

## 基于光谱的春季育苗移栽期番茄苗健壮度量化检测

姬江涛<sup>1,2</sup>, 李鹏阁<sup>1</sup>, 金鑫<sup>1,2\*</sup>, 马湜<sup>1,2</sup>, 李明勇<sup>1</sup>

1. 河南科技大学农业装备工程学院, 河南 洛阳 471003

2. 机械装备先进制造河南省协同创新中心, 河南 洛阳 471003

**摘要** 为了筛选影响春季育苗移栽期番茄穴盘苗健壮程度的关键指标, 并实现其快速无损检测, 测定了5项秧苗指标, 经向量归一化预处理并采用独立性权系数法确定各指标权重, 并根据权重结果挑选出包含信息较全面, 影响较大的两个指标: 叶绿素和干质量。两项指标所组成的简化秧苗评价价值可以近似表示综合评价价值, 相关系数  $r$  为 0.92, 大大减少了品质检测所需的指标量, 并可以很好的表征春季育苗移栽期番茄苗的健壮度。提取了各穴盘苗的可见-近红外光谱数据, 经去噪和多元散射校正(MSC)预处理, 消除了由光散射等带来的光谱干扰信息, 相较原始光谱信息更具可利用性。随后采用光谱-理化值共生距离(SPXY)算法对本集进行划分, 利用波段值和评价价值两种变量同时计算样本间距离, 以最大化表征样本分布, 提高样本差异性和代表性。采用竞争性自适应重加权算法(CARS)和无信息变量消除连续投影算法(UVE-SPA)优选光谱特征波数, 降低光谱数据维度, 得到了更能体现光谱特征的简化光谱信息, 减少了冗余信息对建立模型精准度和分析速度的影响。最后应用偏最小二乘支持向量机(LS-SVM)和基于 U-Net 模型改造的卷积神经网络(CNN), 以预处理后的光谱数据和提取特征波长后的光谱数据分别作为模型的输入, 建立了光谱数据与综合评价价值的非线性映射模型, 并进行对比选优。结果显示: 应用 UVE-SPA 预处理方法筛选出的波段, 光谱信息更加丰富有效; 两种预处理后的优选波段所建模型回归效果整体优于全波段建立的模型; CNN 模型的建模效果整体优于 LS-SVM 模型, 且 UVE-SPA-CNN 模型对光谱数据和秧苗评价价值的回归分析效果最好, 其建模集和预测集的相关系数  $r$  分别为 0.988 和 0.946, 均方根误差 RMSE 分别为 0.085 和 0.025, 为直接利用光谱数据获取融合了多种因素的番茄秧苗评价价值, 从而判别秧苗健壮度提供了理论依据。

**关键词** 可见-近红外; 番茄穴盘苗; 评价价值; 卷积神经网络

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)06-1741-08

### 引言

穴盘苗的培育和移栽是蔬菜生产的重要环节, 也是农业增收、确保农产品供给安全的精耕细作重要农艺措施。目前国内工厂化育苗的形式正在逐渐推广, 温室大棚在进行番茄穴盘苗培育过程中, 考虑到不同季节条件下的大田移栽环境, 适宜的番茄秧苗培育苗龄也会发生改变, 其中春季移栽的适宜苗龄为 55~60 d。为了提高培育效率, 需先将种子播种在较高密度的穴盘内, 经过一段时间生长后, 幼苗间会出现互相挤占生长空间的情况, 导致其生长缓慢和发育失衡, 所以需要在它们生长至适当的叶面积指数时, 将其移栽到较低密度的穴盘中<sup>[1]</sup>。在该育苗期移栽的过程中, 还需对其健

壮程度进行评判, 筛选剔除不合格的幼苗和补入优苗, 以保证最终移栽期的幼苗质量。因此, 需要建立育苗移栽期幼苗的健壮度评价指标和实现其快速检测。

传统的壮苗指标包括株高、茎粗、鲜重、干重等单项指标及其组合而成的(茎粗/株高+根重/冠重)×苗干重、茎粗/株高×苗干重/茎的复合指标, 应用广泛但主要以形态指标为基础且穴盘育苗密度大、营养面积小和根际环境变化大, 这同传统育苗方式相比有很大差异, 使得评价结果难以体现实际情况<sup>[2]</sup>。叶绿素是叶片进行光合作用, 积累有机物的重要化学色素指标, 作为主要的光合作用场所, 叶片的叶绿素含量与幼苗的质量评价关系密切<sup>[3-4]</sup>。因此结合形态指标和色素指标, 对育苗移栽期的番茄苗进行健壮度评价的结果更加全面和准确。

收稿日期: 2021-06-04, 修订日期: 2021-07-30

基金项目: 国家自然科学基金项目(51975186, 51875175)资助

作者简介: 姬江涛, 1965年生, 河南科技大学农业装备工程学院教授 e-mail: jjt0907@163.com

\* 通讯作者 e-mail: jx.771@163.com

检测体现健壮特征各项指标的手段主要包括机器视觉和光谱检测, 相关的研究较为广泛。杨斯等<sup>[5]</sup>利用机器视觉识别快速分析和判断苗高信息, 所得苗高结果的平均测量误差为 2.30 mm, 平均测量相对误差为 7.69%; 丁永军等<sup>[6]</sup>利用线性四点内插法估测番茄叶片叶绿素含量, 根据 4 个波段 (670, 700, 740 和 780 nm) 的叶片反射率计算红边位置, 所建模型效果良好, 其相关系数  $r$  为 0.875; David 等<sup>[7]</sup>在 395~885 nm 波长范围内对冷胁迫幼苗进行扫描, 并用支持向量机 (SVM) 建立了分类模型, 结果显示, 以 590 和 873 nm 为特征波长的分类效果更佳。以上研究表明可见-近红外光谱技术在番茄幼苗元素含量检测和环境胁迫方面具有重要意义, 但针对春季育苗移栽期的番茄穴盘苗, 能反映多信息融合的健壮度指标的光谱检测尚未见报道。

以春季育苗移栽期番茄苗为研究对象, 测定了多个单项指标, 利用数学分析的方法筛选出了代表性强, 评价效果较好的两项指标, 并建立了综合评价指数; 对样品进行光谱检测, 分别用全波段和特征波段光谱变量与秧苗评价价值建立了定量分析模型, 旨在为春季育苗移栽期穴盘苗健壮度的快速检测提供科学依据。

## 1 实验部分

### 1.1 材料

番茄穴盘育苗于 2020 年 12 月—2021 年 2 月在河南洛阳城研辣椒研究所温室大棚环境中培养, 白天温度控制在 20~25 °C, 夜间温度控制在 10~15 °C。选用的番茄品种为毛粉 802, 种子经消毒处理后, 采用一孔一种的播种方法, 将其播种于 4 个 72 孔穴盘中, 穴盘育苗基质为包含泥炭、椰糠和珍珠岩的混合基质 (比例为 3:1:1)。番茄秧苗日历苗龄为 45 d 时, 处于育苗移栽期, 此时穴盘中有 276 株幼苗 (缺苗 12 株), 对其进行各项指标的测定和光谱检测。对番茄叶片光谱反射率的采样使用海洋光学 USB4000 光纤光谱仪, 其波长范围为 200~1100 nm, 光谱间隔为 0.21 nm, 试验选择在晴朗无云或少云的天气进行, 用 OceanView 软件对光谱数据进行采集。

### 1.2 方法

按照以下顺序采集所需数据: 将番茄苗取出洗净并用滤纸擦干, 用直尺测定株高; 用游标卡尺测定茎粗; 采用万分之一电子天平称量全株鲜质量; 将光谱仪预热 30 分钟后采集光谱数据; 每个活体样本由多个叶片组成; 光谱仪探头垂直向下, 距叶片 2 cm 左右; 每次采集目标光谱前后进行参考板校正, 每点反复测量 3 次取平均值作为该样本光谱测量结果。用分光光度法测定叶绿素 a、叶绿素 b 和总叶绿素含量; 用叶绿素仪测定叶片叶绿素相对含量 SPAD; 最后将整株装于信封袋内置于鼓风干燥箱 105 °C 杀青 30 min, 80 °C 下烘至恒质量, 并称量其全株干质量。

### 1.3 数据处理

试验数据采用 SPSS 26.0 进行分析处理, 光谱预处理和特征变量的选择是基于 MATLAB 2019b 实现的, 评价价值的定量分析模型基于 Python 语言的 TensorFlow 框架实现, 绘

图工作在 Origin 2019 上完成。

## 2 影响秧苗健壮度的重要指标和评价方法

### 2.1 评价指标的分类和预处理

试验采集的各评价指标均为定量指标, 其中株高为区间型指标, 即它在某一范围内时, 秧苗质量较佳, 偏离这一范围不论是过长或是偏少都不利于秧苗质量。其他指标均为极大型指标, 即这些指标没有明显的上下限, 随着指标的增长与秧苗质量成正相关。

为了构建评价指标体系, 便于后续分析, 需要对各指标进行一致化和归一化处理。其中, 先将区间型指标转化为极大型指标 [式(1)], 然后将其他极大型指标进行向量归一化处理 [式(2)], 得到每个秧苗各单项指标的评价价值。

$$\hat{x}_i = \begin{cases} 1 - \frac{a - x_i}{M}, & x_i < a \\ 1, & a \leq x_i \leq b \\ 1 - \frac{x_i - b}{M}, & x_i > b \end{cases}$$

$$M = \max\{a - \min\{x_i\}, \max\{x_i\} - b\} \quad (1)$$

$$x_i^* = \frac{\hat{x}_i}{\sqrt{\sum_{i=1}^n \hat{x}_i^2}} \quad (2)$$

式中,  $a$  为株高指标最佳区间的下限,  $b$  为株高指标最佳区间的上限,  $x_i$  为转化前的指标值,  $\hat{x}_i$  为转化后的指标值,  $x_i^*$  为最终无量纲化后的评价价值。

### 2.2 单项指标权重系数的确定

采用独立性权重系数法, 根据各指标与其他指标之间的共线性强弱来确定指标权重。首先建立幼苗的某项指标  $x_i$  与其他指标  $x_1, x_2, \dots, x_n$  的线性关系方程, 其中  $b_0, \dots, b_n$  为常数,  $\hat{x}_i$  为因变量。

$$\hat{x}_i = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \quad (3)$$

指标  $x_i$  与其他指标的复相关系数  $R_i$  越大, 则说明它们之间的共线性关系越强, 重复信息越多, 则该指标的权重也就应该越小。

$$R_i = \frac{\sum (x_i - \bar{x}_i)(\hat{x}_i - \bar{\hat{x}}_i)}{\sqrt{\sum (x_i - \bar{x}_i)^2 \sum (\hat{x}_i - \bar{\hat{x}}_i)^2}} \quad (i \leq 5) \quad (4)$$

每个指标的复相关系数的倒数  $\frac{1}{R}$  作为得分  $S$ , 再经归一化处理得到权重系数  $S^*$ 。所得结果如表 1 所示。

$$S^* = \frac{S}{\sqrt{\sum_{i=1}^n S^2}} \quad (5)$$

由表 1 可知, 干质量和总叶绿素的权重较大, 即这两个指标所包含的信息量较多, 可以结合起来反映春季育苗移栽期的番茄苗健壮程度。其他指标的权重指数较为接近且均明显低于前两项, 可能是因为在育苗期间温室大棚的水分、温湿度、光照、养分等统一苗期管理中, 由经验对传统的壮苗指标予以干预, 如为防止徒长施加矮壮素等, 使得它们差异较小, 对健壮苗的评判结果贡献度低。

表 1 春季育苗移栽期番茄苗单项指标权重系数

Table 1 Weight coefficient of individual index of tomato seedlings in spring transplanting period

| 指标   | 权重    |
|------|-------|
| 株高   | 0.135 |
| 茎粗   | 0.128 |
| 鲜质量  | 0.117 |
| 干质量  | 0.336 |
| 总叶绿素 | 0.284 |

2.3 番茄苗健壮度的评价方法

根据上述各项指标评价值和权重的确定，每株番茄健壮度的评价值  $V$  可表示如式(6)

$$V = 0.135x_1^* + 0.128x_2^* + 0.117x_3^* + 0.336x_4^* + 0.284x_5^* \quad (6)$$

式(6)中,  $x_1^*$ ,  $x_2^*$ ,  $x_3^*$ ,  $x_4^*$  和  $x_5^*$  分别为各幼苗的株高、茎粗、鲜质量、干质量、总叶绿素的评价值。将所有被测秧苗的评价值  $V$  汇总后, 得到其范围在 0.12~0.89 之间, 如图 1 所示。这表明试验苗样范围广, 包含各种优劣苗, 能够作为评判春季育苗移栽期番茄苗健壮度的依据。

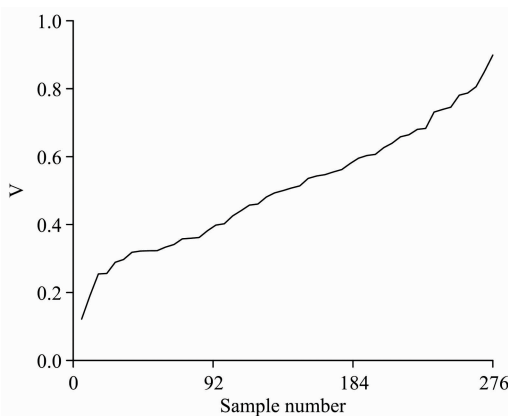


图 1 春季育苗移栽期番茄苗健壮度评价值

Fig. 1 The robustness evaluation value of tomato seedlings in the spring seedling transplanting period

作为对秧苗评价值  $V$  影响最大的两个因素: 干质量和总叶绿素, 若仅用这两项指标来反映春季育苗移栽期番茄苗健壮度, 则该简化后的秧苗评价值  $V_s$  可表示为式(7)

$$V_s = 0.542x_4^* + 0.458x_5^* \quad (7)$$

将简化后的秧苗评价值  $V_s$  与评价值  $V$  进行对比分析, 如图 2 所示。可以看出两者线性度高, 决定系数  $R^2 = 0.920$ , 即简化后的秧苗评价值  $V_s$  可以很好的表征春季育苗移栽期番茄苗的健壮度。

2.4 番茄苗的品质分类

将样本简化后的秧苗评价值  $V_s$  作为分类依据, 将番茄苗分为三种类型: 优质苗、合格苗, 劣质苗, 对其进行聚类划分, 同时, 应考虑育苗移栽的实际情况, 对其进行优化, 避免因标准过高而导致的资源浪费。分类结果如表 2 所示, 分类样本情况如图 3 所示, 其中图 3(a)为所培育 72 穴

穴盘苗的整体情况, 图 3(b), (c)和(d)分别为基于该分类依据所筛选出的优质苗, 合格苗和劣质苗。基于该依据的分类结果相对于传统经验, 对健壮程度的划分更加精细。

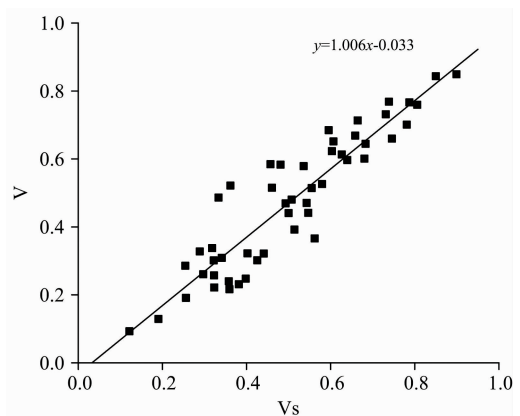


图 2 番茄简化评价值在评价值范围内的分布

Fig. 2 The distribution of tomato simplified evaluation value within the evaluation value range

表 2 番茄苗的品质分类结果

Table 2 Quality classification results of tomato seedlings

| 类别  | 聚类样本数 | 聚类评价值 $V_s$        | 最终样本数 | 最终评价值 $V_s$        | 最终占比/% |
|-----|-------|--------------------|-------|--------------------|--------|
| 优质苗 | 77    | $\geq 0.623$       | 88    | $\geq 0.600$       | 32     |
| 合格苗 | 94    | $0.366 \sim 0.623$ | 127   | $0.300 \sim 0.600$ | 46     |
| 劣质苗 | 105   | $\leq 0.366$       | 61    | $\leq 0.300$       | 22     |



图 3 春季育苗移栽期番茄苗分类情况

(a): 番茄穴盘苗整体情况; (b): 优质苗; (c): 合格苗; (d): 劣质苗

Fig. 3 Classification of tomato seedlings in spring seedling transplanting period

(a): Overall condition of tomato plug seedlings; (b): High quality seedlings; (c): Qualified seedlings; (d): Poor quality seedlings

### 3 番茄苗健壮度的光谱检测

#### 3.1 不同品质穴盘苗的可见-近红外光谱图

在构建了春季育苗移栽期番茄穴盘苗的评价体系的基础上,为了获取其健壮度信息,可以发挥光谱的无损检测优势,用化学计量学的方法构建光谱数据与反映秧苗品质的两个重要指标及秧苗评价价值  $V_s$  的定量分析模型,以实现质量鉴别。样本的原始光谱曲线如图 4(a)所示,是典型的绿色植物光谱曲线,可以清晰的看出图像两端有较大噪声。将光谱信号两端噪声较大的波段截掉后,得到的光谱曲线如图 4(b)所示。为了消除由于颗粒分布不均匀及颗粒大小不同产生的散射对光谱的影响,增强有用的光谱信息,采用多元散射校正(MSC)的预处理方法,得到的光谱曲线如图 4(c)所示。

#### 3.2 光谱特征的重要波长选择

利用竞争性自适应重加权算法(CARS)和无信息变量消除-连续投影算法(UVE-SPA)对番茄苗的光谱特征进行重要波长的选择,可以减少冗余的光谱特征信息,便于构建高效的健壮苗检测模型,并提高后期数据分析的计算速率。

CARS是通过选择模型中回归系数绝对值大的波长,去掉权重小的波长,再经内部交叉验证选出 RMSECV 值最低的子集,可有效的找到最优波长组合<sup>[8]</sup>。经过多次试验的结果如图 5 所示,得到最佳迭代次数为 25 次,并据此筛选出了 58 个光谱重要波长,作为简化后番茄苗评价的判别依据。

采用 UVE 算法对全波长下的 2 604 个光谱变量进行选择,UVE 中产生的随机变量个数同样设置为 2 604 个,决定被选变量的阈值选择标准为随机变量稳定性最大值的 99%,阈值外的波长变量将被保留作为特征波长<sup>[9]</sup>。经过 UVE 的变量选择,得到了 240 个波长,可以看出经 UVE 选择得到的特征波长数量仍较为庞大,因此采用 SPA 对其选择后的波长变量做进一步选择。验证集经过多元线性回归分析(MLR),得到了最小的验证标准偏差(RMSEV),其对应的变量即为所需的光谱重要波长。最终选择结果为 104 个特征波长。

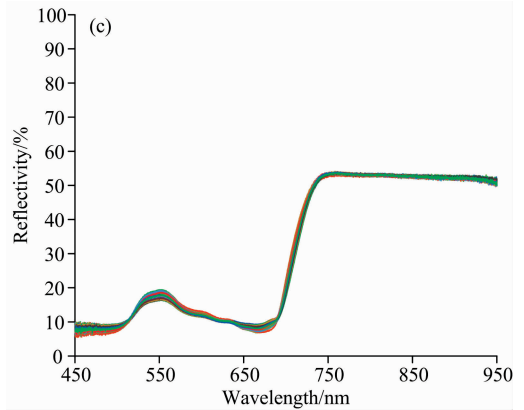
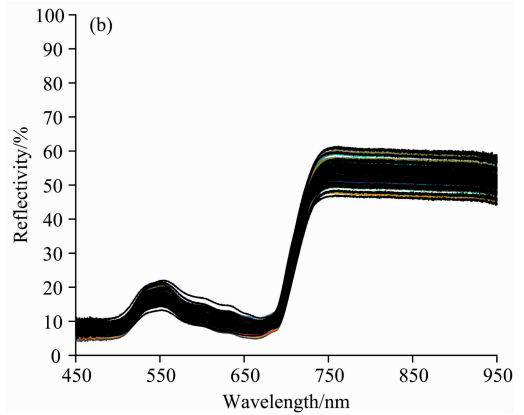
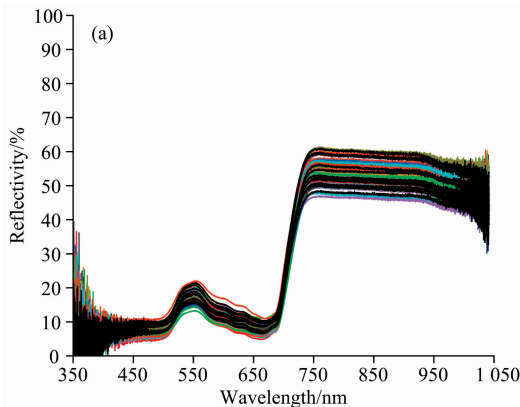


图 4 春季育苗移栽期番茄苗的光谱

(a): 原始光谱; (b): 去噪后光谱;

(c): MSC 预处理后光谱

Fig. 4 Spectra of tomato seedlings in the spring nursery and transplant period

(a): Raw spectral curve;

(b): Spectral curve after denoising;

(c): Spectral curve after MSC pretreatment

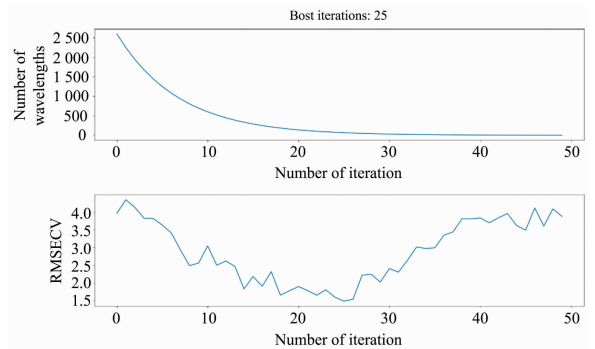


图 5 CARS 算法挑选特征波长的过程

Fig. 5 Key variables selection results by CARS

通过 CARS 和 UVE-SPA 两种方式选择出的特征波长在全谱中的分布情况如图 6 和图 7 所示。可以看出,两种方法得出的结果在 450~520, 620~720 和 750~950 nm 波段下分布较为类似,后者较前者多筛选出的特征波长大部分集中于 530~610 nm 波段。其中 450~520, 530~610 和 620~

720 nm 三个波段与组成秧苗评价值的叶绿素含量相关性很强，这与丁永军<sup>[10]</sup>等的研究结果吻合；750~950 nm 波段与秧苗评价值中干质量相关性较强，这与王洋等人研究结果相似<sup>[11]</sup>；干物质在可见光波段作用不明显，在近红外-短波红外区域影响较大，干物质含量越高，叶片反射能力越弱。

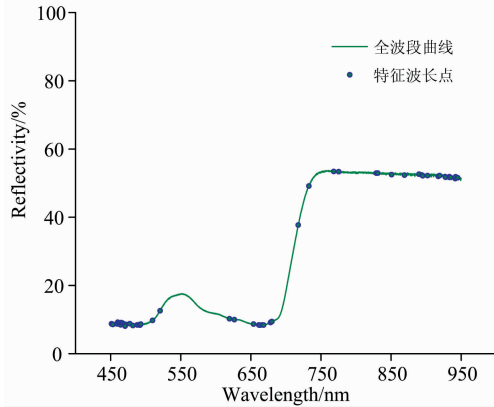


图 6 CARS 挑选的特征波长在全波段的分布

Fig. 6 The distribution of characteristic wavelengths selected by CARS in the whole band

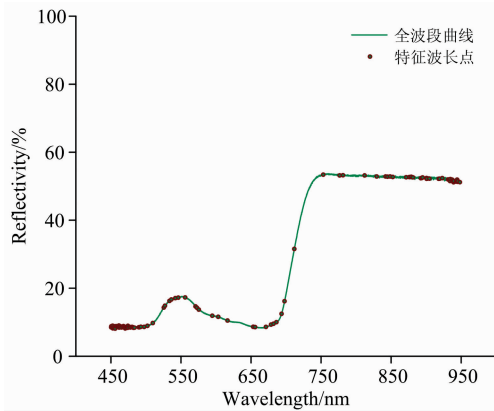


图 7 UVE-SPA 挑选的特征波长在全波段的分布

Fig. 7 The distribution of characteristic wavelengths selected by UVE-SPA in the whole band

### 3.3 样本集划分

采用 SPXY 算法进行样本集划分<sup>[12]</sup>，相较于传统的 KS 算法，它以光谱变量为  $x$  变量，秧苗评价值为  $y$  变量，利用两种变量同时计算样品间距离以保证最大程度表征样本分布，有效地覆盖多维向量空间，增加了样本间的差异性和代表性。最终以 2 : 1 的比例将样本集划分为建模集和预测集，其中建模集样本有 184 个，预测集样本有 92 个，保证了训练集中的样本按照空间距离分布均匀。样本间距离表示如式(8)

$$d_x(p, q) = \sqrt{\sum_{j=1}^N [x_p(j) - x_q(j)]^2}; p, q \in [1, N]$$

$$d_y(p, q) = \sqrt{(y_p - y_q)^2} = |y_p - y_q|; p, q \in [1, N]$$

$$d_{xy}(p, q) = \frac{d_x(p, q)}{d_x(p, q)_{\max}} + \frac{d_y(p, q)}{d_y(p, q)_{\max}}; p, q \in [1, N]$$

(8)

把所有的样本都看作建模集候选样本，依次从中挑选样本进入建模集。首先选择距离最远的两个样本进入建模集，其后通过计算剩下的每一个样品到建模集内每一个已知样品的距离，找到拥有最大最小距离的待选样本放入建模集，以此类推，直到达到所要求的样本数目。样本集划分的结果见表 3，可以看出建模集样本的秧苗评价值涵盖了预测集样本的秧苗评价值范围，且平均值和标准偏差均接近，表明样品集划分合理。

表 3 样本集划分

Table 3 Sample set division

| 统计参数 | 秧苗评价值 |       |
|------|-------|-------|
|      | 建模集   | 预测集   |
| 最小值  | 0.093 | 0.117 |
| 最大值  | 0.849 | 0.814 |
| 平均值  | 0.483 | 0.462 |
| 标准偏差 | 0.083 | 0.109 |

### 3.4 建模方法与评估

由于秧苗评价值在光谱曲线上的响应会受多种因素影响，而简单的线性回归模型很难处理非线性、随机性等复杂的问题，因此在模型的选择上应考虑非线性模型。选择 LS-SVM 和 CNN 两种非线性模型，并通过比较分析选取最优组合。

最小二乘支持向量机(least squares support vector machine, LS-SVM)是对 SVM 的改进，它能有效的处理回归问题，在解决小样本、非线性和高维模式识别等问题中优势显著<sup>[13]</sup>。其优势主要表现在将经典 SVM 中复杂的二次优化问题改为了线性方程组的问题，降低了复杂度，从而大大加快计算速度，在 LS-SVM 算法模型中，通过选择合适核参数  $\sigma$  和惩罚系数  $\gamma$ ，可以实现模型复杂度和模型准确度的平衡折中。本研究选择径向基函数(RBF)作为核函数，并用网格搜索和交叉验证法自优化模型的  $\sigma$  和  $\gamma$ 。

卷积神经网络(convolutional neural networks, CNN)是一类包含卷积计算且具有深度结构的前馈神经网络，其中 U-Net 网络模型因其结构优势应用广泛。它基于全卷积网络进行改进，不需要大量的样本和计算资源，就可以利用数据增强对一些比较少样本的数据进行训练<sup>[14-15]</sup>。基于 U-Net 网络进行改造，最终建立的模型结构如图 8 所示。

结构左侧部分为收缩路径，可视为有 4 个子模块的编码器，每个子模块有两个卷积核(核函数为 ReLU，大小为  $1 \times 3$ ，填充为 0，步长为 1)组成的卷积层，用于特征提取，后加一个最大池化层(大小为  $1 \times 2$ ，步长为 2)实现下采样，降低数据维度。除此之外，卷积过滤器的深度从 64 逐层增长到 1 024，使得每一次降采样都会增加一倍的通道数。右侧为扩展路径，与收缩路径组成类似，可视为实现上采样的解码器，以恢复数据维度。特别的是，各模块完成一次反卷积后，就会将收缩路径中对应步骤的特征图进行修剪，使其数据尺寸能与反卷积的结果拼接，以完成特征融合。为了适应回归需求，将 U-Net 原网络中最后用于输出的卷积层(大小为 1

×1) 替换成了两个全连接层, 使其最终的输出为秧苗评价价值。训练过程中, 设置学习率为  $10^{-3}$ , 训练次数为 1 000 次。

将预处理后的光谱数据和提取特征波长后的光谱数据分别作为模型的输入, 且所建立光谱模型优劣采用以下参数进行评估: 建模和预测集的相关系数( $r$ ), 建模均方根误差 (root mean square error for calibration, RMSEC) 以及预测均

方根误差 (root mean square error for prediction, RMSEP)。所建模型的相关系数越高, 均方根误差越低, 说明模型效果越好。同时, RMSEC 和 RMSEP 的绝对值的差越小, 表明所建立的模型的稳定性越好, 鲁棒性越强。所得结果如表 4 所示。

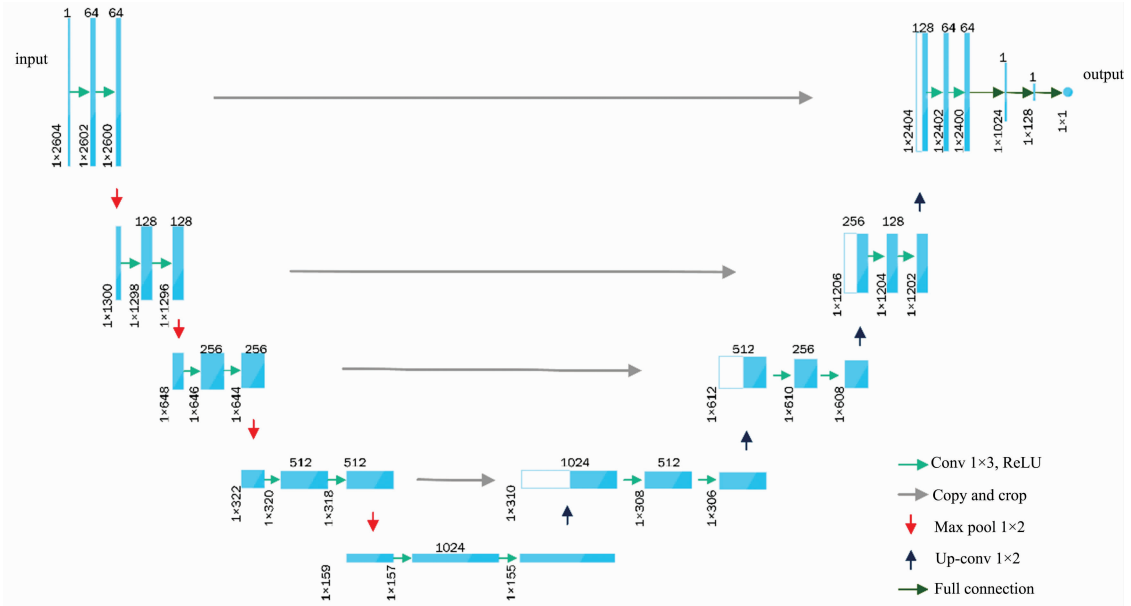


图 8 基于 U-Net 模型的卷积神经网络结构

Fig. 8 CNN structure based on U-Net model

表 4 不同光谱变量的各模型效果对比

Table 4 Comparison of model effects of different spectral variables

| 模型             | 变量数   | 建模集   |       | 预测集   |       |
|----------------|-------|-------|-------|-------|-------|
|                |       | $R_c$ | RMSEC | $R_p$ | RMSEP |
| LS-SVM         | 2 604 | 0.783 | 0.131 | 0.566 | 0.179 |
| CNN            | 2 604 | 0.905 | 0.086 | 0.873 | 0.098 |
| CARS-LS-SVM    | 58    | 0.900 | 0.141 | 0.723 | 0.091 |
| CARS-CNN       | 58    | 0.954 | 0.084 | 0.907 | 0.064 |
| UVE-SPA-LS-SVM | 104   | 0.937 | 0.130 | 0.770 | 0.086 |
| UVE-SPA-CNN    | 104   | 0.988 | 0.085 | 0.946 | 0.025 |

由表 4 可以看出, 基于全波段建立的 LS-SVM 和 CNN 模型, 建模集和预测集的结果差于其他经 CARS 和 UVE-SPA 预处理后波段所建模型, 可能是由于太多无用的波段信息对建模产生了干扰, 从而降低了最终回归的准确率。CNN 模型整体比 LS-SVM 模型的建模效果好, 说明训练的光谱信息数据量较大时, 神经网络的非线性回归效果更好。经 CARS 筛选出的特征波段数比 UVE-SPA 的波段数少 46 个, 但在 CNN 模型的回归结果中, 基于 UVE-SPA 预处理的建模集和预测集,  $r$ , RMSEC, RMSEP 的表现效果均好于 CARS 的预处理结果, 说明, CARS 相对于 UVE-SPA 在简化光谱变量输入的同时, 过滤了一部分重要波段信息, 导致模型精度下降。

上述分析可知, 对于采集的春季移栽期番茄穴盘苗可见光-近红外光谱信息, 采用 UVE-SPA-CNN 模型对光谱数据和秧苗评价价值的回归分析效果最好, 该模型既保留了原始光谱大量的有用信息, 又将输入变量由 2 604 个降为 104 个波长信息, 建模速度快且精度高。其中建模集和的相关系数  $r_c$  和建模均方根误差 RMSEC 分别为 0.988 和 0.085, 预测集的相关系数  $r_p$  和预测均方根误差 RMSEP 分别为 0.946 和 0.025。

### 4 结 论

以春季育苗移栽期番茄穴盘苗为研究对象, 采集了形态数据和光谱信息, 用叶绿素和干质量两种因素建立了能表征秧苗健壮度的定量模型—秧苗评价价值  $V_s$ , 并利用 CARS 和 UVE-SPA 方法对光谱数据进行降维, 选取特征波长, 在此基础上, 对比分析了 LS-SVM 和 CNN 两种不同非线性回归模型对秧苗评价价值的反演精度, 得出如下结论:

- (1) 简化后的秧苗评价价值  $V_s$  与评价价值  $V$  的相关系数  $r$  高达 0.920, 表明简化后的秧苗评价价值  $V_s$  可以很好的表征春季育苗移栽期番茄苗的健壮度。
- (2) CARS 相对于 UVE-SPA 能够提取更少的特征波数, 但同时也丧失了更多的有用波段, 导致后续回归效果变差。
- (3) CNN 模型的反演精度高于 LS-SVM, 且整体上 UVE-SPA-CNN 模型的效果最好, 其建模集和预测集的相关

系数  $r$  分别为 0.988 和 0.946, 均方根误差 RMSE 分别为 0.037 和 0.025。可为直接利用光谱数据获取融合了多种因素的番茄秧苗评价值, 从而判别其健壮度提供科学依据。

## References

- [ 1 ] HAN Lü-hua, MAO Han-ping, HU Jian-ping, et al(韩绿化, 毛罕平, 胡建平, 等). Transactions of The Chinese Society for Agricultural Machinery(农业机械学报), 2016, 47(11): 59.
- [ 2 ] CAO Ling-ling, TIAN Ya-nan, ZHAO Li-qun et al(曹玲玲, 田雅楠, 赵立群, 等). Agricultural Engineering Technology(农业工程技术), 2017, 37(19): 15.
- [ 3 ] GONG Bin-bin, WANG Ning, ZHANG Tie-jun, et al(宫彬彬, 王 宁, 章铁军, 等). Transactions of the Chinese Society for Agricultural Engineering(农业工程学报), 2019, 35(8): 237.
- [ 4 ] HE Shi-xing, HE Di, XU Chun-lin, et al(何诗行, 何 堤, 许春林, 等). Transactions of The Chinese Society for Agricultural Machinery(农业机械学报), 2017, 48(12): 319.
- [ 5 ] YANG Si, GAO Wan-lin, MI Jia-qi, et al(杨 斯, 高万林, 米家奇, 等). Transactions of The Chinese Society for Agricultural Machinery(农业机械学报), 2019, 50(S1): 128.
- [ 6 ] DING Yong-jun, ZHANG Jing-jing, LI Xiu-hua, et al(丁永军, 张晶晶, 李修华, 等). Transactions of The Chinese Society for Agricultural Machinery(农业机械学报), 2016, 47(3): 292.
- [ 7 ] Xie Chuanqi, Yang Ce, Moghimi Ali. Proc. SPIE, 2017, 10203: 1021305.
- [ 8 ] SHAO Yuan-yuan, WANG Yong-xian, XUAN Guan-tao, et al(邵园园, 王永贤, 玄冠涛, 等). Transactions of The Chinese Society for Agricultural Machinery(农业机械学报), 2020, 51(8): 344.
- [ 9 ] Song Xiangzhong, Huang Yue, Tian Kuangda, et al. Optik, 2020, 203: 164019.
- [10] ERSBAT Abulaiti, BAIDENGSHA Maimaitiaili, MAMAT Sawut(依尔夏提·阿不来提, 白灯莎·买买提艾力, 买买提·沙吾提, 等). Acta Optica Sinica(光学学报), 2019, 39(9): 372.
- [11] WANG Yang, XIAO Wen, ZOU Huan-cheng, et al(王 洋, 肖 文, 邹焕成, 等). Journal of Shenyang Agricultural University(沈阳农业大学学报), 2018, 49(1): 121.
- [12] Wei Xuan, He Jincheng, Zheng Shuhe, et al. Infrared Physics and Technology, 2020, 105: 103099.
- [13] GUAN Shan, PANG Hong-yang, SONG Wei-jie, et al(关 山, 庞弘阳, 宋伟杰, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2018, 34(14): 61.
- [14] XIE Zhong-hong, XU Huan-liang, HUANG Qiu-gui, et al(谢忠红, 徐焕良, 黄秋桂, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2019, 35(13): 277.
- [15] Zhang Bin, Miao Yanjie, Tian Yongzhi, et al. Journal of Nuclear Science and Technology, 2021, 58(7): 787.

# Study on Quantitative Detection of Tomato Seedling Robustness in Spring Seedling Transplanting Period Based on VIS-NIR Spectroscopy

JI Jiang-tao<sup>1,2</sup>, LI Peng-ge<sup>1</sup>, JIN Xin<sup>1,2\*</sup>, MA Hao<sup>1,2</sup>, LI Ming-yong<sup>1</sup>

1. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China

2. Henan Collaborative Innovation Center for Advanced Manufacturing of Mechanical Equipment, Luoyang 471003, China

**Abstract** To screen the key indicators that affect the robustness of tomato plug seedlings during the spring nursery and transplanting period and realize its rapid non-destructive testing, this paper measured 5 seedling indicators, then used vector normalization and the independence weight coefficient method to determine each indicator. According to the weighting results, two indicators containing more comprehensive information and greater influence are selected: chlorophyll and dry quality. The simplified seedling evaluation value composed of the two indicators can approximate the comprehensive evaluation value. The correlation coefficient is 0.92, which greatly reduces the number of indicators required for quality testing, and can well represent the robustness of tomato seedlings during the spring seedling transplanting period. At the same time, the visible-near infrared spectrum data of each plug seedling is extracted and pre-processed by denoising and multi-scattering correction (MSC). This way, it can eliminate the spectral interference information caused by light scattering and make it more usable than the original spectral information. Subsequently, the spectrum-physical and chemical value symbiosis distance (SPXY) algorithm is used to divide the sample set. The distance between the samples is calculated using two variables of the band value and the evaluation value to maximize the characterization of the sample distribution to improve the difference and representativeness of the samples. Secondly, the competitive adaptive weighting algorithm (CARS) and the uninformative variable elimination-successive projections algorithm (UVE-SPA) are used to optimize the spectral feature wave number, reduce the spectral data dimension and obtain simplified spectral information that can better reflect the spectral characteristics and reduce redundancy. Finally, partial least squares-support vector machine (LS-SVM) and convolutional neural network (CNN) based on U-Net model transformation are applied. After extracting the characteristic wavelength, the preprocessed spectral data and the spectral data are respectively used as the input of the model and established a non-linear mapping model of spectral data and comprehensive evaluation value. We can carry out comparison and selection. The results show that the spectral information of the bands filtered by the UVE-SPA preprocessing method is more abundant and effective. The regression effects of the models built for the two preprocessed optimal bands are overall better than the models built for the full bands; the modeling effect of the CNN model is overall. It is better than the LS-SVM model, and the UVE-SPA-CNN model has the best effect on the regression analysis of spectral data and seedling evaluation values. The correlation coefficients of the modeling set and prediction set are 0.988 and 0.946, respectively, and the values of the root mean square error are 0.085 and 0.025, respectively, which provide a theoretical basis for directly using spectral data to obtain the evaluation value of tomato seedlings that incorporate multiple factors, thereby judging the robustness of the seedlings.

**Keywords** Visible-near infrared; Tomato plug seedlings; Evaluation value; Convolutional neural network (CNN)

(Received Jun. 4, 2021; accepted Jul. 30, 2021)

\* Corresponding author