶高出10%左右,甚至优于采用全部叶片的结果,表明不同叶类间下部叶的差异较小,而中、上部叶展示了更明显的叶形变化,可以捕捉到更多的品种特征。相较于开花期的叶类判别结果,团棵期烟草叶类判别结果也都表现出了以上特征(数据未列出)。但整体上团棵期叶片的判别精度较低,标志点数据的判别精度范围仅为49%~55%。

由于烟草叶形有典型类和非典型类之分,一些具有非典型叶形的烟草品种仅靠肉眼很难分辨出具体叶形类别,因此人工叶形分类标准具有模糊性,采用该标准作为真值会导致判别精度下降。去除12类非典型烟草品种,对剩余的典型品种进行基于标志点数据的机器学习判别,使用上部叶、中部叶、下部叶分别进行判别,最高精度分别为77%、74%、71%,判别效果得到明显改善。

3 讨论

3.1 标志点数据与常用叶形指标的对比

在标志点法的研究中,如何选点极为重要。研

表1 基于机器学习的开花期烟草叶形判别精度

Table 1 Discriminant accuracies of tobacco leaf shapes based on machine learning at the flowering stage

叶位 Leaf position	叶形指标 Leaf shape indicator	机器学习算法 Machine learning method			不同指标均值 Means of different indicators
		决策树 DT	随机森林 RF	支持向量机 SVM	
全部叶片 All leaves	常用指标 Commonly used indicator	0.52	0.51	0.54	0.52
	标志点 Landmark	0.52	0.60	0.62	0.58
	典型品种标志点 Typical variety landmark	0.64	0.69	0.73	0.69
上部叶 Superior leaves	常用指标 Commonly used indicator	0.58	0.55	0.59	0.57
	标志点 Landmark	0.60	0.65	0.67	0.64
	典型品种标志点 Typical variety landmark	0.69	0.75	0.77	0.74
中部叶 Medial leaves	常用指标 Commonly used indicator	0.60	0.61	0.61	0.61
	标志点 Landmark	0.61	0.65	0.64	0.63
	典型品种标志点 Typical variety landmark	0.69	0.74	0.74	0.72
下部叶 Inferior leaves	常用指标 Commonly used indicator	0.52	0.54	0.54	0.53
	标志点 Landmark	0.46	0.54	0.58	0.53
	典型品种标志点 Typical variety landmark	0.67	0.68	0.71	0.68

究人员大多借助MorphoJ^[23]、ImageJ^[24]等软件手动选择标志点。该方法主要适合一些棱角特征明显的叶形,对于烟草这种近似椭圆形的叶片,很难从中选择出具有代表性的合适点位,并且采用手动标点也存在着标点误差的问题^[9]。有研究者为解决难以提取出具有同源性标志点的问题,提出了半标志点法^[25],即等间距地在样本边缘获取标志点。但是,该方法是一种单纯基于数学的形态模拟,缺少生物学支撑。本研究以烟草叶片形态特征为基础,自动获取标志点,不仅实现了不同叶片间采样标准的统一,还具有方便、快捷、精准等优势,为椭圆形叶片的标志点选取提供了新方法。此外,选点数量的不同也会对结果的精度造成影响。增加标志点的数量可以更细致地描述出存在于叶柄、叶耳等部位的表型特征,对叶片形状进行更为精确的刻画。

方差分析结果表明烟草叶形间存在的细微差异可以通过标志点数据捕捉。相对于各种常用的叶形指标,标志点数据对叶形的描述更为全面。基于标志点数据的主成分分析结果表明,不同品种的烟草叶片之间存在着很多差异,最主要为叶片宽和叶片最大宽位置、叶片扭转程度以及叶柄部的差

异。叶片的宽度对应叶片长宽比这一指标,是最常用、最基本的叶片表型信息^[26]。但是有研究表明,尽管叶片的长宽比是叶片最主要的差异来源,但只采用长宽比对叶片进行描述会丢失约40%的叶片轮廓信息^[27]。机器学习判别结果也表明,目前常用的叶形指标对烟草叶形的鉴别能力较标志点数据更低。而叶片最大宽位置以及叶片扭转程度、叶柄差异等叶形指标需要人工测量,费时费力。采用标志点法可以自动提取这些信息,不仅可以替代人眼识别,还可以将其量化为更精确的描述。

3.2 烟草叶形特征分析

由标志点的方差分析结果可知,烟草叶片的叶形在不同品种和不同叶位间存在差异和互作效应,表明对于不同叶位的叶片,其品种差异的表现强度不同,进一步印证了本研究中对烟草叶片进行判别分析时不同叶位叶片的分类效力有所区别。由机器学习的结果可知,本研究中,无论是对于数量大、特征明显的第一朵中心花开放时期数据,还是数量少、特征较模糊的团棵期数据,支持向量机方法的判别效果均最佳,该结果与前人研究^[5]相似。而决策树法效果最差,原因是决策树法在每个节点上都

会产生分枝,在离异对象数量较多的情况下判别准确率较低。采用上部叶和中部叶进行建模预测的精度更高,原因是烟草下部叶中普遍存在叶片较宽的现象,从而掩盖了其他叶片表型特征^[26]。对比团棵期和开花期的烟草叶片数据,发现团棵期数据的判别精度普遍较差,这是由于团棵期叶片正处于旺盛的营养生长期,其叶形特征还未显露完全。因此在对烟草叶片进行特征提取时,应尽量采用处于现蕾期后上、中部的叶片。在后续研究中,我们会加入其他生长时期以及不同年份的烟草叶形数据,对烟草叶片发育规律进行更加深入的探讨。

由以上分析可知,对烟草叶片进行标志点的测量可以捕捉到叶片存在的大部分轮廓信息,但就机器学习的结果来说,判别分析的精度并不高,原因是本研究采用的人为分类标准对非典型品种很容易产生模糊的分类结果,从而影响机器学习结果的准确度。去除一些非典型品种后,机器学习判别精度得到了很大提升。由此可见,发展更精确、更全面的叶形描述方法十分必要。标志点法不仅可以捕捉到人眼可以看到的差异,还可以将差异转为数值特征进行精确输出,结合叶片的纹理信息^[3]、三维株型信息^[28-29]等其他表型指标,将可以实现对作物品种的精确定位,从而成为未来育种和表型分析的新指标。

4 结论

本研究将几何形态学的方法应用到作物表型研究领域,采用标志点法对处于团棵期和开花期(第一朵中心花开放的时期)的39个烟草品种进行叶片形状信息提取。结论如下:

1)不同叶形类别和不同叶位的烟草叶片形状信息存在显著性差异,且两者间存在互作效应。对所有烟草叶片标志点数据进行主成分分析,前3个主成分分别解释了叶片宽和最大宽位置、叶片扭转程度以及叶柄处的不同所造成的叶片差异。其中叶片宽和最大宽位置是造成不同烟草叶形差异的最主要因素。

2)分别采用决策树、随机森林、支持向量机3类机器学习算法对烟草叶类进行判别。采用开花期全部叶片标志点数据进行判别的精度范围为52%~62%,高于常用叶形指标的判别精度(51%~54%)。相较于其他2种机器学习判别方法,支持向量机的

判别效果最优,精度约为65%。采用上部叶、中部叶的判别精度要高于下部叶10%左右。开花期烟草叶片的判别精度明显优于团棵期叶片,能展示出更明显的品种表型特征。在去除12种非典型烟草品种后,标志点数据判别精度上升至77%。本研究结果可为自动化提取叶片形状信息提供新思路。

参考文献(References):

- [1] 张宁,刘文萍.基于图像分析的植物叶片识别技术综述[J]. 计算机应用研究,2011,28(11):4001-4007. DOI:10.3969/j.issn.1001-3695.2011.11.001
ZHANG N, LIU W P. Plant leaf recognition technology based on image analysis[J]. *Application Research of Computers*, 2011, 28(11): 4001-4007. (in Chinese with English abstract)
- [2] 王晓峰,黄德双,杜吉祥,等.叶片图像特征提取与识别技术的研究[J]. 计算机工程与应用,2006(3):190-193.
WANG X F, HUANG D S, DU J X, et al. Feature extraction and recognition for leaf images[J]. *Computer Engineering and Applications*, 2006(3): 190-193. (in Chinese with English abstract)
- [3] WÄLDCHEN J, MÄDER P. Plant species identification using computer vision techniques: a systematic literature review[J]. *Archives of Computational Methods in Engineering*, 2018, 25(2): 507-543. DOI:10.1007/s11831-016-9206-z
- [4] LEE S H, CHAN C S, MAYO S J, et al. How deep learning extracts and learns leaf features for plant classification[J]. *Pattern Recognition*, 2017, 71: 1-13. DOI:10.1016/j.patcog.2017.05.015
- [5] LARESE M G, NAMÍAS R, CRAVIOTTO R M, et al. Automatic classification of legumes using leaf vein image features[J]. *Pattern Recognition*, 2014, 47(1): 158-168. DOI: 10.1016/j.patcog.2013.06.012
- [6] MUNISAMI T, RAMSURN M, KISHNAH S, et al. Plant leaf recognition using shape features and colour histogram with *k*-nearest neighbour classifiers[J]. *Procedia Computer Science*, 2015, 58: 740-747. DOI:10.1016/j.procs.2015.08.095
- [7] 郑一力,钟刚亮,王强,等.基于多特征降维的植物叶片识别方法[J]. 农业机械学报,2017,48(3):30-37. DOI:10.6041/j.issn.1000-1298.2017.03.004
ZHENG Y L, ZHONG G L, WANG Q, et al. Method of leaf identification based on multi-feature dimension reduction[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(3): 30-37. (in Chinese with English abstract)
- [8] 白明,杨星科,李静,等.几何形态学:关于形态定量比较的科学计算工具[J]. 科学通报,2014,59(10):887-894. DOI:10.1360/972012-1561
BAI M, YANG X K, LI J, et al. Geometric morphometrics, a super scientific computing tool in morphology comparison [J]. *Chinese Science Bulletin*, 2014, 59(10): 887-894. (in Chinese with English abstract)

- [9] 刘媛,吕佳,宋佳,等.基于几何形态测量学的植物鉴定方法[J].植物科学学报,2017,35(6):894-899. DOI:10.11913/PSJ.2095-0837.2017.60894
LIU Y, LÜ J, SONG J, et al. Plant species delimitation method based on geometric morphometrics[J]. *Plant Science Journal*, 2017, 35(6): 894-899. (in Chinese with English abstract)
- [10] CHITWOOD D H, OTONI W C. Morphometric analysis of *Passiflora* leaves: the relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade[J]. *GigaScience*, 2017, 6(1): giw008. DOI:10.1093/gigascience/giw008
- [11] FELDMANN M J, HARDIGAN M A, FAMULA R A, et al. Multi-dimensional machine learning approaches for fruit shape phenotyping in strawberry[J]. *GigaScience*, 2020, 9(5): giaa030. DOI:10.1093/gigascience/giaa030
- [12] LIU Y, LI Y J, SONG J L, et al. Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae)[J]. *Annals of Forest Science*, 2018, 75(4): 90. DOI:10.1007/s13595-018-0770-2
- [13] 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.
- [14] OTSU N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1): 62-66.
- [15] FIRMANSYAH Z, HERDIYENI Y, SILALAH B P, et al. Landmark analysis of leaf shape using polygonal approximation [C]. *IOP Conference Series: Earth and Environmental Science*, 2016, 31(1): 012018. DOI:10.1088/1755-1315/31/1/012018
- [16] ROHLF F J, SLICE D. Extensions of the Procrustes method for the optimal superimposition of landmarks[J]. *Systematic Zoology*, 1990, 39(1): 40-59.
- [17] VISCOSI V, CARDINI A. Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners [J]. *PLoS ONE*, 2011, 6(10): e25630. DOI: 10.1371/journal.pone.0025630
- [18] DRYDEN I L. *Shapes: Statistical Shape Analysis, with Application in R Package Version 1.2.5*[CP]. Chichester, UK: Wiley, 2019. <https://CRAN.R-project.org/package=shapes>
- [19] 柴宏红,邵科,于超,等.基于三维点云的甜菜根表型参数提取与根型判别[J].农业工程学报,2020,36(10):181-188. DOI:10.11975/j.issn.1002-6819.2020.10.022
CHAI H H, SHAO K, YU C, et al. Extraction of phenotypic parameters and discrimination of beet root types based on 3D point cloud[J]. *Transactions of the CSAE*, 2020, 36(10): 181-188. (in Chinese with English abstract)
- [20] BREIMAN L, FRIEDMAN J H, OLSHEN R A, et al. *Classification and Regression Trees*[M]. New York, USA: Routledge, 2017. DOI:10.1201/9781315139470
- [21] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [22] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [23] KLINGENBERG C P. MorphoJ: an integrated software package for geometric morphometrics[J]. *Molecular Ecology Resources*, 2011, 11(2): 353-357. DOI:10.1111/j.1755-0998.2010.02924.x
- [24] MALOOF J N, NOZUE K, MUMBACH M R, et al. LeafJ: an ImageJ plugin for semi-automated leaf shape measurement [J]. *Journal of Visualized Experiments*, 2013(71): 50028. DOI: 10.3791/50028
- [25] ADAMS D C, ROHLF F J, SLICE D E. A field comes of age: geometric morphometrics in the 21st century[J]. *Hystrix, the Italian Journal of Mammalogy*, 2013, 24(1): 7-14. DOI: 10.4404/hystrix-24.1-6283
- [26] 招启柏,顾世梁,俞惠梅,等.烟草叶片生长的动态分析[J].中国烟草学报,2004,10(2):12-20.
ZHAO Q B, GU S L, YU H M, et al. A dynamic analysis of tobacco leaf growth[J]. *Acta Tabacaria Sinica*, 2004, 10(2): 12-20. (in Chinese with English abstract)
- [27] MIGICOVSKY Z, LI M, CHITWOOD D H, et al. Morphometrics reveals complex and heritable apple leaf shapes[J]. *Frontiers in Plant Science*, 2018, 8: 2185. DOI: 10.3389/fpls.2017.02185
- [28] XIAO S F, CHAI H H, SHAO K, et al. Image-based dynamic quantification of aboveground structure of sugar beet in field[J]. *Remote Sensing*, 2020, 12(2): 269. DOI: 10.3390/rs12020269
- [29] HUI F, ZHU J Y, HU P C, et al. Image-based dynamic quantification and high-accuracy 3D evaluation of canopy structure of plant populations[J]. *Annals of Botany*, 2018, 121(5): 1079-1088. DOI:10.1093/aob/mcy016