

基于 THz 成像和集成学习的番茄根长表型提取及预测

罗 慧，刘星语，韦 骁，吴嘉仪，余 赞，卢 伟

(南京农业大学人工智能学院，南京 210095)

摘 要：为检测番茄根系表型，该研究基于 THz (Terahertz) 成像和集成学习提出一种根系检测技术。首先，对 20 天生长过程中番茄根系进行多次 THz 成像。其次，对最优重构后的根系 THz 伪彩色图去除根系重叠和主根区域的噪声数据。再次，采用 Rosenfeld 细化算法和滑动窗口遍历法计算根系长度。最后，提取根系有效区域中 THz 时域光谱和折射率光谱，由 Stacking 集成模型对番茄根长进行预测。由 THz 成像计算的番茄根长结果误差小，平均相对误差仅为 4.16%；由 THz 时域数据预测的根长与计算得到的根长之间最大决定系数为 0.999，最小均方根误差为 0.743 cm；由折射率光谱数据预测根长的最大决定系数为 0.998，最小均方根误差为 0.976 cm。该方法不仅能根据 THz 图像准确地计算出番茄根系的长度，还能由番茄根系的 THz 光谱有效地预测番茄根长表型，该研究为根系表型检测方法提供了理论依据。

关键词：番茄；THz 成像；根系表型；Stacking 集成学习

doi: 10.11975/j.issn.1002-6819.202405178

中图分类号: S24; TP391.4

文献标志码: A

文章编号: 1002-6819(2024)-18-0176-08

罗慧, 刘星语, 韦骁, 等. 基于 THz 成像和集成学习的番茄根长表型提取及预测[J]. 农业工程学报, 2024, 40(18): 176-183. doi: 10.11975/j.issn.1002-6819.202405178 <http://www.tcsae.org>

LUO Hui, LIU Xingyu, WEI Xiao, et al. Extracting and predicting tomato root length phenotype using THz imaging and ensemble learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 176-183. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202405178 <http://www.tcsae.org>

0 引 言

作物主要通过根系吸收土壤中的养分，发达的根系不仅能固定支持植株、合成植株生长代谢所需的生理活性物质，还能促进植株的生长^[1-3]。番茄富含维生素、蛋白质等多种化学物质，是最受消费者青睐的蔬菜之一^[4-6]，番茄根系表型检测对筛选优势基因型番茄种子、改良番茄品种、提高番茄果实的产量和质量具有重要意义。

根系表型检测有二维根系检测技术和三维根系检测技术^[7-9]。二维根系表型检测技术通过相机、扫描仪及图像传感器等设备获取根系的二维图像，结合图像处理方法进行根系表型分析。孙国祥等基于阵列式 ESP32-CAM 对番茄根系表型进行原位测量，提取根系长度、面积、平均直径、根深和根宽^[10]。KANG 等采用语义分割模型对复杂土壤背景中的棉花根系图像进行分割处理^[11]。TERAMOTO 等使用卷积神经网络对沟槽剖面的根系图像进行语义分割，可快速量化沟槽剖面图下根系的分布状况^[12]。这类方法可高通量、自动化地对根系进行成像，但土壤的遮蔽会导致根系成像的不完整^[13-14]。三维根系表型检测技术采用探地雷达、磁共振成像、计算机断层扫描等三维检测装备获取作物根系的三维结构特征，HOU 等利用 X 射线断层扫描对土壤核心部分进行三维重建检测土壤中的根结构^[15]。DUSSCHOTEN 等利用磁共振成像技术对土壤中的根系进行三维成像，确定根长、

生长角度及尖端数量等根系表型参数^[16]，BAGNALL 等基于低场磁共振成像无损地呈现土壤中高粱根系的形态和结构^[17]。三维成像检测技术可获取根系的三维空间分布特征，能更加全面地观测根系生长特点，但该技术通常使用的试验设备成本高昂，获取的数据维度高，数据处理较为复杂^[18-19]。

光谱成像能够将光学成像和光谱分析相结合，与传统的二维平面根系成像相比，该方法既可展现根系的二维图像特征，又可全面地展现根系的内部信息和微观结构。目前为止，热红外成像、荧光光谱、高光谱已应用到小麦、玉米等作物的根系表型检测中^[20-22]。THz 成像是一种光谱成像技术，能够同时获取样本的 THz 图像和 THz 光谱，其波长在 0.03~3 mm 之间，振荡频率介于 0.1~10 THz^[23]，与其他光谱成像技术相比，THz 波穿透力强，可穿透根系表面获取根系的内部信息，THz 波的光子能量低，不会因电离而损伤根系细胞结构，可重复对根系样本进行 THz 成像，THz 波具有指纹谱特性，可用于定性和定量鉴定根系的内部特征^[24-27]。

本文研究一种基于 THz 成像和集成算法的番茄根系表型检测技术，其目的是通过 THz 图像能提取出番茄根系长度，同时建立 THz 光谱数据和根系长度的集成学习模型，实现番茄根系长度的有效预测。

1 材料与方 法

1.1 样本制备

本文选择番茄幼苗根系作为测试样本，为增加番茄幼苗根系生长的差异性，在三组不同浓度营养液浇灌的沙质土壤基质中培育 3 棵番茄幼苗，每隔 5 d 采集一次

收稿日期: 2024-05-24 修订日期: 2024-08-15

基金项目: 国家自然科学基金项目 (No. 32071896)

作者简介: 罗 慧, 博士, 硕士生导师, 副教授, 研究方向为农业无损及智能检测。Email: lh821005@njau.edu.cn

根系样本, 20 d 内共进行 4 组采样, 获得 12 个根系样本, 分别标记为 N11、N12、N13、N14、N21、N22、N23、N24、N31、N32、N33、N34, 其中下标第一位表示不同浓度营养液下的根系样本, 下标第二位表示采样的次数。为了便于番茄根系的 THz 成像, 番茄幼苗种植在 60° 倾角具有开合页的透明种植盒中, 具有倾斜角度的种植盒可确保番茄幼苗的根系尽量沿着种植盒表面生长, 如图 1 所示, 在进行 THz 成像前, 将番茄根系清洗干净, 使用吸水纸吸干番茄根系表面水分, THz 成像完成后番茄幼苗继续种回种植盒。

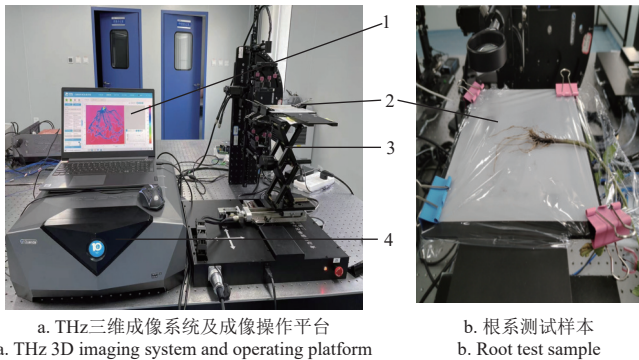


图 1 番茄种植

Fig.1 Tomato planting

1.2 番茄根系 THz 成像的数据采集

试验中 THz 成像系统为青源峰达 THz 科技有限公司自主研发的 THz 三维成像检测仪 (QT-TO1000), 工作模式为透射成像。QT-TO1000 成像系统的最大成像速度为 60 像素/s, 光谱检测宽度范围为 0.1~5.5 THz, 频谱分辨率为 8 GHz, 试验中通过扫描平台采集番茄根系的 THz 成像数据, 采样点为 1 687, 采样时间区间为 0~90 ps, 番茄根系的 THz 成像试验如图 2 所示。



1. THz 成像软件 2. 测试样本 3. 移动测试平台 4. THz 三维成像系统
1. THz imaging software 2. Test sample 3. Mobile test platform 4. THz 3D Imaging system

图 2 根系 THz 图谱数据采集

Fig.2 Root THz (Terahertz) mapping data acquisition

1.3 番茄根系 THz 光学参数的计算

折射率、吸收系数等光学信息可表征样本对 THz 波的吸收、散射等性质, 本文根据 DORNEY 等^[28-29]提出的 THz 波段光学参数计算式计算折射率及吸收系数。设 $E_{ref}(\omega)$ 为目标样本的 THz 频域信号, $E_{sam}(\omega)$ 为参考 THz 频域信号, 式 (1)~(3) 分别为折射率 $n(\omega)$ 、消光系数 $k(\omega)$ 及吸收系数 $\alpha(\omega)$ 的计算式。

$$n(\omega) = c \frac{\varphi(\omega)}{\omega d} + 1 \quad (1)$$

$$k(\omega) = \frac{c}{\omega d} \ln \frac{4n(\omega)}{A(\omega)[n(\omega)+1]^2} \quad (2)$$

$$\alpha(\omega) = \frac{2k(\omega)\omega}{c} = \frac{2}{d} \ln \frac{4n(\omega)}{A(\omega)[n(\omega)+1]^2} \quad (3)$$

式中 d 为样本厚度, mm; $A(\omega)$ 表示样本信号与背景信号的幅值比; $\varphi(\omega)$ 为样本信号与参考信号的相位差; ω 为角频率, rad/s; c 为真空中的光速, m/s。

1.4 基于 THz 伪彩色图的根系去重叠

番茄根系在生长过程中会产生重叠根现象, 同时番茄主根明显比侧根粗大, 根系的重叠和主根厚度的增加在 THz 成像时会直接影响 THz 波的吸收特性。由图 3 所示重叠根系和主根系的 THz 时域光谱波形的毛刺噪声很多, 高频噪声会淹没番茄根系真实 THz 时域信息。THz 伪彩色图的颜色信息能够直接反映根系对 THz 波的吸收特性, 其中颜色栏值取值越大, 表示此区域根系对 THz 波的吸收强度越弱。本文根据根系的 THz 伪彩色图去除重叠根系和主根系区域中含有噪声的 THz 光谱数据, 提取完整的根系轮廓, 并将有效根系区域中 THz 时域光谱作为后续集成模型训练和测试的数据集。基于 THz 伪彩色图的根系去重叠具体步骤如下:

1) 采用 THz 时域信号峰值对根系进行 THz 伪彩色图成像。

2) 对根系 THz 伪彩色图进行 R、G、B 三通道分量。其中 R 通道为 THz 伪彩色图的红色分量, 保留了测试背景、重叠根系和主根系的红色信息, 其余颜色信息均转变为黑色。G 通道对应其绿色分量, 由于该分量表征根系重叠的特征不明显, 本试验中未做处理。B 通道对应蓝色分量, 保留了根系的蓝色轮廓信息, 其余颜色信息均变为黑色。

3) 将 R 通道及 B 通道的图像进行灰度化和二值化处理, 分别获取非重叠根系与重叠根系的轮廓图, 对两种轮廓图进行异或和取反运算, 图像重构后即可获得去重叠后根系的轮廓图。

1.5 番茄的根长表型提取

根系表型有多种参数指标如根系长度、面积等, 本文主要研究番茄根长表型的提取, 对去噪后的根系 THz 伪彩色图进行骨架提取, 计算番茄根系在动态生长过程中的根长值, 番茄根系总根长的计算步骤如下:

1) 对番茄根系三维 THz 图谱数据进行 THz 时域最大值重构成像, 获得清晰的 THz 成像;

2) 根据 THz 伪彩色图的颜色信息去除重叠根系和

主根系区域中含噪声 THz 光谱数据。

3) 对番茄的 THz 图像进行灰度化、二值化等图像

处理, 采用 Rosenfeld 细化算法^[30] 获取番茄根系的骨架图, 根据滑动窗口遍历法计算根系像素点的长度。

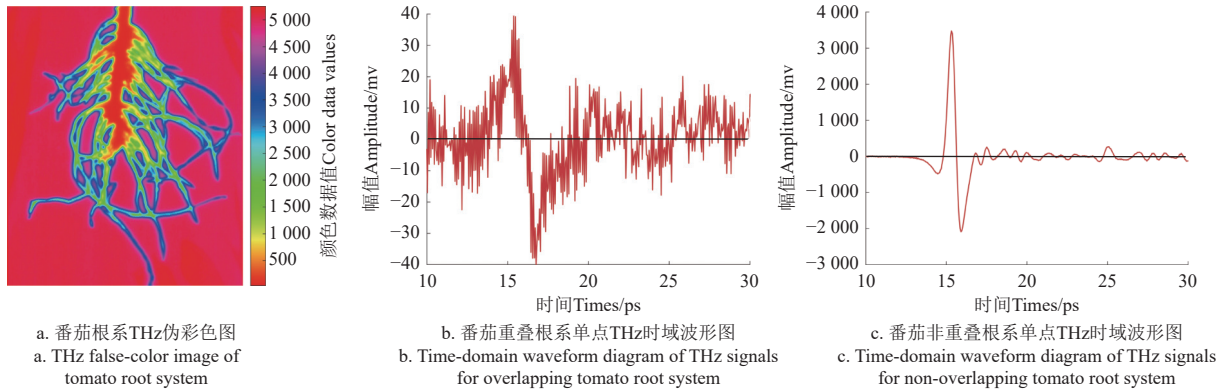


图3 番茄根系伪彩色图及 THz 光谱

Fig.3 False-color image and THz spectrum of tomato root system

1.6 Stacking 集成算法

与传统二维图像成像不同, THz 成像除了能获得根系的二维空间信息, 其光谱数据可表征根系的内部信息, 本文分别用 THz 时域数据和折射率数据预测番茄根系的根长, 根长的预测模型采用 Stacking 集成算法, 由决定系数判断 THz 光谱数据与根长的相关性, 根据均方根误差值的大小判断预测结果的准确性。

Stacking 集成算法利用多个学习器进行数据的训练, 将基模型的结果作为元学习器的输入, 以元学习器的预测结果作为 Stacking 模型的最终预测结果^[31]。相较于单一的机器学习模型, Stacking 集成算法可以充分利用基学习器的优势, 以提高自身的拟合效果和抗干扰能力, Stacking 集成算法在处理复杂的非线性数据时具有较好的泛化能力和预测准确性^[32-33]。

采用的 Stacking 集成模型第一层集成了 GBDT

(gradient boosting decision tree)、XGBoost (eXtreme gradient boosting)、Catboost (categorical boosting)、Adaboost (adaptive boosting) 4 种基模型, 第二层采用线性回归作为元学习器。为防止数据的过拟合采用 5 折交叉验证法训练基模型。将数据集按照 7:3 的比例分为训练集和测试集, 并将训练集进行 5 等分, 分别记作 A1、A2、A3、A4、A5。首先, 将 A1、A2、A3、A4、A5 分别作为验证集, 4 组基模型各训练 5 次, 将生成的 5 组结果记为 B1、B2、B3、B4、B5, 4 组基模型共生成 4 组预测结果。其次, 集成 4 组模型的预测结果分别记作 C1、C2、C3、C4, 同时测试集数据经基模型训练后也会生成 4 组预测结果, 分别记作 D1、D2、D3、D4。最后, 将 C1、C2、C3、C4 作为元学习器的训练集, D1、D2、D3、D4 作为元学习的测试集进行预测, 得到集成模型的最终预测结果, 集成训练过程如图 4 所示。

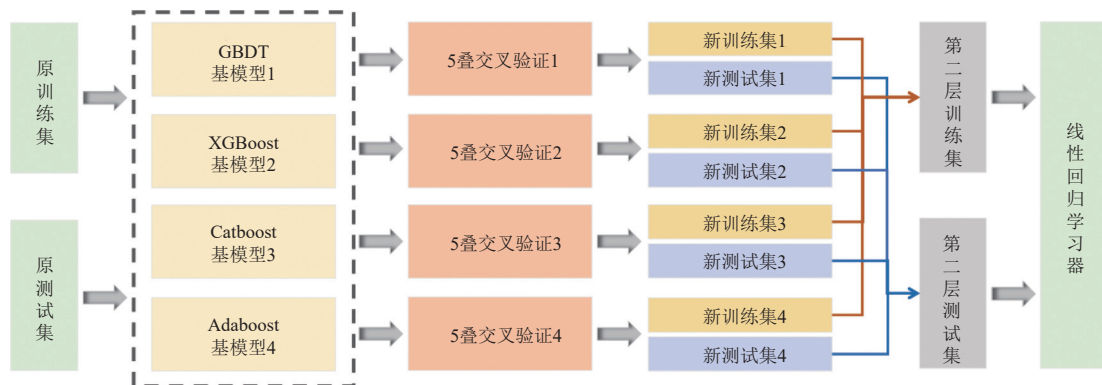


图4 Stacking 集成模型及 5 折交叉验证过程

Fig.4 Stacking integration model and the process of 5-fold cross-validation

2 结果与分析

2.1 番茄根系的 THz 成像

采集的 THz 成像数据为三维结构, 包含 THz 图像及 THz 光谱数据, 三维数据大小为 $1687 \times 229 \times 196$, 其中 1687 代表根系有 1687 张 THz 成像, 229×196 代表样本的光谱数据量, 每个光谱数据对应一条时域信号波形。

图 5 为番茄根系 N11、N21、N31 采集的 THz 三维数据根据时域光谱最大值重构后的 THz 成像及根系有效区域内的 THz 光谱, 由图 5 可知 THz 成像能清晰地呈现水稻根系生长特征, 伪彩色图的颜色值可直观地观测根系生长的粗细及重叠情况, 三种根系样本太赫兹谱的时域最大值接近。

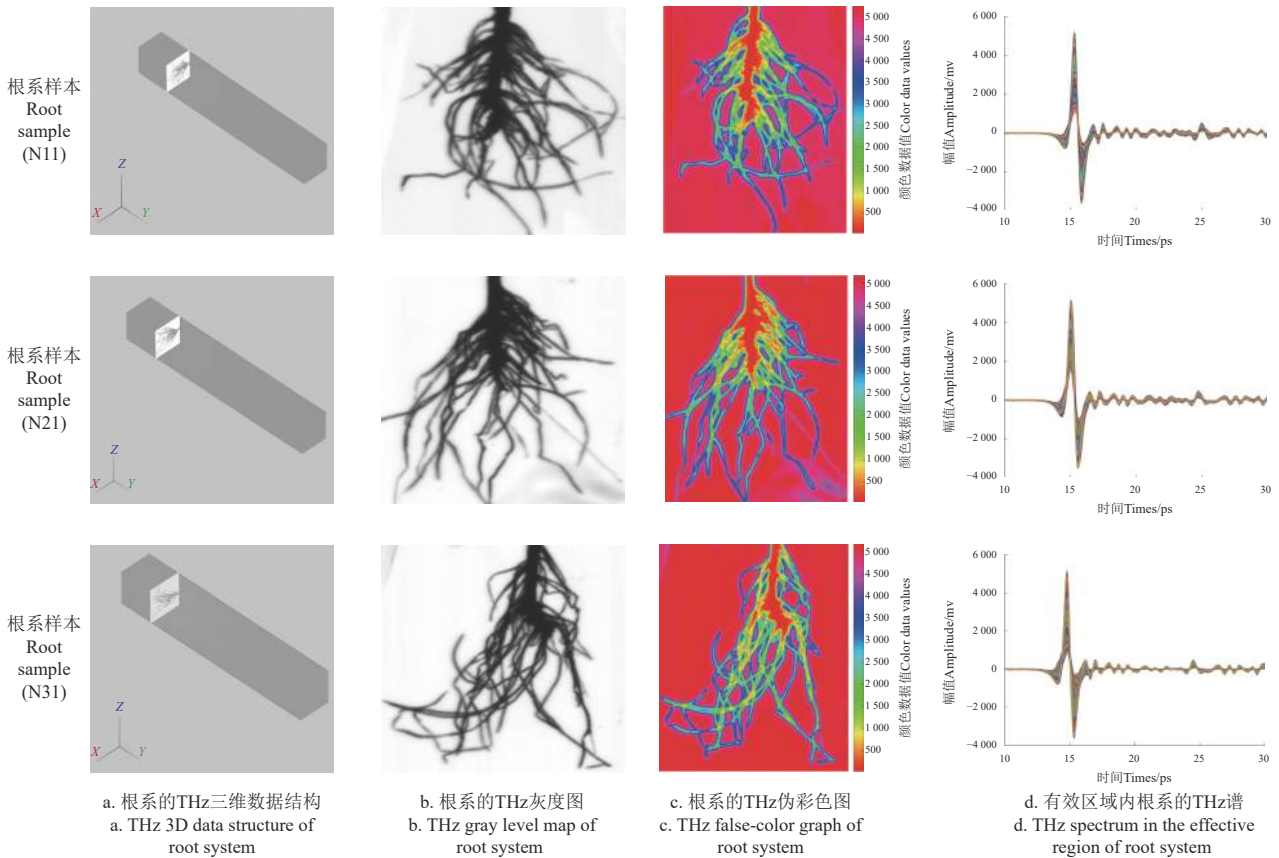


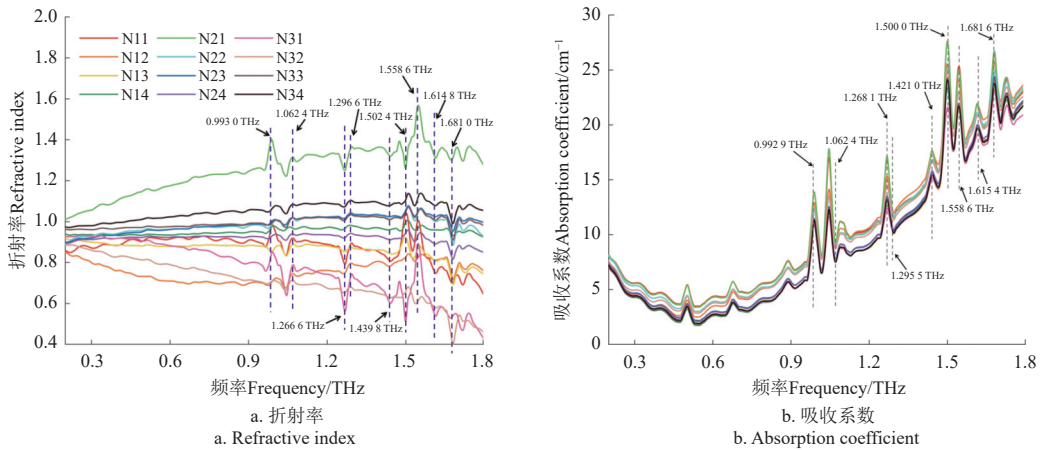
图 5 番茄根系 THz 成像的预处理

Fig.5 Preprocessing of tomato root THz imaging

2.2 光学参数计算

针对番茄去重叠根系 THz 时域数据，本文根据式 (1)

及 (3) 分别计算 12 组番茄根系在 0.2~1.8 THz 波段的平均折射率及吸收系数如图 6 所示。



注：N11~N14 曲线为第一株番茄幼苗根系的 4 次 THz 成像计算得到的折射率；N21~N24 曲线为第二株番茄幼苗根系的 4 次 THz 成像计算得到的折射率；N31~N34 曲线为第三株番茄幼苗根系的 4 次 THz 成像计算得到的折射率。
 Note: The N11-N14 curves are the refractive index calculated by four THz imaging of the root system of the first tomato seedling. The N21-N24 curves are the refractive index calculated by the fourth THz imaging of the root system of the second tomato seedling. The N31-N34 curves are the refractive index calculated by the fourth THz imaging of the root system of the third tomato seedling.

图 6 折射率及吸收系数

Fig.6 Refractive index and absorption coefficient

12 组平均折射率及吸收系数波形基本一致，其中折射率曲线幅值明显不同，但在几个特征频率点处有共同的特征峰，吸收系数波形一致呈上升趋势，在相同频率点附近同样存在明显的吸收峰，根据式 (1) 及式 (3)，两组光学参数的特征频率值略有不同，但几乎都在相同

的频率点附近，本文根据折射率及吸收系数产生波峰或波谷的频率点的平均值作为样本的特征频率点，各特征频率点分别为 0.993 0、1.062 4、1.267 4、1.296 1、1.430 4、1.501 2、1.558 6、1.615 1、1.681 3 THz。试验中将这 7 组特征频率点与根长进行线性建模，试验结果显示根系

的 THz 特征频率点与根长无明显的线性关系, 由于折射率和吸收系数可相互计算, 后续试验中将折射率光谱数据作为根系长度的预测数据集之一。

2.3 根系去重叠结果分析

图 7 为番茄根系样本的 THz 伪彩色图经 RGB 三通道分离的结果, 其中 R 通道仅包含黑色根系轮廓信息与红色测试背景及根系重叠区域信息, B 通道仅包含蓝色根系轮廓与黑色的背景、重叠根系区域, G 通道表示的根系轮廓出现明显的断根现象, 后续试验中不处理 G 通道的图片信息。

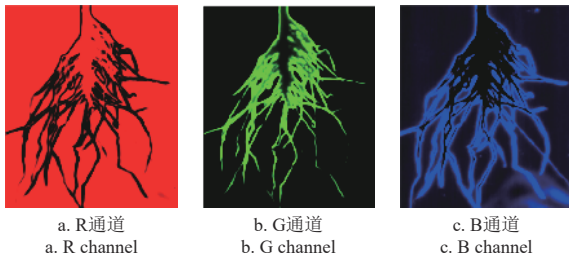


图 7 RGB 三通道分离的根系图

Fig.7 Root diagram of RGB three-channel separation

图 8 中分别将 R 通道及 B 通道的图像进行灰度化、二值化操作, 获取非重叠根系与重叠根系的轮廓图。然后, 对番茄根系未去重叠根系和重叠根系图像进行异或和取反操作, 图像重构后获得完整番茄根系。

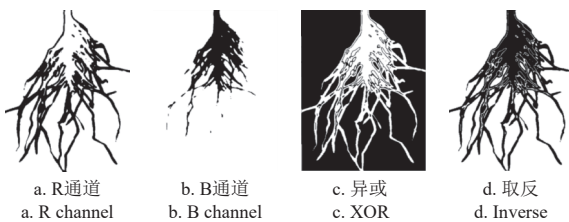


图 8 异或和取反的根系图

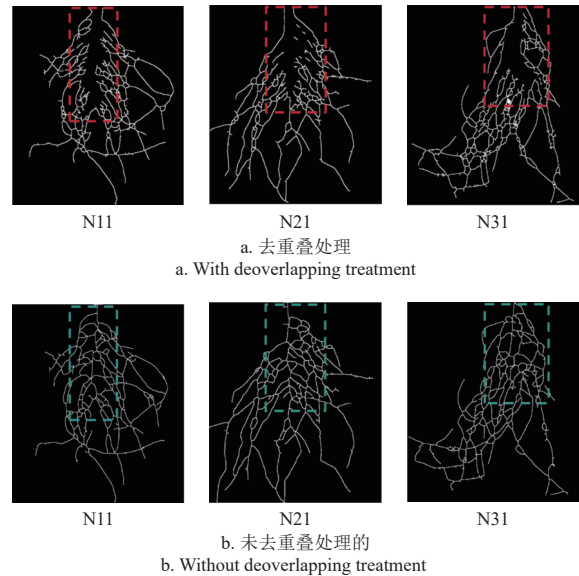
Fig.8 Root diagram operated by exclusive or and inverse

图 9 表示采用 Rosenfeld 细化算法获得的番茄根系骨架图, 对比两组图像结果可知, 未去重叠处理的根系骨架图在主根部位和重叠根系部位的根系骨架混乱, 会直接影响根系根长计算的结果, 而经 RGB 三通道分离后的根系骨架完整的显示了主根区域以及剔除了重叠根系骨架, 减少了根长计算的误差。

2.4 根系表型参数计算

试验中采用 Rosenfeld 细化算法结合八邻域滑动窗口遍历法计算番茄根长值, 为了验证本文由 THz 图像计算番茄根长结果的可信程度, 本文将采集的二维根系图像由 Image J 软件计算的番茄根长结果作为参考值, 计算两组测量数据之间的相对误差, 通过计算本文结果值与软件计算结果值之间的比例关系, 判断本文计算结果的准确性。由表 1 可知, 本文计算的根长结果与 Image J 软件计算的根长结果相近, 计算的平均根长分别为 69.252 cm 和 68.999 cm, 平均相对误差仅为 4.16%。图 10 为本文方法计算的根长与 Image J 软件计算根长的线性拟合结果图, 线性拟合模型的决定系数为 0.967。试验结果表明,

由本文方法计算番茄根长的结果是准确的, 番茄根系的 THz 图像可直接准确的计算番茄根系根长的参数。



注: 虚线框表示根系主根及重叠根系区域

Note: The dashed box indicates the root main root and overlapping root areas

图 9 采用 Rosenfeld 细化算法的根系骨架图

Fig.9 Root skeleton map using Rosenfeld thinning algorithm

表 1 番茄根长计算结果对比

样本 Sample	根长 Root length/cm		
	本文方法 Proposed method	Image J 软件 Image J software	相对误差 Relative error/%
N11	83.598	83.991	0.47
N12	57.142	63.950	11.91
N13	52.223	55.965	7.16
N14	34.314	35.166	2.48
N21	88.663	87.590	1.21
N22	68.984	64.711	6.19
N23	62.980	57.046	9.42
N24	54.528	52.631	3.48
N31	104.742	107.052	2.21
N32	81.057	80.977	0.10
N33	75.544	72.621	3.87
N34	67.247	66.294	1.42
平均值 Average	69.252	68.999	4.16

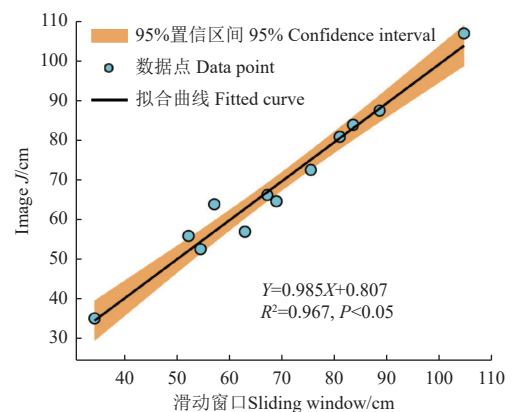


图 10 Image J 与滑动窗口计算根长的线性拟合

Fig.10 The linear fitting of the calculation root length of Image J and Sliding window

2.5 番茄根系表型参数预测结果及分析

试验中将番茄根系的 THz 时域光谱数据和 THz 折射

率光谱数据作为数据集进行根长的预测建模, 两类光谱数据输入预测模型之前采用 WD (wavelet denoising, WD) 去噪算法和 SPE (spatial proximity embedding, SPE) 降维算法进行预处理。Stacking 集成模型采用决定系数 (R^2) 及均方根误差 (root mean square error, RMSE) 作为评估指标, 决定系数越大、均方根误差越小表示 Stacking 集成模型对番茄根长参数的预测结果越好。

Stacking 集成模型对番茄根系 THz 时域数据 (time-domain data, TD) 和折射率数据 RI (refractive index, RI) 的预测结果如表 2 所示。由表 2 可知, 所有时域数据的预测结果中, 即使采用 3 种不同的光谱数据预处理方法, Stacking 集成模型对番茄根长的预测结果均优于各子模型, 说明 Stacking 集成模型能有效的提高预测精度, 其中采用 WD 预处理方法的预测结果最优, 取得了 0.999 最大的决定系数和 0.743 cm 最小的均方根误差。此外, 除 Adaboost 模型外, 集成模型和其他三种子模型获得的决定系数均大于 0.94, 说明番茄根系的 THz 光谱数据与根长具有较大的相关性。

表 2 THz 时域数据及折射率数据预测的番茄根长结果
Table 2 Predicting results of tomato root length by THz time-domain data and refractive index data

数据类型 Data type	模型 Model	预处理方法 Preprocessing method	R^2	RMSE/cm
THz 时域数据 THz time-domain data	Stacking	TD	0.962	4.001
		WD	0.999	0.743
		WD-SPE	0.988	2.237
	GBDT	TRD	0.956	4.332
		WD	0.998	0.903
		WD-SPE	0.981	2.768
	XGBoost	TD	0.948	4.711
		WD	0.994	1.636
		WD-SPE	0.987	2.309
	Catboost	TD	0.956	4.332
		WD	0.998	0.903
		WD-SPE	0.981	2.768
Adaboost	TD	0.885	7.002	
	WD	0.991	1.926	
	WD-SPE	0.856	7.687	
折射率数据 Refractive index data	Stacking	RI	0.857	7.740
		WD	0.998	0.976
		WD-SPE	0.995	1.410
	GBDT	RI	0.784	9.523
		WD	0.994	1.621
		WD-SPE	0.984	2.561
	XGBoost	RI	0.730	10.645
		WD	0.995	1.514
		WD-SPE	0.988	2.204
	Catboost	RI	0.784	9.523
		WD	0.994	1.621
		WD-SPE	0.984	2.561
Adaboost	RI	0.613	12.759	
	WD	0.992	1.844	
	WD-SPE	0.954	4.390	

同样, 根据表 2 可知, 使用 RI 数据直接对根长参数的预测结果最差, Stacking 及其子模型的决定系数均低于 0.86, 对 RI 进行 WD 去噪及 WD-SPE 降维均明显提高模型的预测性能, 且 WD 去噪后的数据对根长参数的预测结果最优, 和根系 THz 时域数据预测的结果一致。所有预测模型中, Stacking 对根长表型参数的预测结果均优于各子模型, 采用 WD 去噪后数据集的最优预测决定系数为 0.998, 均方根误差为 0.976 cm。

综合两组预测结果可知, 采用 Stacking 集成模型和 4 种基模型均能有效的预测出番茄根系的长度, Stacking 集成模型能有效的提高预测精度, 结合 WD 去噪方法可获得最优的测试结果。因此, 试验结果证明番茄根系的 THz 光谱数据能有效的预测番茄根长表型, 原因可能和根系在动态生长过程中根系的长度和宽度不断的改变, 导致样本测试时 THz 波接触面积不断的增大, 同时根系变粗会直接影响 THz 波的吸收强度, 因此 THz 光谱可反映出根长表型特征发生的变化。另外, 根系在生长过程中内部的营养成分也会发生变化, 这也会影响根系 THz 光谱的特性发生变化。

3 结 论

本文基于 THz 成像技术研究了一种新颖的根系表型检测方法, 并应用到番茄根系的根长表型检测中。THz 成像可同时获得根系的 THz 图像信息和 THz 光谱信息, 根据试验结果得到以下结论:

1) THz 伪彩色图的颜色信息能够直接反映根系对 THz 波的吸收特性, 采用 RGB 三通道分离方法可有效地去除重叠根系, 删除含有噪声的光谱数据, 完整地显示根系骨架, 减少根长计算的误差。

2) 根据 THz 图像可准确地计算番茄根系的长度。THz 伪彩色图计算的番茄根长值与 Image J 软件计算的根长值之间的平均相对误差仅为 4.16%, 两类数据结果的线性拟合决定系数达到 0.967。

3) 番茄根系有效区域内的 THz 光谱和折射率数据结合 Stacking 集成模型可有效地预测番茄根长表型。Stacking 集成模型的预测结果优于子模型, 采用 WD 去噪处理后的 THz 数据对番茄根长的预测结果最佳, 两组数据预测的最优决定系数分别是 0.999 和 0.998, 最小均方根误差分别是 0.743 和 0.976 cm。

基于 THz 成像技术的研究方法为根系表型检测提供了一种研究思路和技术方法, 该方法还处于初步研究阶段, 如何无损地检测完整根系表型参数以及根据 THz 光谱数据检测根系内部特征如含水量、氮素含量等还有待进一步的研究。

[参 考 文 献]

- [1] 王传宇, 郭新宇, 温维亮, 等. 基于玉米根系图像的表型指标获取方法[J]. 农业工程学报, 2021, 37(8): 169-176.
WANG Chuanyu, GUO Xinyu, WEN Weiliang, et al. Phenotyping index acquisition method based on maize root images[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(8): 169-176. (in Chinese with English abstract)
- [2] JIA Z, GIEHL R F H, WIRÉN N V. Nutrient-hormone relations: Driving root plasticity in plants[J]. Molecular Plant, 2022, 15: 86-103.
- [3] RONG Y, ZHAO Z, FENG X, et al. Computational morphomechanics of growing plant roots[J]. Journal of the Mechanics and Physics of Solids, 2023, 178: 105346.

- [4] 罗珍, 宁初光, 李祚祺, 等. 番茄、野樱莓提取物抗光老化功效及机制研究[J]. 食品工业科技, 2023, 44(16): 395-402.
LUO Zhen, NING Chuguang, LI Zuoqin, et al. Anti-photoaging effect and mechanism of tomato and aronia melanocarpa extract[J]. Science and Technology of Food Industry, 2023, 44(16): 395-402. (in Chinese with English abstract)
- [5] 向凌潇, 张俊威, 李建明. 灌溉量与灌溉频率对番茄根系生长、产量和营养元素吸收的影响[J]. 西北农林科技大学学报(自然科学版), 2024, 52(5): 80-92, 123.
XIANG Lingxiao, ZHANG Junwei, LI Jianming. Effects of irrigation amount and frequency on root growth, yield and nutrient element absorption of tomato[J]. Journal of Northwest A&F University (Natural Science Edition), 2024, 52(5): 80-92, 123. (in Chinese with English abstract)
- [6] AVDIKOS I D, TAGIAKAS R, TSOUVALTZIS P, et al. Comparative evaluation of tomato hybrids and inbred lines for fruit quality traits[J]. Agronomy, 2021, 11(3): 0609.
- [7] 何勇, 李禧尧, 杨国峰, 等. 室内高通量种质资源表型平台研究进展与展望[J]. 农业工程学报, 2022, 38(17): 127-141.
HE Yong, LI Xiyao, YANG Guofeng, et al. Research progress and prospect of indoor high-throughput germplasm phenotyping platforms[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(17): 127-141. (in Chinese with English abstract)
- [8] 吴茜, 张伟欣, 张玲玲, 等. 植物根系表型信息获取技术研究进展[J]. 江苏农业科学, 2021, 49(5): 31-37.
- [9] 李龙, 李超男, 毛新国, 等. 作物根系表型鉴定评价方法的现状与展望[J]. 中国农业科学, 2022, 55(3): 425-437.
LI Long, LI Chaonan, MAO Xinguo, et al. Advances and perspectives of approaches to phenotyping crop root system[J]. Scientia Agricultura Sinica, 2022, 55(3): 425-437. (in Chinese with English abstract)
- [10] 孙国祥, 蔡嘉奇, 周新竹, 等. 基于阵列式 ESP32-CAM 的番茄根系表型原位测量方法[J]. 农业工程学报, 2023, 39(18): 172-182.
SUN Guoxiang, CAI Jiaqi, ZHOU Xinzhu, et al. In-situ measuring tomato root phenotype using array ESP32-CAM[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(18): 172-182. (in Chinese with English abstract)
- [11] KANG J, LIU L, ZHANG F, et al. Semantic segmentation model of cotton roots in-situ image based on attention mechanism[J]. Computers and Electronics in Agriculture, 2021, 189: 106370.
- [12] TERAMOTO S, UGA Y. A deep learning-based phenotypic analysis of rice root distribution from field images[J]. Plant Phenomics, 2020, 2020: 3194308.
- [13] CHANDNAN R, QIN T, YE H, et al. Application of an improved 2-Dimensional high-throughput soybean root phenotyping platform to identify novel genetic variants regulating root architecture traits[J]. Plant Phenomics, 2023, 5: 0097.
- [14] TRIPATHI P, ABDULLAH J S, KIM J, et al. Investigation of root morphological traits using 2D-imaging among diverse soybeans (*Glycine max L.*) [J]. Plants, 2021, 10(11): 2535.
- [15] HOU L, GAO W, BOM F. Use of X-ray tomography for examining root architecture in soils[J]. Geoderma, 2022, 405: 115405.
- [16] DUSSCHOTEN D V, METZNER R, KOCHS J, et al. Quantitative 3D analysis of plant roots growing in soil using magnetic resonance imaging[J]. Plant Physiology, 2016, 170(3): 1176-1188.
- [17] BAGNALL G C, KOONJOO N, ALTOBELLI S, et al. Low-field magnetic resonance imaging of roots in intact clayey and silty soils[J]. Geoderma, 2020, 370: 114356.
- [18] 朱荣胜, 李帅, 孙永哲, 等. 作物三维重构技术研究现状及前景展望[J]. 智慧农业(中英文), 2021, 3(3): 94-115.
ZHU Rongsheng, LI Shuai, SUN Yongzhe, et al. Research advances and prospects of crop 3D reconstruction technology[J]. Smart Agriculture, 2021, 3(3): 94-115. (in Chinese with English abstract)
- [19] 刘玲玲, 房焕, 魏晓庆, 等. 利用 X 射线 CT 技术研究根系三维构型的进展[J]. 植物营养与肥料学报, 2024, 30(2): 375-384.
LIU Lingling, FANG Huan, WEI Xiaqing, et al. Research advances on using X-ray CT in the study of 3-D root architecture[J]. Journal of Plant Nutrition and Fertilizers, 2024, 30(2): 375-384. (in Chinese with English abstract)
- [20] WASSON A, BISCHOF L, ZWART A, et al. A portable fluorescence spectroscopy imaging system for automated root phenotyping in soil cores in the field[J]. Journal of Experimental Botany, 2016, 67(4): 1033-1043.
- [21] 卢伟, 韩钊, 蹇兴亮, 等. 基于热红外成像和断根修复算法的玉米根系表型检测方法[J]. 光谱学与光谱分析, 2020, 40(9): 2845-2850.
LU Wei, HAN Zhao, JIAN Xingliang, et al. Maize root phenotype detection method based on thermal infrared imaging and root break repair algorithm[J]. Spectroscopy and Spectral Analysis, 2020, 40(9): 2845-2850. (in Chinese with English abstract)
- [22] BODNER G, NAKHFOROOSH A, ARNOLD T et al. Hyperspectral imaging: A novel approach for plant root phenotyping[J]. Plant Methods, 2018, 14: 84.
- [23] 王与焯, 李海滨, 葛梅兰, 等. THz 成像技术及其应用[J]. 激光与光电子学进展, 2023, 60(18): 46-61.
WANG Yuyue, LI Haibin, GE Meilan, et al. Terahertz imaging technology and its application[J]. Laser & Optoelectronics Progress, 2023, 60(18): 46-61. (in Chinese with English abstract)
- [24] 刘翠玲, 王少敏, 吴静珠, 等. 基于 THz 时域透射成像技术的葵花籽内部品质无损检测研究[J]. 光谱学与光谱分析, 2020, 40(11): 3384-3389.
LIU Cuiling, WANG Shaomin, WU Jingzhu, et al. Study on internal quality nondestructive detection of sunflower seed based on terahertz time-domain transmission imaging technology[J]. Spectroscopy and Spectral Analysis, 2020, 40(11): 3384-3389. (in Chinese with English abstract)
- [25] SEI N, SAKAI T, HAYAKAWA Y, et al. Observation of terahertz coherent edge radiation amplified by infrared free-electron laser oscillations[J]. Scientific Reports, 2021, 11:

- 3433(2021).
- [26] LEI T, SUN D. Introducing the THz time domain CT system for evaluating kernel weight and plumpness of sunflower seed[J]. *Journal of Food Measurement and Characterization*, 2023, 17: 3616-3624.
- [27] HOSSAIN S, MOLLAH M A, HOSAIN M K, et al. Designing of hollow core grapefruit fiber using cyclo olefin polymer for the detection of fuel adulteration in terahertz region[J]. *Polymers*, 2023, 15(1): 0151.
- [28] DORNEY T D, BARANIUK R G, MITTLEMAN D M. Material parameter estimation with terahertz time-domain spectroscopy[J]. *Journal of the Optical Society of America-A Optics Image Science and Vision*, 2001, 18(7): 1562-1571.
- [29] DUVILLARET L, GARET F, COUTAZ J L. A reliable method for extraction of material parameters in Terahertz time-domain spectroscopy[J]. *IEEE Journal of Selected Topics in Quantum Electronics*. 2002, 2(3): 739-746.
- [30] XIE F, XU G, CHENG Y, et al. Human body and posture recognition system based on an improved thinning algorithm[J]. *IET Image Processing*, 2011, 5(5): 420-428.
- [31] ZHANG H, ZHU T. Stacking Model for photovoltaic-power-generation prediction[J]. *Sustainability*, 2022, 14(9): 5669.
- [32] 吴彤, 李勇, 葛莹, 等. 利用 Stacking 集成学习估算柑橘叶片氮含量[J]. *农业工程学报*, 2021, 37(13): 163-171.
- WU Tong, LI Yong, GE Ying, et al. Estimation of nitrogen contents in citrus leaves using Stacking ensemble learning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(13): 163-171. (in Chinese with English abstract)
- [33] 陈一凡, 李璇, 王绍雯, 等. 集成学习结合多源数据预测河南省冬小麦单产[J]. *农业工程学报*, 2024, 40(4): 183-191.
- CHEN Yifan, LI Xuan, WANG Shaowen, et al. Predicting winter wheat yield per unit area in Henan Province of China using ensemble learning and multi-source data[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(4): 183-191. (in Chinese with English abstract)

Extracting and predicting tomato root length phenotype using THz imaging and ensemble learning

LUO Hui , LIU Xingyu , WEI Xiao , WU Jiayi , YU Yun , LU Wei

(College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: Here rapid and non-destructive detection of tomato root phenotypes was realized using THz (Terahertz) imaging and ensemble learning. Firstly, three groups of tomato seedlings were grown in three sandy soil substrates watered with different concentrations of nitrogen nutrient solution. 12 groups of tomato root seedlings were collected for THz imaging during 20 days of growth. Secondly, the THz pseudo-color map was imaged on the root system after the optimal reconstruction of time-domain peaks. Color value of pseudo-color map was directly related to the intensity of the root system's THz absorption. The noisy data were removed from the overlapping and main root region in the THz maps of the root system, according to the color value of pseudo-color map. Again, the Rosenfeld refinement was used to obtain the skeleton map of tomato root system. The lengths of the root system pixel points were calculated using the sliding window traversal method. Finally, THz time-domain data and refractive index data were extracted from the effective feature region of the root system. The tomato root length was predicted by the Stacking ensemble model. Among them, the first layer of Stacking ensemble model was integrated with the four base models, namely, GBDT (gradient boosting decision tree), XGBoost (eXtreme gradient boosting), Catboost (categorical boosting), and Adaboost (adaptive boosting). The second layer was employed the linear regression as a meta-learner, in order to prevent the over data fitting. A 5-fold cross-validation was used to train the base models. The extraction of root skeleton showed that the RGB three-channel separation was effectively removed the overlapping roots and the spectral data containing noise, in order to fully display the root framework. Therefore, the calculation error of root length was reduced significantly. Only 4.16% was found in the average relative error of tomato root length value calculated by between the THz false-color image and Image-J software. The linear fitting determination coefficient of two types was 0.967. The THz time domain and refractive index of Stacking model were effectively predicted the root length, indicating the better performance than that of the sub-models. The optimal prediction of tomato root length was obtained using the THz data after WD denoising. The optimal determination coefficient of THz time-domain data prediction was 0.999, and the minimum root-mean-square error was 0.743. The optimal determination coefficient of THz refractive index data prediction was 0.998, and the root-mean-square error was 0.976 cm. The length of tomato root system was accurately calculated to predict the root length phenotypes using THz spectra images. The finding can provide a theoretical basis for the rapid and nondestructive detection of root system phenotypes. Further research can be required to nondestructively detect the complete parameters of root phenotype, and then detect the root internal characteristics using THz spectral data.

Keywords: tomato; THz imaging; root phenotype; stacking ensemble learning