

基于无人机影像与机器学习的柑橘产量估测研究

吴立峰 徐文浩 裴青宝
(南昌工程学院水土保持学院, 南昌 330099)

摘要: 为了准确、快速地预测柑橘产量以准确指导果园生产管理,通过大疆多光谱版无人机获取柑橘果实成熟期的遥感影像数据,并从图像中提取了可见光和多光谱波段指数作为特征变量,采用极端梯度提升(eXtreme gradient boosting, XGB)、随机森林(Random forest, RF)以及支持向量机(Support vector machine, SVM)模型分别构建柑橘果实有无分类模型、果实数量和质量估测模型。结果表明:通过XGB模型对特征变量进行筛选分析,柑橘果实有无的分类中超红指数 ExR 最重要,而数量和质量估测中改进超绿指数 $MExG$ 最重要。组合建模中3个模型均在组合4的情况下精度较好。对于分类模型,最优模型为SVM模型(AUC为0.969,准确率为0.919),而对于数量和质量估测模型,最优模型为XGB模型(数量: $R^2=0.79$, RMSE为466个;质量: $R^2=0.79$, RMSE为19.51 kg)。最后利用Shapley additive explanations(SHAP)方法揭示了植被指数特征在产量估测模型构建时的重要性,并阐明了SHAP值排在前四的特征交互影响。本研究结果可为无人机遥感在柑橘产量方面的研究提供应用参考和理论依据。

关键词: 柑橘产量; 无人机; 多光谱; 机器学习; SHAP

中图分类号: S127

文献标识码: A

文章编号: 1000-1298(2024)12-0294-12

OSID:



Citrus Yield Estimation by Integrating UAV Imagery and Machine Learning

WU Lifeng XU Wenhao PEI Qingbao

(School of Soil and Water Conservation, Nanchang Institute of Technology, Nanchang 330099, China)

Abstract: In order to accurately and rapidly predict citrus yield to precisely guide orchard production management, remote sensing image data of citrus fruit ripening stage was obtained by DJI multispectral version of UAV, and visible and multispectral band indices were extracted as feature variables from the images. The eXtreme gradient boosting (XGB), random forest (RF) and support vector machine (SVM) model were used to construct citrus fruit presence and absence classification model, fruit number and quality estimation model, respectively. The results showed that the excess red index was the most important in the classification of citrus fruit presence and absence while the modified excess green index was the most important in the estimation of number and quality through the screening analysis of feature variables by the XGB model. All three models in combination modeling had better accuracy in combination 4. For the classification model, the optimal model was the SVM model with AUC of 0.969 and accuracy of 0.919. While the XGB model was the best model for estimating both number and quality, with the number estimation model's R^2 value being 0.79 and RMSE being 466, and the quality estimation model's R^2 value being 0.79 and RMSE being 19.51 kg. Finally, the Shapley additive explanations (SHAP) method was utilized to reveal the importance of vegetation index features in the construction of the yield estimation model and to elucidate the interaction effects of the features with the top four SHAP values. The research results can provide an application reference and theoretical basis for the research of UAV remote sensing in citrus yield.

Key words: citrus yield; UAV; multispectral; machine learning; SHAP

收稿日期: 2024-08-07 修回日期: 2024-09-07

基金项目: 江西省教育厅科技项目(GJJ180925)和江西省科技厅重点研发计划重点项目(20212BDH80016)

作者简介: 吴立峰(1985—),男,讲师,博士,主要从事农业资源高效利用研究, E-mail: china.sw@163.com

0 引言

柑橘在我国广泛种植,其果实具有较高的营养价值,对人类的身体健康有益,属于较为重要的经济作物^[1]。江西省橘类产品资源丰富,是我国柑橘生产的主产区和原产地之一,其中南丰蜜橘较为著名^[2]。由于我国柑橘产业多集中于丘陵山地环境,传统的人工测产费时费力且具有破坏性。因此,准确、快速、无损地估测南丰蜜橘的产量,可促进当地柑橘产业的发展,提高当地果农的经济收入。

近些年,无人机遥感技术凭借其操作方便、机动灵活、低成本且具有较高的时空分辨率等特点,已在精准农业中广泛应用^[3]。诸多学者已开展了基于无人机遥感技术对果树的研究^[4-5]。及时准确地预测粮食产量能够为农民的粮食种植和市场交易提供正确指导,对农业管理和粮食政策制定具有重要的意义^[6]。目前,关于作物产量的无人机遥感估测研究已有一定进展^[7-8]。以上研究表明,利用无人机遥感技术和机器学习来估测农作物的产量具有很好的适用性及准确性。

在先前的农业领域无人机遥感研究中,机器学习算法已被证明在反演作物生长生理参数方面表现良好,但由于其特有的“黑箱”性质,无法解释各个变量对预测结果的贡献程度^[9]。为此,部分学者在建模前利用特征筛选方法确定输入特征变量的重要性,以提高模型预测能力^[10-11]。SHAP 是一种机器学习事后可解释性方法,它通过 SHAP 值对模型的预测结果进行解释^[12]。文献[13]利用 6 种机器学习算法建立预测土壤中除草剂阿特拉津降解效率的最佳模型,并应用 SHAP 方法对输入特征进行解释,发现土壤湿度和时间是主要因素。文献[14]对全球 9 种豆科植物的产量影响因素进行研究,并通过 SHAP 方法发现各种作物对不同影响因素的反应并不相同。

过往的研究大多利用无人机遥感技术来估测小麦、玉米和油菜等农作物的产量,而对于柑橘果树的产量估算研究较少。此外,SHAP 方法在利用机器学习解决实际问题的应用研究中已被使用^[15-16]。但目前,较少有学者在农作物特别是柑橘果树的产量估测上使用该方法。因此本文通过无人机遥感技术来解决柑橘产量估测费时费力问题,以南丰蜜橘为对象,采集果实成熟期的无人机图像,分别计算可见光波段的植被指数和多光谱波段的植被指数,将植被指数作为特征变量,利用极端梯度提升、随机森林以及支持向量机 3 个机器学习模型构建柑橘果树果实有无分类模型、果实数量和质量估测模型,通过

对比和分析筛选出最优模型,并引入 SHAP 方法对最优模型的输入特征变量进行解释性分析,以期探索出一种利用无人机图像估测柑橘产量的方法,为快速准确地获取柑橘产量提供理论参考。

1 材料与方法

1.1 研究区概况

研究区位于江西省抚州市南丰县江西省水利科学院农村水利科研示范基地(27°12'N,116°45'E)。该地属于中亚热带季风气候区,地处赣东南山地丘陵区,温和湿润,雨量充足,四季分明,年平均气温为 19.8℃,年平均降水量为 1 791.8 mm,土壤类型为红壤土,适合种植柑橘类水果。试验柑橘品种为“南丰蜜橘”。

1.2 数据获取

1.2.1 地面数据获取

为了保证选取果树长势差异,果树采收前在试验区内随机选取 118 棵样本树,并按照顺序挂上带有数字标签,同时记录每棵树有果无果情况,最后用警戒线将样本树圈住防止被当地果农误采。本试验于 2023 年 11 月 19、20 日对柑橘果树产量数据(果实数量和质量)进行采集。表 1 为样本树果实数量和质量统计特征。

表 1 柑橘单株果实数量和质量描述性统计
Tab.1 Descriptive statistics for number and quality of citrus single fruit

参数	最大值	最小值	中值	均值	标准差
数量/个	3 890	0	960	1 041	1 005
质量/kg	151.0	0	44.5	45.4	42.2

1.2.2 无人机影像获取与预处理

在样本树进行采摘的前一天,为保证影像质量,选择天气晴朗、无风少云的时间段进行无人机多光谱图像的拍摄工作。采用 DJI MAVIC3 型多光谱版无人机(M3M)获取柑橘果树冠层多光谱影像数据,该设备集成 1 个 2 000 万像素可见光相机及 4 个 500 万像素的多光谱相机(绿光、红光、红边和近红外)。此外,M3M 顶部集成光强传感器并搭配 RTK 模块,可对影像数据进行光照补偿,实现厘米级高精度定位。采集无人机影像前,由遥控器上的 DJI Pilot 软件规划飞行航线,设置飞行高度 50 m,航向重叠率和旁向重叠率均为 80%,采用等时间拍照方式并开启仿地飞行模式。飞行前手动起飞无人机拍摄辐射定标板,以便后期进行影像辐射标定。使用 Pix4Dmapper 软件进行影像拼接和辐射校正,获得 4 个单波段正射反射率影像。使用 ENVI 软件进行图

像剪裁得到试验区的影像数据并进行影像合成,以便进一步选取感兴趣区域获取相应数据。

1.3 植被指数选取

植被指数能够简单且有效地估测农作物产量等特征,基于前人的研究结果,挑选 38 个植被指数进行机器学习建模^[17-46],其中包括 12 个可见光波段指数(超绿指数 *ExG*、超红指数 *ExR*、超蓝指数 *ExB*、超绿超红差分指数 *ExGR*、改进超绿指数 *MExG*、归一化绿红差异指数 *NGRDI*、绿红比值指数 *GRRI*、红绿蓝植被指数 *RGBVI*、可见光大气阻抗指数 *VARI*、可见光差异植被指数 *VDVI*、植被指数 *VEG*、植被颜色提取指数 *CIVE*)和 26 个多光谱波段指数(绿色叶绿素指数 *CIg*、红边叶绿素指数 *CIre*、*DATT* 指数 *DATT*、差值植被指数 *DVI*、绿色差值植被指数 *GDVI*、归一化绿色差值植被指数 *GNDVI*、*Meris* 陆地叶绿素指数 *MTCI*、改进非线性指数 *MNLI*、改进简单比值指数 *MSR*、归一化红边植被指数 *NDRE*、归一化叶绿素指数 *NDCI*、归一化植被指数 *NDVI*、归一化绿色指数 *NGI*、归一化红色指数 *NRI*、归一化近红外指数 *NNIRI*、归一化红边指数 *NREI*、再归一化植被指数 *RDVI*、红边差值植被指数 *REDVI*、红边转换型植被指数 *RETVI*、比值植被指数 *RVI*、土壤调节植被指数 *SAVI*、优化土壤调节植被指数 *OSAVI*、光谱植被指数 *SVI*、三角植被指数 *TVI*、优化植被指数 *Vlopt*、宽范围动态植被指数 *WDRVI*)。

1.4 模型的构建

建模方法包括极端梯度提升模型(XGB)、随机森林模型(RF)、支持向量机模型(SVM)。本文基于 R 语言 4.3.0 平台,利用 *Sample* 函数将数据集分为 70% 建模集和 30% 验证集,进行柑橘果树果实有无的分类研究以及果实数量和质量的估测建模。其中,XGB 是一种基于梯度提升决策树(GBDT)框架的机器学习算法,通过构建和组合多个决策树来提高预测精度,每一棵树都是在前一棵树的基础上进行优化,以最小化损失函数。本研究使用 XGB 模型的 *Importance* 函数对 38 个植被指数进行特征筛选,将排在前 15 的植被指数筛选出来并分成 4 组进行建模分析,即组合 1 为重要性前 3 的植被指数,组合 2 为重要性前 5 的植被指数,组合 3 为重要性前 10 的植被指数,组合 4 为重要性前 15 的植被指数。为了方便表示,在模型后面标上“1”、“2”、“3”、“4”表示组合 1、组合 2、组合 3、组合 4。

1.5 SHAP 方法理论

SHAP 为机器学习模型提供了一种事后解释的方法,属于根据预期表示一致性和局部准确性的加性特征归因方法^[12]。SHAP 为模型的每一个特征变

量分配一个特定的预测重要性值,SHAP 值是博弈论中用于衡量合作博弈中每个参与者对于最终收益的贡献度,机器学习中 SHAP 值被用来衡量每个特征对于模型预测的贡献度^[47]。

本文利用 R 语言的 *shapviz* 包来量化 3 个机器学习模型中各个特征变量的重要性。

1.6 模型精度评价指标

采用 ROC (Receiver operating characteristic curve) 曲线下的面积值(AUC)和混淆矩阵计算的分类结果准确率来评价分类模型的分类精度和可靠性。对于分类模型,以上两种评价指标值越接近 1,表明模型分类性能越好。采用决定系数 R^2 和均方根误差(RMSE)评价模型估测精度和准确性。

2 结果分析

2.1 柑橘果实有无分析

在分析柑橘果实有无状况时,新增 84 株只判断了果实有无状况的果树参与建模,其中有果实编码为“1”,无果实编码为“0”。

2.1.1 柑橘果实有无的特征筛选

图 1 中纵坐标表示特征增益值,值越高表示对模型贡献度越大。后 3 个特征 *ExR*、*NGRDI*、*NRI* 特征增益值都在 0.1 以上,*ExR* 增益值最大达到 0.267 5,其余特征均在 0.1 以下。由此可以说明这 3 个植被指数重要性较大。

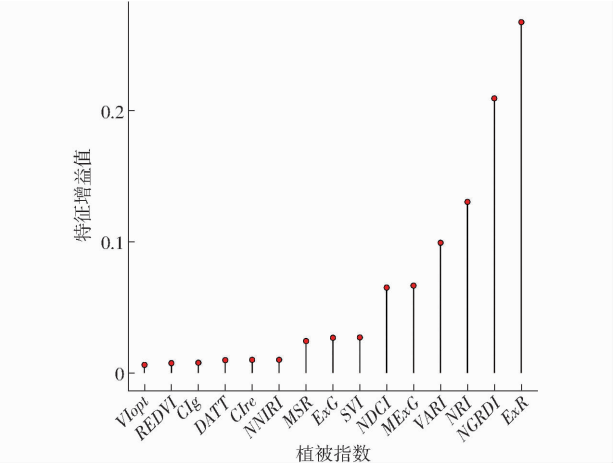


图 1 柑橘果实有无中植被指数的特征重要性排序

Fig.1 Characteristic importance ranking of vegetation indices in citrus fruit presence and absence

2.1.2 柑橘果实有无组合建模

ROC 曲线(图 2)中横坐标表示假阳性率,即模型将无果样本错误分类为有果的比例;纵坐标表示真阳性率,即模型将有果样本正确分类的比例。在 AUC 值方面,3 个模型最大 AUC 值均出现在组合 4 中,分别为 0.948、0.938、0.969。由此可以看出,3 个模型在组合 4 情况下均表现出较好的果实有无分

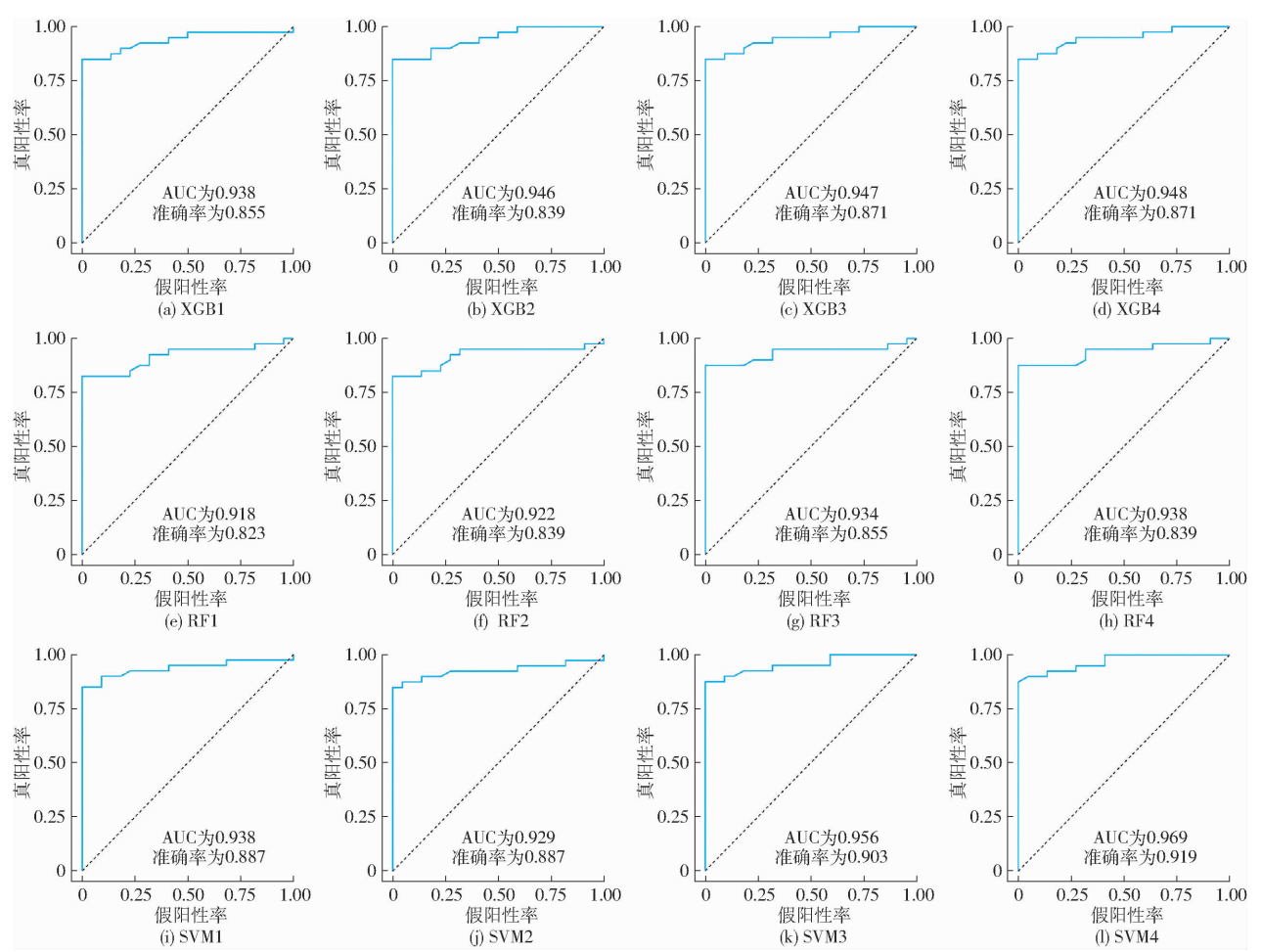


图 2 基于不同模型 4 种组合的柑橘果实有无 ROC 曲线

Fig. 2 ROC curves of citrus fruit presence or absence based on four combinations of different models

类性能,但 SVM 模型为最佳果实有无分类的模型。随着模型输入特征数量增加,XGB 模型 AUC 提升 1.06%,RF 模型 AUC 提升 2.18%,表明增加合适的特征能提高分类性能。而对于 SVM 模型,组合 2 AUC(0.929)略低于组合 1(0.938),特征 *VARI* 和 *ME_xG* 的加入并未提升模型性能。

在准确率方面,XGB 模型在组合 3、4 的情况下表现较好,准确率均为 0.871,说明 *CI_{re}*、*DATT*、*CI_g*、*REDVI*、*VI_{opt}* 这 5 个特征对模型的贡献较小。对比组合 1,*VARI* 和 *ME_xG* 这两个特征在组合 2 中的加入可能对准确率产生了负向影响。RF 模型在组合 1、组合 2 和组合 3 的情况下整体上表现出随着特征数量的增多,准确率逐渐增大。同时组合 2 和组合 4 准确率均为 0.839,这表明特征数量的增加在一定程度上有助于提升模型的分类效果。SVM 模型在组合 1、2 情况下准确率均为 0.887,说明 *VARI* 和 *ME_xG* 这两个特征的增加对模型的分类性能没有提升效果。其次组合 4 下的模型表现稍好,准确率为 0.919。

2.2 柑橘果实数量分析

2.2.1 柑橘果实数量特征筛选

从图 3 可以看出, *ME_xG* 特征特征增益值为

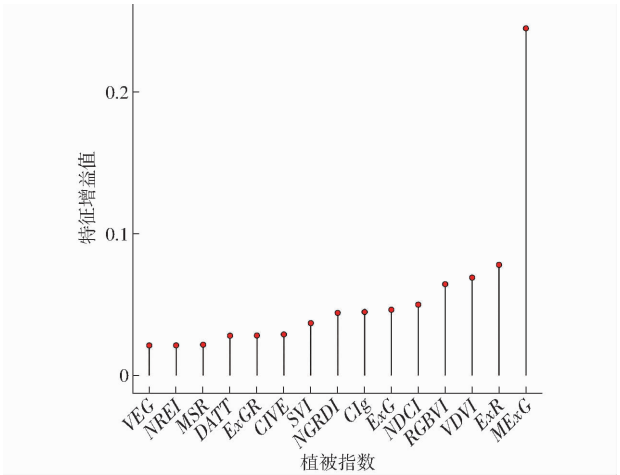


图 3 柑橘果实数量中植被指数特征重要性排序

Fig. 3 Characteristic importance ranking of vegetation indices in number of citrus fruits

0.245 1,明显大于其他 14 个特征,说明该特征对模型的贡献度最大。其余 14 个特征的特征增益值为 0.021 2~0.078 1,且均处于 0.1 以下,说明这 14 个特征对模型的贡献度都较小且相近。

2.2.2 柑橘果实数量的组合建模

根据图 4,3 个模型在组合 4 中表现较好,而 3

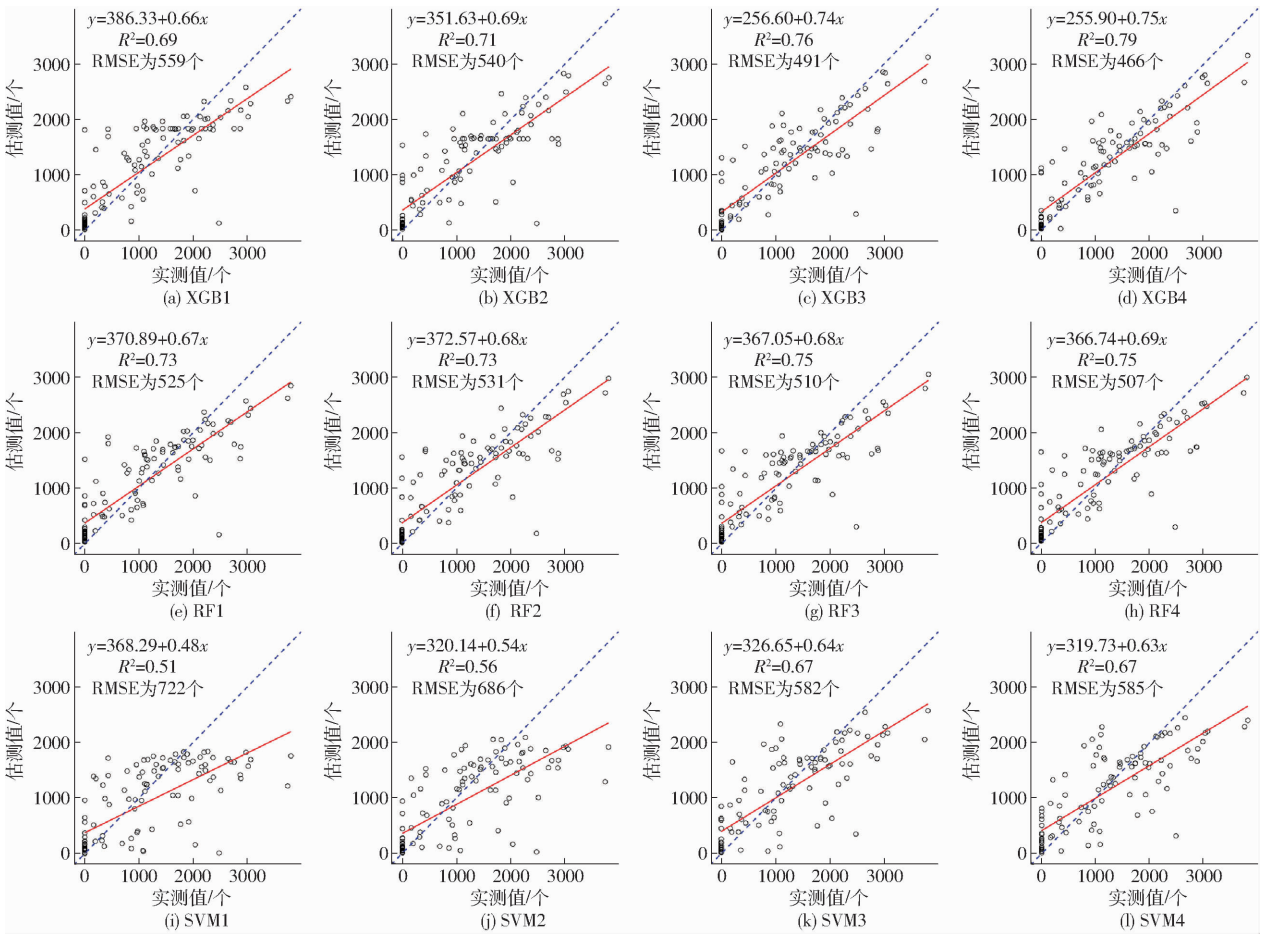


图 4 基于不同模型 4 种组合的柑橘果实数量散点图

Fig. 4 Scatter plots of number of citrus fruits based on four combinations of different models

个模型中 XGB 模型的表现较好 ($R^2 = 0.79$, RMSE 为 466 个)。相较于 RF 模型, XGB 模型 R^2 增大 5.33%, RMSE 减少 8.09%。相较于 SVM 模型, XGB 模型 R^2 增大 17.91%, RMSE 减少 20.34%。其次, SVM 模型在 3 个模型中较差 ($R^2 = 0.67$, RMSE 为 585 个)。相较于 SVM 模型, RF 模型 R^2 增大 11.94%, RMSE 减少 13.33%。随着模型输入特征数量增加,模型 R^2 增大, RMSE 减小,说明增加合适的特征进行建模可以提高模型估测单株果实数量性能。其中 SVM 模型性能提高较多, R^2 增大 31.37%, RMSE 减小 19.39%。RF 模型性能提高较少, R^2 增大 2.74%, RMSE 减小 3.43%。其次从 R^2 来看,对于 RF 模型和 SVM 模型,组合 3 和组合 4 的结果一样,说明特征 $ExGR$ 、 $DATT$ 、 MSR 、 $NREI$ 和 VEG 对两个模型的估测性能提升作用有限。XGB 模型在组合 3 到组合 4 的 R^2 仅提高 0.03,进一步证实了这一点。上述研究表明,XGB 模型在组合 4 的情况下估测单株果实数量时表现更好。

2.3 柑橘果实质量分析

2.3.1 柑橘果实质量特征筛选

从图 5 中可以看出, $MExG$ 特征特征增益值为

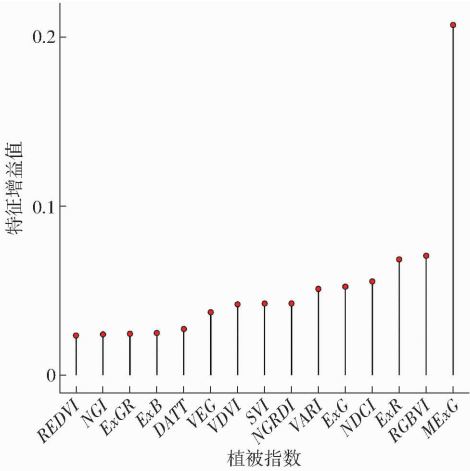


图 5 柑橘果实质量中植被指数特征重要性排序

Fig. 5 Characteristic importance ranking of vegetation indices in quality of citrus fruits

0.207 2,明显大于其他 14 个特征,说明该特征对模型贡献度最大。其余 14 个特征特征增益值为 0.023 6~0.070 7 之间,且均处于 0.1 以下,说明这 14 个特征对模型的贡献度都较小且相近。

2.3.2 柑橘果实质量组合建模

根据图 6,3 个模型在组合 4 中的表现较好,而 3 个模型中 XGB 模型表现较好 ($R^2 = 0.79$, RMSE 为

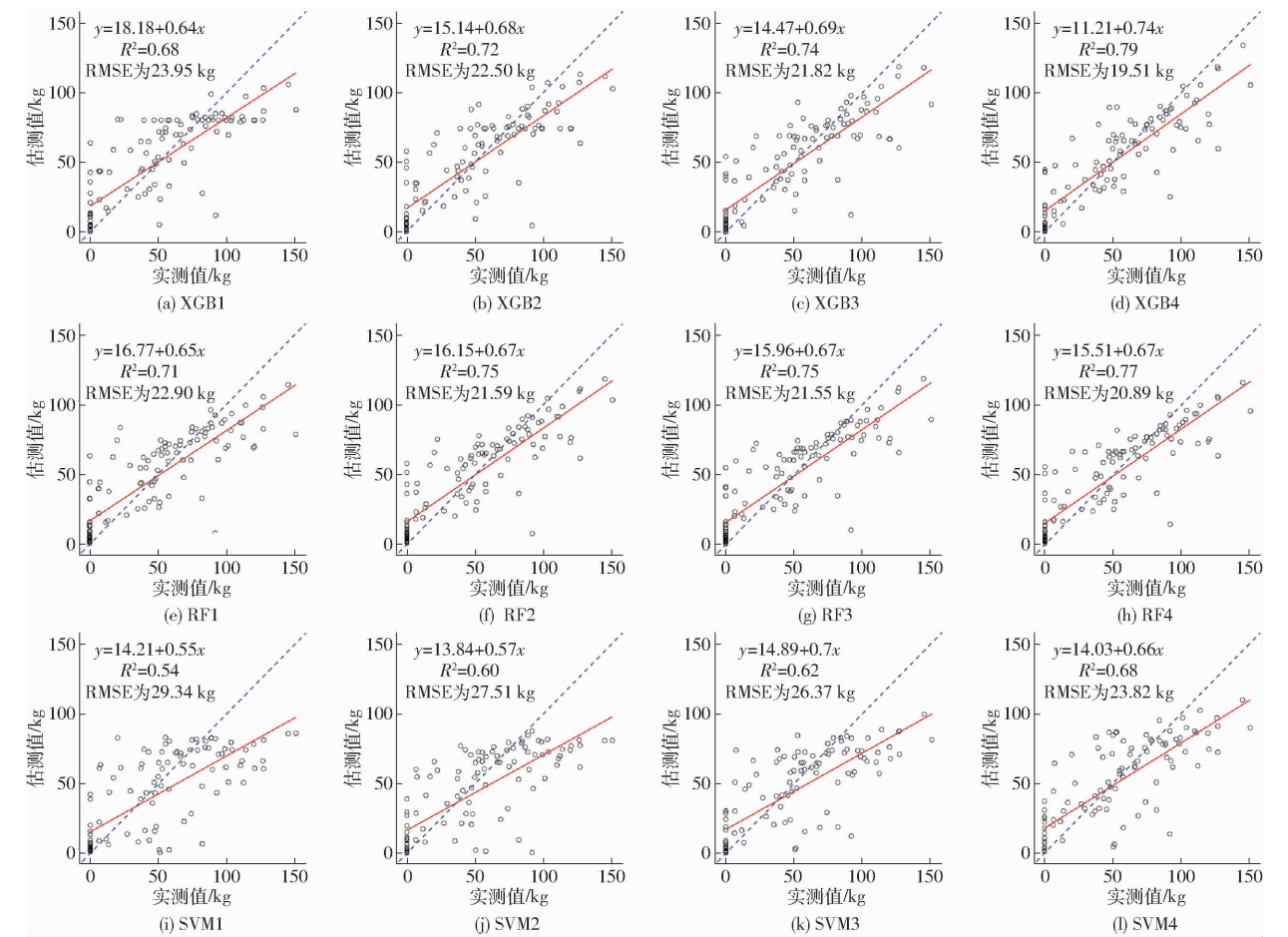


图 6 基于不同模型 4 种组合的柑橘果实质量散点图

Fig. 6 Scatter plots of quality of citrus fruits based on four combinations of different models

19.51 kg)。相较于 RF 模型, XGB 模型 R^2 增大 2.60%, RMSE 减少 6.61%。相较于 SVM 模型, XGB 模型 R^2 增大 16.18%, RMSE 减少 18.09%。其次, SVM 模型在 3 个模型中较差 ($R^2 = 0.68$, RMSE 为 23.82 kg)。相较于 SVM 模型, RF 模型 R^2 增大 13.24%, RMSE 减少 12.31%。随着模型输入特征数量的增加,模型 R^2 增大, RMSE 减小,说明增加合适的特征进行建模可以提高模型估测单株果实质量的性能。其中 SVM 模型性能提高较多, R^2 增大 25.93%, RMSE 减小 18.81%。RF 模型性能提高较少, R^2 增大 8.45%, RMSE 减小 8.78%。其次 3 个模型的 R^2 从组合 2 到组合 3 提升幅度都较小, XGB 模型 R^2 增大 2.70%, RF 模型 R^2 没有发生变化, SVM 模型 R^2 增大 3.33%, 说明特征 $VARI$ 、 $NGRDI$ 、 SVI 、 $VDVI$ 和 VEG 对 3 个模型估测性能提升作用有限, 尤其对 RF 模型的估测性能没有提升。上述研究表明, XGB 模型在组合 4 的情况下估测单株果实质量时表现更好。

2.4 基于 SHAP 的可解释性分析

2.4.1 基于 SHAP 的全局解释

根据前面的分析, XGB 模型为估测柑橘单株果

实数量和质量的最优模型, 且柑橘果实有无分类中 SVM 模型已达到相对较高精度。其次果实数量和质量分布具有更多的变化性。为了更准确探究 38 个植被指数对柑橘单株果实数量和质量的重要性, 使用 SHAP 方法对输入 XGB 模型的特征变量进行可解释性分析。

分布图中的每个点都是一个特征值和一个 SHAP 值, 以 SHAP 值取零为中间分界线, 处在下侧的样本对估测值表现为负向作用, 处在上侧的样本表现为正向作用, 颜色表示对应特征值的高低程度。图 7a 中, $MExG$ 取值越小对模型输出的正影响越大; 取值越大对模型输出的负影响越大。然而 ExR 取值越大对模型输出的正影响越大; 取值越小对模型输出的负影响越大。此外, TVI 取值越大对模型只有负影响, 取值越小对模型既有正影响也有负影响。同样的图 7b 中, ExR 取值越大对模型输出的正影响越大; 取值越小对模型输出的负影响越大。然而 $MExG$ 和 $RGBVI$ 取值越小对模型输出的正影响越大; 取值越大对模型输出的负影响越大。

特征重要性图中每个特征在所有样本中的平均绝对值被当作该特征的全局重要性。由图 7c、7d 可

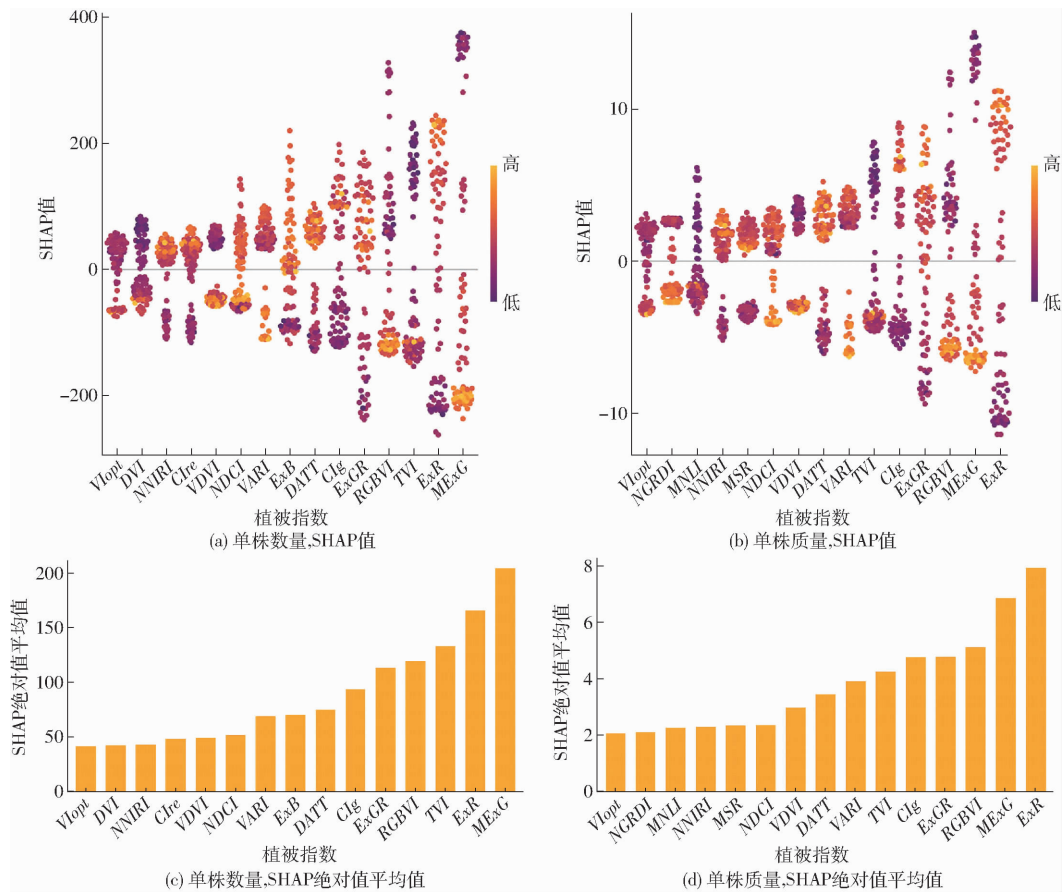


图 7 模型输入特征 SHAP 值分布图和特征重要性
Fig. 7 SHAP value of model input feature distribution and feature importance diagram

知, *MExG*、*ExR* 和 *TVI* 植被指数对单株果实数量估测模型的重要性排在前三, 其 SHAP 绝对平均值大于 125; *ExR*、*MExG* 和 *RGBVI* 植被指数对单株果实质量估测模型重要性排在前三, 其 SHAP 绝对平均值大于 5。2.2 节和 2.3 节利用 XGB 模型特征筛选, 同样发现模型估测单株果实数量和质量时重要性排在前三的有 *MExG* 和 *ExR* 特征。这不仅说明 *MExG* 和 *ExR* 特征对模型估测单株果实数量和质量的影响显著, 还证明 XGB 模型算法在模型构建时计算特征重要性的合理性。

2.4.2 基于 SHAP 的特征交互影响

为深入分析两两特征之间对 XGB 模型估测结果的交互影响, 根据 SHAP 多特征依赖图可进一步分析两个特征值之间的相互作用对估测结果的影响, 横轴表示特征的实际值, 纵轴表示特征的 SHAP 值, 数据点的颜色表示第 2 个特征值。

从图 8a 可以看出, *MExG* 与 *NGRDI* 呈正相关, 即 *MExG* 增大时, 对应的 *NGRDI* 也增大。此外, *MExG* < 1 时, *NGRDI* 越小, 其 SHAP 值较大, 说明较小的 *MExG* 和 *NGRDI* 可以提高模型估测果实数量的性能。从图 8b 可以看出, *ExR* 与 *VARI* 呈负相关, 即 *ExR* 增大时, 对应的 *VARI* 随之减小。此外,

ExR > 0.1 时, *VARI* 越小, 其 SHAP 值较大, 说明较大的 *MExG* 和较小的 *VARI* 可以提高模型估测果实数量的性能。图 8c 展示出当 *TVI* < 11 时, SHAP 值均大于 0; 当 *TVI* ≥ 11 时, SHAP 值开始变小为负值。图 8d 显示了 *RGBVI* 值在 0.25 附近, SHAP 值发生了突变, 之后 SHAP 值随着 *ExB* 的减小而减小。

从图 9a 可以看出, *ExR* 与 *DATT* 没有明显的正相关和负相关关系, 当 *ExR* > 0.1 时, 其 SHAP 值较大且均处于 5 以上。从图 9b 可以看出, *MExG* 与 *WDRVI* 呈正相关, 即 *MExG* 增大时, 对应的 *WDRVI* 也增大。此外, *MExG* > 0.1 时, *WDRVI* 越大, 其 SHAP 值较小, 说明较大的 *MExG* 和 *WDRVI* 会降低模型估测果实质量的性能。图 9c 显示了 *RGBVI* 在 0.25 附近, SHAP 值同样也发生了突变, 之后 SHAP 值随着 *ExB* 的减小而减小。图 9d 展示出 *ExGR* 处于 -0.050 ~ -0.025 之间时, SHAP 值的变化为先增大后减小, 说明此时 *ExGR* 特征对模型性能的影响由负向转为正向, 又由正向转为负向并在 -0.025 处发生了突变。

3 讨论

传统的果树测产主要通过实地测量, 这种测量

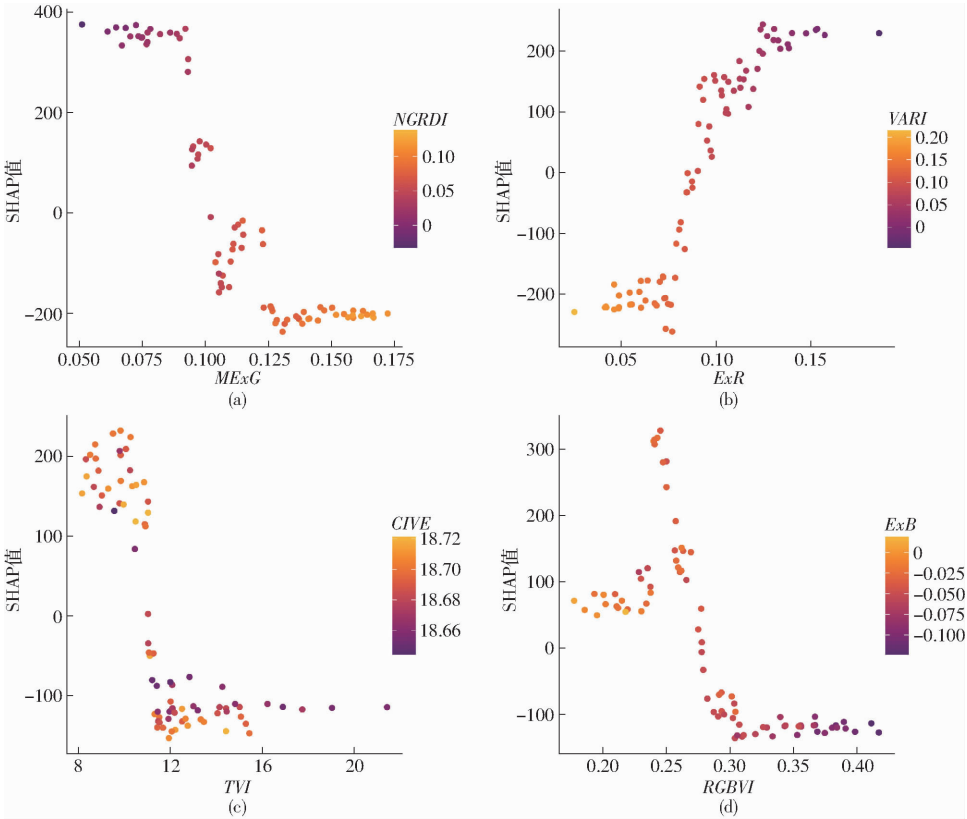


图 8 估测果实数量时 $MExG$ 、 ExR 、 TVI 和 $RGBVI$ 特征的依赖图

Fig. 8 Dependency diagrams of $MExG$, ExR , TVI and $RGBVI$ features for estimating fruit number

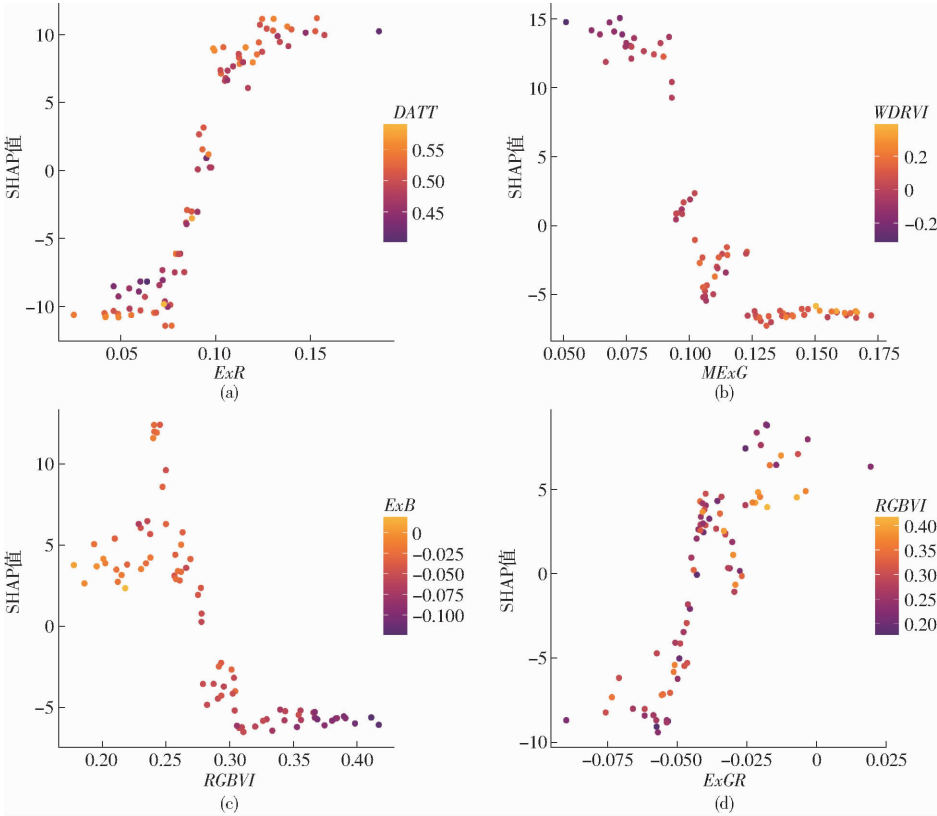


图 9 估测果实质量时 ExR 、 $MExG$ 、 $RGBVI$ 和 $ExGR$ 特征的依赖图

Fig. 9 Dependency diagrams of ExR , $MExG$, $RGBVI$ and $ExGR$ features for estimating fruit quality

方式费时费力,为了解决这一问题,高效准确地测量果树产量就变得尤为重要。本研究借助于无人机和机器学习模型建立基于植被指数的柑橘果实有无分类模型、果实数量和质量的估测模型,取得了较好的

效果,为柑橘园产量的快速估测提供了新方法。在柑橘果实有无的分类任务中,SVM 模型(AUC 为 0.969)展现出了优越的性能,这可能归功于其在处理线性可分问题时的高效性。此外,SVM 对于数据的高维特性具有很强的适应能力,使得它能够充分利用遥感影像中的多光谱信息,从而提高分类的准确性。然而,当涉及到更为复杂的柑橘果实数量和质量估测时,这类估测通常涉及多个变量之间的复杂相互作用以及非线性关系,XGB 模型($R^2=0.79$)因其灵活性和对特征交互的捕捉能力,提供了更为准确的估测结果。而文献[48]构建的小麦产量预测模型 R^2 可达 0.87,这可能因为柑橘为多年生常绿果树、具有全年不落叶所形成的大而厚的树冠,这对无人机精确获取冠层光谱信息造成困难,导致建模精度偏低^[49]。其次选择植被指数建模是因为其已被证明不仅与果园水分状况和果树光合活性有关,还与果实产量有关^[50]。

在实际应用中,不同特征变量对机器学习模型完成任务的贡献不同。特征选择通过去除不相关和冗余的特征,可以减少计算时间,提高模型精度,并有助于更好地理解模型^[51]。基于 XGB 模型的 importance 函数对 38 个植被指数特征进行重要性排序,发现 ExR 是在柑橘果实有无分类中最重要的特征,而在柑橘果实数量和质量估测中, $MExG$ 特征则显得尤为重要。成熟柑橘果实相较于叶片,在红光波段反射更多光线, ExR 指数通过增强这一信号差异,能有效区分果实与叶片;同时, ExR 对果实从绿转黄或橙的成熟颜色变化敏感,有助于果实成熟度监测及植株有无果实状态的识别^[52]。 $MExG$ 指数与叶绿素含量密切相关,反映植物健康与光合作用效率,能预示更高产量潜力;同时, $MExG$ 可提供植被结构信息,该信息往往与产量相关,对单株产量估测至关重要^[53]。其次组合建模发现模型在组合 4 表

现较好,表明多个植被指数组合有信息互补的优势,可提高模型精度。这一发现与文献[54]研究结果相符。

为了探讨特征对模型的作用以及特征之间的关系,引入了 SHAP 方法。SHAP 分析结果不仅识别出了影响柑橘产量估测模型的主要特征,还揭示了较小的 $MExG$ 和 $NGRDI$ 可以提高模型估测果实单株数量性能,较大的 $MExG$ 和 $WDRVI$ 会降低模型估测单株质量性能。这一发现为山地丘陵区果树产量研究在特征选择和模型优化上提供了新的视角。尽管本研究取得了一定的成果,但仍存在一些局限性。数据集可能受限于特定的地理和气候条件,模型泛化能力需要在不同地区和条件下进行验证,在未来研究中可以结合当地土壤条件和气候条件进行建模。收获时柑橘处于果实成熟期,柑橘园果树冠层覆盖度较大,可能会导致植被指数出现易饱和现象^[55]。为了改善该问题,后续研究可以增加纹理、株高和冠层面积等特征。

4 结 论

(1)基于 XGB 模型的特征重要性筛选,柑橘果实有无分类中重要性最大的指数为 ExR ,而柑橘果实数量和质量估测中重要性最大的指数均为 $MExG$ 。

(2)SVM 模型是柑橘果实有无分类中表现最优的模型,且最佳组合为组合 4。相比于果实有无分类研究,XGB 模型在柑橘果实数量和质量上具有较好的估测效果,且最佳组合均同样为组合 4。

(3)引入 SHAP 方法解释柑橘果实数量和质量模型的最优结果,估测果实数量时,重要性前三的特征为 $MExG$ 、 ExR 、 TVI ;估测果实质量时,重要性前三的特征为 ExR 、 $MExG$ 、 $RGBVI$ 。特征之间的相互作用对模型估测影响各不相同。

参 考 文 献

[1] SAINI R K, RANJIT A, SHARMA K, et al. Bioactive compounds of citrus fruits: a review of composition and health benefits of carotenoids, flavonoids, limonoids, and terpenes[J]. *Antioxidants*, 2022, 11(2): 239.

[2] 汤雨晴, 刘心澄, 杨惠栋, 等. 南丰蜜橘不同品种(系)果实化渣性比较[J]. *南方农业学报*, 2023, 54(12): 3657 - 3664.

TANG Yuqing, LIU Xincheng, YANG Huidong, et al. Comparison of fruit mastication trait among different *Citrus reticulata* Blanco cv. Kinokuni varieties (lines)[J]. *Journal of Southern Agriculture*, 2023, 54(12): 3657 - 3664. (in Chinese)

[3] YANG G J, LIU J G, ZHAO C J, et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives[J]. *Frontiers in Plant Science*, 2017, 8: 1111.

[4] 李美炫, 朱西存, 白雪源, 等. 基于无人机影像阴影去除的苹果树冠层氮素含量遥感反演[J]. *中国农业科学*, 2021, 54(10): 2084 - 2094.

LI Meixuan, ZHU Xicun, BAI Xueyuan, et al. Remote sensing inversion of nitrogen content in apple canopy based on shadow removal in UAV multi-spectral remote sensing images[J]. *Scientia Agricultura Sinica*, 2021, 54(10): 2084 - 2094. (in Chinese)

[5] ZHANG Y M, TA N, GUO S, et al. Combining spectral and textural information from UAV RGB images for leaf area index monitoring in kiwifruit orchard[J]. Remote Sensing, 2022, 14(5): 1063.

[6] 程千, 徐洪刚, 曹引波, 等. 基于无人机多时相植被指数的冬小麦产量估测[J]. 农业机械学报, 2021, 52(3): 160 – 167.
CHENG Qian, XU Honggang, CAO Yinbo, et al. Grain yield prediction of winter wheat using multi-temporal UAV based on multispectral vegetation index[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(3): 160 – 167. (in Chinese)

[7] 向友珍, 安嘉琪, 赵笑, 等. 基于无人机多光谱遥感的大豆生长参数和产量估算[J]. 农业机械学报, 2023, 54(8): 230 – 239.
XIANG Youzhen, AN Jiaqi, ZHAO Xiao, et al. Soybean growth parameters and yield estimation based on UAV multispectral remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(8): 230 – 239. (in Chinese)

[8] 杨楠, 周萌, 陈欢, 等. 基于无人机 RGB 影像的小麦叶面积指数与产量估算研究[J]. 麦类作物学报, 2023, 43(7): 920 – 932.
YANG Nan, ZHOU Meng, CHEN Huan, et al. Estimation of wheat leaf area index and yield based on UAV RGB images[J]. Journal of Triticeae Crops, 2023, 43(7): 920 – 932. (in Chinese)

[9] MI J X, LI A D, ZHOU L F. Review study of interpretation methods for future interpretable machine learning[J]. IEEE Access, 2020, 8: 191969 – 191985.

[10] 付虹雨, 王薇, 卢建宁, 等. 基于无人机多光谱遥感和机器学习的苧麻理化性状估测[J]. 农业机械学报, 2023, 54(5): 194 – 200, 347.
FU Hongyu, WANG Wei, LU Jianning, et al. Estimation of ramie physicochemical property based on UAV multi-spectral remote sensing and machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(5): 194 – 200, 347. (in Chinese)

[11] LIU J K, ZHU Y J, TAO X Y, et al. Rapid prediction of winter wheat yield and nitrogen use efficiency using consumer-grade unmanned aerial vehicles multispectral imagery[J]. Frontiers in Plant Science, 2022, 13: 1032170.

[12] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 4768 – 4777

[13] 李香伶, 陈奉献, 陈希娟. 基于 XGBoost 模型的土壤中阿特拉津降解预测[J]. 应用生态学报, 2024, 35(3): 789 – 796.
LI Xiangling, CHEN Fengxian, CHEN Xijuan. Prediction of atrazine degradation in soil based on XGBoost model[J]. Chinese Journal of Applied Ecology, 2024, 35(3): 789 – 796. (in Chinese)

[14] LI Yadong, LI Rujia, JI Rongbiao, et al. Research on factors affecting global grain legume yield based on explainable artificial intelligence[J]. Agriculture, 2024, 14(3): 438.

[15] 王志强, 任金哥, 韩硕, 等. 基于可解释性机器学习的建筑物物化阶段碳排放量预测研究[J]. 安全与环境学报, 2024, 24(6): 2454 – 2466.
WANG Zhiqiang, REN Jin'ge, HAN Shuo, et al. Interpretable machine learning-based carbon emission prediction in the materialization stage of buildings[J]. Journal of Safety and Environment, 2024, 24(6): 2454 – 2466. (in Chinese)

[16] 申采玉, 王帅, 周锐盈, 等. 慢性心力衰竭合并肺部感染患者院内死亡风险预测: 基于可解释性机器学习方法[J]. 南方医科大学学报, 2024, 44(6): 1141 – 1148.
SHEN Caiyu, WANG Shuai, ZHOU Ruiying, et al. Prediction of risk of in-hospital death in patients with chronic heart failure complicated by lung infections using interpretable machine learning[J]. Journal of Southern Medical University, 2024, 44(6): 1141 – 1148. (in Chinese)

[17] GUIJARRO M, PAJARES G, RIOMOROS I, et al. Automatic segmentation of relevant textures in agricultural images[J]. Computers and Electronics in Agriculture, 2011, 75(1): 75 – 83.

[18] 吴兰兰, 熊利荣, 彭辉. 基于 RGB 植被指数的大田油菜图像分割定量评价[J]. 华中农业大学学报, 2019, 38(2): 109 – 113.
WU Lanlan, XIONG Lirong, PENG Hui. Quantitative evaluation of in-field rapeseed image segmentation based on RGB vegetation indices[J]. Journal of Huazhong Agricultural University, 2019, 38(2): 109 – 113. (in Chinese)

[19] 丁艳玲, 赵凯, 李晓峰, 等. 吉林西部盐碱地数字图像植被覆盖度的自动提取[J]. 地球信息科学学报, 2013, 15(4): 618 – 624.
DING Yanling, ZHAO Kai, LI Xiaofeng, et al. An automatic extraction approach to fractional vegetation cover of saline land with digital images[J]. Journal of Geo-information Science, 2013, 15(4): 618 – 624. (in Chinese)

[20] HUNT E R, CAVIGELLI M, DAUGHTRY C S T, et al. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status[J]. Precision Agriculture, 2005, 6(4): 359 – 378.

[21] 万亮, 岑海燕, 朱姜蓬, 等. 基于纹理特征与植被指数融合的水稻含水量无人机遥感监测[J]. 智慧农业(中英文), 2020, 2(1): 58 – 67.
WAN Liang, CEN Haiyan, ZHU Jiangpeng, et al. Using fusion of texture features and vegetation indices from water concentration in rice crop to UAV remote sensing monitor[J]. Smart Agriculture, 2020, 2(1): 58 – 67. (in Chinese)

[22] BENDIG J, YU K, AASEN H, et al. Combining UAV-based plant height from crop surface models, visible, and near infrared

vegetation indices for biomass monitoring in barley[J]. International Journal of Applied Earth Observation and Geoinformation, 2015, 39: 79 – 87.

[23] GITELSON A A, KAUFMAN Y J, STARK R, et al. Novel algorithms for remote estimation of vegetation fraction[J]. Remote Sensing of Environment, 2002, 80(1): 76 – 87.

[24] 汪小钦, 王苗苗, 王绍强, 等. 基于可见光波段无人机遥感的植被信息提取[J]. 农业工程学报, 2015, 31(5):152 – 159.

WANG Xiaoqin, WANG Miaomiao, WANG Shaoqiang, et al. Extraction of vegetation information from visible unmanned aerial vehicle images[J]. Transactions of the CSAE, 2015, 31(5):152 – 159. (in Chinese)

[25] LEE M K, GOLZARIAN M R, KIM I. A new color index for vegetation segmentation and classification [J]. Precision Agriculture, 2021, 22(1): 179 – 204.

[26] GITELSON A A, VIÑA A, CIGANDA V, et al. Remote estimation of canopy chlorophyll content in crops[J]. Geophysical Research Letters, 2005, 32(8): L08403.

[27] DATT B. Visible/near infrared reflectance and chlorophyll content in *Eucalyptus* leaves[J]. International Journal of Remote Sensing, 2010, 20(14): 2741 – 2759.

[28] TUCKER C J. Red and photographic infrared linear combinations for monitoring vegetation [J]. Remote Sensing of Environment, 1979, 8(2): 127 – 150.

[29] GITELSON A A, KAUFMAN Y J, MERZLYAK M N. Use of a green channel in remote sensing of global vegetation from EOS – MODIS[J]. Remote Sensing of Environment, 1996, 58(3): 289 – 298.

[30] DASH J, CURRAN P J. Evaluation of the MERIS terrestrial chlorophyll index (MTCI) [J]. Advances in Space Research, 2007, 39(1): 100 – 104.

[31] GONG P, PU R L, BIGING G S, et al. Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41(6): 1355 – 1362.

[32] CHEN J M. Evaluation of vegetation indices and a modified simple ratio for boreal applications [J]. Canadian Journal of Remote Sensing, 2014, 22(3): 229 – 242.

[33] PENG Y, GITELSON A A. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content[J]. Remote Sensing of Environment, 2012, 117: 440 – 448.

[34] MISHRA S, MISHRA D R. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-*a* concentration in turbid productive waters[J]. Remote Sensing of Environment, 2012, 117: 394 – 406.

[35] LIU H Q, HUETE A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise[J]. IEEE Transactions on Geoscience & Remote Sensing, 1995, 33(2): 457 – 465.

[36] SRIPADA R P, HEINIGER R W, WHITE J G, et al. Aerial color infrared photography for determining early in-season nitrogen requirements in corn[J]. Agronomy Journal, 2006, 98(4): 968 – 977.

[37] CAO Q, MIAO Y X, WANG H Y, et al. Non-destructive estimation of rice plant nitrogen status with crop circle multispectral active canopy sensor[J]. Field Crops Research, 2013, 154: 133 – 144.

[38] ROUJEAN J L, BREON F M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements[J]. Remote Sensing of Environment, 1995, 51(3): 375 – 384.

[39] HUANG S Y, MIAO Y X, YUAN F, et al. Potential of RapidEye and WorldView – 2 satellite data for improving rice nitrogen status monitoring at different growth stages[J]. Remote Sensing, 2017, 9(3): 227.

[40] BIRTH G S, MCVEY G R. Measuring the color of growing turf with a reflectance spectrophotometer[J]. Agronomy Journal, 1968, 60(6): 640 – 643.

[41] HUETE A R. A soil-adjusted vegetation index (SAVI) [J]. Remote Sensing of Environment, 1988, 25(3): 295 – 309.

[42] RONDEAUX G, STEVEN M, BARET F. Optimization of soil-adjusted vegetation indices[J]. Remote Sensing of Environment, 1996, 55(2): 95 – 107.

[43] PERRY JR C R, LAUTENSCHLAGER L F. Functional equivalence of spectral vegetation indices[J]. Remote Sensing of Environment, 1984, 14(1 – 3): 169 – 182.

[44] BROGE N H, LEBLANC E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density[J]. Remote Sensing of Environment, 2001, 76(2): 156 – 172.

[45] REYNIERS M, WALVOORT D J J, DE BAARDEMAAKER J. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat[J]. International Journal of Remote Sensing, 2007, 27(19): 4159 – 4179.

[46] GITELSON A A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation[J]. Journal of Plant Physiology, 2004, 161(2): 165 – 173.

[47] 潘梦瑶, 任瑛, 王思源, 等. 基于梯度提升算法和 SHAP 的石家庄 PM2.5 和臭氧浓度预测及影响因素分析[J]. 环境科学学报, 2024, 44(7): 1 – 8.

PAN Mengyao, REN Ying, WANG Siyuan, et al. Prediction of PM2.5 and ozone concentration in Shijiazhuang and analysis of influencing factors based on gradient boosting algorithm and SHAP[J]. Acta Scientiae Circumstantiae, 2024, 44 (7): 1 – 8. (in Chinese)

[48] 朱志畅,葛焱,臧晶荣,等. 基于无人机图像和 SHAP 特征筛选的小麦田间产量预测方法研究[J/OL]. 麦类作物学报, <http://link.cnki.net/urlid/61.1359.S.20240509.1045.026>.

[49] ALI A,IMRAN M. Remotely sensed real-time quantification of biophysical and biochemical traits of citrus (*Citrus sinensis* L.) fruit orchards—a review[J]. Scientia Horticulturae, 2021, 282: 110024.

[50] CHEN R Q, ZHANG C J, XU B, et al. Predicting individual apple tree yield using UAV multi-source remote sensing data and ensemble learning[J]. Computers and Electronics in Agriculture, 2022, 201: 107275.

[51] CAI J, LUO J W, WANG S L, et al. Feature selection in machine learning: a new perspective[J]. Neurocomputing, 2018, 300: 70 – 79.

[52] UPENDAR K, AGRAWAL K. N, CHANDEL N. S, et al. Greenness identification using visible spectral colour indices for site specific weed management[J]. Plant Physiology Reports, 2021, 26(1): 179 – 187.

[53] KAZEMI F,GHANBARI PARMEHR E. Evaluation of RGB vegetation indices derived from UAV images for rice crop growth monitoring[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2023, 10: 385 – 390.

[54] FAN X, GAO P, ZHANG M, et al. The fusion of vegetation indices increases the accuracy of cotton leaf area prediction[J]. Frontiers in Plant Science, 2024, 15: 1357193.

[55] XU W C, YANG F F, MA G C, et al. Multiscale inversion of leaf area index in citrus tree by merging UAV LiDAR with multispectral remote sensing data[J]. Agronomy, 2023, 13(11): 2747.

(上接第 256 页)

[19] 张德胜,杨港,赵旭涛,等. 基于 BP 神经网络的立式离心泵导叶与蜗壳优化设计[J]. 农业机械学报,2022,53(4):130 – 139.
ZHANG Desheng,YANG Gang,ZHAO Xutao,et al. Optimization design of vane diffuser and volute in vertical centrifugal pump based on back propagation neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(4): 130 – 139. (in Chinese)

[20] 杨魏,王福军,王宏. 离心风机叶片三维反问题优化设计[J]. 农业机械学报,2012,43(8):105 – 109.
YANG Wei,WANG Fujun,WANG Hong. Aerodynamic optimization design of centrifugal fan blades based on 3D inverse design method[J]. Transactions of the Chinese Society for Agricultural Machinery,2012,43(8):105 – 109. (in Chinese)

[21] 王春林,叶剑,曾成,等. 基于 NSGA-II遗传算法高比转速混流泵多目标优化设计[J]. 农业工程学报,2015,31(18):100 – 106.
WANG Chunlin,YE Jian,ZENG Cheng,et al. Multi-objective optimum design of high specific speed mixed-flow pump based on NSGA-II genetic algorithm[J]. Transactions of the CSAE,2015,31(18):100 – 106. (in Chinese)

[22] 王春林,冯一鸣,叶剑,等. 基于 RBF 神经网络与 NSGA-II 算法的渣浆泵多目标参数优化[J]. 农业工程学报,2017, 33(10):109 – 115.
WANG Chunlin,FENG Yiming,YE Jian,et al. Multi-objective parameters optimization of centrifugal slurry pump based on RBF neural network and NSGA-II genetic algorithm[J]. Transactions of the CSAE,2017,33(10):109 – 115. (in Chinese)

[23] 王春林,胡蓓蓓,冯一鸣,等. 基于径向基神经网络与粒子群算法的双叶片泵多目标优化[J]. 农业工程学报,2019, 35(2):25 – 32.
WANG Chunlin,HU Beibei,FENG Yiming,et al. Multi-objective optimization of double vane pump based on radial basis neural network and particle swarm[J]. Transactions of the CSAE,2019,35(2):25 – 32. (in Chinese)

[24] 袁建平,范猛,GIOVANNI P,等. 高比转速轴流泵正交优化设计研究[J]. 振动与冲击,2018,37(22):115 – 121.
YUAN Jianping,FAN Meng,GIOVANNI P,et al. Orthogonal optimum design method for high specific-speed axial-flow pumps [J]. Journal of Vibration and Shock,2018,37(22):115 – 121. (in Chinese)

[25] 龙云. 喷水推进泵水力优化设计方法及空化研究[D]. 上海:上海交通大学,2018.
LONG Yun. Study on the hydraulic optimization design method and cavitation of waterjet pump[D]. Shanghai:Shanghai Jiao Tong University, 2018. (in Chinese)

[26] SMOLA A J, BARTLETT P. Sparse greedy Gaussian process regression[C]// Neural Information Processing Systems. MIT Press, 2000.