

基于电学参数的稻谷含水率和出糙率机器学习预测模型

邓 安，徐永阳，邱伟强，李 立，金银哲*

(1. 上海海洋大学食品学院, 上海 201306; 2. 上海海洋大学食品热加工工程中心, 上海 201306)

摘 要: 收购现场稻谷出糙率和含水率的快速检测有助于稻谷的快速收购和分级管理, 从而减少稻谷在运输、储藏和加工过程中的损耗。研究使用 LCR 测量仪测试 1~8MHz 频率下含水率为 9.94%~23.51% 的稻谷的介电常数、电容、介电损耗、损耗系数、导电率和电导。将获得的数据用于建立二次回归模型和机器学习模型, 机器学习模型分别为神经网络、决策树、随机森林和支持向量机。结果表明, 频率为 1MHz 时, 导电率与稻谷的含水率和出糙率回归模型决定系数均最高, 分别为 0.960 和 0.929。此外, 使用神经网络在预测稻谷的含水率和出糙率时获得了更好的预测效果, 决定系数分别为 0.987 和 0.935, 这说明了神经网络模型可以更加灵活的获取和表达电学参数与稻谷含水率和出糙率之间的非线性关系。这种便捷、高效的模型可以为稻谷品质无损检测提供理论参考。

关键词: 机器学习; 含水率; 电学参数; 回归模型; 稻谷; 出糙率

doi: 10.11975/j.issn.1002-6819.202312197

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2024)-15-0245-08

邓安, 徐永阳, 邱伟强, 等. 基于电学参数的稻谷含水率和出糙率机器学习预测模型[J]. 农业工程学报, 2024, 40(15): 245-252. doi: 10.11975/j.issn.1002-6819.202312197 <http://www.tcsae.org>
DENG An, XU Yongyang, QIU Weiqiang, et al. Predicting moisture content and husked rice yield using electrical parameters and machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(15): 245-252. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202312197 <http://www.tcsae.org>

0 引 言

稻谷是世界三大粮食作物之一, 种植面积占世界粮食作物总面积的 25%^[1-2]。预计到 2030 年, 全球稻米产量将增加 58 t, 达到 5.67 亿 t^[3]。中国是一个水稻种植大国, 水稻总产量连续 11 年稳定在 2 亿 t 以上^[4]。面对巨大的稻谷产量, 稻谷的收购存在收购周期短、运输距离长的难题, 而且采收地区气候和稻谷成熟度的差异会显著影响稻谷的含水率。在微生物的作用下, 水分越高的稻谷在运输、储藏过程中越容易发生霉变从而导致经济损失^[5-6]。在稻谷收购现场, 收购人员会根据稻谷的含水率和出糙率来评估稻谷的品质并进行定价。稻谷的出糙率也会受到含水率的影响^[7]。因此, 在收购现场准确测定稻谷的含水率至关重要。然而, 传统稻谷含水率检测的方法主要有烘箱法、甲苯蒸馏法、卡尔费休法、减压干燥法^[8]。这些方法普遍存在测试时间长、操作繁琐、需要破坏稻谷等缺点, 因此不宜在收购现场使用^[9]。此外, 出糙率的检测一般需要对稻谷进行脱壳处理, 在挑拣糙米的过程中需要花费大量的时间, 而且容易因为主观因素出现误差^[10]。

为了实现粮食含水率的快速检测, 国内外专家基于粮食的物理特性、化学特性和光学特性开发了一些无损

检测的方法, 包括电阻法^[11]、微波法^[12]、电容法^[13]、红外法^[14]和核磁共振法^[15]。其中红外法不易检测物质内部的水分; 微波法的检测设备质量大, 造价高; 而核磁共振法的使用成本高^[9,12,14-16]。因此, 这 3 种方法并没有被广泛使用。目前, 使用较多是通过电学特性(电容、介电特性等)来间接预测粮食的含水率^[17]。郭文川等^[18]使用 LCR 测量仪测定了含水率为 12%~21% 燕麦的介电特性, 通过回归模型建立了含水率与介电特性之间的联系, 并通过试验证实了含水率预测模型的可行性。刘志壮等^[13]使用平行电容传感器将电容值转化为含水率值, 并建立回归模型准确预测了稻谷含水率。俞世钢等^[19]通过稻谷的电阻抗特性建立的预测含水率模型相关系数达到 0.933 4。这些研究证明了使用谷物的电学参数可以实现水分的快速检测。此外, 作为一种高效预测工具, 机器学习也被应用于食品检测领域。商亮等^[20]基于介电特性和人工神经网络提出了一种快速检测油桃内部品质的方法。KUBO 等^[21]开发了人工神经网络模型, 并成功用于建立果汁的温度、频率和含糖量与介电特性之间的关系。然而, 目前尚未有使用电学参数来预测稻谷出糙率的研究。

因此, 本研究拟使用 LCR 测量仪获取稻谷的 6 个电学参数(介电常数、电容、介电损耗、损耗系数、导电率和电导)。将获取的数据用于建立二次回归模型和机器学习模型, 机器学习模型包括神经网络、决策树、支持向量机和随机森林。然后通过模型获得电学参数、含水率和出糙率之间的相互关系, 以期实现收购现场稻谷含水率和出糙率的快速检测, 为稻谷的分级收购、运输、储藏和加工提供试验数据和理论依据。

收稿日期: 2023-12-26 修订日期: 2024-06-29

基金项目: 上海市粮食和物资储备局项目(D-8028-23-1124)

作者简介: 邓安, 研究方向为食品热加工。Email: deng_an@163.com

*通信作者: 金银哲, 博士, 副教授, 主要从事食品热加工以及食品分析。
Email: yzjin@shou.edu.cn

1 材料制备与方法

1.1 材料与仪器

新鲜稻谷, 购自广东湛江 (2023 年 7 月)。

DHG-9245A 型鼓风干燥箱, 上海慧泰仪器制造有限公司; FA2204 型电子天平, 上海良平仪器仪表有限公司; IM 3536 型 LCR 测量仪, 日本日置电机株式会社; JLG-2018 型砻谷机, 济宁市翰业机械设备有限公司。

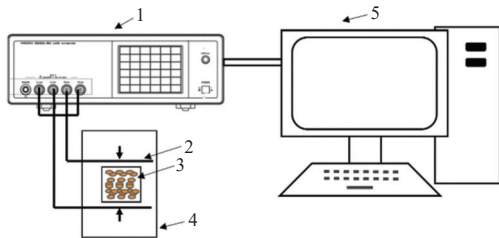
1.2 试验方法

1.2.1 稻谷样品的制备

在制备样品前, 先去除稻谷中的虫子、杂物、护颖上的稻须。将处理好的稻谷混合均匀后分成质量相同的 31 份, 每份质量为 (150 ± 0.5) g。每一份新鲜稻谷使用透明密封袋密封, 做好标记为 0~30 号。根据文献 [22] 测定初始样品 (0 号) 的含水率为 24.6%。使用同样的干燥方法分别处理剩余的 30 份稻谷 (分别编号 1~30), 获得了含水率为 9.94%~23.51% 的稻谷样品。将这些不同编号的稻谷装入透明密封袋中定期晃动确保水分均匀分布。在 2 h 后, 将每个编号的稻谷样品均分为 3 等份, 得到总共 90 份样品用于电学参数和出糙率的测试。

1.2.2 电学参数的测定

本研究使用的试验设备简图如图 1。在测定电学参数前, 启动 LCR 测量仪并进行校准。将稻谷装入模具, 使用 LCR 测量仪测定稻谷的介电常数, 并将获得的电学参数输入电脑进行数据分析。测试模具为定制的聚四氟乙烯正方体容器, 容器壁厚度 4 mm。模具内部长度、宽度和高度都为 20 mm, 可用于装稻谷。模具壁相对的 2 个壁面各放置了一个平行电极板通过夹具连接 LCR 测量仪。LCR 测量仪测试电学参数的方法参考文献 [23]。使用 LCR 测量仪获得稻谷在 1~8 MHz 条件下 8 个频率点的电学参数, 每个频率点对应 6 项电学参数 (包括介电常数、介电损耗、损耗系数、导电率、电导和电容), 共计 48 个电学参数。这些电学参数常用于食品品质的检测 [17-19], 已经得到许多学者的认可。



1. LCR 测量仪 2. 平行板电极板 3. 稻谷 4. 屏蔽箱 5. 电脑
1. LCR meter 2. Parallel electrode plate 3. Paddy rice 4. Shielding box 5. Computer

图 1 稻谷 LCR 测试的试验简图

Fig.1 Schematic of LCR testing for paddy rice

1.2.3 稻谷出糙率的测定

出糙率的测试方法参照文献 [24]。在开始剥壳之前, 先打开砻谷机, 根据稻谷的粒形调节砻谷机的辊轴距离和风速。使用少量测试稻谷进行脱壳处理, 直到碎米数量最少, 且糙米和谷壳的分离效果最佳时进行正式试验。

如果经过一次脱壳后有少量稻谷未剥壳成功, 则进行二次处理。

根据式 (1) 计算不同含水率稻谷的出糙率:

$$H = \frac{(m_1 + m_2) - (m_1 + m_3)/2}{m} \times 100\% \quad (1)$$

式中 H 表示稻谷的出糙率, %; m_1 表示生芽粒糙米的质量, g; m_2 表示砻谷机脱壳糙米质量, g; m_3 表示人工挑选的不完善粒糙米质量, g; m 表示净稻谷质量, g。

2 模型构建

2.1 数据预处理

将试验获得的 90 份稻谷的含水率、出糙率和电学参数进行汇总, 共计获得 90 组数据集, 每个数据集包含了稻谷的水分含量、出糙率和 1~8 MHz 的电学参数 (包括介电常数、介电损耗、损耗系数、导电率、电导和电容)。依据文献 [19] 选取 1 MHz 条件下的电学参数作为回归模型的分析数据。在该频率下获得的电学参数更加稳定, 有利于分析电学参数与稻谷品质 (含水率和出糙率) 之间的关系。此外, 为了提高预测稻谷品质的准确性, 研究将获得的全部数据集用来建立机器学习模型, 其中包括神经网络、随机森林、支持向量机和决策树模型。这 4 种机器学习模型在进行大量数据预测时获得更好的效果 [25]。每组数据集中具有 48 个电学参数, 90 组数据集共计 4320 个数据点作为机器学习模型输入数据, 稻谷品质 (含水率和出糙率) 作为输出数据。数据集的 80% 用于训练模型, 剩下的 20% 用于验证模型。

2.2 二次回归模型建立

在 1 MHz 的条件下, 使用回归模型分析电学参数 (包括介电常数、电容、介电损耗、损耗系数、导电率和电导) 与稻谷品质 (含水率和出糙率) 之间的关系。通过式 (2) 计算:

$$f(x) = a + b \cdot x + c \cdot x^2 \quad (2)$$

式中 x 表示模型输入的自变量 (电学参数); $f(x)$ 表示模型的因变量 (稻谷的含水率或出糙率); a , b 和 c 表示模型中的参数。

2.3 机器学习模型建立

机器学习在进行数据训练和预测的时候, 每个数据集表示为 (X_i, y_i) , 其中 $X_i = [x_{i1}, x_{i2}, \dots, x_{i48}]$ 是输入特征向量, 包含 48 个电学参数, y_i 是对应稻谷的品质指标 (含水率或出糙率)。

2.3.1 人工神经网络

人工神经网络可以自适应地学习复杂的模式和规律, 使其在处理非线性问题时表现出色 [21]。本研究中神经网络是由输入层的 6 个电学参数 (介电常数、电容、介电损耗、损耗系数、导电率和电导)、隐含层的 15 个神经元和输出层的 2 个神经元 (含水率和出糙率) 组成的。

1) 前向传播

输入层到隐含层:

$$z^{(1)} = X_i \cdot W^{(1)} + b^{(1)} \quad (3)$$

$$a^{(1)} = \sigma(z^{(1)}) \quad (4)$$

隐含层到输出层：

$$z^{(2)} = a^{(1)} \cdot W^{(2)} + b^{(2)} \quad (5)$$

$$\hat{y}_{NN} = a^{(2)} = z^{(2)} \quad (6)$$

式中 $W^{(1)}$ 表示输入层到隐含层的权重矩阵，其中每一行对应一个隐含层神经元，每一列对应一个输入特征； $W^{(2)}$ 表示隐含层到输出层的权重矩阵，其中每一行对应一个输出神经元，每一列对应一个隐含层神经元； $b^{(1)}$ 表示隐含层的偏置向量，其中每个元素对应一个隐含层神经元； $b^{(2)}$ 表示输出层的偏置向量，其中每个元素对应一个输出神经元； σ 表示激活函数 Sigmoid； $z^{(1)}$ 表示输入层到隐含层的加权和，未经过激活函数； $z^{(2)}$ 表示隐含层到输出层的加权和，未经过激活函数； $a^{(1)}$ 表示经过激活函数后的隐含层的输出； $a^{(2)}$ 表示神经网络的最终输出，即模型对稻谷品质指标的预测； $\hat{y}_{NN}$ 表示神经网络模型的输出，表示对稻谷品质指标的预测。

2) 使用均方误差 (mean squared error, M_{SE}) 作为损失函数：

$$M_{SE}(y_i, \hat{y}_{NN}) = \frac{1}{2} (\hat{y}_{NN} - y_i)^2 \quad (7)$$

3) 反向传播

输出层误差：

$$\delta^{(2)} = \frac{\partial M_{SE}}{\partial z^{(2)}} = \hat{y}_{NN} - y_i \quad (8)$$

隐含层误差：

$$\delta^{(1)} = \frac{\partial M_{SE}}{\partial z^{(1)}} = \delta^{(2)} \cdot (W^{(2)})^T \cdot \sigma'(z^{(1)}) \quad (9)$$

式中 σ' 是激活函数的导数； $\delta^{(1)}$ 表示隐含层的误差项，反映了输出误差对隐含层的影响； $\delta^{(2)}$ 表示输出层的误差项，是模型预测与真实值的差异。

4) 参数更新

$$W^{(2)} = W^{(2)} - \sigma \cdot \sigma^{(1)T} \cdot \delta^{(2)} \quad (10)$$

$$b^{(2)} = b^{(2)} - \alpha \cdot \sum_{i=1}^N \delta^{(2)} \quad (11)$$

$$W^{(1)} = W^{(1)} - \alpha \cdot X_i^T \cdot \delta^{(1)} \quad (12)$$

$$b^{(1)} = b^{(1)} - \alpha \cdot \sum_{i=1}^N \delta^{(1)} \quad (13)$$

式中 α 表示梯度下降算法中的一个超参数，用于控制每次参数更新的步长，本研究中设置为 0.01， N 表示样本数， T 表示转置。

2.3.2 决策树

决策树预测模型能有效映射对象属性与对象值之间的复杂关系^[26]。通过对数据的逐步分割，生成了树状结构，其中每个节点代表一个对象，分叉路径表示可能的属性值，而叶节点对应具体的含水率和出糙率。

使用均方误差作为分裂准则：

$$M_{SE}(D_i) = \frac{1}{|D_i|} \sum_{k=1}^{|D_i|} (y_{ik} - \bar{y}_{D_i})^2 \quad (14)$$

$$\operatorname{argmin}_{F_j, t} \left[\frac{|D_{i1}|}{|D_i|} M_{SE}(D_{i1}) + \frac{|D_{i2}|}{|D_i|} M_{SE}(D_{i2}) \right] \quad (15)$$

式中 $|D_i|$ 表示数据集 D_i 中样本的数量； y_{ik} 表示数据集 D_i 中的第 k 个样本的稻谷品质指标； $\bar{y}_{D_i}$ 表示数据集 D_i 中所有样本的稻谷预测的平均值； F_j 表示特征集合，在每次分裂节点时随机选择； t 表示特征的阈值，用于决定分裂。

对于新的输入特征向量 X_{new} ，通过决策树进行预测，表达式如下：

$$\hat{y}_{DT} = \operatorname{Predict}(X_{new}) \quad (16)$$

式中 $\hat{y}_{DT}$ 表示决策树模型对稻谷品质指标的预测结果。

2.3.3 随机森林

随机森林是一种通过集成学习的 Bagging 思想将多棵树集成的算法，其基本单元是决策树^[26]。在处理电学参数和稻谷品质的问题时，随机森林可以通过构建多棵决策树来集成多个分类器，形成一个强大的整体分类器。

随机选择样本，使用自主采样法 (bootstrap sampling)：

$$B_i = \{(X_{i1}, y_{i1}), (X_{i2}, y_{i2}), \dots, (X_{im}, y_{im})\} \quad (17)$$

式中 B_i 表示通过自助采样法获得的随机抽样的数据集，包含若干个样本。

对于回归问题，随机森林的预测是通过对所有决策树的判断结果进行平均得到的，表达式如下：

$$\hat{y}_{RF} = \frac{1}{E} \sum_{j=1}^E \hat{y}_j \quad (18)$$

式中 $\hat{y}_{RF}$ 表示随机森林模型对稻谷品质指标的预测结果， E 表示决策树的数量。

2.3.4 支持向量机

支持向量机模型将电学参数和稻谷品质的实例表示为空间中的点，以便各个电学参数和稻谷品质的实例以尽可能宽的清晰间隙分隔开^[26]。新的电学参数和稻谷品质实例被映射到同一空间，并根据它们落在间隔的哪一侧来预测它们的相应类别。这种方法帮助支持向量机模型更好地理解电学参数与稻谷品质之间的关系，并在预测中提供更准确的结果。

支持向量机的目标是通过解决拉格朗日对偶问题，找到最优的超平面：

$$\max_{\phi} \left[\sum_{i=1}^N \phi_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \phi_i \phi_j y_i y_j K(X_i, X_j) \right] \quad (19)$$

$$\text{subject to } \sum_{i=1}^N \phi_i y_i = 0 \quad (20)$$

$$0 \leq \phi_i \leq C, \text{ for } i = 1, 2, \dots, N \quad (21)$$

式中 ϕ 表示拉格朗日乘子，每个训练样本 X_i 对应一个 ϕ_i ； y_i 表示训练样本 X_i 的标签； $K(X_i, X_j)$ 表示核函数，即 2 个样本之间的相似性； C 表示正则化参数，用于平衡最大化间隔和最小化误差的权重； N 表示数据集的样本数量。

找到满足 $0 \leq \phi_i \leq C$ 的样本点，它们对应的 ϕ_i 值大于零。

对于新的输入特征向量 X_{new} ，通过支持向量的加权 and 进行预测，表达式如下：

$$\hat{y}_{SVM} = Predict(X_{new}) = \sum_{i=1}^N \varphi_i y_i K(X_i, X_{new}) + b \quad (22)$$

式中 $\hat{y}_{SVM}$ 表示支持向量机模型对稻谷品质指标的预测结果。

2.4 模型评估

使用 Origin 2018 软件对 1 MHz 条件下的稻谷的含水率、电学参数和出糙率数据进行拟合分析并绘制图形。使用 MATLAB R2021a 软件建立神经网络、随机森林、支持向量机和决策树模型。这 4 个模型在训练集上进行拟合,并分别使用决定系数(R^2)和均方根误差(root mean squared error, RMSE)来衡量模型的拟合效果和精确程度, R^2 越接近 1,且 RMSE 越小说明模型效果越好[27]。

3 结果与分析

3.1 含水率和频率对电学参数的影响

为了更好地观察和研究频率与含水率对电学参数的影响,图 2 展示了 30 组样品中含水率分别为 9.94%、13.01%、16.97%、19.50%、21.59% 和 23.51% 的电学数据。

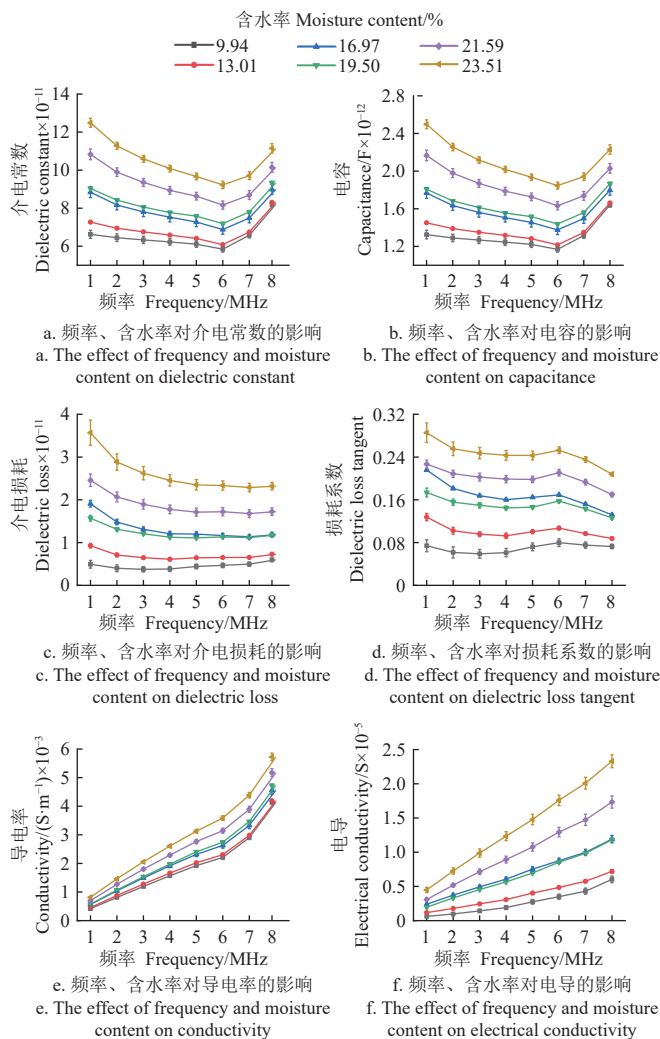


图 2 频率和含水率对于电学参数的影响

Fig.2 Effect of frequency and moisture content on electrical parameters

介电常数反映了食品组分对电磁辐射吸收的能力,其数值与电容呈正相关[13,28]。随着含水率的增加,稻谷吸收微波的能力变得更强,电容值也逐渐提高。此外,

随着频率的提高,介电常数和电容呈现先降低后提高的趋势。在 6 MHz 时达到最小值,该拐点常见于食品的介电特性中,可能是由于在电流作用下介质表面极化导致的。介电损耗表示稻谷极性组分对电磁波的损耗能力[29]。含水率的增加提高了稻谷中极性成分的数量,从而增加了稻谷的介电损耗。随着频率的增加,介电损耗在不断的减少,在接近最高频率时缓慢提高。在更高频率下,分子没有足够的时间来跟随电场的变化,导致分子无法有效地转向以产生热量,导致介电损耗减少。这现象通常被称为“极化效应”或“Maxwell-Wagner 效应”。损耗系数反映了物料对外部作用的能量吸收程度的指标,可以用介电常数和介电损耗的比值来表示[28]。该指标随着频率增加呈现先降低后升高,在 6 MHz 之后又降低的趋势。

导电率和电导两个指标单位不同,都可作为物质对于电流的输送能力的指标。水作为溶剂,可以促进稻谷中可溶性物质的溶解。当稻谷中含水率提高后,会增加稻谷中离子的浓度。此外,新鲜高含水率稻谷的代谢速度更快,使得稻谷内部离子的运动性更高。在进行 LCR 测试的时候电流通过稻谷时遇到的阻碍也会更小。因此,含水率越高的稻谷也将获得更高的导电率和电导。随着频率的提高导电率和电导都在不断提高。这是因为电场频率升高时,趋肤效应导致电流更倾向于集中在导体表面,从而增强了导体的导电性。

3.2 回归模型的拟合效果分析

3.2.1 稻谷含水率与电学参数的回归模型

在 1 MHz 条件下,稻谷含水率与电学参数的回归模型参数及相关性如图 3 所示。含水率与 6 个电学参数的决定系数从大到小分别为导电率 (0.960)、电容 (0.924)、介电常数 (0.923)、电导 (0.907)、损耗系数 (0.865)、介电损耗 (0.858)。导电率、电容和介电常数的拟合效果较好,相关系数均大于 0.920。其中,导电率与含水率拟合获得的决定系数最高 (0.960), RMSE 最小 (0.89%)。这说明稻谷的导电率与稻谷含水率相关性最显著,可以用导电率来准确的获取稻谷的含水率数据。稻谷中的水分越多,导致稻谷的阻值越少,从而提高稻谷的导电率。此外,电容和介电常数也常用于检测食品的品质。刘志壮等[13]基于介电常数开发了检测电容的设备,并发现稻谷的含水率为 8%~23%,电容与含水率呈正相关关系。在对燕麦、苹果和颗粒饲料的介电特性的研究中也发现了类似的规律[18,23,30]。

3.2.2 稻谷含水率与出糙率的回归模型

出糙率是反映稻谷加工品质的重要指标,可用于评价稻谷的生长条件、成熟程度和出米情况[31]。出糙率高的稻谷在进行加工时出米率高,食用品质更好。图 4 表示了含水率与出糙率之间的关系。通过拟合含水率与出糙率之间的数学模型,得到稻谷含水率与出糙率拟合的决定系数为 0.944, RMSE 为 3.05%,说明了模型的准确性较高。由图 4 可以看到,随着含水率的增加,稻谷的出糙率发生了显著的变化。当含水率低于 13.34% 时,含水率越低,

稻谷的出糙率越小。水分低时，会导致米粒变得很脆，在机械脱壳的过程中米粒容易受到外力的作用断裂而降低出糙率。含水率在 13.34% 时，稻谷的出糙率达到最大值为 69.31%。当含水率从 13.34% 提高到接近 24% 时，稻谷的出糙率不断降低。水分的增加会降低稻谷颗粒的流动性，降低稻谷的硬度。在使用砻谷机对稻谷进行脱壳处理时，稻谷更容易产生机械损伤。古汶玉等^[7]也发现过高的水分会导致稻谷的出糙率降低。由此可知，在实际的工业加工环境下，应控制稻谷的水分在合理的范围内，以减少过高的水分导致脱壳过程中造成的损耗。

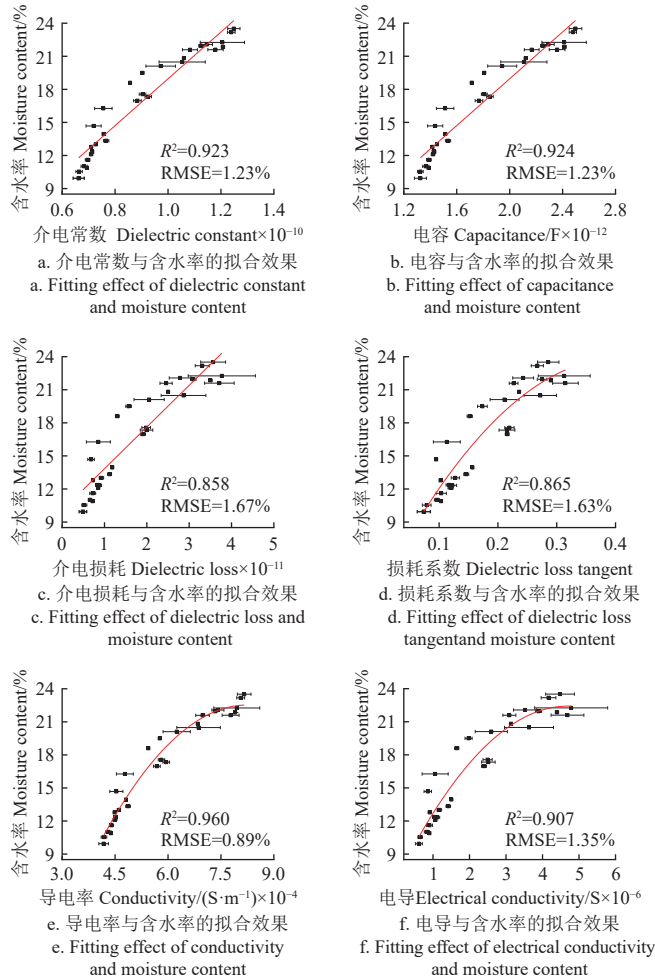


图 3 稻谷电学参数与含水率的拟合关系

Fig.3 Fitting relationship between paddy electrical parameters and moisture content

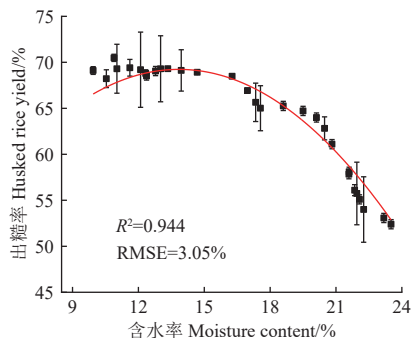


图 4 稻谷含水率与出糙率之间的拟合关系

Fig.4 Fitting relationship between paddy moisture content and husked rice yield

3.2.3 稻谷电学参数与出糙率之间的关系

为了获得电学参数与出糙率之间的数学关系，本研究分别建立了 1 MHz 条件下 6 个电学参数（介电常数、电容、介电损耗、损耗系数、导电率和电导）与出糙率之间的回归模型。由图 5 可知，6 个电学参数和出糙率的拟合决定系数从高到低分别为导电率（0.929）、电导（0.826）、介电常数（0.808）、电容（0.805）、损耗系数（0.708）、介电损耗（0.645）。相较于其他 5 个电学参数与出糙率的拟合模型，导电率与出糙率的拟合效果较好，决定系数最大为 0.929，RMSE 最小为 3.08%。由图 5 拟合曲线也可看到，稻谷的出糙率随着 6 个电学参数增加而不断减少。这可能是因为电学参数增加反映了稻谷的含水率也在不断增加，从而导致稻谷的出糙率持续降低。

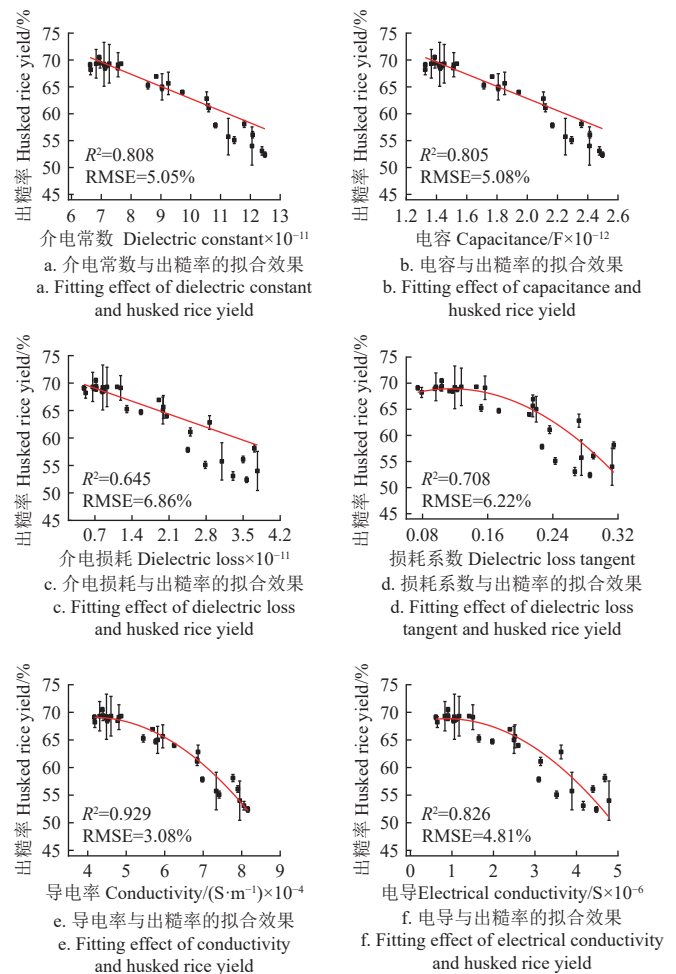


图 5 稻谷电学参数与出糙率之间的拟合关系

Fig.5 Fitting relationship between paddy electrical parameters and husked rice yield

3.3 机器学习模型的预测效果分析

为了提高模型的预测性能并充分地利用电学数据，本研究采用了 4 种机器学习模型（神经网络、决策树、随机森林和支持向量机）用于预测稻谷的含水率和出糙率。图 6 为机器学习模型对含水率的预测效果，可以看到 4 种机器学习算法预测含水率的决定系数从高到低分别为：神经网络（0.987）、决策树（0.956）、随机森林（0.934）、支持向量机（0.899）。4 种算法预测含水率

的 RMSE 由低到高分别为：神经网络（0.87%）、决策树（1.10%）、随机森林（1.43%）、支持向量机（1.78%）。同时，由图 7 可以看到 4 种机器学习模型预测出糙率的决定系数从高到低分别为：神经网络（0.935）、决策树（0.910）、支持向量机（0.847）、随机森林（0.816）。

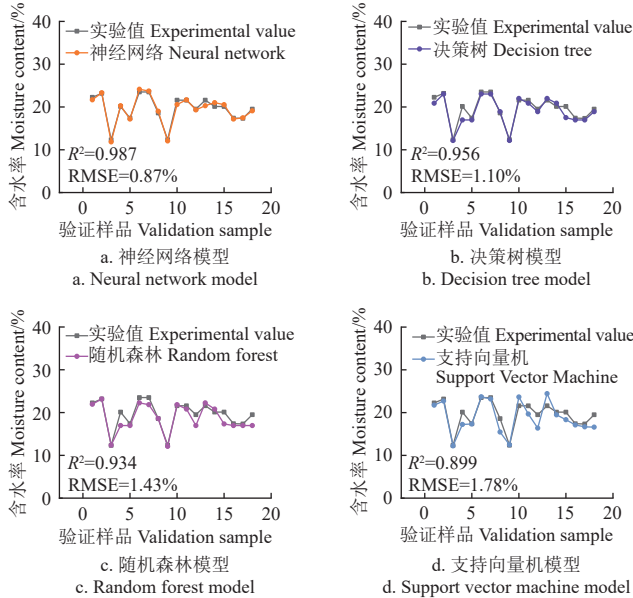


图 6 机器学习模型预测含水率的效果

Fig.6 The effect of machine learning models on predicting moisture content

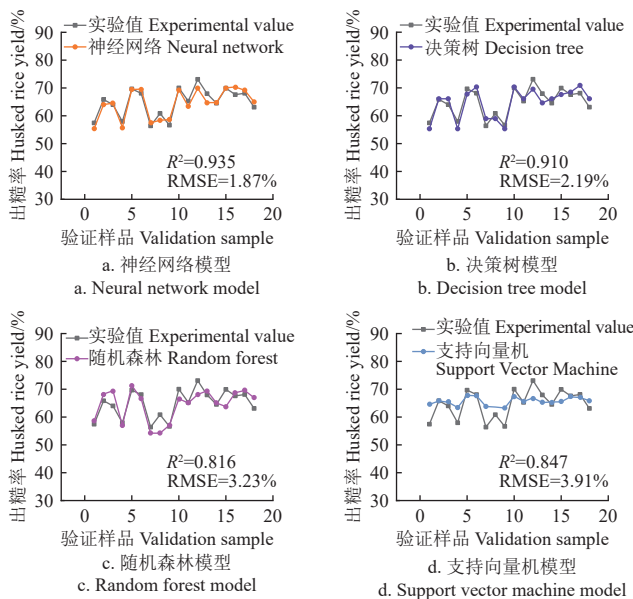


图 7 机器学习模型预测出糙率的效果

Fig.7 The effect of machine learning models on predicting husked rice yield

4 种算法预测出糙率的 RMSE 由低到高分别为：神经网络（1.87%）、决策树（2.19%）、随机森林（3.23%）、支持向量机（3.91%）。可以发现，相对于仅使用 1 MHz 电学参数来预测稻谷含水率和出糙率的回归模型，利用机器学习模型训练数据集获得的决定系数更高。机器学习的优势主要体现在两个方面。首先，机器学习模型能够更灵活地捕捉和表达电学参数与稻谷品质之间的非线性关系。其次，机器学习模型具备自动学

习和提取输入特征关系的能力，无需手动指定具体的函数形式。这使得在处理大量电学特征的情况下，机器学习模型能够自动选择并利用最相关的特征，从而提高模型的性能。机器学习模型相比郭文川等^[18]单独使用介电常数或者介电损耗建立多元方程，更能准确建立电学参数与稻谷品质之间的关系。此外，根据图 6 和图 7 的结果可以发现，相较于决策树、随机森林和支持向量机模型，神经网络模型在预测稻谷含水率和出糙率中表现最佳，其决定系数最高，RMSE 最小。这表明神经网络的强大的学习能力使其能够更好地适应数据的复杂模式，更好地分析了电学参数与稻谷含水率和出糙率之间的复杂非线性关系。此外，已有研究表明借助电学参数可以建立水分检测的传感器，以实现林木和苹果纸条的无损检测^[32-33]。因此，将神经网络模型与传感器相结合则能够更加快速和准确的预测稻谷的含水率和出糙率。

4 结 论

1) 本研究基于 LCR 测量仪，在 1~8 MHz 频率下测试了含水率为 9.94%~23.51% 稻谷的 6 个电学参数（介电常数、介电损耗、损耗系数、导电率、电导和电容），并使用砵谷机对稻谷进行脱壳处理。将获得的电学参数、含水率和出糙率数据集用于建立回归模型和机器学习模型，机器学习模型包括神经网络、决策树、支持向量机和随机森林，从而找出最佳的数学模型用于预测稻谷的含水率和出糙率。

2) 随着频率的增加，介电常数和电容先降低后升高；介电损耗一直在降低；损耗系数则先降低后升高，在 6 MHz 之后又下降；导电率和电导则一直在提高。6 个电学参数在稻谷含水率提高后都提高。在电学参数和含水率建立的回归模型中，含水率均随着 6 个电学参数的增加而提高。含水率与 6 个电学参数的决定系数从大到小分别为：导电率、电容、介电常数、电导、损耗系数、介电损耗。其中，导电率和含水率的拟合模型决定系数最高为 0.960，RMSE 为 0.89%。在含水率和出糙率拟合模型中，模型的决定系数为 0.944，RMSE 为 3.05%。稻谷的出糙率随着含水率先增后减，在含水率为 13.34% 时，出糙率达到最大值为 69.31%。在分析电学参数和出糙率的拟合模型时，可以发现 6 个电学参数与出糙率的拟合决定系数从大到小分别为：导电率、介电常数、电容、电导、损耗系数、介电损耗。其中，导电率与出糙率的拟合效果最佳，决定系数为 0.929，RMSE 为 3.08%。

3) 机器学习模型中神经网络在预测稻谷含水率和出糙率时均获得了最高的效果，拟合的决定系数分别为 0.987 和 0.935，比二次回归模型的效果更好。由此，基于本研究的内容和市场的需求，开发一种可以快速检测稻谷含水率和出糙率的方法将有巨大的应用市场。

【参 考 文 献】

- [1] 徐广超, 谢天. 中外大米质量标准对比分析及发展建议[J]. 粮油食品科技, 2019, 27(6): 35-39.
XU Guangchao, XIE Tian. Comparative analysis of rice quality standards at home and abroad and development suggestions[J].

- Science and Technology of Cereals, Oils and Foods, 2019, 27(6): 35-39. (in Chinese with English abstract)
- [2] ZHAO W X, CHOU J M, LI J N, et al. Impacts of extreme climate events on future rice yields in global major rice-producing regions[J]. *International Journal of Environmental Research and Public Health*, 2022, 19(8): 4437.
- [3] MOHIDEM N A, HASHIM N, SHAMSUDIN R, et al. Rice for food security: revisiting its production, diversity, rice milling process and nutrient content[J]. *Agriculture-Basel*, 2022, 12(6): 741.
- [4] 徐春春, 纪龙, 陈中督, 等. 2021 年我国水稻产业形势分析及 2022 年展望[J]. *中国稻米*, 2022, 28(2): 16-19.
XU Chunchun, JI Long, CHEN Zhongdu, et al. Analysis of China's rice industry in 2021 and the outlook for 2022[J]. *China Rice*, 2022, 28(2): 16-19. (in Chinese with English abstract)
- [5] TIAN T, ZHANG B, CHEN S B, et al. Characterization of differences between microwave and traditional thermal sterilization to prevent fungal spoilage during storage of high-moisture paddy rice[J]. *Cereal Chemistry*, 2020, 98(1): 154-163.
- [6] WANG R L, LIU L L, GUO Y P, et al. Effects of deterioration and mildewing on the quality of wheat seeds with different moisture contents during storage[J]. *RSC Advances*, 2020, 98(1): 154-163.
- [7] 古汶玉, 黄玉, 谢宇霞, 等. 稻谷水分对出糙率和整精米率的影响[J]. *粮食科技与经济*, 2020, 45(4): 80-81.
- [8] 宋琦. 基于电阻抗谱分析的稻谷水分传感器设计[D]. 镇江: 江苏大学, 2018.
SONG Qi. Design of Grain Moisture Sensor Based on Electrical Impedance Spectroscopy[D]. Zhenjiang: Jiangsu University, 2018. (in Chinese with English abstract)
- [9] 赵晶, 黄操军, 李博识. 谷物含水率检测方法综述[J]. *农业科技与信息*, 2018(16): 46-49.
- [10] 吕荣文, 包乌兰, 秦晓亮. 稻谷出糙率和出米率关系探讨[J]. *粮食储藏*, 2022, 51(1): 53-56.
- [11] 李长友. 稻谷干燥含水率在线检测装置设计与试验[J]. *农业机械学报*, 2008, 39(3): 56-59.
LI Changyou. Design and experiment of on-line moisture content metering device for paddy drying process[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2008, 39(3): 56-59. (in Chinese with English abstract)
- [12] 张伟, 杨刚, 雷军波, 等. 基于微波反射法的谷物含水率在线检测装置研制[J]. *农业工程学报*, 2019, 35(23): 21-28.
ZHANG Wei, YAGN Gang, LEI Junbo, et al. Development of on-line detection device for grain moisture content using microwave reflection method[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(23): 21-28. (in Chinese with English abstract)
- [13] 刘志壮, 吕贵勇. 基于电容法的稻谷含水率检测[J]. *农业机械学报*, 2013, 44(7): 179-182.
LIU Zhizhuang, LV Guiyong. Moisture content detection of paddy rice based on capacitance approach[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2013, 44(7): 179-182. (in Chinese with English abstract)
- [14] CHEN N N, CHEN M Q, FU B A, et al. Far-infrared irradiation drying behavior of typical biomass briquettes[J]. *Energy*, 2017, 121: 726-738.
- [15] 金志强, 张锦胜, 钱菲, 等. 利用核磁共振技术研究谷物在储藏过程中的变化[J]. *食品科学*, 2008, 29(9): 66-69.
JIN Zhiqiang, ZHANG Jinsheng, QIAN Fei. Study on changes of grains during storage by NMR[J]. *Food Science*, 2008, 29(9): 66-69. (in Chinese with English abstract)
- [16] ZHANG G, WANG Y, WANG W, et al. Rapid online determination of feed concentration in nitroguanidine spray drying process by near infrared spectroscopy[J]. *Infrared Physics and Technology*, 2020, 109: 103432.
- [17] 王丹阳, 王洁, 张本华, 等. 基于介电特性的稻谷干燥含水率检测模型构建与验证[J]. *沈阳农业大学学报*, 2022, 53(6): 752-758.
WANG Danyang, WANG Jie, ZHANG Benhua, et al. Construction and verification of detection model of moisture content for paddy rice drying based on dielectric properties[J]. *Journal of Shenyang Agricultural University*, 2022, 53(6): 752-758. (in Chinese with English abstract)
- [18] 郭文川, 王婧, 朱新华. 基于介电特性的燕麦含水率预测[J]. *农业工程学报*, 2012, 28(24): 272-279.
GUO Wenchuang, WANG Jing, ZHU Xinhua. Moisture content prediction of oat seeds based on dielectric property[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2012, 28(24): 272-279. (in Chinese with English abstract)
- [19] 俞世钢, 黄仕华. 基于电阻抗特性稻谷含水率测定研究[J]. *山西农业科学*, 2010, 38(12): 19-22.
YU Shigang, HUANG Shihua. Study on the grain moisture rate testing based on impedance characteristics[J]. *Journal of Shanxi Agricultural Sciences*, 2010, 38(12): 19-22. (in Chinese with English abstract)
- [20] 商亮, 谷静思, 郭文川. 基于介电特性及 ANN 的油桃糖度无损检测方法[J]. *农业工程学报*, 2013, 29(17): 257-264.
SHANG Liang, GU Jingsi, GUO Wenchuang. Non-destructively detecting sugar content of nectarines based on dielectric properties and ANN[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(17): 257-264. (in Chinese with English abstract)
- [21] KUBO M T K, CURET S, AUGUSTO P E D, et al. Artificial neural network for prediction of dielectric properties relevant to microwave processing of fruit juice[J]. *Journal of Food Process Engineering*, 2018, 41(6): e12815.
- [22] 国家卫生和计划生育委员会. GB 5009.3—2016 食品安全国家标准 食品中水分的测定[S]. 北京: 中国标准出版社, 2016.
- [23] 黄铝文, 田旭, 任烈弘, 等. 基于电学参数的苹果可溶性固形物含量预测[J]. *农业工程学报*, 2023, 39(2): 252-259.
HUANG Lvwen, TIAN Xu, REN Liehong, et al. Prediction of the soluble solid contents for apple fruit using electrical parameters[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(2): 252-259. (in Chinese with English abstract)
- [24] 国家质量监督检验检疫总局. GB/T 5495—2008 粮油检验 稻谷出糙率检验[S]. 北京: 中国标准出版社, 2008.
- [25] SAEED N, KHAYER Z R, VITALIEVICH P V, et al. Artificial intelligence models and employee lifecycle management: a systematic literature review[J]. *Organizacija*, 2022, 55: 181-198.
- [26] LIN X, LI C, HAO H, et al. Accelerated search for ABO₃-type the electronic contribution of polycrystalline dielectric constants by machine learning[J]. *Computational Materials Science*, 2021, 193: 110404.
- [27] CHANG H, WU Q, TIAN H, et al. Non-destructive identification of internal water core in apples based on online Vis/NIR spectroscopy[J]. *Transaction of the American Society of Agricultural and Biological Engineers*, 2020, 63(6): 1711-1721.
- [28] XU Y Y, XIANG P, QIU W Q, et al. Dielectric properties of the Maillard reaction solution formed between enzymatic

- hydrolysate of Antarctic krill and glucose under microwave heating[J]. *LWT-Food Science and Technology*, 2022, 161: 113355.
- [29] 张保艳, 于海洋, 程裕东, 等. 温度、频率和水分含量对罗非鱼介电特性的影响[J]. *水产学报*, 2012, 36(11): 1785-1792. ZHANG Baoyan, YU Haiyang, CHENG Yudong, et al. Effects of temperature, frequency, and moisture content on the dielectric properties of tilapia[J]. *Journal of Fisheries of China*, 2012, 36(11): 1785-1792. (in Chinese with English abstract)
- [30] 牛智有, 刘芳宏, 刘鸣, 等. 平行极板电容传感器介电式颗粒饲料水分检测仪设计与试验[J]. *农业工程学报*, 2019, 35(18): 36-43. NIU Zhiyou, LIU Fanghong, LIU Ming, et al. Design of dielectric pellet feed moisture detector based on parallel plate capacitance sensor[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(18): 36-43. (in Chinese with English abstract)
- [31] 袁军. 稻谷出糙率检验的研究[J]. *粮食问题研究*, 2022(6): 38-40.
- [32] 赵燕东, 顾家华, 赵玥, 等. 基于活立木介电特性的茎干蓄水量检测传感器研[J]. *农业工程学报*, 2022, 38(3): 38-46. ZHAO Yandong, GU Jiahua, ZHAO Yue, et al. Development of a Stem Water Storage Detection Sensor Using the Dielectric Properties of Living Standing Trees[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(3): 38-46. (in Chinese with English abstract)
- [33] 程相林, 向阳, 颜小飞, 等. 苹果树枝条含水率无损测量传感器研制[J]. *农业工程学报*, 2022, 38(13): 30-37. CHENG Xianglin, XIANG Yang, YAN Xiaofei, et al. Development of the Nondestructive Sensor for Measuring the Moisture Content of Apple Branches[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(13): 30-37. (in Chinese with English abstract)

Predicting moisture content and husked rice yield using electrical parameters and machine learning

DENG An , XU Yongyang , QIU Weiqiang , LI Li , JIN Yinzhe*

(1. College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Engineering Research Center of Food Thermal-Processing Technology, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Rice has been one of the most important crops in the world. Among them, the rapid on-site testing of husked rice yield and moisture content can greatly facilitate to acquisition and grade in the paddy field, thereby reducing the losses during transportation, storage, and processing. This study aims to identify the optimal model for the prediction of the moisture content and husked rice yield of paddy. An LCR meter was used to test six electrical parameters (dielectric constant, capacitance, dielectric loss, dielectric loss tangent, conductivity, and electrical conductivity) of paddy samples with the moisture content ranging from 9.94% to 23.51% at frequencies between 1 and 8 MHz. Subsequently, a rice huller was used to dehusk the paddy samples, in order to determine the husked rice yield. The datasets were then obtained from the electrical parameters, moisture content, and husked rice yield. The electrical parameters at 1 MHz were used to train the regression models, in order to predict the moisture content and husked rice yield of paddy. All electrical parameters at 1-8 MHz were used to train four machine learning models: neural networks, decision trees, support vector machines, and random forests. The models were evaluated for correlation and accuracy using determination coefficients (R^2) and root mean square errors (RMSE). Results indicated that the dielectric constant and capacitance initially decreased and then increased with increasing frequency. While the dielectric loss decreased consistently. The dielectric loss tangent initially decreased, then increased, and finally decreased again after 6 MHz. Conductivity and electrical conductivity increased consistently. All six electrical parameters increased with the rising moisture content. The regression model was also established using electrical parameters and moisture content. Therefore, the moisture content also increased under the condition of 1 MHz, as the six electrical parameters increased. The determination coefficients of moisture content with the six electrical parameters were ranked in descending order of conductivity, capacitance, dielectric constant, electrical conductivity, dielectric loss tangent, and dielectric loss. Among them, the fitting model with the highest determination coefficient for the conductivity and moisture content was 0.960, with an RMSE of 0.89%. In the fitting model of moisture content and husked rice yield, the determination coefficient was 0.944, with an RMSE of 3.05%. The husked rice yield of paddy increased first and then decreased with rising moisture content. Moreover, the husked rice yield reached the maximum of 69.31% at a moisture content of 13.34%. According to the fitting model of electrical parameters and husked rice yield, the determination coefficients of the six electrical parameters with the husked rice yield were ranked in the descending order of the conductivity, dielectric constant, capacitance, electrical conductivity, dielectric loss tangent, and dielectric loss. Notably, the best fitting was achieved in the conductivity and husked rice yield, with a determination coefficient of 0.929 and an RMSE of 3.08%. Additionally, the neural network achieved the highest performance in predicting the paddy moisture content and husked rice yield among the machine learning models, with the determination coefficients of 0.987 and 0.935, and RMSEs of 0.87% and 1.87%, respectively, indicating the best regression. As such, the neural network model can effectively establish the relationship between electrical parameters and rice quality, thus accurately predicting the moisture content and husked rice yield of paddy. Machine learning was highlighted for the non-destructive testing of agricultural products, in terms of speed and efficiency. The rapid and accurate assessment was also performed on the key quality indicators, such as moisture content and husked rice yield. Meanwhile, this convenient and efficient approach can greatly enhance the efficiency of paddy management for cost savings in the supply chain. The valuable theoretical insights can also provide for the non-destructive quality assessment of the paddy, particularly in the broader application of advanced analytical techniques in agricultural science and industry.

Keywords: machine learning; moisture content; electrical parameters; regression model; paddy; husked rice yield