

# 基于根系图像处理的番茄枯萎病检测研究

郑琼洁<sup>1</sup>, 李敬蕊<sup>2</sup>, 籍颖<sup>1,3\*</sup>

<sup>1</sup>河北农业大学信息科学与技术学院, 河北 保定

<sup>2</sup>河北农业大学园艺学院, 河北 保定

<sup>3</sup>河北省农业大数据重点实验室, 河北 保定

收稿日期: 2022年2月20日; 录用日期: 2022年4月13日; 发布日期: 2022年4月20日

## 摘要

番茄枯萎病是番茄病害中最严重的一种, 枯萎病的早期识别具有重要意义。本研究以患枯萎病番茄的根部为实验对象, 通过图像处理技术, 首先将番茄根部用扩展高斯差分(XDoG)进行边缘检测, 在HSV色彩空间中对番茄枯萎病进行检测。对于根系没有颜色变化的样本, 提取与病害相关的根部形状参数, 并结合从根系扫描仪获取的参数, 建立随机森林(RF)检测模型, 识别率为92.64%。为了缩短该方法的运行时间并提高准确率, 引入主成分分析法(PCA), 建立PCA-RF模型, 该模型的运行时间提高了62.13%, 平均识别率提高了2.62%。结果表明, 与常用的识别算法相比, PCA-RF模型具有更高的检测准确率。本研究为番茄枯萎病识别提供了一种高效稳定的方法。

## 关键词

番茄, 枯萎病, 病害检测, 图像处理, PCA-RF

# Research on Tomato Fusarium Wilt Detection Based on Root Image Processing

Qiongjie Zheng<sup>1</sup>, Jingrui Li<sup>2</sup>, Ying Ji<sup>1,3\*</sup>

<sup>1</sup>College of Information Science and Technology, Hebei Agricultural University, Baoding Hebei

<sup>2</sup>College of Horticulture, Hebei Agricultural University, Baoding Hebei

<sup>3</sup>Hebei Key Laboratory of Agricultural Big Data, Baoding Hebei

Received: Feb. 20<sup>th</sup>, 2022; accepted: Apr. 13<sup>th</sup>, 2022; published: Apr. 20<sup>th</sup>, 2022

## Abstract

Tomato Fusarium wilt is one of the most serious tomato diseases, and early identification of to-  
\*通讯作者。

tomato Fusarium wilt is of great significance. In this study, the roots of tomatoes with Fusarium wilt were used as the experimental object. Through image processing technology, the tomato roots were edge detected by using the extended difference of Gaussian (XDoG), and the tomato Fusarium wilt was detected in the HSV color space. For samples with no color change at the root, the root shape parameters related to the disease were extracted, and combined with the parameters obtained from the root system scanner, a Random Forest (RF) detection model was established, and the recognition rate was 92.64%. In order to shorten the running time of the method and improve the accuracy, Principal Component Analysis (PCA) was introduced and a PCA-RF model was established. The running time of the model was increased by 62.13% and the average recognition rate was increased by 2.62%. The results show that the PCA-RF model has higher detection accuracy than the commonly used recognition algorithms. This study provides an efficient and stable method for tomato Fusarium wilt identification.

## Keywords

Tomato, Fusarium Wilt, Disease Monitoring, Image Processing, PCA-RF

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

番茄是世界上第二大经济作物, 目前我国的番茄产量占到世界蔬菜生产总量的 10% 左右[1] [2]。随着近些年番茄种植面积的扩大, 番茄种植户的经济收入明显增加, 但终年连作导致了番茄很多病害的发生[3], 使得番茄减产降质。如今番茄病害已经阻碍了番茄产业的发展, 给农民带来了很大的经济损失。因此番茄病害检测具有非常重要的意义。

根系是植物的重要器官, 它吸收土壤中的养分[4]并将其输送到植物的其他部位, 以保证植物的生长。因此, 研究植物根系的形态参数有助于识别植物的生长状况。近年来, 随着人工智能技术的发展, 机器学习进入了人们的视野, 计算机解析根系图像的方法得到了学者们的重视。

目前, 一些学者综合比较了根系特征对根系生长状况的影响[5], 并对提取根系特征的方法提供了一些指导性的建议[6]。在根系研究方面国外开展得比较早, 一些学者已经将计算机技术应用于根系研究中。目前, 很多国家在该领域进行了探索性的研究, 部分成熟技术已投入使用[7], 比如美国的 GiA 系统[8], 澳大利亚的射线照相技术, 加拿大的 ET-100 根系生态检测系统等, 利用这些技术对根系图像进行处理, 可以快速并准确地测量出根系参数。

在国内, 赵来宾等[9]集成现代光电技术研制出一种三维植物根系扫描仪, 这为之后的研究工作提供了很大的便利。胡秀娟[10]将计算机视觉信息技术和机电一体化技术应用于根系形态分析, 得到了适合根系形态分析的图像平滑算法和增强算法, 进而提出了一种快速有效的获取根系图像的方法。此后, 杨国梁[11]在分析根系形态固有特征的基础上, 设计并实现了一种基于曲线走向最短路径的主根识别算法, 能够快速识别主根和主根上的侧根。植物根部的复杂性以及生长环境的多样化, 使得植物根系图像提取工作变成了一个更加复杂的、具有挑战性的工作。

植物根系图像处理技术的应用目前还未达到成熟状态, 为了更好地应用于实际环境, 还需要进行进一步的研究。目前学者们的研究都是围绕怎么提取根系参数来展开, 如果能将提取的根系参数用于植物病害的识别, 那将更有实用价值。本文提出了一种将提取的根系参数用于判断植物是否患病的新思路。

## 2. 材料与方法

### 2.1. 番茄枯萎病症状

番茄枯萎病[12] [13] [14]是一种根系病害。一般来说,枯萎病病菌会在番茄开花结果的时期率先侵染番茄植株的根部,然后从下到上依次侵染茎部和叶片。番茄患枯萎病时根系的表现症状为根系稀少,根系随着侵染程度的增大会渐渐变黄,严重的甚至会变成褐色;发病早期茎部出现凹陷,有的会呈现开裂症状;随着病情的日益严重,番茄病叶开始慢慢由绿变黄,有的植株叶片萎蔫、焦枯直至坏死。

### 2.2. 数据采集

本实验番茄品种为草莓番茄,实验样品取自河北农业大学园艺学院实验室,主要取患有枯萎病番茄和健康番茄的根部扫描图像。播种期为2021年4月6日,首先将番茄种子播种在穴盘中,每个穴盘种植72株番茄,穴盘中的基质按照7:2:1的比例将草炭、椰糠和蛭石混合配制而成。温室每天进行12小时左右的光照,白天温度在23℃左右,夜间保持在13℃左右,相对湿度为60%。

待幼苗长到两叶一心时,采用浸根法进行枯萎病病菌接种处理。实验一共进行了28天,每天在相同时间段内将番茄从穴盘中取出,然后将番茄根部的基质清洗干净,晾干之后采集番茄根部的扫描图像。一共采集110株植株的数据,其中患枯萎病的番茄根部数据有52幅,对照组的番茄根部数据有58幅。

## 3. 结果与分析

### 3.1. 根系扫描图像边缘提取

为了准确的提取根系特征,首先需要将根系扫描图像进行边缘提取。扩展高斯差分[15] (XDoG)是在高斯差分的基础上进行改进的。XDoG可看做是高斯模糊和高斯差分结果的加权平均,所以XDoG的计算方法可表示成公式(1)。

$$D_{\sigma,k,\tau}(x) = G_{\sigma}(x) - \tau \cdot G_{k,\sigma}(x) = (1-\tau) \cdot G_{\sigma}(x) + \tau \cdot D_{k,\sigma}(x) \quad (1)$$

高斯模糊图像具有与输入图像相同的平均亮度,所以图像的平均亮度将随着 $\tau$ 的增加而降低,增加 $\tau$ 是增加边缘强调线权值的唯一方法。

为了简化XDoG滤波器的参数控制,进行具有以下属性的重新参数化:去除先前参数的紧密相互依赖性;需要调整的参数可以更加符合我们的直觉;可逆性,即可以在新旧参数空间来回转换。通过将公式(1)除以 $\tau-1$ ,就可以找到满足这些需求的参数,使得XDoG滤波器表示为调整后的图像锐化运算符,得到公式(2)。

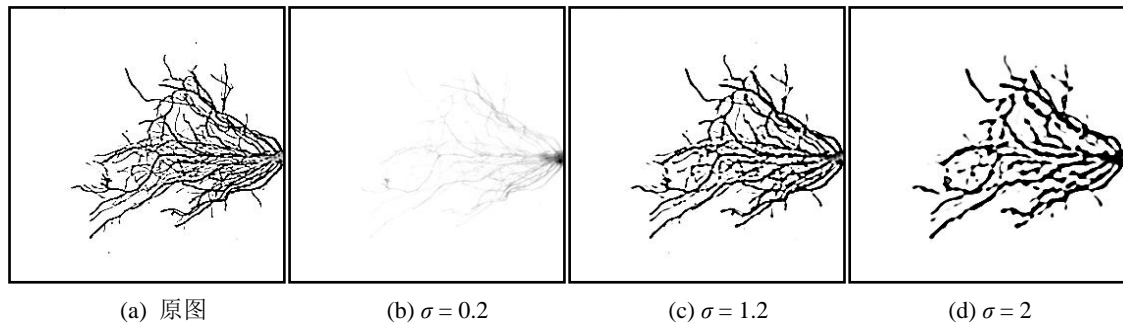
$$S_{\sigma,k,p}(x) = \frac{D_{\sigma,k,p}(x)}{\tau-1} = (1+p) \cdot G_{\sigma}(x) - p \cdot G_{k,\sigma}(x) \quad (2)$$

用 $p$ 替换 $\tau$ 使得可以通过这个变量控制边缘的锐化效果而不会对模板产生影响。 $p$ 是与像素值相关的阈值,可以更加直观的设置相应的参数值。

本文在实验过程中将公式(2)的参数 $\sigma$ 从0~2依次测试,将参数 $\varepsilon$ 从-0.1~0.08依次测试,选择效果最佳且最适合后续图像处理的参数。图1和图2展示了由一些参数的变化而引起的图像变化。

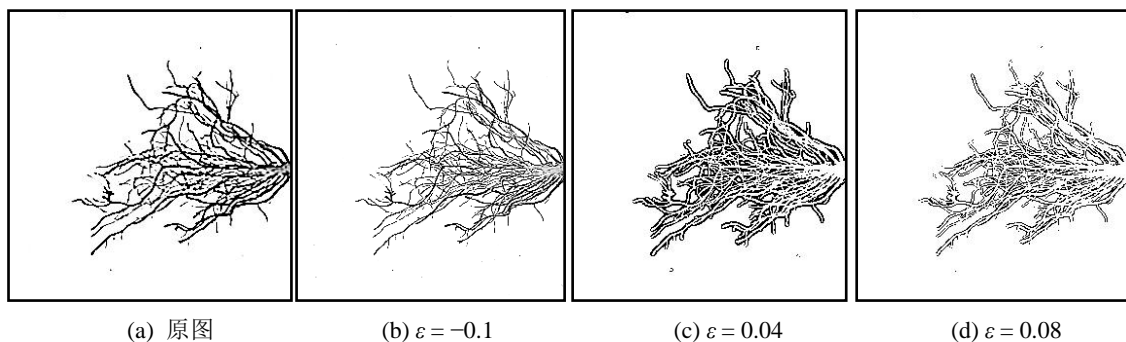
在图1中,(a)图为原图,随着 $\sigma$ 的增大,图像黑色边缘的厚度在随之增大,即图像边缘的形状结构也随之发生变化, $\sigma=0.2$ 和 $\sigma=2$ 的情况下无法准确还原原图根系的结构,当 $\sigma=1.2$ 时,不仅边缘定位准确,而且噪声抑制能力强,所以本文将 $\sigma$ 的值设置为1.2。图2展示了参数 $\varepsilon$ 的变化引起的图像变化,

通过对  $\varepsilon$  的调整可以创建不同的线条外观，根据实验需求，本文选择将  $\varepsilon$  的值设置为  $-0.1$ 。



**Figure 1.** The image change caused by the parameter  $\sigma$  from 0.2 to 2

**图 1.** 参数  $\sigma$  从 0.2 到 2 引起的图像变化



**Figure 2.** The image change caused by the parameter  $\varepsilon$  from  $-0.1$  to  $0.08$

**图 2.** 参数  $\varepsilon$  从  $-0.1$  到  $0.08$  引起的图像变化

### 3.2. 基于 HSV 色彩空间的番茄枯萎病检测

番茄枯萎病是根系病害，枯萎病先从番茄植株的根部开始发病，然后从下到上依次慢慢地侵染番茄植株的茎部和叶片。一般来说，番茄植株感染枯萎病后其根系受损，从颜色方面来看，患病严重的番茄根部颜色变深褐色，而健康番茄根部颜色理论上为白色。所以本文首先通过颜色特征来进行番茄枯萎病检测。由于样品是从基质中取出的，洗干净后的番茄根部不可避免的附着有基质，基质中的草炭是浅褐色的，所以对对照组的番茄根部为浅褐色。而患枯萎病的番茄表现症状为深褐色，在 RGB 空间难以区分，所以本文使用 HSV [16] (Hue Saturation Value) 颜色空间对番茄根系扫描图像进行分析。

HSV 颜色空间是基于颜色直观特性的，相比于面向硬件的 RGB 空间，HSV 颜色空间更面向用户。HSV 空间是由 A. R. Smith 创建的一种以人眼视觉的直观反映为依据的颜色空间，HSV 模型中的参数分别代表：色调、饱和度以及亮度。色调  $H$  是指在不同的波长的光照下，人眼感受的颜色的不同；饱和度  $S$  是指色彩的纯度，在不同种类的色彩模型中，饱和度有不同的量化模式；亮度  $V$  是指的色彩的亮度，黑色最暗，白色最亮。将颜色从 RGB 颜色空间转换为 HSV 颜色空间如公式(3)、(4)和(5)所示：

$$V = \max(R, G, B) \quad (3)$$

$$S = \begin{cases} \frac{V - \min(R, G, B)}{V}, & V \neq 0 \\ 0, & V = 0 \end{cases} \quad (4)$$

$$H = \begin{cases} \frac{60(G-B)}{V - \min(R,G,B)}, & V = r \\ 120 + \frac{60(B-R)}{V - \min(R,G,B)}, & V = g \\ 240 + \frac{60(R-G)}{V - \min(R,G,B)}, & V = b \end{cases} \quad (5)$$

公式(3)、(4)和(5)中，R、G、B 分别为红、绿、蓝分量。H、S、V 分别为色调、饱和度、亮度分量，其中 max 为 R、G、B 中的最大值，min 为 R、G、B 中的最小值。

本文所采用的检测方法主要分为两大部分：一是进行 HSV 色彩空间变换，对亮度通道进行同态滤波处理，对饱和度通道进行自适应直方图均衡处理，将处理之后的特征灰度图按照一定的权重逆转换回 RGB 色彩空间；二是对变换之后的图像进行阈值分割。

图 3 和图 4 分别展示了对照组和枯萎病番茄根系在 RGB 模型和 HSV 模型下的图像。由图 3 和图 4 的对比可知，在 RGB 模型下难以区分的颜色经过 HSV 模型的增强，特征更加突出。经过阈值分割检测出了三株番茄患有枯萎病。分析其原因为：在枯萎病病情加重时番茄根系会表现出颜色变褐色，发病初期颜色特征并不明显，所以将检测结果显示正常的 107 个番茄根系样本进行下一步的检测。

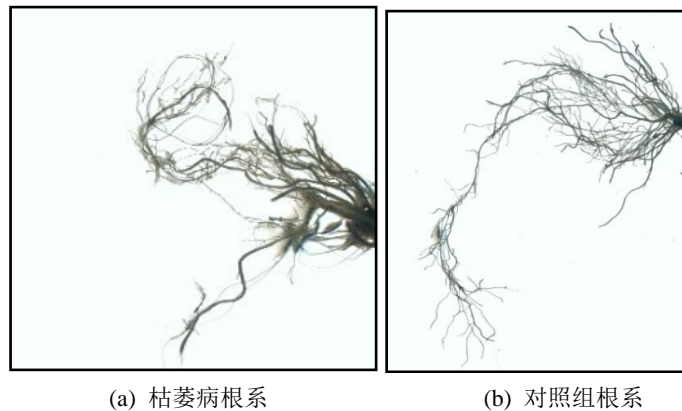


Figure 3. Root scan image under RGB model  
图 3. RGB 模型下的根系扫描图像

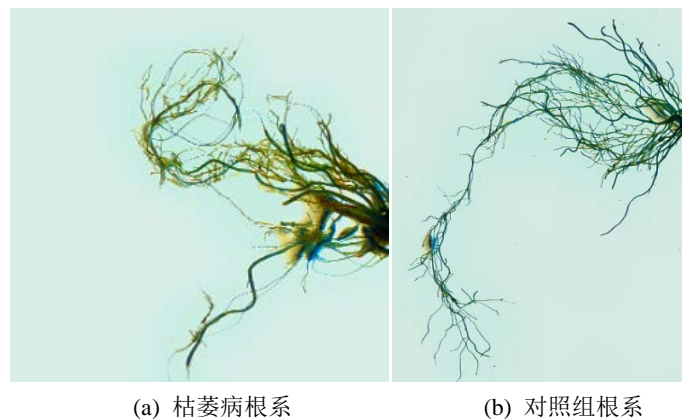


Figure 4. Root scan image under HSV model  
图 4. HSV 模型下的根系扫描图像

### 3.3. 根系参数提取

根据枯萎病发病规律和特征, 本文选择在根系扫描图像中提取像素面积、分叉数、平均长度、根系轮廓特征即最小外接矩形, 并以最小外接矩形的中心为圆心画圆, 提取圆外面积来实现枯萎病的检测。

#### 3.3.1. 根系像素面积提取

在数字图像中, 图像的面积与其所占的总像素数呈完全直线关系, 所以图像的面积可以用图像所占像素数表示。在相同的拍摄条件下, 即相同的角度和焦距等, 物体面积越大, 则在图像中所占的像素就越多。由于实验的样本都是扫描图像, 不存在拍照设备和样本距离以及角度不同造成的误差, 所以可以用像素点的个数来近似于表示面积。

#### 3.3.2. 根系分叉数和平均长度提取

分叉数顾名思义就是根系分叉点的个数, 本文采用遍历搜索的方式来计算根系交叉数。首先从图像中找到目标区域, 即确定白色像素值的位置, 将目标像素值的位置按照行递增的顺序存入矩阵中。遍历矩阵中的元素, 当某像素的右边、下边或右下位置的像素至少有一个在矩阵中时, 说明该像素为一个分叉点, 当搜索到行数大于该像素行数加一时, 停止搜索。对每一个像素都依次进行上述操作, 即可获得分叉点的个数即分叉数。

由于番茄根系结构复杂, 有的根系存在闭合, 所以本文用目标区域的像素与分叉数的比值来近似根系平均长度。

#### 3.3.3. 根系最小外接矩形提取

本文采用旋转卡壳算法来计算最小外接矩形[17]。设某凸多边形的切线为两对与  $x$  与  $y$  方向四个端点相切的线, 这四条线确定了多边形的一个外接矩形, 当多边形有一条垂直或水平的边时, 不能用这个方法处理。这种情况的解决方法是: 通过旋转线直到条件满足, 输入一个凸多边形的  $n$  个顶点后, 旋转卡壳法的步骤如下所示:

- 1) 计算多边形的四个端点, 分别为  $x_{minP}$ ,  $x_{maxP}$ ,  $y_{minP}$ ,  $y_{maxP}$ 。
- 2) 通过第一步的四个端点来构造  $P$  的四条切线。
- 3) 如果切线与一条边重合, 那四条线刚好可以确定一个矩形, 记录下此时矩形的面积, 并且将其作为最小值保存下来, 否则最小值为无穷大。
- 4) 顺时针旋转线, 直到其中一条线和多边形的一条边重合。
- 5) 计算新矩形的面积, 并且和当前最小值比较。如果小于当前最小值则更新, 并保存最小值的矩形信息。
- 6) 重复步骤 4 和步骤 5, 直到线旋转过的角度大于  $90$  度。

根据上述步骤求最小外接矩形, 每旋转一次矩形, 计算一次矩形的面积, 比较计算出来的面积, 其中面积最小的外接矩形就是通过旋转卡壳算法求得的最小外接矩形, 效果图如图 5 所示。

#### 3.3.4. 根系圆外面积提取

计算出最小外接矩形的几何中心, 然后以几何中心为中心画圆。经过反复调整半径值的大小来确定圆的大小。当圆的半径为 132 时, 效果较好, 患枯萎病番茄的根部基本上都在圆内, 对照组的根部在圆外的部分很多。即圆基本能覆盖枯萎病番茄的根部, 无法覆盖对照组的根部, 效果图如图 6 和图 7 所示。

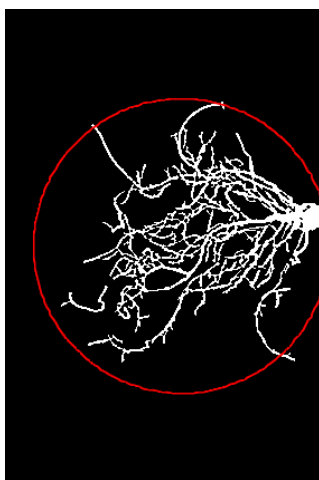
将所有样本的圆内像素都变为黑色, 计算其圆外面积。

通过根系扫描仪可得到的参数一共有 9 个, 分别为: 长度、投影面积、表面积、体积、连接数、节点数、根尖数、交叉数、分形维数。将根系扫描仪分析出来的 9 个参数和之前从图像中提取出来的 6 个

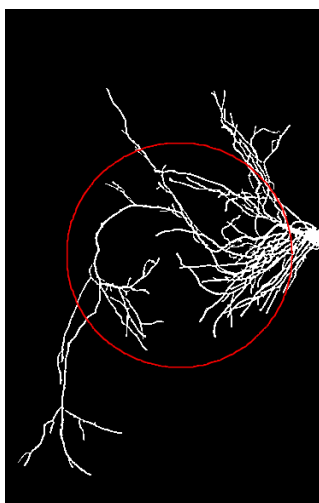
参数整合到同一个表中，这些数据将作为分类器的输入。



**Figure 5.** The smallest enclosing rectangle renderings  
**图 5.** 最小外接矩形效果图



**Figure 6.** Circling the root of tomato with Fusarium wilt  
**图 6.** 对枯萎病番茄根部画圆



**Figure 7.** Circling the root of the control group  
**图 7.** 对对照组番茄根部画圆

### 3.4. 随机森林检测模型的构建

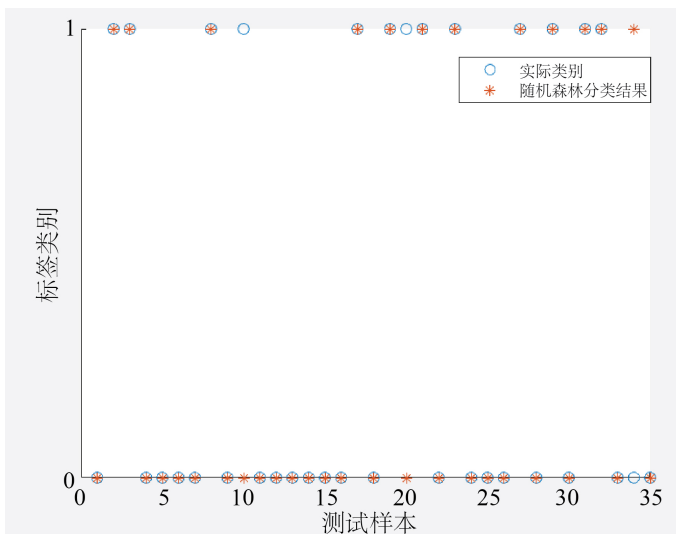
为了将患有枯萎病的番茄从样本中筛选出来，本文使用随机森林建立了番茄根系检测模型。将数据分为两类，分别是枯萎病番茄和对照组番茄，将 3.3 节提取的 15 个参数作为随机森林[18] (RF)的输入变量，分类标签作为输出变量。表 1 为分类的标签。

**Table 1.** Classification label

**表 1.** 分类标签

样本	标签
枯萎病	1
对照组	0

对训练集中的 72 条数据构建随机森林模型，图 8 为五次测试中 Test1 的随机森林算法对测试集的分类结果，图中用“o”来表示样本的实际类别，用“\*”来表示随机森林分类结果，由图 8 可以看出，有两个枯萎病样本被误判为是对照组，有一个对照组样本被误判为枯萎病。



**Figure 8.** Test sample label category

**图 8.** 测试样本标签类别

表 2 为五次测试中，使用随机森林模型进行根系检测的识别率。

**Table 2.** The accuracy of random forest model classification

**表 2.** 随机森林模型分类的准确率

测试组	识别率/%
Test1	92.11
Test2	94.74
Test3	92.11
Test4	92.11
Test5	92.11
average	<b>92.64</b>



从表 2 可以看出, 使用随机森林算法进行根系检测的五次测试的平均准确率为 92.64%, 随机森林核心算法平均运行时间为 28.5367 秒, 如果数据量足够大, 算法运行耗时将非常大, 下节将对此缺点进行改进。

### 3.5. PCA-RF 检测模型的构建

通过主成分分析[19], 将初始变量转换为 2 个主成分, 以减少维度。当累计贡献率达到要求时, 将主成分作为随机森林的输入, 将是否为患枯萎病的根作为随机森林的输出, 构建 PCA-RF 的番茄根系检测模型。该模型通过将 15 个根系特征参数转换为 2 个主成分, 并将主成分输入到随机森林来实现根系分类, 结果表明主成分分析的 2 个主成分可以更好地取代原来的 15 个特征指标的信息, 从而将 15 维信息减少到 2 维信息。

为了方便表示, 将圆外面积像素个数记为 A, 最小外接矩形面积记为 B, 最小外接矩形周长记为 C, 长度记为 L, 投影面积记为 PA, 像素面积记为 P, 表面积记为 SA, 体积记为 V, 平均直径记为 D, 连接数记为 NC, 节点数记为 N, 根尖数记为 TN, 分叉数记为 BN, 交叉数记为 CN, 分形维数记为 FD。记下累积贡献率大 85% 的特征值的序号, 计算主成分荷载和得分。最后得到的结果是选取了两个主成分, 分别为 PC1 和 PC2, 根据 PC1 和 PC2 的特征向量得到主成分数学模型。

$$PC1 = 0.96 * A + 0.92 * B + 0.92 * C + 0.97 * L + 0.95 * PA + 0.973 * P + 0.94 * SA + 0.91 * V + 0.84 * D + 0.92 * NC + 0.87 * N + 0.58 * TN + 0.90 * BN + 0.88 * CN + 0.77 * FD;$$

$$PC2 = -0.11 * A - 0.21 * B - 0.21 * C - 0.03 * L - 0.216 * PA - 0.15 * P - 0.23 * SA - 0.25 * V - 0.23 * D + 0.30 * NC + 0.44 * N + 0.70 * TN + 0.28 * BN + 0.08 * CN + 0.15 * FD;$$

$$F = 12.1 * PC1 + 1.2 * PC2 + 0.6 * PC3 + 0.5 * PC4 + 0.2 * PC5 + 0.1 * PC6 + 0.09 * PC7 + 0.08 * PC8 + 0.03 * PC9 + 0.02 * PC10 + 0.009 * PC11 + 0.006 * PC12 + 0.002 * PC13 + 0.0013 * PC14 + 0.0012 * PC15。$$

表 3 为主成分分析的特征值、贡献率和累计贡献率。

**Table 3.** Eigenvalue and contribution rate of principal component analysis  
**表 3.** 主成分分析的特征值和贡献率

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8
特征值	12.1000	1.2000	0.6000	0.5000	0.2000	0.1000	0.0900	0.0800
贡献率	80.46%	8.24%	3.94%	3.13%	1.63%	0.99%	0.59%	0.53%
累积贡献率	<b>80.46%</b>	<b>88.70%</b>	92.64%	95.77%	97.40%	98.39%	98.98%	99.51%

**Table 4.** Eigenvalue and contribution rate of principal component analysis (Continued Table)  
**表 4.** 主成分分析的特征值和贡献率(续表)

	PC 9	PC 10	PC 11	PC 12	PC 13	PC 14	PC 15
特征值	0.0300	0.0200	0.0090	0.0060	0.0020	0.0013	0.0012
贡献率	0.22%	0.13%	0.06%	0.04%	0.02%	0.01%	0.01%
累积贡献率	99.73%	99.86%	99.92%	99.96%	99.98%	99.99%	100.00%

表 3 和表 4 可以看出, 特征值 1 和特征值 2 大于 1, 两个主成分的累计贡献率达到 88.7%, 说明两个

主成分可以更好地替换原始 15 个特征指标的信息，从而将 15 维信息简化为二维，把 PC1 和 PC2 作为随机森林的输入，分类标签作为输出，构建随机森林模型。

本实验一共有 107 条数据，其中包括患枯萎病的番茄根部数据有 49 条，对照的番茄根部数据有 58 条，将这 107 条数据按照 2:1 的比例随机分成训练集和测试集，对训练集中的 69 条数据构建 PCA-RF 模型。图 9 为五次测试中 Test4 的 PCA-RF 对测试集的分类结果，图中用“o”来表示样本的实际类别，用“+”来表示 PCA-RF 分类结果，由图 9 可以看出，有一个对照组样本被误判为枯萎病。

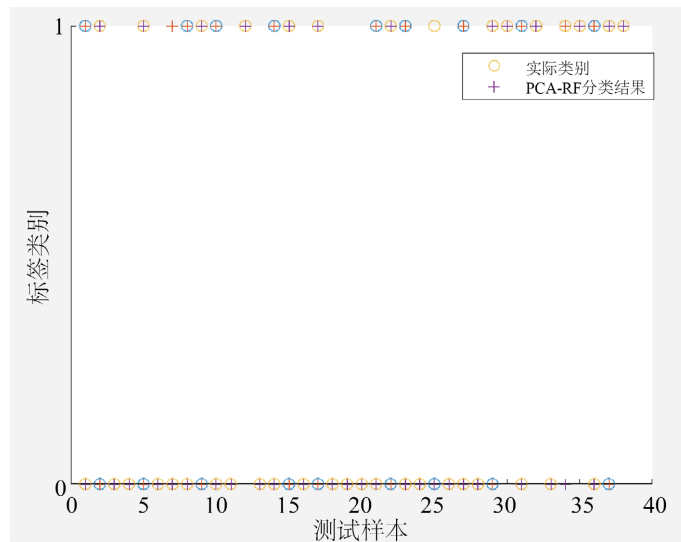


Figure 9. PCA-RF test results  
图 9. PCA-RF 测试结果

### 3.6. 结果

选取识别率和 Kappa 系数来评价模型对根系病害检测的影响。识别率是指样本正确分类的个数占样本总数的百分比；Kappa 指标主要用于检验一致性，Kappa 的计算公式如公式(6)和公式(7)所示：

$$K = \frac{P_o - P_e}{1 - P_e} \quad (6)$$

$$P_o = \frac{\sum_{i=1}^N a_i}{n} \quad (7)$$

其中  $K$  表示 Kappa 系数， $P_o$  为识别的准确率， $P_e$  为实际和预测样本的乘积之和除以样本总数的平方， $n$  为样本总数， $N$  为类数， $a_i$  为样本中  $i$  类正确分类个数。

选择识别率和 Kappa 评价 PCA-RF 模型的检测效果，并与随机森林(RF)、支持向量机[20] (SVM)和 K 近邻算法[21] (KNN)进行比较，得到表 5。

表 5 可以看出，PCA-RF 模型在所有 5 次测试中检测准确率在 94% 以上，平均准确率 94.74%，在识别率和 Kappa 系数上，PCA-RF 模型在测试集上明显优于其他三种算法。且使用随机森林算法进行根系检测的核心算法平均运行时间为 28.537 秒，使用 PCA-RF 模型进行根系检测的核心算法平均运行时间为 10.806 秒，运行时间提高了 62.13%。结果表明，主成分分析可以对根系病害参数进行敏感筛选，基于主成分的随机森林模型具有根系病害分类能力。

**Table 5.** The comparison of recognition rate and kappa in four models  
**表 5.** 四种模型的识别率和 Kappa 对比

测试组	识别率/%				Kappa			
	RF	SVM	KNN	PCA-RF	RF	SVM	KNN	PCA-RF
Test1	92.11	84.21	81.58	94.74	0.70	0.40	0.35	0.8
Test2	94.74	86.84	84.21	94.74	0.50	0.50	0.40	0.75
Test3	92.11	84.21	81.58	94.74	0.50	0.40	0.35	0.80
Test4	92.11	81.58	81.58	97.36	0.40	0.35	0.35	0.85
Test5	92.11	81.58	81.58	94.74	0.70	0.35	0.35	0.80
average	92.64	83.68	82.11	<b>95.26</b>	0.56	0.40	0.36	<b>0.80</b>

#### 4. 结论与讨论

番茄根系病害的快速检测对番茄及时诊断和防治具有重要意义。本文用 PCA-RF 模型实现病害根部的分类, 从中检测出枯萎病样本。与随机森林模型相比, PCA-RF 模型的平均识别率提高了 2.62%, 核心算法平均运行时间提高了 62.13%。将 PCA-RF 模型与支持向量机和 K 近邻算法作对比, 在识别率和 Kappa 系数上, PCA-RF 模型在测试集明显优于其他算法。本研究证明, PCA-RF 模型简化了复杂的问题, 继承了随机森林的优势, 具有很强的泛化能力和鲁棒性, 可以作为检测的有效解决方案。由于样本数量有限, 本文采用了机器学习的方法将病害类型识别分类, 在今后的工作中, 应该增大番茄病害样本引入深度学习机制。

#### 基金项目

- 1) 河北省科技厅农业节水科技创新专项(项目编号: 21326903D);
- 2) 鲜食型口感番茄绿色生产与品质提升关键技术研究示范(项目编号: 20326901D)。

#### 参考文献

- [1] 金博诣. 高山番茄优质高产栽培技术[J]. 农业开发与装备, 2021(10): 219-220.
- [2] McGovern, R.J. (2015) Management of Tomato Diseases Caused by *Fusarium oxysporum*. *Crop Protection*, **3**, 78-92. <https://doi.org/10.1016/j.cropro.2015.02.021>
- [3] 王云霞. 番茄枯萎病的发生原因及综合防治[J]. 农业技术与装备, 2019(6): 89-90.
- [4] 邓蕊. 植物根系图像的边缘检测方法的研究与实现[D]: [硕士学位论文]. 重庆: 重庆大学, 2015.
- [5] Sakaguchi, R., Morton, K.D., Collins, L.M. and Torriane, P. (2017) A Comparison of Feature Representations for Explosive Threat Detection in Ground Penetrating Radar Data. *IEEE Transactions on Geoscience & Remote Sensing*, **55**, 6736-6745.
- [6] Reichman, D., Collins, L. and Malof, J. (2018) On Choosing Training and Testing Data for Supervised Algorithms in Ground-Penetrating Radar Data for Buried Threat Detection. *IEEE Transactions on Geoscience and Remote Sensing*, **56**, 497-507.
- [7] Pierret, A., Moran, C.J., Mclachlan, C.B. and Kirby, J.M. (2000) Measurement of Root Length Density in Intact Samples Using X-Radiography and Image Analysis. *Image Analysis & Stereology*, **19**, 145-149. <https://doi.org/10.5566/ias.v19.p145-149>
- [8] Galkovskiy, T., Mileyko, Y., Bucksch, A., Moore, B., Symonova, O., Price, C.A., et al. (2012) GiA Roots: Software for the High Throughput Analysis of Plant Root System Architecture. *BMC Plant Biology*, **12**, Article No. 116. <https://doi.org/10.1186/1471-2229-12-116>
- [9] 赵来宾. 三维植物根系扫描系统的研究与实现[D]: [硕士学位论文]. 武汉: 华中科技大学, 2012.
- [10] 胡秀娟. 植物根系形态参数计算机视觉检测系统的研究[D]: [硕士学位论文]. 哈尔滨: 东北林业大学, 2004.

- 
- [11] 杨国梁. 计算机视觉在根系形态分析中的应用研究[D]: [硕士学位论文]. 北京: 首都师范大学, 2006.
- [12] 彭威勋. 棚室番茄枯萎病发生原因及综合防治措施探讨[J]. 现代农业研究, 2019(12): 50-51.
- [13] 曹俊库. 棚室番茄枯萎病发生原因及综合防治措施探讨[J]. 农业与技术, 2017, 37(12): 18.
- [14] 邓光道. 粉果圣女小番茄青枯病与枯萎病的症状区别及防治方法[J]. 现代农业科技, 2018(5): 115+117.
- [15] Winnemöller, H., Kyprianidis, J.E. and Sven, C. (2012) XDoG: An Extended Difference-of-Gaussians Compendium Including Advanced Image Stylization. *Computers & Graphics*, **36**, 740-753 <https://doi.org/10.1016/j.cag.2012.03.004>
- [16] 吴俊涛, 曹开颜, 付明慧, 曹雨萌. HSV 色彩空间和 OTSU 阈值分割的盲道识别算法设计[J]. 电子世界, 2021(12): 101-103.
- [17] 黄超, 茅健, 徐斌, 周玉凤. 基于最小外接矩形和 Hough 变换的定位算法[J]. 组合机床与自动化加工技术, 2021(8): 66-71.
- [18] Ebrahim, S., Wassu, M. and Tessfaye, A. (2021) Genetic Diversity based on Cluster and Principal Component Analyses in Potato (*Solanum tuberosum* L.) for Yield and Processing Attributes. *Journal of Horticulture*, **8**, Article No. 298.
- [19] 窦世卿, 宋莹莹, 徐勇, 苗林林, 陈治宇, 郑贺刚. 基于随机森林的高分影像分类及土地利用变化检测[J]. 无线电工程, 2021, 51(9): 901-908.
- [20] 顾天宇, 严壮志, 蒋皆恢. 基于支持向量机的中风病中医证候分类[J]. 中医药信息, 2021, 38(9): 1-3.
- [21] 屠菁, 刘登胜, 钟雪景. KNN 分类算法在停车场车牌识别系统中的应用[J]. 淮南师范学院学报, 2021, 23(2): 141-148.