doi:10.6041/j.issn.1000-1298.2025.12.043

基于近红外透射光谱与卷积神经网络的蜜柚枯水病无损检测方法

陈俊杰[1,2]　朱文杰[1,2]　赵泽华[2]　李佳琪[2]　王庆艳[2,3]　黄文倩[2,3]

(1. 江苏大学农业工程学院, 镇江 212013; 2. 北京市农林科学院智能装备技术研究中心, 北京 100097;
3. 国家农业智能装备工程技术研究中心, 北京 100097)

摘要: 枯水病是蜜柚采后的常见病害,枯水病出现时会影响蜜柚果实口感,甚至使其失去食用价值。作为一种大型厚皮水果,难以通过肉眼及果实外表判断蜜柚内部是否存在枯水病害。本研究基于近红外透射光谱检测方法结合一维卷积神经网络(One-dimensional convolutional neural networks, 1D-CNN)建立蜜柚枯水病快速无损检测模型。采用实验室自主研发的可见-近红外全透射光谱测量系统,以红肉蜜柚为研究对象,采用短积分模式采集样本的近红外透射光谱波长范围(562.03~1 110.16 nm)数据进行分析,在没有光谱数据预处理情况下采用 Kennard-Stones 算法将样本划分训练集和测试集,随后建立一维卷积神经网络蜜柚枯水判别模型,经测试,模型对于枯水程度分类准确率达到 98.15%。与采用原始光谱数据与经多项式平滑处理(Savitzky-Golay, SG)和使用竞争性自适应重加权算法(Competitive adapative reweighted sampling, CARS)筛选特征变量后的反向传播神经网络(Backpropagation neural network, BPNN)、核极限学习机(Kernel extreme learning machine, K-ELM)、极限梯度提升树(Extreme gradient boosting, XGBoost)等传统判别模型的判别结果进行对比,最后使用数据增强方法对原始光谱数据进行扩充,以验证卷积神经网络模型稳定性。结果表明,本文提出的卷积神经网络模型在不使用复杂预处理以及特征筛选方法的情况下,对蜜柚内部枯水以及枯水程度的判别效果最佳,且能够在光谱处理过程选择有效信息,极大地提升了模型判别效率,对实现蜜柚内部枯水程度快速、准确地无损判别具有借鉴意义。

关键词: 无损检测; 红肉蜜柚; 枯水病; 一维卷积神经网络; 近红外光谱

中图分类号: O657.33　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)12-0470-09　　**OSID**:

Non-destructive Detection of Pomelo Granulation Disease Based on Near-infrared Transmission Spectroscopy and Convolutional Neural Networks

CHEN Junjie[1,2]　ZHU Wenjie[1,2]　ZHAO Zehua[2]　LI Jiaqi[2]　WANG Qingyan[2,3]　HUANG Wenqian[2,3]

(1. *School of Agricultural Engineering*, *Jiangsu University*, *Zhenjiang* 212013, *China*
2. *Intelligent Equipment Research Center*, *Beijing Academy of Agriculture and Forestry Science*, *Beijing* 100097, *China*
3. *National Research Center of Intelligent Equipment for Agriculture*, *Beijing* 100097, *China*)

Abstract: Granulation is a common post-harvest disease of citrus fruits, particularly in pomelos. When this disease occurs, it will affect the taste of pomelo fruits and even make them lose their edible value. Pomelos are large, thick-skinned fruits which are difficult to visually detect their internal granulation. The rapid, non-destructive detection of granulation of pomelos was explored by using near-infrared transmission spectroscopy combined with one-dimensional convolutional neural networks. A self-developed visible-near-infrared full-transmission spectroscopy measurement system was used to collect data from red-flashed pomelos. Data were collected in the 562.03~1 110.16 nm range under short integration mode. Without spectral preprocessing, the Kennard-Stones algorithm divided samples into training and testing sets. Subsequently, a one-dimensional convolutional neural network model for pomelo dryness identification was established. The network had a seven-layer structure. Testing showed that the model achieved a classification accuracy of 98.15% for granulation levels. Compared with traditional

收稿日期: 2024-08-20　修回日期: 2024-10-23

基金项目: 国家重点研发计划项目(2022YFD2002202)和国家西甜瓜产业技术体系项目(CARS-25-2024-G23)

作者简介: 陈俊杰(2000—),男,硕士生,主要从事水果品质无损检测研究,E-mail: cjunjie09@163.com

通信作者: 黄文倩(1980—),男,研究员,博士生导师,主要从事农产品品质无损检测技术与装备研究,E-mail: huangwq@nercita.org.cn

models such as backpropagation neural network (BPNN), kernel extreme learning machine (K-ELM), and extreme gradient boosting tree (XGBoost), which used raw spectral data and polynomial smoothing processing (SG) and competitive adaptive reweighted sampling (CARS) to extract feature variables. Finally, data augmentation further validated the model's stability. The proposed convolutional neural network model effectively identified internal granulation without complex preprocessing. It can select effective information during the spectral processing, greatly improving the model identification efficiency. The researh provided a reference for the application of deep learning methods to achieve rapid and accurate non-destructive identification of the internal granulation level of pomelo.

Key words: non-destructive detection; red-fleshed honey pomelo; granulation; one-dimension convolutional neural network; near-infrared spectroscopy

0　引言

柚子是中国的传统水果,为柑橘类水果,营养价值较高。枯水是柑橘类果实常见的一种生理性病害,表现为果肉汁胞出现不规则膨大、纤维化、木栓化、失水变硬、干枯,颜色由深变浅,失去原有口感,严重影响果实商品性[1]。柑橘类水果发生汁胞粒化的原因之一是汁液囊泡中次生壁明显发育,细胞壁显著增厚。收获时间也对蜜柚果实内部品质有重要影响,收获时间较晚可能会加重蜜柚的汁胞粒化程度[2-3]。作为一种复杂的生理现象,目前对蜜柚枯水相关研究比较有限。传统蜜柚枯水判别主要是靠人工筛查,误判率高且依赖经验。探索一种无损、快速的蜜柚枯水病检测方法,可大幅降低人力物力,蜜柚枯水病的研究对提高果农收益和促进我国水果产业优化发展具有重要意义。

近红外光谱技术已被广泛应用于水果品质内部检测,其可对小型薄皮水果的可溶性固形物含量[4]、水果成熟度[5]或水果一些常见病害如柑橘黄龙病[6]、苹果霉心病[7]等进行检测。然而在大型厚皮水果检测应用上仍有较多问题未攻克。蜜柚果实体积大、皮厚,且有着较厚的海绵层,使得获取中心部位果肉信息的透射光谱变得非常困难。国内外对相关水果品质检测主要应用如机器视觉、近红外光谱、高光谱、X 射线等技术[8-10],PUANGSOMBUT 等[11]采用可见近红外光谱在透射模式下探讨了检测鲜切柚子可溶性固形物含量和总酸度的可能性,并采用每旋转 45°的平均光谱建立了预测模型,可溶性固形物含量预测集相关系数(Correlation coefficient of prediction set, R_p)为 0.841,均方根误差(Root mean square error of prediction, RMSEP)为 0.508%,总酸度 R_p 为 0.693,RMSEP 为 0.084%,但实际上这种方法已属有损检测,不能实现无损预测。LI 等[12]采用近红外光谱技术结合模式识别方法对 4 个品种柚子进行判别,准确率达到 97.92%。LIU 等[13]采集蜜柚图像,对树上的柚果进行成熟度判别,准确率达到 93.5%。王贤达等[14]发现采用核磁共振成像技术发现可以对蜜柚汁胞粒化进行有效检测。刘燕德等[15]采用近红外高光谱实现了对脐橙的粒化检测,发现使用无信息变量消除法(Uniformative variable elimination, UVE)后的最小二乘支持向量机(Least squares support vector machine, LS-SVM)模型效果最佳,对粒化程度误判率最低为 0.78%。YU 等[16]采用堆栈自编码机(Stacked auto-encoders, SAE)处理梨的可见/近红外高光谱图像(380~1 030 nm)并结合全连接神经网络(Full connect neural network, FNN)实现了果实可溶性固形物含量预测并且取得了较好的效果,决定系数为 0.921。TIAN 等[17]提出了将直径矫正方法(Diameter correction method, DCM)与 1D-CNN 相结合用于柑橘早期冻害的识别,早期冻害橙子与未冻橙子最优模型召回率分别为 95.15% 和 88.54%。上述柚子内外品质检测方法相关研究中,关于近红外光谱技术结合深度学习方法实现蜜柚枯水病无损检测相关文献较有限。

针对蜜柚枯水需依赖人工判别,以及常规检测方法多采用长积分时间和较强光源[18-21],难以获取蜜柚有效光谱信号并用于在线检测等问题,本文采用自主研发的水果全透射无损检测在线检测设备,研制核心光谱仪,通过优化光路结构,显著提升通光量和灵敏度,大幅缩短积分时间,确保在单个 150 W 功率灯杯照明和短积分时间 10 ms 条件下能够获得有效的光谱信号,以红肉蜜柚为研究对象,探究利用近红外光谱技术结合深度学习方法判别蜜柚枯水病可行性,建立不同判别模型并确定最优模型。

1　试验材料与处理方法

1.1　试验样本

样本选自福建省平和县李添元家庭农场的红肉蜜柚以及库存蜜柚,大小均匀且果质量为 1.1~1.9 kg。试验前,将所有样本置于室内阴暗通风环境 24 h,放置至室温以消除温度对检测结果的影

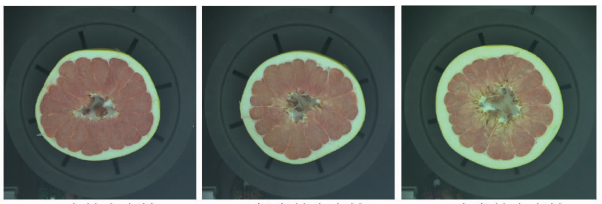
响,然后剔除有明显缺陷的样本。图 1 为部分蜜柚样本。采集光谱信息前对每个蜜柚进行编号,采集完光谱信息后,将蜜柚切开,观察枯水情况。邀请 7 位感官评价人员对不同枯水程度的样本进行打分,根据样本中蜜柚枯水程度不同,将样本分为 4 类,其中发生枯水面积为 0 为未枯水蜜柚记为 1 分,发生枯水面积在 0 ~ 25% 为轻度枯水蜜柚记为 2 分,发生枯水面积在 25% ~ 50% 为中度枯水蜜柚记为 3 分,发生枯水面积在 50% 以上为重度枯水记为 4 分[22]。由于重度枯水蜜柚枯水面积较大失去了食用价值,且能够通过外观判别,故不进行研究。



图 1 部分蜜柚样本

Fig. 1 Some samples of pomelos

图 2 为蜜柚横切剖面图,成熟的红肉蜜柚果实一般从囊瓣近蒂端汁胞发生枯水,汁胞异常膨大,变硬,木质化。红肉蜜柚正常汁胞显得饱满多汁,且颜色为均匀的红色,汁胞之间紧密结合,难以分离;处于枯水状态的汁胞相对松散,颜色稍浅,呈浅红色,汁胞一端呈现黄色,汁胞较硬,汁胞壁增厚,可见红肉蜜柚果实发生枯水时,汁胞形态结构发生了较明显变化[23]。



(a) 未枯水蜜柚 (b) 轻度枯水蜜柚 (c) 中度枯水蜜柚

图 2 蜜柚横切剖面图

Fig. 2 Cross sectional views of pomelo

1.2 近红外光谱数据采集及光谱处理

1.2.1 仪器设备

所涉及主要试验仪器包括北京市农林科学院自主研发的可见-近红外全透射光谱测量系统、图像采集系统。自主研发的短积分全透射光谱在线检测装置如图 3 所示,此设备由光源卤素灯(FUJI, JCR, 150 W, 15 V)、高灵敏度光谱仪(Online NIR711 型,采集样本光谱范围 562.03 ~ 1 110.16 nm,积分时间为10 ms,光谱间隔 0.5 nm)、传送带、光电传感器和工业计算机(Advantech610H, Intel(R) Core(TM) i5 - 2400 CPU@ 3.10 GHz, RAM 8 GB)等组成。光源和光谱仪分别呈直线式分布于样本两侧,光谱仪采集样本全透射光谱信号,以便消除水果镜面反射干扰。水果传送通道采用双层遮光帘以避免外界环境光的干扰。

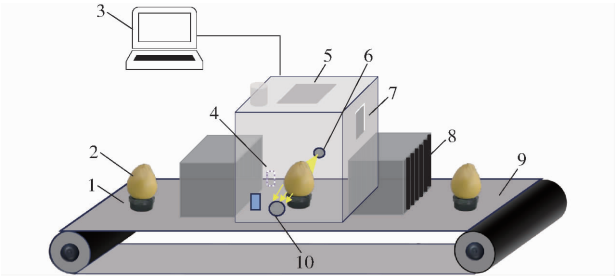


图 3 全透射近红外光谱在线检测设备示意图

Fig. 3 Schematic of online detection equipment for full transmission near-infrared spectroscopy

1. 自由果托 2. 蜜柚样本 3. 工业计算机 4. 光电传感器 5. 盖板 6. 光源 7. 风扇 8. 遮光帘 9. 传送带 10. 光谱仪

1.2.2 光谱数据采集

试验开始前,启动近红外光谱透射检测系统,等待 10 min 至其能够稳定运行。对蜜柚光谱采集时,试验人员将样本放置传送带上,传送带输送蜜柚以恒定速度 0.5 m/s 移动。当被测样本前端到达光电传感器位置时,系统会接收到 1 个触发信号,结合编码器检测到的移动速度,可准确连续地采集被测样本的透射光谱数据,如图 4 所示,蜜柚此摆放状态下光谱数据采集。为了保证样本在光谱采集过程的稳定移动,将样本放置在黑色自由果托上,每旋转 90° 进行一次光谱数据采集,最终取光谱平均值作为当前样本的光谱数据。

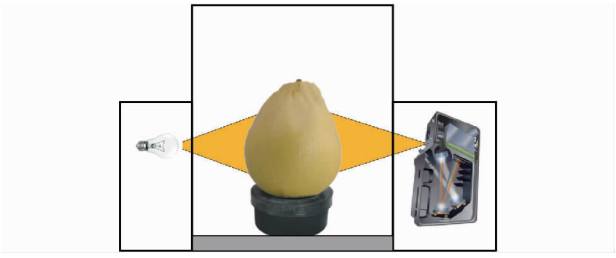


图 4 蜜柚摆放状态

Fig. 4 Pomelo placement status

1.2.3 光谱数据处理

光谱数据处理均基于 Tensorflow 和 Keras 框架使用 Python 3.11 进行编写,同时为了对模型性能进行评估并比较不同模型之间性能,系统配置了 Pandas、Numpy、Scikit-learn、Scipy、Matplotlib 等第三方软件库更直观地表现模型分类结果以及对模型效果进行优化。

1.3 近红外透射光谱数据分析

对采集的波长为 562.03 ~ 1 110.16 nm 光谱数据共 2 048 个波段进行分析,图 5 为所有红肉蜜

柚样品原始光谱曲线。从图 5 可知，光谱曲线在 928 nm 附近出现了波峰且覆盖波段范围较大，很可能是因为果实内部水分和碳水化合物 C—H、O—H 和 N—H 化学键，不同物质的官能团特征谱带重叠形成。图 6 为未枯水蜜柚、轻度枯水蜜柚、中度枯水蜜柚平均光谱曲线，可以看出不同枯水程度蜜柚光谱曲线变化趋势大体一致，在近红外光谱区域（750 ~ 1 050 nm），此处光谱曲线的差距较大，说明不同枯水程度蜜柚在此处成分含量存在差异，重叠吸收峰出现在约 928 nm 处，960 nm 附近吸收峰为水分吸收带[24]，枯水过程中果汁囊胞损失了更多水分，所以枯水样本吸收峰均低于健康样本。

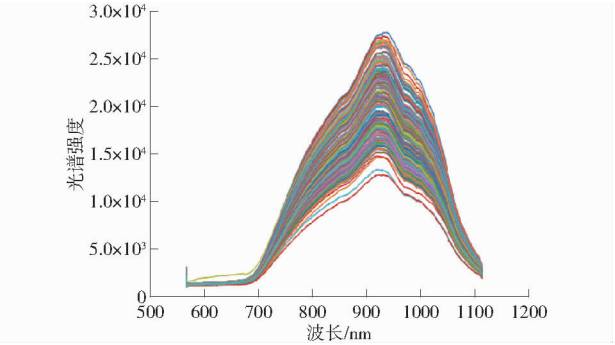


图 5　蜜柚光谱曲线

Fig. 5　Spectral curves of pomelos

1.4　光谱数据建模方法

1.4.1　一维卷积神经网络模型

卷积神经网络是一种基于前馈神经网络的深度学习模型，现已广泛应用于图像识别、计算机视觉和语音识别等领域。在使用卷积神经网络对图像进行处理时选取的图像多为二维和三维数据。由于近红外光谱数据包含波段中物质化学成分和含量相关信息的特点，所以可将连续的近红外光谱曲线看作一维方向上相关联的像素点，对一维光谱数据进行卷积处理，通过一维卷积神经网络（1D－CNN）实现光谱特征、光谱有效信息的提取以及非线性光谱数据的处理。1D－CNN 网络结构如图 7 所示，包括 1 个输入层、2 个卷积层、1 个批归一化层、1 个激活层、2 个最大池化层、1 个全连接层和 1 个输出层。1D－CNN 结构描述如下：

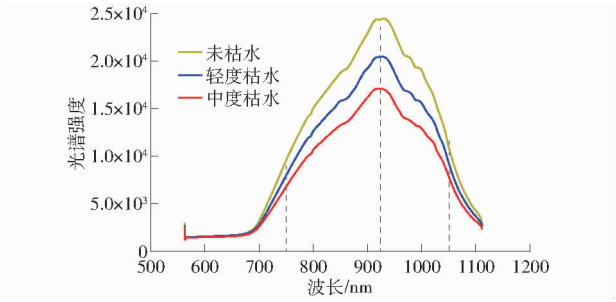


图 6　蜜柚平均光谱曲线

Fig. 6　Average original spectral curves of pomelos

（1）输入层：由于近红外光谱为一维数据，所以将光谱矩阵数据宽度和通道数均设置为 1，数据以 2 048 × 1 矩阵形式输入。

（2）卷积层 1：卷积层由多个卷积核构成，对输入数据进行卷积操作，从而提取目标更深层次特征，卷积运算公式为

$$X_i^l = f\left(\sum_{i \in M_j} X_i^{l-1} \boldsymbol{W}_{ji}^l + b_j \right) \tag{1}$$

式中　f——激活函数

l——输入长度

X_i^{l-1}——输入数据待卷积区域

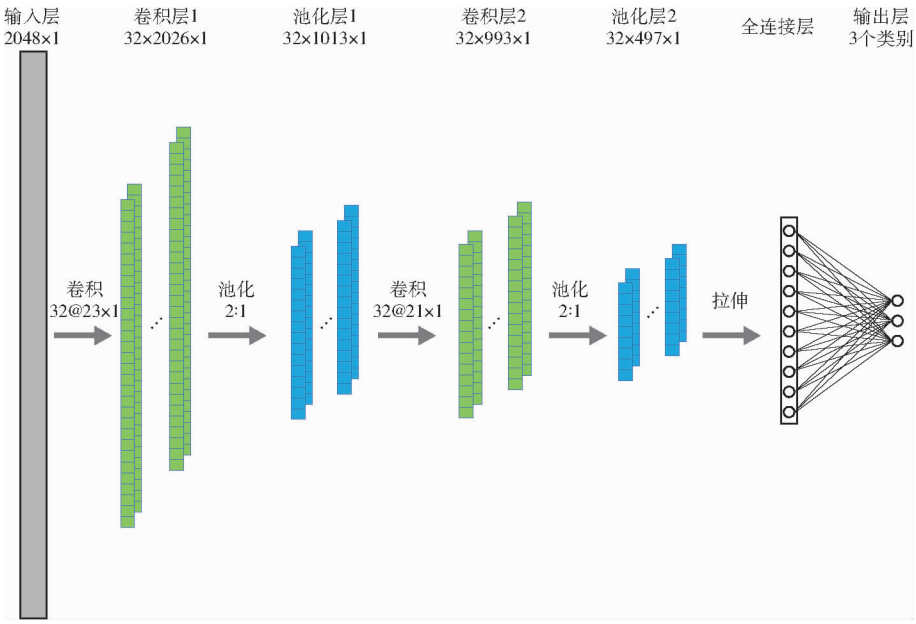


图 7　1D－CNN 网络结构图

Fig. 7　Structure of one-dimensional convolutional neural network

$\boldsymbol{W}_{ji}^{l}$——卷积核矩阵

b_j——卷积核偏置系数

采用 32 个 25 × 1 大小的卷积核构建第 1 个卷积层进行数据特征的提取。

(3)批归一化层:使用批归一化层对神经网络训练进行加速,并提升网络对输入数据的适应性和泛化能力,通过对每个 mini-batch 进行归一化操作使输入数据分布更加稳定。

(4)最大池化层 1:池化层负责对卷积层的输出进行下采样,将光谱数据各区域代表特征提取出来,从而降低卷积层输出的特征维度,有助于减少模型参数量和计算复杂度,同时又保留重要特征信息,最大池化层对信号进行下采样的计算公式为

$$f(X_k) = \max_{a \in X_k} \{a_1, a_2, \cdots, a_s\} \tag{2}$$

将卷积层运算所得到的特征划分成多个不重叠区域 $X_k(k = 1, 2, \cdots, K)$,区域大小为 $s \times 1$。将池化层窗口定位 2 × 1,使用 2 × 1 步长。

(5)卷积层 2:使用 32 个 23 × 1 大小的卷积核构建第 2 个卷积层。

(6)最大池化层 2:采用窗口大小为 2 × 1、步长为 2 × 1 对输出特征继续降维。

(7)全连接层:全连接层对数据进行全局特征分析,全连接层节点数设置为 128。

(8)输出层:输出层由 softmax 层和分类层组成,前者将神经网络的原始输出转换为概率分布,使类别的预测概率为 0 ~ 1 之间,后者计算分类损失并度量分类精度,最终得到分类精确率。

1D - CNN 模型主要参数如表 1 所示。

表 1 1D - CNN 模型参数

Tab. 1 Parameters of 1D - CNN model

网络层	参数及数值	激活函数
输入层	原始近红外透射光谱数据	
卷积层 1	卷积核数量:32,卷积核大小:25	Leakey ReLU
最大池化层 1	池化窗口:2	
卷积层 2	卷积核数量:32,卷积核大小:23	Leakey ReLU
最大池化层 2	池化窗口:2	
全连接层	神经元个数:128	Leakey ReLU
输出层	3 个类别	

1.4.2 反向传播神经网络模型

反向传播神经网络(BPNN)具有很好的非线性映射能力,由多层感知器(Multilayer Perceptro, MLP)演变而来,通过前向传递过程将数据输入网络,数据依次通过隐含层与输出层并进行相关计算[25]。

1.4.3 极限学习机模型

核极限学习机模型(K - ELM)是由极限学习机(Extreme learning machine, ELM)改进的一种机器学习模型,ELM 是一种在单隐层前向神经网络(Single-hidden layer feed forward network, SLFN)基础上的算法,能够随机分配隐藏层并分析确定 SLFN 输出权重。由于 ELM 在训练前随机选择参数,难以保证稳定性和持续性,所以引入核函数形成 K - ELM 算法,无需使用随机参数,能够有效增加模型的鲁棒性和稳定性[26]。

1.4.4 XGBoost 模型

XGBoost 模型是一种高效且强大的梯度提升树(Gradient boosting tree, GBT)算法,广泛用于机器学习和数据挖掘任务。采用梯度提升方法可减少模型的预测误差,逐步改善模型性能,并且能够在训练过程自动处理缺失值,在一些回归和分类问题上取得了较好效果[27]。

1.4.5 模型评估标准

为评估 1D - CNN 枯水病定性判别效果,使用 BPNN、K - ELM、XGBoost 等分类方法分别建立枯水病蜜柚判别模型与 1D - CNN 模型效果进行对比。以近红外光谱数据作为模型输入数据,类别标签作为模型输出数据。

对于蜜柚枯水定性判别模型,基于训练集数据训练判别模型,使用测试集数据对模型性能进行评估,采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数(F1score)作为模型评价指标。

2 模型建立与结果

2.1 样本集划分

共筛选出蜜柚样本 180 个,其中未枯水蜜柚为 60 个且类别标签为 0,轻度枯水蜜柚为 60 个且类别标签为 1,中度枯水蜜柚为 60 个且类别标签为 2。利用 Kennard - Stones 算法将样本划分为训练集和测试集,其中训练集为样品总数的 70% 并用于神经网络的训练,30% 作为测试集用于测试训练后的模型性能。训练集共有样本 126 个,其中未枯水蜜柚、轻度枯水蜜柚、中度枯水蜜柚均为 42 个;测试集共有样本 54 个,其中未枯水蜜柚、轻度枯水蜜柚、中度枯水蜜柚均为 18 个。

2.2 基于卷积神经网络模型的训练与结果

1D - CNN 训练采用 Adam 优化器(Adaptive moment estimation, Adam)。Adam 算法为一阶优化算法,可替代传统的随机梯度下降(Stochastic gradient descent, SGD)方法,能够在训练过程中根据数据逐步调整网络权重。对于激活函数,选择 Leakey ReLU 作为 1D - CNN 模型激活函数,可有效

避免梯度消失。模型训练参数设置最大迭代次数为200，初始化学习率(Learning rate)为0.000 01，学习率下降因子为0.1。为了使该模型能够更好地收敛，样本被分成不同批次(batch)进行训练，每个批次大小(batch size)设置为80。为了防止模型过拟合，在第2个卷积层、池化层以及全连接层后加入Dropout层，并设置为0.5。为了提升模型泛化能力，在每次训练前将样本进行打乱。将蜜柚光谱数据输入至1D－CNN模型进行训练，训练过程中训练集和测试集的损失值变化曲线如图8所示，训练集及测试集损失值随着迭代次数增加均稳定下降。模型训练后在测试集进行测试，结果如图9所示，可观察到只有中度枯水样本中有1个被误识别为轻度枯水。

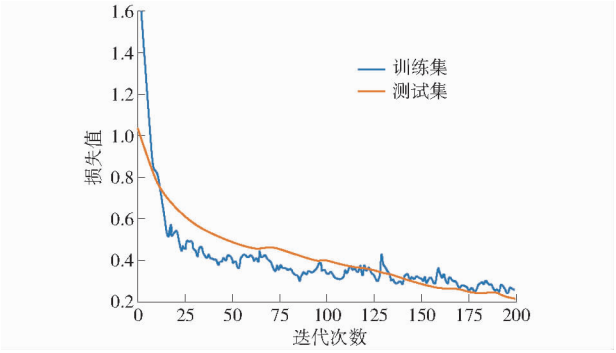


图8　训练过程训练损失变化曲线

Fig. 8　Changing trend of training loss in training process

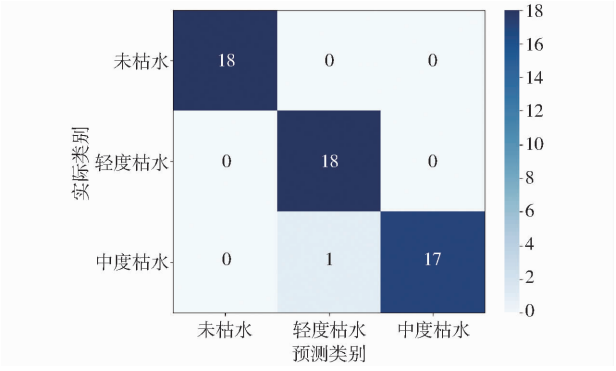


图9　1D－CNN模型混淆矩阵

Fig. 9　Confusion matrix of 1D－CNN model

表2为1D－CNN模型对蜜柚枯水程度判别的性能结果。由表2可以看出，测试集中对未枯水样本判别完全正确，精确率、召回率和F1分数均达到100%，而对于轻度枯水和中度枯水样本，测试集精确率、召回率和F1分数分别为100%、88.89%、94.12%和90%、100%、94.74%，由此可见1D－CNN模型对蜜柚枯水程度区分时表现较优。

表2　1D－CNN模型性能结果

Tab. 2　Performance of 1D－CNN model　%

枯水程度	准确率	精确率	召回率	F1分数
未枯水	100	100	100	100
轻度枯水	88.89	100	88.89	94.12
中度枯水	100	90.00	100	94.74

2.3　不同判别模型判别效果

采用1D－CNN、BPNN、K－ELM、XGBoost建立蜜柚枯水判别模型并进行对比，模型建立过程中训练集占比为70%，测试集占比为30%。由于光谱数据特征的复杂性，需要对传统分类算法进行预处理以及光谱特征提取，因此需要剔除光谱数据中非目标信息，过滤光谱噪声并进行特征变量选择后对比不同模型定性判别效果。

将经过SG处理和CARS特征提取后的光谱数据输入到BPNN、K－ELM、XGBoost模型，CARS共选择350条波长，测试集判别结果如图10所示。由图10可看出，经SG－CARS处理后的光谱数据在BPNN模型中有4个轻度枯水样本被误判为中度枯水，未枯水和中度枯水样本判别完全正确。K－ELM模型和XGBoost模型则共出现3个误判样本，相比之下K－ELM和XGBoost模型误判个数较少。

为了进一步比较不同判别模型判别效果，根据混淆矩阵判别结果对SG－CARS－BPNN、SG－CARS－K－ELM、SG－CARS－XGBoost模型性能指标进行计算，判别结果如表3所示。由表3可以看出，在未枯水蜜柚判别效果上，SG－CARS－BPNN和SG－CARS－K－ELM模型均有较好表现，测试集

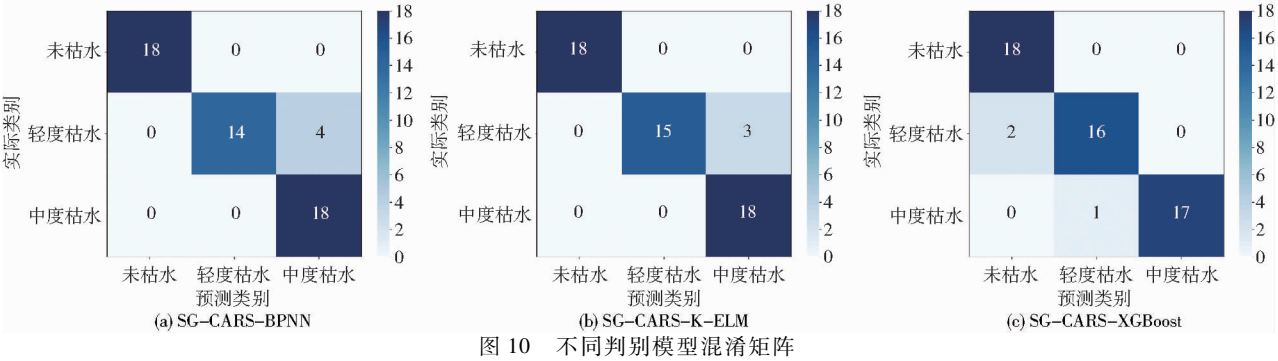


图10　不同判别模型混淆矩阵

Fig. 10　Confusion matrix of different models

表 3 不同判别模型性能对比

Tab. 3 Performance of different models %

分类模型	枯水程度	准确率	精确率	召回率	F1 分数
SG－CARS－BPNN	未枯水	100	100	100	100
	轻度枯水	77.78	100	77.78	87.50
	中度枯水	100	81.82	100	90.00
SG－CARS－K－ELM	未枯水	100	100	100	100
	轻度枯水	83.33	100	83.33	90.91
	中度枯水	100	85.71	100	92.31
SG－CARS－XGBoost	未枯水	100	90.00	100	94.74
	轻度枯水	88.89	94.12	88.89	91.43
	中度枯水	94.44	100	94.44	97.14

精确率、召回率和 F1 分数均为 100%，对于轻度枯水判别，SG－CARS－XGBoost 表现最优，测试集的精确率、召回率和 F1 分数为 94.12%、88.89%、91.43%。对于中度枯水，SG－CARS－K－ELM 和 SG－CARS－BPNN 表现相当，而 SG－CARS－XGBoost 模型表现最差，精确率、召回率和 F1 分数为 100%、94.44%、97.14%。综合来看 K－ELM 和 XGBoost 模型在经过 SG 和 CARS 处理后在枯水 3 种类别判别上具有较优表现。

将 4 种分类算法总体准确率进行比较，结果如表 4 所示。从表 4 可以看出，使用原始光谱数据作为输入，1D－CNN 模型对蜜柚枯水程度的判别效果最佳，准确率达到 98.15%，经过 SG 和 CARS 特征提取后作为输入 BPNN、K－ELM 和 XGBoost 模型准确率都有一定提升，BPNN 模型由 83.33% 提升至 92.59%，K－ELM 模型由 92.59% 提升至 94.44%，XGBoost 模型由 90.74% 提升至 94.44%。由此可见，原始近红外光谱数据在经过平滑处理并进行数据特征提取后，对传统判别方法训练也能有效提升判别模型的准确率。而 1D－CNN 模型由于其结构及特点，本身就能够对数据进行降维和特征提取，同时还能保证判别时具有较高准确率。因此不需要对光谱数据进行预处理和特征提取便可使用原始光谱数据对模型进行训练，有效节省了计算时间。

表 4 不同判别算法准确率对比

Tab. 4 Comparison of accuracy rates of different classification algorithms %

分类算法	原始光谱准确率	SG＋特征选择后光谱准确率
BPNN	83.33	92.59
K－ELM	92.59	94.44
XGBoost	90.74	94.44
1D－CNN	98.15	

试验结果表明，1D－CNN 在进行光谱数据训练时能够有效地提取到相关性很高的特征信息，避免了在处理过程微小特征峰的丢失，在对光谱数据处理的复杂度、计算时间以及判别模型的准确度方便有较大优势，在对仅采用原始光谱数据进行训练后的蜜柚枯水程度区分上具有较高准确率，1D－CNN 表现出模型更为强大的判别及特征学习能力。反观 BPNN、K－ELM 以及 XGBoost 模型则对原始光谱数据的平滑处理以及特征变量提取等步骤有着不同程度依赖，相比之下 1D－CNN 模型在对蜜柚枯水程度的定性判别方面性能最优。

2.4 数据增强与模型稳定性验证

针对深度学习模型需要大量样本训练问题，且为了增强模型的泛化能力和鲁棒性，采用平移法和添加随机噪声作为数据增强方法，对原始近红外光谱数据进行扩充[28]。平移法主要通过将原始光谱数据随机向左或向右平移 1～6 nm 实现对光谱数据扩充。添加高斯噪声法主要通过向光谱数据添加原始数据标准差 1%～5% 范围内的随机高斯噪声实现。最终通过 2 种方法均将样本总数由 180 个扩充至 600 个，扩充后光谱数据曲线如图 11 所示，并将数据增强后的光谱数据与原始光谱数据混合使用 Kennard－Stones 划分数据集，训练集共 420 个样本，测试集共 180 个样本，其中 3 个类别在训练集中个

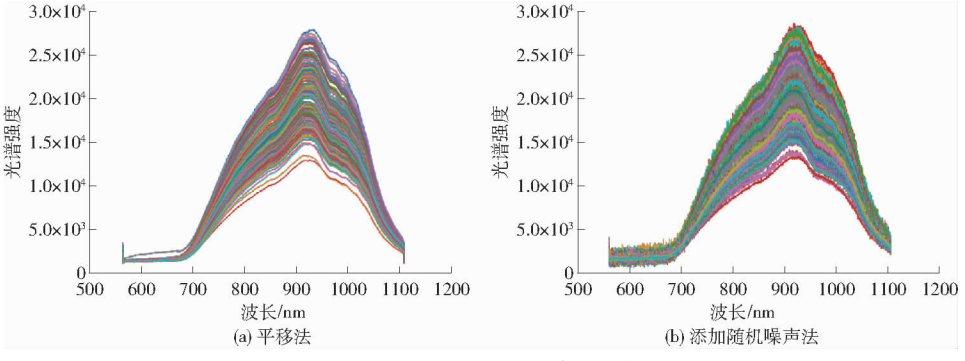


图 11 数据增强后的光谱数据曲线

Fig. 11 Spectral data curves after data augmentation

数均为 140 个，在测试集中个数均为 60 个。将划分后的数据作为模型输入，最终得到测试集结果如表 5 所示。

表 5　数据增强后的 1D－CNN 模型性能

Tab. 5　Performance of data-enhanced 1D－CNN model

%

数据增强方法	枯水程度	准确率	精确率	召回率	F1 分数
平移法	未枯水	100	96.77	100	98.36
	轻度枯水	96.67	96.67	96.67	96.67
	中度枯水	96.67	100	96.67	98.31
添加高斯噪声法	未枯水	100	100	100	100
	轻度枯水	93.33	98.25	93.33	95.73
	中度枯水	98.33	93.65	98.33	95.93

从表 5 可以看出，经平移法和添加噪声法进行数据增强后，1D－CNN 模型准确率平均值分别为 97.78% 和 97.22%，与原始光谱数据建模结果几乎持平，显示出该模型在处理数据集时的鲁棒性。且经过数据增强处理后模型各项指标如精确率、召回率以及 F1 分数依然保持在较高水平。结果表明，在原始数据以及数据增强后条件下，1D－CNN 模型都能够有效地捕捉关键特征，从而准确区分蜜柚不同的枯水程度，体现出该模型优异的泛化能力以及在复杂数据环境下的可靠性。

3　结论

（1）采用自主研发的水果全透射无损在线检测设备，区别于一般的光谱采集方式，采用短积分时间采集方法，在积分时间 10 ms 条件下获得有效的光谱信号。

（2）提出了一种基于卷积神经网络的蜜柚枯水判别模型，确定了最优的网络结构参数和训练的参数，结果表明该卷积神经网络模型能够精准、快速实现蜜柚不同枯水程度的鉴别，最终判别准确率达到 98.15%。

（3）采用 SG 平滑方法和 CARS 特征变量筛选后，3 种判别模型对蜜柚枯水程度判别准确率均有一定程度提升，BPNN、K－ELM、XGBoost 测试集准确率分别提升至 92.59%、94.44% 和 94.44%，可以看出 SG－CARS 能够有效地剔除光谱数据中的噪声及无关信息并筛选出相关性较高的特征变量，提升了模型的检测精度，说明预处理以及特征变量筛选对模型检测精度的必要性。

（4）对比了 1D－CNN、BPNN、K－ELM、XGBoost 模型在原始光谱下和经预处理和特征筛选后的判别结果，在原始光谱下 1D－CNN 判别准确率可达 98.15%，高于经预处理和特征筛选后的 K－ELM 与 XGBoost 模型判别准确率（均为 94.44%），此外1D－CNN 在计算成本与时间上更具优势，可见 1D－CNN 对光谱数据的强大学习能力，表现出其在蜜柚枯水程度检测中的高效性和准确性。

（5）对原始光谱数据进行了数据增强，并与原始数据混合建模以验证模型稳定性。结果表明，经平移法和添加噪声法数据增强后，1D－CNN 模型在测试集上平均准确率分别达到 97.78% 和 97.22%，与原始光谱数据的判别模型结果无显著差异。验证了该模型在蜜柚枯水程度判别方面具备较好的稳定性和泛化能力。

参　考　文　献

[1] 胡位荣，刘顺枝，江月玲，等. 红肉蜜柚汁胞枯水特性分析[J]. 食品科学，2015，36(10)：233－238.
HU Weirong, LIU Shunzhi, JIANG Yueling, et al. Characteristics of granulated juice sacs of red-fleshed sweet pomelo[J]. Food Science, 2015, 36(10): 233－238. (in Chinese)

[2] 姚世响，李秋雨，曹琦，等. 柑橘枯水研究进展[J]. 食品与发酵工业，2020，46(18)：259－263.
YAO Shixiang, LI Qiuyu, CAO Qi, et al. Granulation of citrus fruit: research progress and future directions[J]. Food and Fermentation Industries, 2020, 46(18): 259－263. (in Chinese)

[3] 汪妮娜，黄宏明，陈东奎，等. 几个柚品种果实贮藏期的生理生化及粒化指数差异[J]. 中国南方果树，2021，50(2)：40－43.

[4] 赵娟，沈懋生，浦育歌，等. 基于近红外光谱与多品质指标的苹果出库评价模型研究[J]. 农业机械学报，2023，54(2)：386－395.
ZHAO Juan, SHEN Maosheng, PU Yuge, et al. Out-of-warehouse evaluation and prediction model of apple based on near-infrared spectroscopy combined with multiple quality indexes[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 386－395. (in Chinese)

[5] 刘文政，周雪健，平凤娇，等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测[J]. 农业机械学报，2024，55(2)：372－383.
LIU Wenzheng, ZHOU Xuejian, PING Fengjiao, et al. Detection of key indicators of ripening quality in table grapes based on visible-near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 372－383. (in Chinese)

[6] 刘燕德，肖怀春，邓清，等. 柑桔黄龙病近红外光谱无损检测[J]. 农业工程学报，2016，32(14)：202－208.
LIU Yande, XIAO Huaichun, DENG Qing, et al. Nondestructive detection of citrus greening by near infrared spectroscopy[J]. Transactions of the CSAE, 2016, 32(14): 202－208. (in Chinese)

[7] 孙通，李翰林，孔令飞，等. 光源光斑直径对苹果霉心病近红外检测的影响[J]. 农业工程学报，2023，39(24)：298－305.

SUN Tong, LI Hanlin, KONG Lingfei, et al. Effects of light source spot diameter on the near-infrared detection of mold heart diseases in apples[J]. Transactions of the CSAE, 2023,39(24):298 - 305. (in Chinese)
[8] 李勋兰,易时来,何绍兰,等. 高光谱成像技术的柚类品种鉴别研究[J]. 光谱学与光谱分析,2015,35(9):2639 - 2643.
LI Xunlan, YI Shilai, HE Shaolan, et al. Identification of pummelo cultivars based on hyperspectral imaging technology[J]. Spectroscopy and Spectral Analysis, 2015,35(9):2639 - 2643. (in Chinese)
[9] 许骞,蔡健荣,杜灿,等. 基于软 X 射线成像技术的柑橘内部浮皮和枯水检测[J]. 智慧农业,2021,3(4):53 - 65.
XU Qian, CAI Jianrong, DU Can, et al. Detection of peel puffing and granulation in citrus based on soft X-ray imaging technology[J]. Smart Agriculture, 2021,3(4):53 - 65. (in Chinese)
[10] 张欣欣,李尚科,李跑,等. 近红外光谱的不同产地柑橘无损鉴别方法[J]. 光谱学与光谱分析,2021,41(12):3695 - 3700.
ZHANG Xinxin, LI Shangke, LI Pao, et al. A nondestructive identification method of producing regions of citrus based on near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2021,41(12):3695 - 3700. (in Chinese)
[11] PUANGSOMBUT A, PATHVEERAT S, TERDWONGWORAKUL A, et al. Evaluation of internal quality of fresh-cut pomelo using Vis/NIR transmittance[J]. Journal of Texture Studies, 2012,43(6): 445 - 452.
[12] LI Xunlan, YI Shilai, HE Shaolan, et al. Identification of pummelo cultivars by using Vis/NIR spectra and pattern recognition methods[J]. Precision Agriculture, 2016, 17: 365 - 374.
[13] LIU Tianhu, EHSANI R, TOUDESHKI A, et al. Identifying immature and mature pomelo fruits in trees by elliptical model fitting in the Cr - Cb color space[J]. Precision Agriculture, 2019,20: 138 - 156.
[14] 王贤达,陈添峰,黄镜浩,等. 琯溪蜜柚汁胞粒化的磁共振成像无损检测[J]. 中国南方果树,2017,46(6):21 - 24.
[15] 刘燕德,李茂鹏,胡军,等. 近红外高光谱的脐橙粒化检测研究[J]. 光谱学与光谱分析,2022,42(5):1366 - 1371.
LIU Yande, LI Maopeng, HU Jun, et al. Detection of citrus granulation based on near-infrared hyperspectral data[J]. Spectroscopy and Spectral Analysis, 2022,42(5):1366 - 1371. (in Chinese)
[16] YU Xinjie, LU Huanda, WU Di. Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging[J]. Postharvest Biology and Technology, 2018,141: 39 - 49.
[17] TIAN Shijie, WANG Shuai, XU Huirong. Early detection of freezing damage in oranges by online Vis/NIR transmission coupled with diameter correction method and deep 1D - CNN[J]. Computers and Electronics in Agriculture, 2022, 193: 106638.
[18] TIAN Hao, XU Huirong, YING Yibin. Can light penetrate through pomelos and carry information for the non-destructive prediction of soluble solid content using Vis - NIRS? [J]. Biosystems Engineering, 2022,214: 152 - 164.
[19] 彭彦昆,赵芳,李龙,等. 利用近红外光谱与 PCA - SVM 识别热损伤番茄种子[J]. 农业工程学报,2018,34(5):159 - 165.
PENG Yankun, ZHAO Fang, LI Long, et al. Discrimination of heat-damaged tomato seeds based on near infrared spectroscopy and PCA - SVM method[J]. Transactions of the CSAE,2018,34(5):159 - 165. (in Chinese)
[20] 曾一凡,刘春生,孙旭东,等. 可见/近红外光谱技术无损检测果实坚实度的研究[J]. 农业工程学报,2008,24(5):250 - 252.
ZENG Yifan, LIU Chunsheng, SUN Xudong, et al. Nondestructive measurement of firmness of pear using visible and near-infrared spectroscopy technique[J]. Transactions of the CSAE, 2008,24(5):250 - 252. (in Chinese)
[21] 徐惠荣,李青青. 皇冠梨糖度可见/近红外光谱在线检测模型传递研究[J]. 农业机械学报,2017,48(9):312 - 317.
XU Huirong, LI Qingqing. Calibration model transfer between visible/NIR spectrometers in sugar content on-line detection of crown pears[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017,48(9):312 - 317. (in Chinese)
[22] 王淼,张晶,贺妍,等. 基于低场核磁共振的柑橘汁胞粒化评级[J]. 农业工程学报,2016,32(7):290 - 295.
WANG Miao, ZHANG Jing, HE Yan, et al. Evaluation of juicy sac granulation in citrus with low field nuclear magnetic resonance[J]. Transactions of the CSAE, 2016,32(7):290 - 295. (in Chinese)
[23] 赖长鸿,王寒雪,许艺嘉,等. 蜜柚果实粒化汁胞结构特征解析[J]. 中国南方果树,2024,53(4):44 - 49.
[24] 孙潇鹏,刘灿灿,陆华忠,等. 基于可见-近红外透射光谱的蜜柚检测中影响因素分析[J]. 包装与食品机械,2022,40(4):1 - 7.
SUN Xiaopeng, LIU Cancan, LU Huazhong, et al. Analysis of influencing factors in the detection of honey pomelo based on visible-near infrared transmittance spectroscopy[J]. Packaging and Food Machinery, 2022,40(4):1 - 7. (in Chinese)
[25] 常洪娟,蒙庆华,吴哲锋,等. 基于反向传播神经网络和高光谱成像的芒果可溶性固形物含量检测[J]. 食品安全质量检测学报,2024,15(2):141 - 148.
CHANG Hongjuan, MENG Qinghua, WU Zhefeng, et al. Detection of soluble solids content in mango based on backpropagation algorithm neural network and hyperspectral imaging[J]. Journal of Food Safety and Quality, 2024,15(2): 141 - 148. (in Chinese)
[26] 陈文丽,王其滨,路皓翔,等. 最小角回归结合核极限学习机的近红外光谱对柑橘黄龙病的鉴别[J]. 分析测试学报,2020,39(10):1267 - 1273.
CHEN Wenli, WANG Qibin, LU Haoxiang, et al. Identification of citrus Huanglongbing by near infrared spectroscopy with least angle regression and kernel extreme learning machine[J]. Journal of Instrumental Analysis, 2020,39(10): 1267 - 1273. (in Chinese)
[27] 许晓栋,张慧敏,刘佳乐,等. 基于 XGBoost 算法的鱼粉中微塑料的红外光谱识别研究[J]. 光谱学与光谱分析,2024,44(7):1835 - 1842.
XU Xiaodong, ZHANG Huimin, LIU Jiale, et al. Study on infrared spectral recognition of microplastics in fishmeal based on XGBoost algorithm[J]. Spectroscopy and Spectral Analysis, 2024,44(7):1835 - 1842. (in Chinese)
[28] 黄铝文,谦博,关非凡,等. 基于小波变换与卷积神经网络的羊脸识别模型[J]. 农业机械学报,2023,54(5):278 - 287.
HUANG Lüwen, QIAN Bo, GUAN Feifan, et al. Goat face recognition model based on wavelet transform and convolutional neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(5):278 - 287. (in Chinese)