

基于机器学习的机械压实对大豆产量的影响预测研究

周修理¹ 秦娜¹ 王开宇¹ 孙浩² 王大维² 乔金友²

(1. 东北农业大学电气与信息学院, 哈尔滨 150030; 2. 东北农业大学工程学院, 哈尔滨 150030)

摘要: 为评估农业机械作业对大豆产量的影响, 本文开展不同机型、不同压实次数的拖拉机压实试验, 获取不同压实环境中的土壤物理性质和大豆产量数据, 分别从影响大豆产量的机械因素、土壤因素和复合因素出发, 使用多元线性回归 (Multiple linear regression, MLR)、随机森林 (Random forest, RF)、自适应增强模型 (Adaptive boosting, AdaBoost)、人工神经网络 (Artificial neural network, ANN) 4 种机器学习算法建立大豆产量影响预测模型, 对模型性能及模型特征重要性进行综合分析。研究表明, 机械作业与大豆产量间关系复杂, 集成学习算法 (AdaBoost 和 RF) 所建立的模型具有更好的拟合效果, 模型决定系数更高; 利用复合因素对大豆产量建立的模型拟合度最高, 其次为机械因素和土壤因素, 其中基于 AdaBoost 的复合因素对大豆产量影响模型其拟合程度最优, 其 R^2 为 0.92, MAE 为 1.33%, RMSE 为 1.86%; 机械因素、土壤因素都会影响大豆产量, 其中机械压实次数以及表层和亚表层的土壤坚实度为影响大豆产量的重要因素, 在实际生产中可通过减少机械作业次数、疏松表层及亚表层土壤来改善机械压实影响。

关键词: 机械压实; 大豆产量; 机器学习; 土壤坚实度; 预测模型

中图分类号: TP181; S233.1

文献标识码: A

文章编号: 1000-1298(2023)11-0139-09

OSID:



Effect of Mechanical Compaction on Soybean Yield Based on Machine Learning

ZHOU Xiuli¹ QIN Na¹ WANG Kaiyu¹ SUN Hao² WANG Dawei² QIAO Jinyou²

(1. College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China

2. College of Engineering, Northeast Agricultural University, Harbin 150030, China)

Abstract: Aiming to find a more accurate method to assess the effect of agricultural machinery compaction on soybean yield, data of soil physical properties and soybean yield in different compaction environments were obtained by carrying out different numbers of compaction walks with different types of tractors. Soybean yield forecast models were developed from mechanical factors, soil factors, and composite factors which affected soybean growth, respectively. To find out the differences of models built by different types of machine learning algorithms, multiple linear regression (MLR), random forest (RF), adaptive boosting (AdaBoost), and artificial neural network (ANN) were used in modeling. In addition, the importance of model features was comprehensively analyzed. The results showed that the relationship between mechanical operation and crop yield was complex, and the models built by integrated learning algorithms (AdaBoost and RF) had a better fit and higher coefficient of determination. Among the machine learning algorithms used, the best performance of the models built was AdaBoost, followed by random forest, artificial neural network and multiple linear regression. The model built using composite factors for soybean yield had the best fit, followed by mechanical and soil factors. The AdaBoost-based composite factor for soybean yield forecast model had the optimal fit with R^2 of 0.92, MAE of 1.33% and RMSE of 1.86%. Mechanical factors and soil factors all had an effect on the variation of soybean yield. The number of mechanical compaction, soil penetration resistance in the surface and subsurface layers were the important factors affecting soybean yield. Therefore, the effects from mechanical compaction can be relieved by reducing the number of mechanical operation and

收稿日期: 2023-06-14 修回日期: 2023-08-18

基金项目: 国家重点研发计划项目(2021YFD2000405-2)和财政部和农业农村部: 国家现代农业产业技术体系项目(CARS-04-PS24)

作者简介: 周修理(1966—), 男, 副教授, 主要从事智能检测与控制研究, E-mail: xlzhou@neau.edu.cn

通信作者: 乔金友(1969—), 男, 教授, 主要从事农业机械化生产与管理研究, E-mail: qiaojinyou@163.com

loosening soil penetration resistance of the surface and subsurface soils.

Key words: mechanical compaction; soybean yield; machine learning; soil penetration resistance; prediction models

0 引言

随着农业生产水平的发展,全球农业机械化率逐年提高,农业机械在农业生产各环节中的使用频率大幅增加,在提高生产效率的同时,也带来了机械压实危害^[1]。机械作业带来的压实会导致土壤结构变形,土壤理化性质恶化,破坏作物的最佳生长环境,最终导致作物减产^[2-4]。机械作业导致的土壤压实逐渐成为制约粮食产量提高的重要因素,研究机械作业与产量间的关系,并实现机械压实环境下的大豆产量评估具有重要应用价值。

机械作业、土壤性质及作物产量间存在复杂的作用关系^[5-6]。机械压实对作物产量的影响一般从两方面展开研究,机械作业带来的土壤环境改变进而影响作物生长发育^[7-8],以及机械作业对作物产量的间接影响^[9]。土壤坚实度可以用于表征土壤抵抗外力的能力^[10],并且土壤坚实度的变化对作物的生长发育也存在着重要影响^[11],鉴于此在研究机械作业对土壤性质以及对作物产量的影响时,常选用土壤坚实度作为主要的土壤物理性质。机械压实会造成土壤坚实度增大^[12-14],从而阻碍作物根系的伸长^[15-16],不利于作物对土壤中养分的吸收,降低出苗率^[17],致使作物减产。CORREA等^[18]研究高粱作物在不同土壤环境的生长状况,发现土壤压实对作物根系的影响大于对枝条的影响。DE MORAES等^[19]的研究指出,不同深度土壤坚实度的变化对作物产量的影响也不同。在机械作业对作物产量的间接影响研究中,SIVARAJAN等^[7]和REINTAM等^[20]均指出压实机械、压实次数的不同,对土壤特性及作物产量的影响也不同。MULAZZANI等^[21]指出,评估土壤压实与作物产量间的关系,对改善作物生长环境,调整实际农业生产具有重要意义。

近几年,随着人工智能技术的不断发展,有学者开始将机器学习应用于机械压实领域。ELAoud等^[22]利用土壤含水率、土壤密度、拖拉机质量、压实次数和车轮压力建立土壤坚实度的多元线性回归预测模型。GHADERNEJAD等^[23]利用自适应神经模糊推理系统建立机械作业以及农家肥施用量对土壤坚实度的影响模型。BEN HASSEN等^[24]利用拖拉机质量、压实次数、接地比压等特征构建了基于贝叶斯网络的土壤压实风险评估决策支持系统。

KHANAL等^[25]通过收集压实环境中作物的遥感数据,结合线性回归、随机森林等机器学习算法,建立了基于遥感图像的玉米产量预测模型。李民赞等^[26]设计了基于树莓派和SVM、GRNN的Bagging集成模型预测表土层土壤容重。周修理等^[27]利用GA-RF算法建立了土壤坚实度对大豆产量影响的分类模型。

虽然众多学者已开展大量研究试验,但仍存在如下问题:①目前的研究方法多为数据分析,不能更好地挖掘数据间的关系。②目前,已引入机器学习方法的研究多为对土壤理化性质的研究,鲜有对作物产量的研究,无法评估机械压实对作物产量的影响,并且缺乏对算法广度的研究,未指明哪类算法更适用于机械压实领域。③在研究内容上,多考虑土壤因素对作物的直接影响,忽略了机械本身对作物的间接影响,也未将不同因素对大豆产量的影响进行对比。

本文结合机械压实对作物产量的影响关系,分别从影响作物产量的机械因素、土壤因素和复合因素(包含机械因素和土壤因素)出发,使用线性回归、集成学习和神经网络3类机器学习算法,建立大豆产量评估模型,分析不同因素对大豆产量的影响,对比不同类型算法所建立大豆产量影响模型的预测能力,以期对寒地黑土区大豆科学化种植提供理论参考。

1 材料与方法

1.1 试验区概况

机械压实试验地点位于黑龙江省哈尔滨市东北农业大学向阳试验基地(44°04'N,125°42'E),土壤类型为黑土,土壤质地为粉壤土。试验区处于第一积温带,属于中温带大陆季风性气候,年平均温度为3.5~4.5℃,大于10℃有效积温为2600~2700℃,年均降雨量为400~600mm,集中在7—8月,无霜期为135~145d。

1.2 数据获取

选用试验所在地区具有代表性的小型、中型和大型3类拖拉机(拖拉机基本参数如表1所示)进行土壤压实试验,依据试验地区大豆全程机械化生产工艺过程,每种类型拖拉机分别实施2、4、6、8、10、12次机械压实行走,模拟实际生产中旋耕起垄、播种、封闭除草、中耕追肥、收获等机械化作业。为

降低操作误差,每种压实处理重复 3 次。同时在压实地块两侧设不压实对照组(CK)。试验地块总宽度 49.6 m、总长度 162 m。试验地块播种的大豆品种为东农 252。拖拉机压实作业如图 1 所示。

表 1 压实拖拉机参数
Tab.1 Tractor parameters for compaction test

类型	型号	使用质量/ kg	功率/kW	轮距/m	前轮胎宽/ m	后轮胎宽/ m	前轮接地 面积/m ²	后轮接地 面积/m ²	接地比压/ Pa
小型	JD-280	1 500	20.59	1.30	0.10	0.22	0.030	0.154	81 521.7
中型	JD-904	4 150	66.18	1.95	0.22	0.22	0.176	0.198	110 962.6
大型	CASE-210	7 300	154.41	1.95	0.56	0.60	0.448	0.516	75 726.1



图 1 拖拉机压实作业

Fig.1 Compaction operations of tractors

选用土壤坚实度作为评估机械压实水平和影响作物生长的土壤物理性质代表参数。机械压实后,采用荷兰 Eijkelkamp 公司生产的 PV6.08 型土壤坚实度测定仪,在各试验区测取深度 0~80 cm 的土壤坚实度。待大豆成熟后,在对应土壤坚实度测试小区压实车辙两侧随机选取植株长势均匀且具有代表性的 2 m²测产区域,经收割、装袋、人工脱粒后获取大豆产量样本,采用干燥法测取大豆籽粒含水率,并计算标准含水率条件下的大豆产量^[28]。

1.3 数据预处理

1.3.1 土壤坚实度处理与产量变化率的计算

对试验所获取的土壤坚实度和大豆产量进行处理。将总深度为 80 cm 的土壤坚实度数据每 10 cm 取均值,得到 0~10 cm,10~20 cm,⋯,70~80 cm 共 8 个深度的土壤坚实度特征。不同压实处理下的大豆产量均低于对照组,根据对照组产量平均值,计算出各个测产区域的产量变化率,计算式为

$$c_i = \left| \frac{m_i - m'}{m'} \right| \times 100\% \tag{1}$$

式中 m_i ——地块 i 的大豆产量
 m' ——CK 组大豆平均产量

1.3.2 数据归一化

为消除不同特征的量纲差异对模型预测精度的影响,采用归一化将特征范围变换至[0,1]区间,计算式为

$$Z_{new} = \frac{Z_i - Z_{min}}{Z_{max} - Z_{min}} \tag{2}$$

式中 Z_i ——第 i 个样本
 Z_{max} ——样本最大值
 Z_{min} ——样本最小值

1.3.3 异常值处理

为减少因环境因素或人为因素对数据造成的误差,采用 3 σ 法则对所测取的土壤坚实度数据进行处理。

$$\mu - 3\sigma < X < \mu + 3\sigma \tag{3}$$

式中 μ ——样本均值
 σ ——样本标准差

满足式(3)为正常值,否则为异常值,予以修改或删除。

1.4 预测模型

为评估不同类型机器学习算法所建立压实环境中大豆产量影响模型的差异,探寻预测精度更高的模型,选用农业领域中常用的线性回归、集成学习和深度学习 3 类机器学习算法。根据本文数据特点和算法的优缺点对各类算法进行选择,其中线性回归选用多元线性回归算法,深度学习选用人工神经网络算法。集成学习通过组合多个弱学习器能够显著提高整体的性能、减少模型因数据噪声造成的误差,但因在训练过程中需要多个弱学习器,其实现过程较多元线性回归算法相比更复杂,可解析性更差。集成学习算法常用的集成方式有 Bagging 和 Boosting 两种,分别对应所选用的随机森林算法和 AdaBoost 算法。模型训练过程中采用网格搜索和 5

折交叉验证寻找最优参数。利用不同因素所建模型的输入特征及符号表示如表2所示。

表2 输入特征及符号表示
Tab.2 Input features and symbol representation

类型	数据字符	说明
机械因素	times	压实次数
	pressure	机械接地比压, Pa
	area1	前轮接地面积, m ²
	area2	后轮接地面积, m ²
	weight	机械质量, kg
土壤因素	p_1	0 ~ 10 cm 土层土壤坚实度, MPa
	p_2	10 ~ 20 cm 土层土壤坚实度, MPa
	p_3	20 ~ 30 cm 土层土壤坚实度, MPa
	p_4	30 ~ 40 cm 土层土壤坚实度, MPa
	p_5	40 ~ 50 cm 土层土壤坚实度, MPa
	p_6	50 ~ 60 cm 土层土壤坚实度, MPa
	p_7	60 ~ 70 cm 土层土壤坚实度, MPa
	p_8	70 ~ 80 cm 土层土壤坚实度, MPa
复合因素	area1	前轮接地面积, m ²
	area2	后轮接地面积, m ²
	weight	机械质量, kg
	pressure	机械接地比压, Pa
	times	压实次数
	p_1	0 ~ 10 cm 土层土壤坚实度, MPa
	p_2	10 ~ 20 cm 土层土壤坚实度, MPa
	p_3	20 ~ 30 cm 土层土壤坚实度, MPa
	p_4	30 ~ 40 cm 土层土壤坚实度, MPa
	p_5	40 ~ 50 cm 土层土壤坚实度, MPa
	p_6	50 ~ 60 cm 土层土壤坚实度, MPa
	p_7	60 ~ 70 cm 土层土壤坚实度, MPa
	p_8	70 ~ 80 cm 土层土壤坚实度, MPa

1.4.1 多元线性回归

多元线性回归是线性回归模型的一种,具有简单易用、可解释性强的特点,与本文所选用的其他算法相比,对自变量间的相关性更敏感。多元线性回归在农业研究领域具有广泛应用^[29-30],其一般形式为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (4)$$

式中 Y ——输出变量

X_i ——第 i 个自变量

β_i ——第 i 个自变量对应的系数

1.4.2 随机森林

随机森林是一种 Bagging 类的有监督集成学习算法,通过集成多个决策树以提高预测准确率,具有抗过拟合能力强和预测精度高的特点^[31]。在训练时,首先用自助采用法对数据集进行采样,使每棵决策树有属于自己的一个数据子集,再设置决策树模型参数,在每个数据子集上训练决策树模型。最后对所有决策树模型进行投票,得到结果。

1.4.3 AdaBoost

AdaBoost 是一种 Boosting 类的集成学习算法,其实现原理与随机森林不同^[32]。AdaBoost 以高度自适应的方式依次训练弱学习器,不对数据集进行重采样,而是针对同一数据集迭代训练弱学习器,通过对每一样本赋予不同的权重,使弱学习器的关注度不断地偏向误差大的样本,最终得到强学习器。AdaBoost 算法具预测精度高、自适应能力强的特点,并且对于弱学习器的选择较为灵活。本文选用最常用的决策树作为弱学习器。

1.4.4 人工神经网络

人工神经网络^[33]是仿造生物神经网络进行函数拟合的一种深度学习算法,具有较强的非线性建模能力、并行处理能力、容错性和自动化智能学习的特点。人工神经网络由输入层、隐藏层和输出层3层组成,其中输入层中有一组神经元节点,每个节点代表一个输入变量,输出层为预测对象。隐藏层可以有一个或者多个,隐藏层连接着输入层和输出层,节点间的连接表示通过该路径权重。

1.5 特征重要性

模型中特征的相对重要性可通过评估模型中每个变量的贡献来量化。多元线性回归模型中,特征的相对重要性可由标准化回归系数来表示,其绝对值越大,则该特征越重要^[34]。在随机森林模型中,当一个预测值变量被置换而其他变量不变时,可通过计算袋外样本的预测精度的增加来获得变量的重要性^[35]。AdaBoost 模型在训练时,决策树以基尼系数选取最优划分属性,因此可将基尼系数作为特征重要性的度量标准^[36]。人工神经网络中的特征相对重要性可利用输入层和隐藏层及输出层之间的全连接值计算^[37-38]。为方便对不同模型中特征重要性进行对比分析,将重要性得分按特征数量进行排序,其中排序数越小表示变量越重要,越大表示变量越不重要。

1.6 模型评价

采用均方根误差 (Root mean square error, RMSE)、决定系数 (Coefficient of determination, R^2)、平均绝对值误差 (Mean absolute error, MAE) 和平均绝对百分比误差 (Mean absolute percentage error, MAPE) 作为模型评价指标。均方根误差用于评估模型的准确性,其值越接近于0,模型的预测效果越精确;决定系数反映模型拟合优度,其值越接近1,模型拟合效果越好;平均绝对误差表示预测值和真实值之间绝对误差的平均值,是一种线性分数;平均绝对百分比误差越接近0说明模型越完美,当值大于1时一般认为模型为劣质模型。

2 结果与分析

试验总共收集到 180 条数据,经异常值处理后剩余 175 条,按照比例 4:1 划分为训练集和测试集,使用 MLR、AdaBoost、RF 和 ANN 4 种机器学习算法分别建立机械因素对大豆产量的影响预测模型、土壤因素对大豆产量的影响预测模型以及复合因素对大豆产量的影响预测模型。其中,AdaBoost、RF、ANN 模型寻优后参数如表 3 所示。

表 3 模型参数网格搜索寻优结果
Tab.3 Grid search optimization results of model parameters

参数	寻优范围	基于机 械因素 的大豆 产量影 响预测 模型	基于土 壤因素 的大豆 产量影 响预测 模型	基于复 合因素 的大豆 产量影 响预测 模型
RF	决策树个数	[2,100]	64	18
	决策树最大深度	[2,8]	5	6
	叶子节点最小样本数	[1,8]	4	2
AdaBoost	决策树个数	[2,100]	54	76
	决策树最大深度	[2,8]	5	6
ANN	隐藏层个数	[1,2]	1	1
	隐藏层神经元个数	[2,100]	97	67

2.1 大豆产量影响预测模型结果分析

利用机械因素建立的大豆产量影响预测模型拟合效果如图 2 所示。由图 2 可知,多元线性回归模型在训练集和测试集中预测性能均最弱,决定系数最低(分别为 0.58、0.42),平均绝对误差最高(分别为 3.07%、3.71%),均方根误差最高(分别为 3.86%、4.56%),并且测试集中 MAPE 大于 1。AdaBoost 和 RF 都属于集成算法,决定系数、平均绝

对误差和均方根误差表现均较优秀,但 RF 算法所建立的模型测试集和训练集中 MAPE 均大于 1。人工神经网络较 AdaBoost 和 RF 集成模型相比,决定系数更低且训练集与测试集中的波动更大,预测效果不如上述两集成学习模型。所有模型中 AdaBoost 模型拟合度更高,其最优拟合函数为 $y = 0.921x + 1.034$ 。

基于土壤因素的大豆产量影响预测模型拟合效果如图 3 所示。由图 3 可知,不同模型的测试集决定系数在 0.30~0.64 之间,均方根误差在 3.83%~5.37% 之间,平均绝对误差在 3.28%~4.67% 之间,整体表现均不如使用机械因素所建立的模型。在所有模型中多元线性模型仍然表现最差,其 R^2 为 0.30, RMSE 为 5.37%, MAE 为 4.67%。人工神经网络模型较多元线性回归模型性能有所提升,但其预测结果仍不理想。AdaBoost 和 RF 模型较其他模型相比,性能有较大提升,但模型在测试集中的表现远不如训练集,均存在过拟合现象。

利用复合因素所建立的大豆产量影响预测模型拟合效果如图 4 所示,与机械因素和土壤因素所建模型相比,复合因素所建立各模型的拟合效果都得到了不同程度的提升,其中,基于 AdaBoost 的复合因素对大豆产量的影响预测模型各指标表现最优,MAPE 在所有模型中最低,与利用机械因素所建立的模型相比, R^2 提升 2.22%, RMSE 降低 1.06%, MAE 降低 16.88%;与利用土壤因素所建立的模型相比, R^2 提升 43.75%, RMSE 降低 51.43%, MAE 降低 59.45%。RF 和 ANN 的拟合效果也保持在较高水平,其测试集的 R^2 分别为 0.87 和 0.86, RMSE 分别为 2.31% 和 2.39%。多元线性回归模型较其余模型相比,模型拟合能力仍较差,但也有较大提升。在对机械压实环境下的作物产量进行模型建立时,

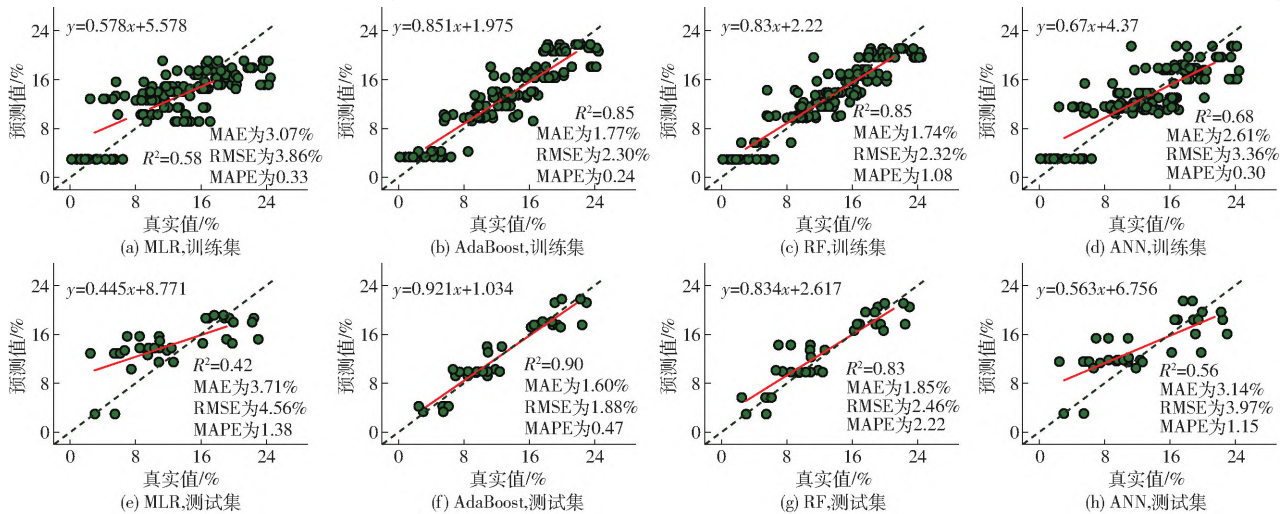


图 2 基于机械因素的大豆产量影响预测模型效果拟合曲线

Fig.2 Effect fitting curves of soybean yield prediction model based on mechanical factors

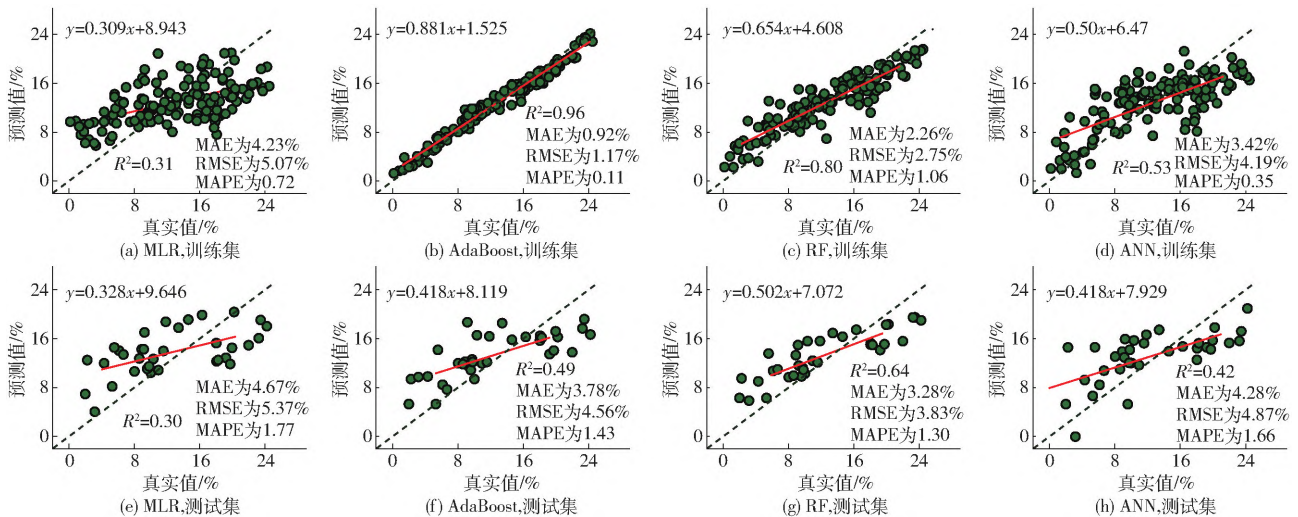


图3 基于土壤因素的大豆产量影响预测模型效果拟合曲线

Fig. 3 Effect fitting curves of soybean yield prediction model based on soil factors

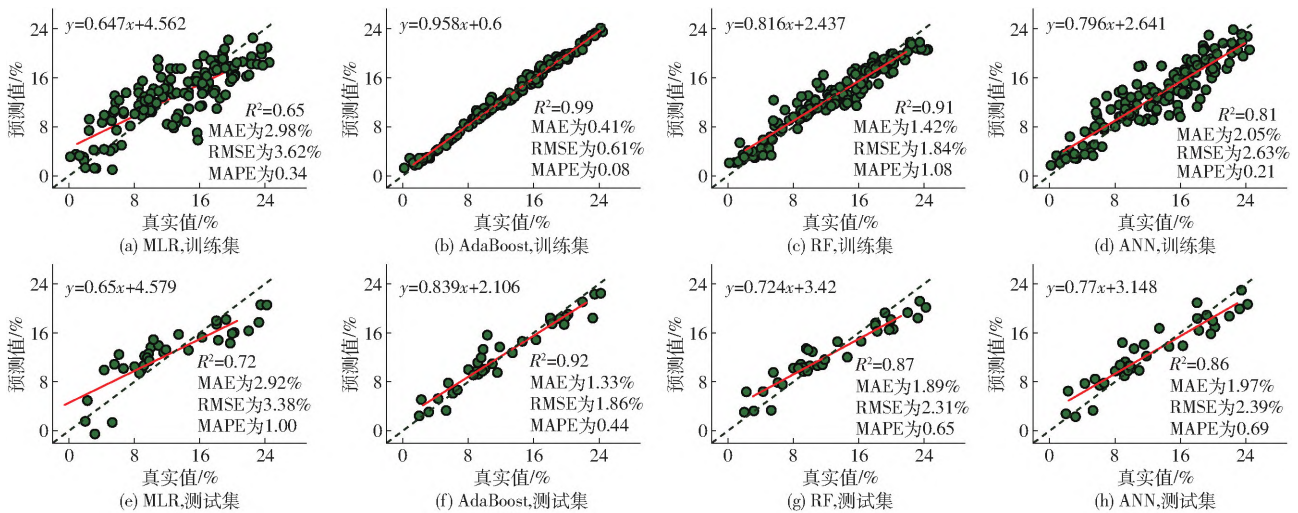


图4 基于复合因素的大豆产量影响预测模型效果拟合曲线

Fig. 4 Effect fitting curves of soybean yield prediction model based on composite factors

同时考虑机械和土壤的影响能够建立更加准确的大豆产量预测模型。

通过综合分析由不同影响因素所建立的大豆产量影响预测模型的性能,可发现机械作业、土壤性质与作物产量间关系复杂,机械因素、土壤因素与大豆产量间并非简单线性关系,无法利用多元线性回归算法建立精确的大豆产量影响模型。ANN作为4种算法中的唯一深度学习算法,其所建立的模型表现并非最优。集成学习算法(包括AdaBoost和随机森林)能够更精准地处理数据间的复杂关系,提高模型的预测性能,所建立的大豆产量影响模型表现更优秀,AdaBoost较RF所建立模型相比,预测精度更高,模型误差更小,模型更优异。所建立的模型中,基于AdaBoost算法的复合因素对大豆产量的影响预测模型预测精度最高,模型拟合效果最好,其 R^2 可达0.92, RMSE为1.86%, MAE为1.33%, MAPE为0.44。通过对比3类机器学习算法可发

现,集成学习算法更适用于本文的机械压实对大豆产量影响的研究。

2.2 模型特征重要性分析

由图5可以看出,从机械因素研究其对大豆产量的影响,所有模型都将压实次数作为影响大豆产量变化的首要因素,这与文献[9]的结论一致,此现象与压实对土壤的影响有关,压实次数同样是影响土壤坚实度变化的重要因素^[15],再次证实了土壤坚实度的增大不利于作物根系的生长,进而影响产量。综合4个模型的特征重要性结果,接地比压对大豆产量的影响高于拖拉机质量的影响。值得注意的是,拖拉机前轮面积对大豆产量的影响高于后轮面积的影响,可为农业机械设计提供参考。

由图6可以看出,在利用土壤坚实度建立大豆产量影响预测模型时,0~40 cm的土壤坚实度对大豆产量的影响高于40~80 cm土壤坚实度对大豆产

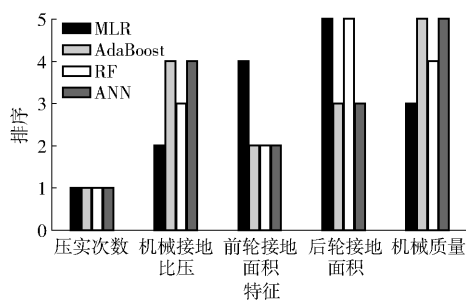


图 5 基于机械因素的建模特征重要性排序

Fig. 5 Feature importance ranking of modeling based on mechanical factors

量的影响, OBOUR 等^[8]的研究也得出了此结论, 这与大豆根系的生长发育有重要关系^[19], 也进一步说明了土壤条件的变化对作物产量的影响主要是通过影响作物根系形成的。RF、AdaBoost、ANN 都将 0 ~ 10 cm 土壤坚实度(p_1)作为影响大豆产量变化的重要特征, 而在多元线性回归模型中该特征重要性排序则比较靠后, 表明土壤坚实度与大豆产量变化并非简单的线性关系。综合不同深度土壤坚实度的特征排序, 应该重视表层和亚表层的土壤疏松工作, 以降低机械压实影响。

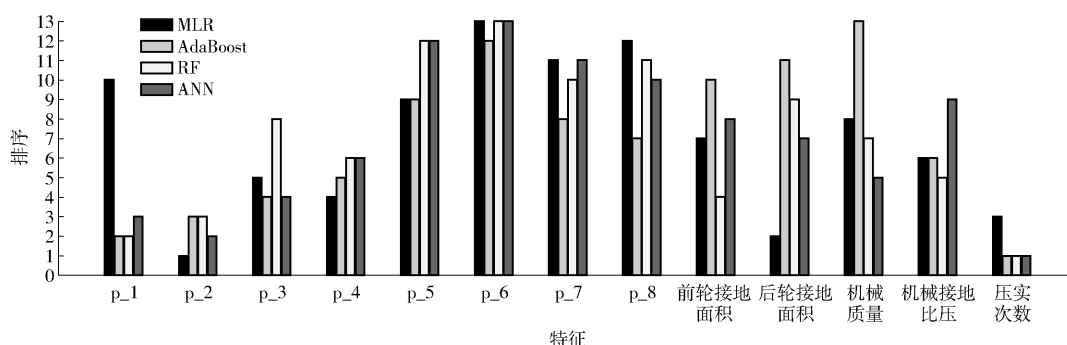


图 7 基于复合因素的建模特征重要性排序

Fig. 7 Feature importance ranking of modeling based on composite factors

通过综合分析不同模型的特征重要性, 发现机械压实环境下的机械因素与土壤因素均对大豆产量存在影响。在研究的所有影响大豆产量的因素中, 机械压实次数以及表层和亚表层的土壤坚实度是影响大豆产量的首要因素。对比不同因素所建立的模型, 使用复合因素所建立的大豆产量预测模型的表现最优, 其次是机械因素, 仅使用土壤因素时表现最差。仅使用机械因素的建模效果优于仅使用土壤因素的建模效果, 表明了压实机械的不同以及压实次数的不同会引起大豆产量较大程度的变动。

3 结论

(1) 使用复合因素建立的大豆产量预测模型最佳, 其次为机械因素、土壤因素。

(2) 与多元线性回归、随机森林、人工神经网络

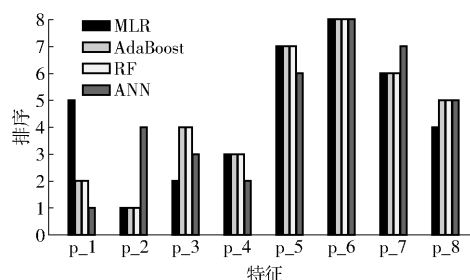


图 6 基于土壤因素的建模特征重要性排序

Fig. 6 Feature importance ranking of modeling based on soil factors

由图 7 可以看出, 同时考虑机械因素和土壤因素研究其对大豆产量变化率的影响时, 不同模型的特征重要性排序有较大差距, 其中所有模型中拖拉机压实次数以及 10 ~ 20 cm 土壤坚实度(p_2)的特征重要性排序较靠前。RF 和 AdaBoost 都为集成学习模型, 但前者的特征排序中机械因素更靠前, 而后者土壤因素更靠前。土壤因素中, 0 ~ 40 cm 土壤坚实度的重要性仍然更高, 这与仅考虑土壤因素时的结论一致。与仅考虑机械因素时相比, 压实次数的重要性仍旧排在前列, 接地比压的重要性相对车轮面积的重要性增强。

算法相比, AdaBoost 算法在对机械压实环境下的大豆产量建立影响预测模型方面表现更优秀, 其中模型中基于 AdaBoost 的复合因素对大豆产量的影响预测模型表现最优异, R^2 为 0.92, MAE 为 1.33%, RMSE 为 1.86%, MAPE 为 0.44。

(3) 机械因素中机械压实次数以及土壤因素中深度 0 ~ 40 cm 的土壤坚实度为影响黑土区大豆产量变化的重要因素。使用机械因素和使用土壤因素所建模型的特征重要性结论与使用复合因素时所得结论基本一致。

(4) 在机械作业过程中尽可能减少压实次数, 同时通过降低拖拉机质量、增加轮胎数量等措施降低机械接地比压, 对于压实后的土壤, 做好表层和亚表层土壤疏松工作, 降低土壤坚实度, 以达到提高作物产量的目的。

参 考 文 献

- [1] 乔金友, 张丹, 张宏彬, 等. 大中小型拖拉机压实对土壤坚实度和大豆产量的影响[J]. 农业工程学报, 2019, 35(21): 26–33.
QIAO Jinyou, ZHANG Dan, ZHANG Hongbin, et al. Effects of soil compaction by various tractors on soil penetration resistance and soybean yields[J]. Transactions of the CSAE, 2019, 35(21): 26–33. (in Chinese)
- [2] SHAH A N, TANVEER M, SHAHZAD B, et al. Soil compaction effects on soil health and crop productivity an overview[J]. Environmental Science and Pollution Research, 2017, 24(11): 10056–10067.
- [3] GREGORICH E G, LAPEN D R, MA B L, et al. Soil and crop response to varying levels of compaction, nitrogen fertilization, and clay content[J]. Soil Science Society of America Journal, 2011, 75(4): 1483–1492.
- [4] 张兴义, 隋跃宇. 土壤压实对农作物影响概述[J]. 农业机械学报, 2005, 36(10): 161–164.
ZHANG Xingyi, SUI Yueyu. Summarization on the effect of soil compaction on crops[J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(10): 161–164. (in Chinese)
- [5] NAWAZ M F, BOURRIE G, TROLARD F. Soil compaction impact and modelling. A review[J]. Agronomy for Sustainable Development, 2013, 33(2): 291–309.
- [6] HEYDARI L, BAYAT H, CASTRIGNANÒ A. Scale-dependent geostatistical modelling of crop-soil relationships in view of precision agriculture[J]. Precision Agriculture, 2023, 24(4): 1261–1287.
- [7] SIVARAJAN S, MAHARLOOEI M, BAJWA S G, et al. Impact of soil compaction due to wheel traffic on corn and soybean growth, development and yield[J]. Soil and Tillage Research, 2018, 175: 234–243.
- [8] OBOUR P B, UGARTE C M. A meta-analysis of the impact of traffic-induced compaction on soil physical properties and grain yield[J]. Soil and Tillage Research, 2021, 211: 105019.
- [9] ZHANG X Y, CRUSE R M, SUI Y Y, et al. Soil compaction induced by small tractor traffic in Northeast China[J]. Soil Science Society of America Journal, 2006, 70(2): 613–619.
- [10] CARTER M R. Penetration resistance to characterize the depth and persistence of soil loosening in tillage studies[J]. Canadian Journal of Soil Science, 1988, 68(4): 657–668.
- [11] LIMA C L R, REINERT D J, REICHERT J M, et al. Yield of crops and soil resistance to penetration of the Alfisol under different management[J]. Pesquisa Agropecuária Brasileira, 2010, 45(1): 89–98.
- [12] 杨敏丽, 彭健, 金剑, 等. 不同农用轮胎对东北黑土区土壤压实的影响[J]. 农业机械学报, 2023, 54(6): 85–93.
YANG Minli, PENG Jian, JIN Jian, et al. Influence of agricultural tire technology innovation on soil compaction in black soil region Northeast China[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 85–93. (in Chinese)
- [13] DA LUZ F B, GONZAGA L C, CASTIONI G A F, et al. Controlled traffic farming maintains soil physical functionality in sugarcane fields[J]. Geoderma, 2023, 432: 116427.
- [14] ZHUO Z, XING A, CAO M, et al. Identifying the position of the compacted layer by measuring soil penetration resistance in a dryland farming region in Northeast China[J]. Soil Use and Management, 2020, 36(3): 494–506.
- [15] KELLER T, SANDIN M, COLOMBI T, et al. Historical increase in agricultural machinery weights enhanced soil stress levels and adversely affected soil functioning[J]. Soil and Tillage Research, 2019, 194: 104293.
- [16] 王宪良, 王庆杰, 李洪文, 等. 免耕条件下轮胎压实对土壤物理特性和作物根系的影响[J]. 农业机械学报, 2017, 48(6): 168–175.
WANG Xianliang, WANG Qingjie, LI Hongwen, et al. Effect of type induced soil compaction on soil properties and crop root growth under no-tillage system[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(6): 168–175. (in Chinese)
- [17] TASER O F, KARA O. Silage maize (*Zea mays* L.) seedlings emergence as influenced by soil compaction treatments and contact pressures[J]. Plant Soil and Environment, 2005, 51(7): 289–295.
- [18] CORREA J, POSTMA J A, WOJCIECHOWSKI T. Phenotypic response to soil compaction varies among genotypes and correlates with plant size in sorghum[J]. Plant and Soil, 2022, 472(1–2): 59–76.
- [19] DE MORAES M T, DEBIASI H, FRANCHINI J C, et al. Soil compaction impacts soybean root growth in an Oxisol from subtropical Brazil[J]. Soil and Tillage Research, 2020, 200: 104611.
- [20] REINTAM E, TRÜKMANN K, KUHT J, et al. Soil compaction effects on soil bulk density and penetration resistance and growth of spring barley (*Hordeum vulgare* L.)[J]. Acta Agriculturae Scandinavica Section B-Soil and Plant Science, 2009, 59(3): 265–272.
- [21] MULAZZANI R P, GUBIANI P I, ZANON A J, et al. Impact of soil compaction on 30-year soybean yield simulated with CROPGRO–DSSAT[J]. Agricultural Systems, 2022, 203: 103523.
- [22] ELAOU D A, HASSEN H B, SALAH N B, et al. Modeling of soil penetration resistance using multiple linear regression (MLR)[J]. Arabian Journal of Geosciences, 2017, 10(20): 1–8.
- [23] GHADERNEJAD K, SHAHGHOLI G, MARDANI A, et al. Prediction effect of farmyard manure, multiple passes and moisture content on clay soil compaction using adaptive neuro-fuzzy inference system[J]. Journal of Terramechanics, 2018,

- 77: 49 – 57.
- [24] BEN HASSEN H, ELAOU D A, MASMOUDI K. Modeling of agricultural soil compaction using discrete Bayesian networks [J]. *International Journal of Environmental Science and Technology*, 2020, 17(5): 2571 – 2582.
- [25] KHANAL S, KLOPFENSTEIN A, KUSHAL K C, et al. Assessing the impact of agricultural field traffic on corn grain yield using remote sensing and machine learning[J]. *Soil and Tillage Research*, 2021, 208: 104880.
- [26] 李民赞, 任新建, 杨玮, 等. 基于树莓派的农田表土层土壤容重检测系统研究[J]. *农业机械学报*, 2021, 52(增刊): 329 – 335, 376.
- LI Minzan, REN Xinjian, YANG Wei, et al. Soil bulk density detection system of farmland topsoil based on raspberry[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(Supp.): 329 – 335, 376. (in Chinese)
- [27] 周修理, 张萍萍, 秦娜, 等. 基于 GA – RF 模型土壤坚实度对黑土区大豆产量的影响[J]. *东北农业大学学报*, 2022, 53(10): 67 – 75.
- ZHOU Xiuli, ZHANG Pingping, QIN Na, et al. Effects of soil penetration resistance on soybean yield in black soil region based on GA – RF model[J]. *Journal of Northeast Agricultural University*, 2022, 53(10): 67 – 75. (in Chinese)
- [28] GB 5009.3—2016 食品安全国家标准 食品中水分的测定[S].
- [29] 段丹丹, 刘仲华, 赵春江, 等. 基于优化植被指数的多生育期茶叶游离氨基酸含量估算[J]. *农业机械学报*, 2022, 53(2): 393 – 400, 420.
- DUAN Dandan, LIU Zhonghua, ZHAO Chunjiang, et al. Estimation of free amino acid content in fresh tea leaves at multiple growth periods based on optimized vegetation index[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(2): 393 – 400, 420. (in Chinese)
- [30] 张唯, 伍霞, 卢灿炯, 等. 基于多元回归的栅格水系阈值计算模型[J]. *农业机械学报*, 2016, 47(10): 131 – 138, 107.
- ZHANG Wei, WU Xia, LU Canjiong, et al. Determination of flow accumulation threshold based on multiple regression model in raster river networks extraction[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(10): 131 – 138, 107. (in Chinese)
- [31] 冯建英, 苏允汇, 龚劭齐, 等. 基于集成学习的农业生产技术效率评价方法[J]. *农业机械学报*, 2021, 52(增刊): 148 – 155.
- FENG Jianying, SU Yunhui, GONG Shaoqi, et al. Evaluation method of agricultural production technical efficiency based on ensemble learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(Supp.): 148 – 155. (in Chinese)
- [32] 吴静珠, 张乐, 李江波, 等. 基于高光谱与集成学习的单粒玉米种子水分检测模型[J]. *农业机械学报*, 2022, 53(5): 302 – 308.
- WU Jingzhu, ZHANG Le, LI Jiangbo, et al. Detection model of moisture content of single maize seed based on hyperspectral image and ensemble learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(5): 302 – 308. (in Chinese)
- [33] 齐建东, 黄金泽, 贾昕. 基于 XGBoost – ANN 的城市绿地净碳交换模拟与特征响应[J]. *农业机械学报*, 2019, 50(5): 269 – 278.
- QI Jiandong, HUANG Jinze, JIA Xin. Simulation of NEE and characterization of urban green-land ecosystem responses to climatic controls based on XGBoost – ANN[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(5): 269 – 278. (in Chinese)
- [34] HUANG Q, XUE H, LIU J, et al. A new method for spectral wavelength selection based on multiple linear regression combined with ant colony optimization and genetic algorithm[J]. *Journal of Spectroscopy*, 2022, 2022: 2440518.
- [35] LIAW A, WIENER M. Classification and regression by random forest[J]. *R news*, 2002, 2(3): 18 – 22.
- [36] GRÖMPING U. Variable importance assessment in regression: linear regression versus random forest[J]. *The American Statistician*, 2009, 63(4): 308 – 319.
- [37] GARSON G D. Interpreting neural-network connection weights[J]. *AI Expert*, 1991, 6(4): 46 – 51.
- [38] GOH A T C. Back-propagation neural networks for modeling complex systems[J]. *Artificial Intelligence in Engineering*, 1995, 9(3): 143 – 151.