

融合特征优选与随机森林算法的GF-6影像 东北一季稻遥感提取

张悦琦, 任鸿瑞

太原理工大学 矿业工程学院, 太原 030024

摘要: 为寻求高效、高精度的东北一季稻种植面积提取方法, 该研究以辽宁省盘锦市为研究区, 利用覆盖水稻关键物候期的6景GF-6 WFV单时相影像和时序影像, 构建光谱特征、植被指数、水体指数和红边指数4类特征变量, 采用平均不纯度减少的方法进行重要性排序并通过袋外误差方法选择最优输入特征, 建立基于特征优选的随机森林模型, 对2020年盘锦市水稻种植分布进行提取。结果表明: (1) 基于水稻不同物候期的单时相影像, 总体分类精度均在94%以上, 以处于水稻移栽期影像分类结果最佳, 其总体精度、F1值(水稻)、Kappa系数与实地验证点精度分别为97.67%、98.84%、0.97和97.22%; (2) 与单时相影像相比, 利用时序影像进行土地覆被分类和水稻信息提取能够有效提高分类精度, 其总体精度、F1值(水稻)、Kappa系数与实地验证点精度分别为99.33%、100.00%、0.99和97.22%; (3) 对有无红边信息参与的水稻提取结果进行对比分析, 红边波段和红边指数的引入可使分类精度有所提高; (4) 引入紫边与黄边波段能够提高分类精度, 但分类结果精度提高效果次于红边信息。该研究证明, 基于特征优选的随机森林模型, 利用水稻移栽期的单时相影像提取水稻种植分布可满足实际应用精度需求, 但利用时序影像可进一步提高分类精度。此外, GF-6卫星的新增波段均能够提高水稻分类精度, 显示出GF-6卫星在作物精细提取方面具有巨大应用潜力。

关键词: 遥感, 随机森林, 红边波段, 特征优选, 高分六号, 水稻, 紫边波段, 黄边波段

中图分类号: TP79/S127/P2

引用格式: 张悦琦, 任鸿瑞. 2023. 融合特征优选与随机森林算法的GF-6影像东北一季稻遥感提取. 遥感学报, 27(9): 2153-2164
Zhang Y Q and Ren H R. 2023. Remote sensing extraction of paddy rice in Northeast China from GF-6 images by combining feature optimization and random forest. National Remote Sensing Bulletin, 27(9): 2153-2164 [DOI: 10.11834/jrs.20221338]

1 引言

中国是世界上最大的水稻生产国和消费国, 其中, 东北三省(黑龙江、吉林、辽宁)一直是水稻的重要产区, 种植面积约占全国的17%, 为中国粮食生产和农业经济的快速发展作出巨大贡献(方福平和程式华, 2018; 陆娣等, 2020)。及时准确获取水稻种植分布信息, 对估算水稻产量、制定农业政策和保障国家粮食安全具有重要作用(李志鹏等, 2014)。

高分六号(GF-6)卫星是中国首颗设置红边波段的光学遥感卫星, 于2018年在酒泉卫星中心成功发射, 具有窄幅(PMS)和宽幅(WFV)2种

不同分辨率影像。其中, GF-6 WFV影像在传统蓝、绿、红、近红外4个波段的基础上新增红边I、红边II、紫边和黄边波段, 具有更加丰富的波段信息。已有研究表明, 红边波段对作物的微小变化敏感, 在作物识别与分类方面发挥重要作用(Elvidge和Chen, 1995; 苏伟等, 2019; 李文杰等, 2020; 孙敏轩等, 2020; Kim和Yeom, 2014; Mehdaoui和Anane, 2020)。李前景等(2021)提出一种适用于GF-6 WFV影像红边波段的卷积神经网络模型算法, 成功对作物种植区进行分类识别并验证了GF-6 WFV红边波段对作物分类的敏感性; 梁继等(2020)基于GF-6影像构建红边植被指数, 组合多个影像波段后利用支持向量机框

收稿日期: 2021-05-31; 预印本: 2022-02-21

基金项目: 山西省省筹资金资助回国留学人员科研项目(编号:2022-055)

第一作者简介: 张悦琦, 研究方向为生态环境遥感。E-mail: yueqizhang1997@126.com

通信作者简介: 任鸿瑞, 研究方向为植被遥感。E-mail: renhongrui@tyut.edu.cn

架对松嫩平原北部作物进行分类, 结果表明红边特征能够显著提高作物识别精度。

随着计算机技术的快速发展, 机器学习近年来广泛应用于水稻、玉米、大豆等作物种植分布提取研究, 其主要包括支持向量机、神经网络、随机森林等算法。神经网络具有较强的自适应、自学习能力, 但分类时间复杂度高、参数难以确定, 且存在过拟合和局部极值问题; 支持向量机较神经网络分类精度高、泛化能力强, 但在核函数的选择与参数设置方面缺乏理论依据(闫利和江维薇, 2016; 贾坤等, 2011)。因此, 具有训练速度快、分类精度高、不易过拟合、可处理高维数据等优点的随机森林算法在作物种植提取研究中更受青睐(侯蒙京等, 2020; 陈安旭和李月臣, 2020; 张鹏和胡守庚, 2019; Immitzer等, 2016)。黄健熙等(2017)利用多时相GF-1 WFV数据并采用随机森林算法成功提取嫩江县玉米与大豆种植分布, 结果表明随机森林算法较支持向量机算法分类精度更高; 王李娟等(2020)基于Sentinel-2数据生成光谱特征、植被指数、红边指数和纹理特征, 根据特征重要性排序构建7种特征组合方案, 分别采用随机森林算法与支持向量机算法对农耕地进行土地覆被分类, 结果表明, 随机森林分类精度优于相同特征变量下的支持向量机算法, 总体精度可达88.24%。

截止目前, 利用GF-6 WFV影像提取水稻种植分布的研究中, 多注重于红边波段对分类精度影响的探讨与分析, 而紫边与黄边波段对水稻分类精度的影响尚未得知。此外, 由于时序影像包含信息多, 其分类精度通常要优于单时相影像, 但时序影像分类速度较慢、效率较低。因此, 为寻找一种高效、高精度的水稻提取方法, 探索GF-6 WFV新增波段对水稻分类精度的影响, 本研究以辽宁省盘锦市为研究区, 以GF-6 WFV影像为数据源, 利用6景覆盖水稻关键物候期的单时相影像及其组合的时序影像, 通过构建光谱特征、植被指数、水体指数和红边指数4类特征变量, 基于特征优选的随机森林模型开展水稻种植分布提取研究。根据验证样本和实地验证点对提取结果进行精度验证, 对比分析后获得提取水稻信息的最优特征和最佳时相, 为国产GF-6卫星的推广应用提供更多理论方案。

2 研究区与数据

2.1 研究区概况

盘锦市地处辽宁省西南部, 介于 $40^{\circ}39'N$ — $41^{\circ}27'N$, $121^{\circ}25'E$ — $122^{\circ}30'E$, 下辖有盘山县、双台子区、大洼区和兴隆台区(图1), 是中国重要的商品粮食生产基地和大米出口基地。研究区属暖温带大陆性半湿润季风气候, 具有光照充足、昼夜温差大、四季分明等特点, 地势平坦, 平均海拔高度约4 m, 且土壤略偏碱性。根据2019盘锦统计年鉴, 该区年均气温 $8^{\circ}C$ — $11^{\circ}C$, 年均降水618.62 mm。因其独特的地理环境和气候条件, 盘锦市适宜农作, 土地覆盖类型以耕地为主, 其中约85%为灌溉水田, 约14%为旱地, 因此水稻在全市粮食生产中占有重要地位。结合2020版30 m全球地表覆盖数据GlobeLand30(Chen等, 2015), 水稻、旱地等耕地占研究区总面积约70%, 除耕地外的主要土地覆盖类型有水体、建筑用地和天然湿地。水体与天然湿地分布较为集中, 主要位于盘锦市南部与西南部; 建筑用地遍布全市, 呈破碎化分布, 占研究区总面积约15%。

结合Google Earth高分辨率影像进行目视解译, 根据研究区土地覆盖情况共获取样本点2000个(图1(b))。样本点均匀分布于研究区内, 包括1000个水稻样本点, 250个水体样本点, 300个天然湿地样本点, 150个旱地作物样本点, 300个建筑用地样本点。其中, 70%用作训练样本, 30%用作验证样本。此外, 通过实地调查获得36个水稻实地验证点(图1(b)), 并于每个点上拍摄实地照片。

2.2 遥感数据及预处理

本研究选用GF-6卫星WFV传感器, 获取6景2020年5—8月覆盖盘锦市的L1A级遥感影像, 分别对应三叶期、移栽期、返青期、孕穗期和抽穗期等水稻关键物候期, 成像时间分别为2020年5月11日、5月25日、6月1日、6月6日、7月20日与8月22日, 空间分辨率为16 m。其中, 返青期由6月1日与6月6日2景影像共同覆盖。为便于表述, 以5.11、5.25、6.01、7.20、8.22分别代表三叶期、移栽期、返青期、孕穗期和抽穗期影像。GF-6 WFV影像具有蓝波段(0.45 — $0.52\ \mu m$)、绿波段(0.52 — $0.59\ \mu m$)、红波段(0.63 — $0.69\ \mu m$)、近红

外波段 ($0.77\text{--}0.89\ \mu\text{m}$)、红边 I 波段 ($0.69\text{--}0.73\ \mu\text{m}$)、红边 II 波段 ($0.73\text{--}0.77\ \mu\text{m}$)、紫边波段 ($0.40\text{--}0.45\ \mu\text{m}$) 和黄边波段 ($0.59\text{--}0.63\ \mu\text{m}$) 8 个波段 (陆春玲等, 2021)。

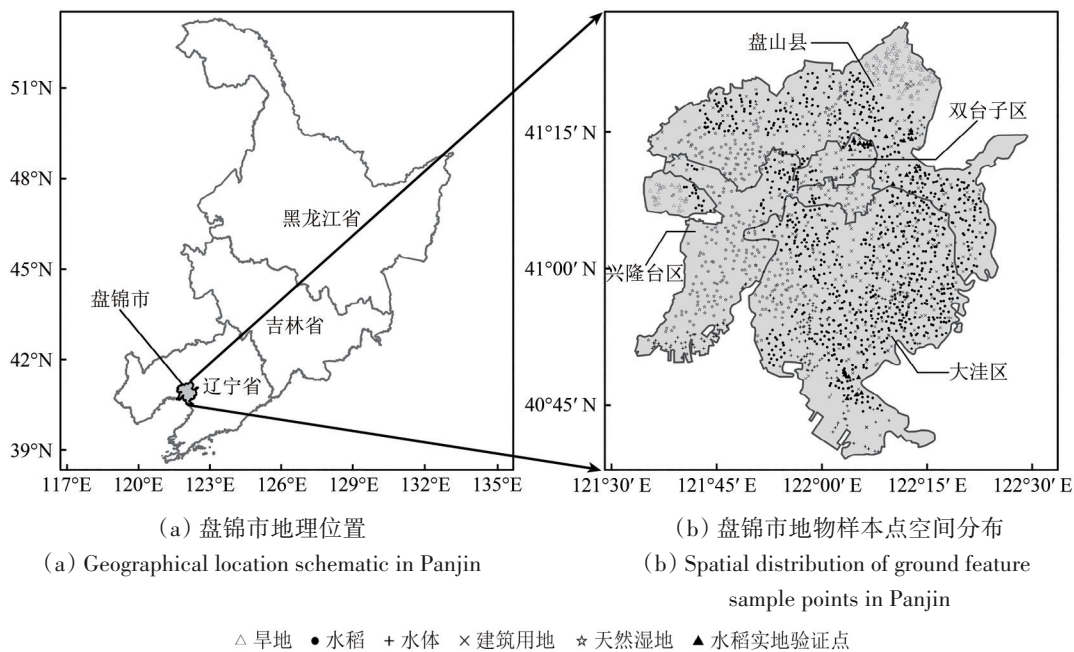


图 1 盘锦市地理位置和样本点空间分布

Fig. 1 Geographical location schematic and spatial distribution of ground feature sample points in Panjin

在 ENVI 5.3 遥感图像处理软件中进行影像预处理, 包括辐射定标、大气校正、正射校正和影像裁剪。遥感影像、绝对辐射定标系数和波谱响应函数可从中国资源卫星应用中心 ([http://www.cresda.com/CN/\[2021-05-31\]](http://www.cresda.com/CN/[2021-05-31])) 下载。

3 研究方法

3.1 特征提取

本研究选取光谱特征、植被指数、水体指数和红边指数共 20 个特征变量 (表 1)。

在影像预处理的基础上, 选取 GF-6 WFV 影像的 8 个波段作为光谱特征。反映土地覆盖类型时, 植被指数和水体指数比单波段信息更为稳定, 可在一定程度上提高分类精度 (何云等, 2019), 故选取具有代表性的 7 种植被指数和 1 种水体指数作为随机森林识别的特征指数波段。红边波段多指植被反射率在近红外波段与红光波段之间快速变化的区域, 是指示绿色植物生长状况的敏感性波段 (Horler 等, 1983)。鉴于此, 本研究将 GF-6 WFV 影像的 2 个红边波段分别代替归一化植被指数 NDVI (Normalized Difference Vegetation Index) 中的近红外波段, 得到 2 个红边指数 (梁继等, 2020), 即归一化植被指数 710 即 NDVI710 (Normalized

Difference Vegetation Index 710) 和归一化植被指数 750 即 NDVI750 (Normalized Difference Vegetation Index 750)。此外, 借鉴了 Sentinel-2 卫星的 2 种红边指数 (Clevers 和 Gitelson, 2013; Dash 和 Curran, 2004), 即归一化差异红边 1 指数 NDRE1 (Normalized Difference Red Edge 1 Index) 和地面叶绿素指数 MTCI (Terrestrial Chlorophyll Index)。

3.2 随机森林分类

随机森林算法是由 Breiman (2001) 提出的以决策树为基本分类器, 结合 Bagging 集成学习理论与随机子空间方法的一种集成学习方法。随机森林通过自助法 (Bootstrap) 从原始训练集中有放回地随机选取 N 个样本作为训练样本, 因此会有约 37% 的样本未被抽取到, 这些样本被称为袋外 OOB (Out of Bag) 数据, 可用于评估随机森林模型性能 (Breiman, 1996)。经 N_{tree} 次样本抽取和训练可得到 N_{tree} 个决策树模型, 在每棵决策树的每个节点处随机选取 M_{try} ($M_{\text{try}} < M$, M 为总特征数) 个特征并进行内部节点划分, 最后对所有决策树模型结果进行集成, 采用多数投票方式决定最终分类结果。已有研究表明, 决策树个数为 100 时 OOB 误差最为稳定 (黄建文等, 2021), 因此本文最终将 N_{tree} 设置为 100, M_{try} 设置为输入特征总数的平方根。

表1 特征变量信息和遥感指数计算公式

Table 1 Information of characteristic variables and formula for calculating remote sensing indices

类别	特征变量	特征简称	描述
光谱特征	B1	B1	蓝波段
	B2	B2	绿波段
	B3	B3	红波段
	B4	B4	近红外波段
	B5	B5	红边 I 波段
	B6	B6	红边 II 波段
	B7	B7	紫边波段
	B8	B8	黄边波段
植被指数	Normalized Difference Vegetation Index	NDVI	$\frac{b4 - b3}{b4 + b3}$
	Ratio Vegetation Index	RVI	$\frac{b4}{b3}$
	Enhanced Vegetation Index	EVI	$\frac{2.5 \cdot (b4 - b3)}{b4 + 6 \cdot b3 - 7.5 \cdot b1 + 1}$
	Difference Vegetation Index	DVI	$b4 - b3$
	Green Index	GI	$\frac{b2}{b3}$
	Normalized Difference Green Index	NDGI	$\frac{b1 - b3}{b1 + b3}$
	Triangular Vegetation Index	TVI	$\sqrt{NDVI + 0.5}$
水体指数	Normalized Difference Water Index	NDWI	$\frac{b2 - b4}{b2 + b4}$
红边指数	Normalized Difference Red Edge 1 Index	NDRE1	$\frac{b6 - b5}{b6 + b5}$
	Normalized Difference Vegetation Index 710	NDVI710	$\frac{b5 - b3}{b5 + b3}$
	Normalized Difference Vegetation Index 750	NDVI750	$\frac{b6 - b3}{b6 + b3}$
	Terrestrial Chlorophyll Index	MTCI	$\frac{b6 - b5}{b5 - b3}$

注: b1、b2、b3、b4、b5 与 b6 分别代表蓝波段、绿波段、红波段、近红外波段、红边 I 波段和红边 II 波段的反射率。

随机森林中决策树构建时采用基尼指数进行节点分裂时的特征选择。基尼指数表示样本集合中随机样本被错分的概率, 基尼指数越小则集合的纯度越高, 被错分的概率越小; 反之, 集合越不纯。集合 D 的基尼指数 $Gini(D)$ 如式 (1) 所示:

$$Gini(D) = 1 - \sum_{k=1}^K P_k^2 \quad (1)$$

式中, K 为训练样本中的种类数; P_k 为集合 D 中随机选取的样本属于类别 K 的概率。

若集合 D 根据特征 A 是否取某一值 α 被分为 D_1 和 D_2 两部分, 则在特征 A 的条件下, 集合 D 的基尼指数 $Gini(D, A)$ 如式 (2) 所示:

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (2)$$

式中, $|D|$ 表示集合 D 中的样本数, $|D_1|$ 表示集合 D_1 中的样本数, $|D_2|$ 表示集合 D_2 中的样本数。由此看出, 随机森林模型构建时, 若通过某特征划分后的基尼指数减少的程度越大, 则划分后集合变纯的程度越大, 该特征在模型中越重要。因此, 采

用该方法即平均不纯度减少的方法可评估特征重要性, 其定义如式 (3) 所示:

$$\Delta Gini = \frac{1}{N_{tree}} \sum_{t=1}^{N_{tree}} (Gini_t(D) - Gini_t(D, A)) \quad (3)$$

式中, N_{tree} 为随机森林模型中决策树个数; $Gini_t(D)$ 、 $Gini_t(D, A)$ 分别为第 t 棵决策树经特征 A 划分前后集合 D 的基尼指数。

由于所有特征参与分类将导致信息冗余, 从而降低分类精度和分类速度, 因此本研究在 Python 平台下通过调用 Scikit-learn 库建立随机森林模型, 采用平均不纯度减少的方法计算模型中各特征重要性并按照从高到低的顺序进行排序。根据重要性排序, 第 1 次选择首位特征, 第 2 次选择前 2 位特征, 依次类推可得到单时相 20 个不同特征组合、时序影像 100 个不同特征组合的随机森林模型, 分别计算其 OOB 误差, 对模型精度和模型复杂度综合考虑后确定最优特征组合, 从而在保证分类精度的条件下降低模型复杂度。

3.3 精度分析

利用验证样本获得混淆矩阵，计算总体精度 OA (Overall Accuracy)、Kappa 系数和 F1 值 (F1_score)，从而对特征优选的随机森林分类结果进行精度评价，计算公式如式 (4)~式 (6) 所示。OA 为验证样本集上分类器预测正确的概率。Kappa 系数是检验验证样本与预测结果吻合度的指标，取值范围为 [0, 1]，越接近 1 表示模型吻合度越高 (冯锐等, 2017)。F1 值是制图精度和用户精度的调和均值，可用于表征模型输出的好坏，取值范围为 [0, 1]，越接近 1 表示模型输出效果越好 (杨建宇等, 2019)。

$$OA = \frac{\sum_{i=1}^n m_i}{N} \quad (4)$$

$$Kappa = \frac{N \cdot \sum_{i=1}^n m_i - \sum_{i=1}^n (G_i \cdot C_i)}{N^2 - \sum_{i=1}^n (G_i \cdot C_i)} \quad (5)$$

$$F1 = \frac{2 \cdot P \cdot U}{P + U} \quad (6)$$

式中， m_i 为第 i 类的分类正确样本数； C_i 为第 i 类的分类像元总数； G_i 为第 i 类的真实像元总数； N 为分类像元总数； n 为分类数； P 和 U 分别为制图精度 PA (Producer Accuracy) 和用户精度 UA (User Accuracy)，计算公式如式 (7) 所示：

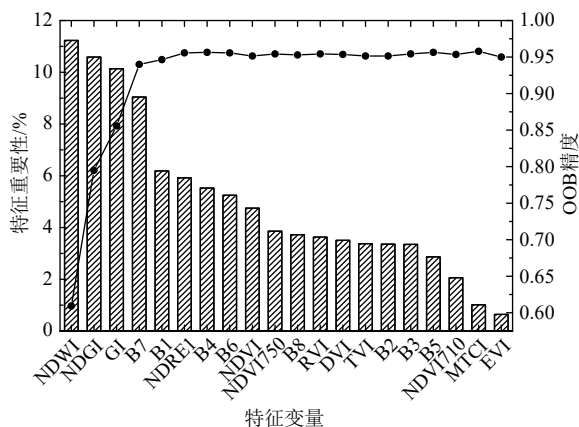
$$P = \frac{m_i}{C_i}, U = \frac{m_i}{G_i} \quad (7)$$

此外，实地验证点均为水稻且未在研究区均匀分布，故仅计算总体精度进行说明。

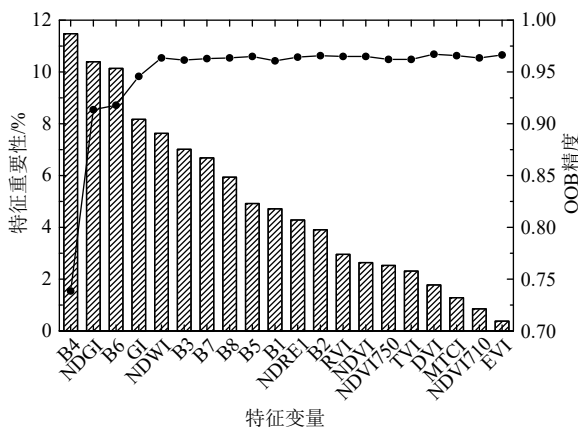
4 结果与分析

4.1 特征重要性评估

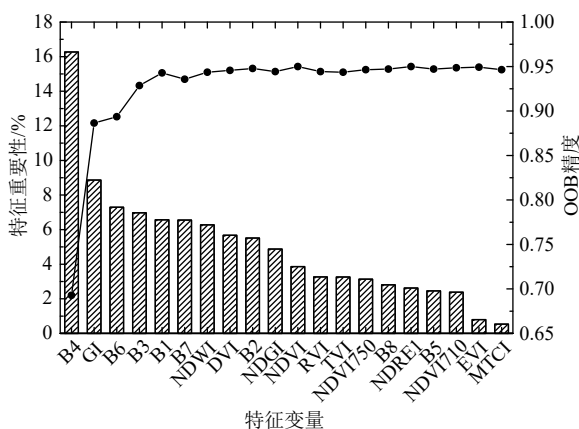
采用平均不纯度减少的方法，分别基于单时相 GF-6 WFV 影像和时序影像计算模型中各特征重要性 (图 2)。由于基于时序影像的特征数多达 100 个，此处仅列出前 20 位。



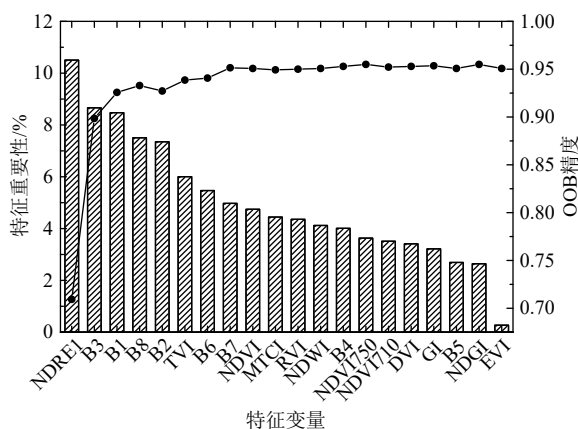
(a) 三叶期影像特征重要性及 OOB 精度
(a) Feature importance and OOB accuracy of trefoil stage image



(b) 移栽期影像特征重要性及 OOB 精度
(b) Feature importance and OOB accuracy of transplanting stage image



(c) 返青期影像特征重要性及 OOB 精度
(c) Feature importance and OOB accuracy of returning green image



(d) 孕穗期影像特征重要性及 OOB 精度
(d) Feature importance and OOB accuracy of booting stage image

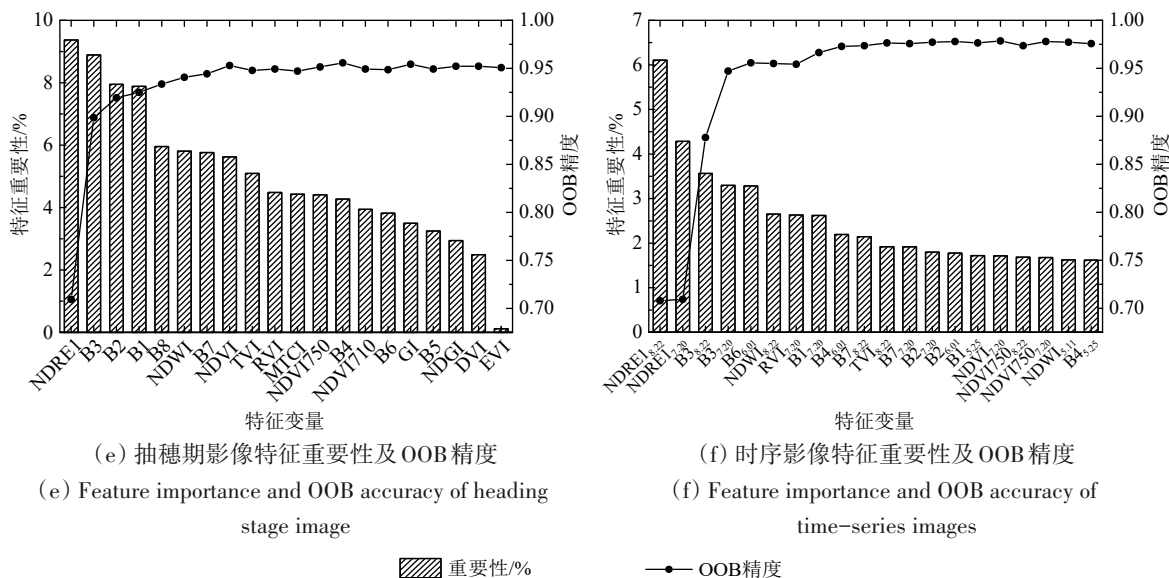


图2 特征重要性评估结果和不同特征组合模型 OOB 精度

Fig. 2 Feature importance assessment results and OOB accuracy of different feature combination models

从水稻三叶期至抽穗期，即2020年5—8月，水体指数和植被指数重要性排名逐渐降低，红边指数的重要性排名逐渐提高（图2），表明红边指数对土地覆盖分类贡献度逐渐增大。2个红边波段中，B6波段重要性相对较高；在NDRE1、MTCI、NDVI710和NDVI750等4个红边指数中，NDRE1的重要性较高，在孕穗期和抽穗期影像中，重要性排名均为首位。究其原因，研究区水稻覆盖范围较大，且5—6月稻田需进行泡田和移栽，呈现出以水体为主的特点，与其他地物类型光谱特征差异较大，归一化水体指数NDWI（Normalized Difference Water Index）对水体信息提取效果较好，故5—6月NDWI的重要性较高。NDRE1用红边的峰和谷来代替NDVI的红光波段与近红外波段，属于窄带绿度指数，相比于宽带绿度指数对叶绿素含量、叶冠层的微小变化等更加敏感，能够提高对作物的识别能力（张沁雨等，2019）。随时间的推移，水稻、天然湿地和旱地作物逐渐生长，7—8月体现出高覆盖植被特点，故NDRE1对土地覆盖分类的贡献度逐渐增大。基于时序影像的重要性排名中，前20位以孕穗期和抽穗期影像的特征量居多，其中光谱特征11个，红边指数4个，植被指数3个，水体指数2个，前2位分别为抽穗期影像和孕穗期影像的红边指数NDRE1，即NDRE1_{8,22}}和NDRE1_{7,20}}。

4.2 最优特征组合

对单时相影像20个不同特征组合和时序影像

100个不同特征组合进行OOB估计，得到OOB精度（图2）。因时序影像特征数过多，图2中仅显示前20位。以三叶期影像为例，特征数从1增加到7，即特征从仅有NDWI到特征包含重要性排序前7位，OOB精度逐渐增加；特征数为7时，OOB精度达到0.956；特征数从7增加到20，OOB精度整体上有微小的下降趋势，但在特征数为19时，OOB精度达到最高，为0.957。由于特征个数过少会导致分类精度较低，特征个数过多会导致信息冗余且增加运行成本，故对模型精度和模型复杂度综合考虑后选择特征重要性排名前7的特征作为随机森林模型的最优输入特征。根据相同特征选择方法，对移栽期、返青期、孕穗期、抽穗期和时序影像进行特征选择，分别选取重要性排名前8、前5、前8、前8和前16的特征作为最优输入特征。

4.3 基于特征优选的随机森林结果

利用单时相影像和时序影像，基于特征优选的随机森林模型对2020年盘锦市水稻种植分布进行提取，获得6种水稻分布结果（图3）。从目视效果来看，6种结果较为一致，水稻主要分布于盘锦市北部、东部与中南部，以大洼区和盘山县为主。较于单时相水稻分类结果，基于时序影像的水稻分类图斑更为规整，一定程度缓解了“椒盐现象”，分类效果最佳。三叶期、移栽期和返青期影像对应的分类结果中，盘锦市西南部水体错分现象明显，原因为此时水稻需进行泡田、移栽，稻田以水体为主，相似的光谱特征导致部分水体

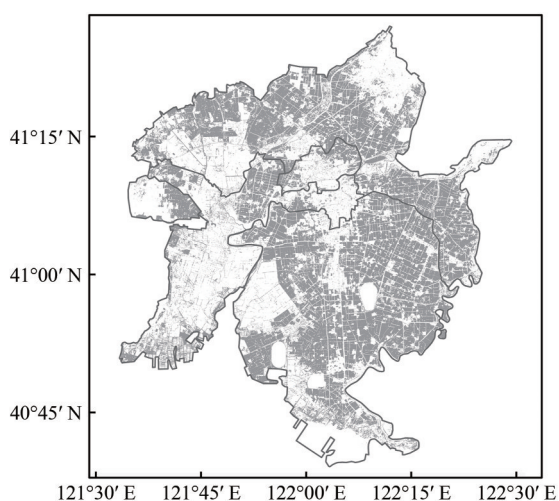
错分为水稻。水稻孕穗期与抽穗期时稻田体现出植被特征，水体错分为水稻现象明显减少，但盘锦市西北与东北部旱地作物错分为水稻现象有所增加，且“椒盐现象”更加明显。

根据混淆矩阵分别计算总体精度、Kappa系数和各土地覆盖类型的F1值，根据水稻实地验证点计算实地验证点精度（表2）。单时相影像分类结果中，移栽期影像分类结果精度最高，其总体精度、F1值（水稻）、Kappa系数和实地验证点精度分别为97.67%、98.84%、0.97与97.22%，较抽穗期影像分类结果分别高出1.00%、1.34%、0.02和2.78%。综合目视情况，单时相分类结果中，移栽期影像分类效果最佳，原因在于此时水稻与研究区内天然湿地、旱地作物物候特征差异较大，易于识别。与单时相影像分类结果相比，基于时序

影像的水稻分类精度显著提高，水稻错分与漏分现象有所减少，总体精度、F1值（水稻）、Kappa系数与实地验证点精度高达99.33%、100.00%、0.99和97.22%，但其分类速度较慢、效率较低。因此，实际应用中为减少运算时间和运行成本，达到高效、高精度的要求，可直接利用GF-6 WFV水稻移栽期影像提取水稻种植信息。

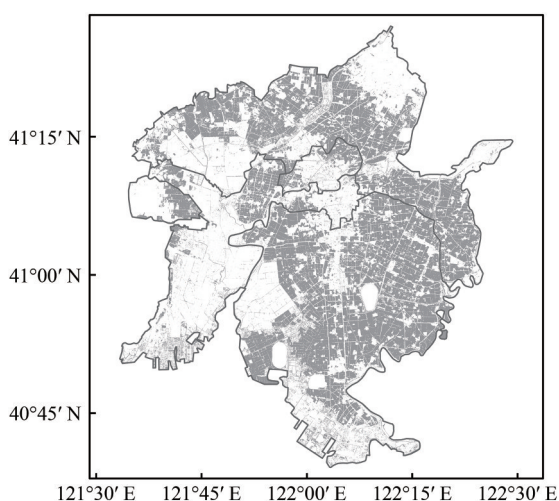
4.4 红边信息对分类精度的影响

GF-6卫星新增2个红边波段，为直观反映红边信息对土地覆盖类型分类精度的影响，将红边波段B5、B6和红边指数NDRE1、MTCI、NDVI710、NDVI750共6个特征去除，使用剩余14个特征并利用单时相影像和时序影像，基于特征优选的随机森林模型对2020年盘锦市水稻种植分布进行提取，根据验证样本和水稻实地验证点进行精度验证（表3）。



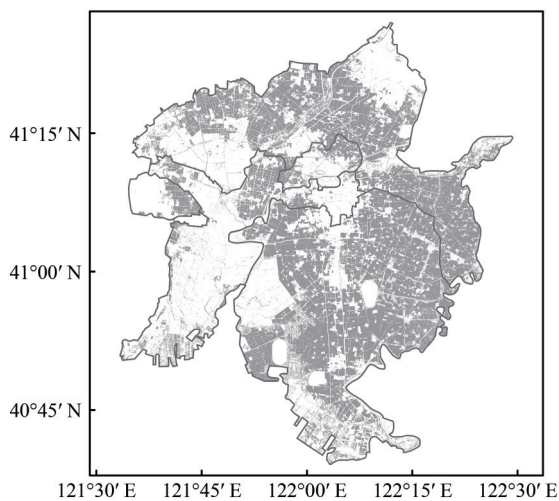
(a) 三叶期影像水稻分布图

(a) Paddy rice distribution of trefoil stage image



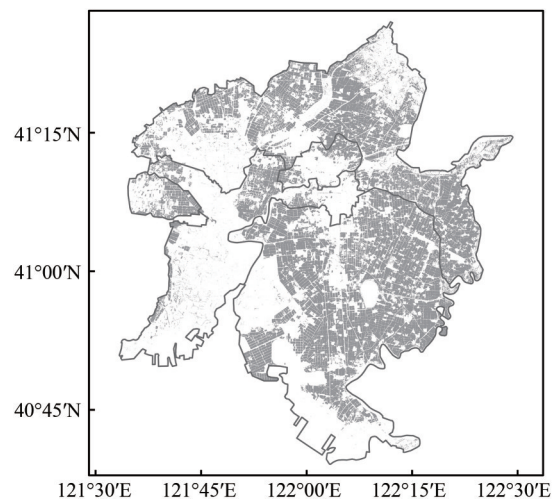
(b) 移栽期影像水稻分布图

(b) Paddy rice distribution of transplanting stage image



(c) 返青期影像水稻分布图

(c) Paddy rice distribution of returning green image



(d) 孕穗期影像水稻分布图

(d) Paddy rice distribution of booting stage image

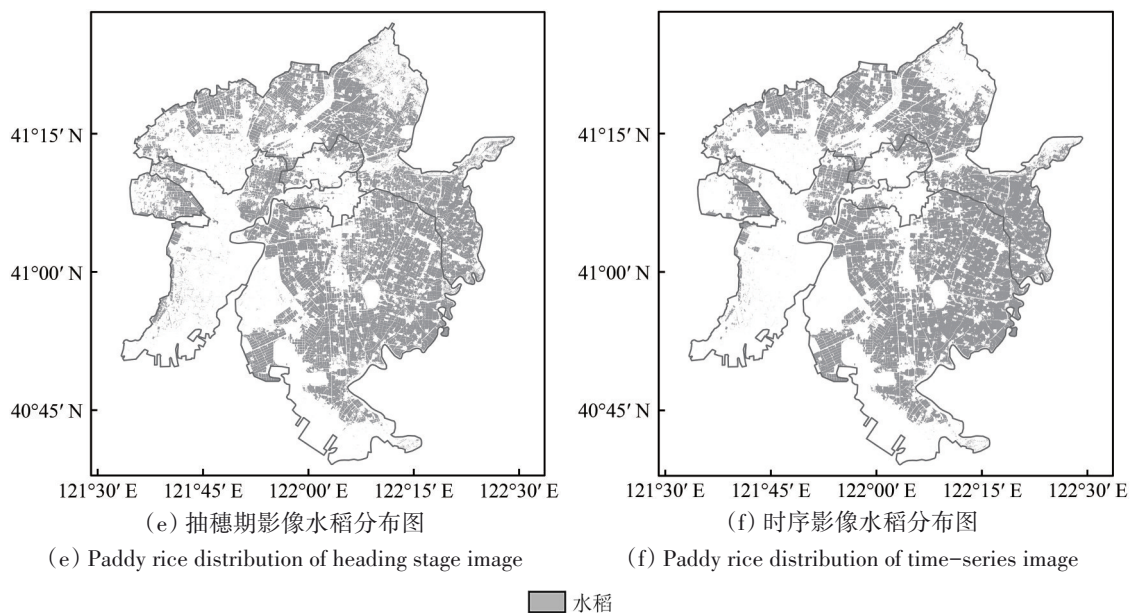


图3 基于不同物候期影像的2020年盘锦市水稻分布图

Fig. 3 Paddy rice distributions based on different images of phenological phases in Panjin City in 2020

表2 基于特征优选随机森林模型的2020年盘锦市水稻分类结果精度评价

Table 2 Paddy rice classification accuracy based on feature-optimized random forest in Panjin City in 2020

遥感影像	影像对应物候期					
	三叶期	移栽期	返青期	孕穗期	抽穗期	时序
建筑用地	96.09	98.31	93.33	98.89	98.34	98.34
天然湿地	93.85	96.13	96.00	96.13	95.60	98.88
F1值/%						
旱地	93.33	95.74	83.87	82.22	86.67	96.70
水稻	97.38	98.84	97.21	97.16	97.50	100.00
水体	91.55	95.24	88.11	100.00	98.65	100.00
总体精度/%	95.67	97.67	94.33	96.50	96.67	99.33
Kappa系数	0.94	0.97	0.92	0.95	0.95	0.99
实地验证点精度/%	91.67	97.22	97.22	94.44	94.44	97.22
特征个数	7	8	5	8	8	16

结果表明, 三叶期和返青期影像的有红边信息与无红边信息提取结果精度较为一致; 移栽期影像的无红边信息分类结果较有红边信息时精度有所下降, F1值(水稻)与总体精度分别降低0.33%和0.84%; 孕穗期和抽穗期影像的无红边信息分类结果精度显著下降, F1值(水稻)分别降低0.47%和0.82%, 总体精度分别降低1.67%和1.84%; 时序影像的无红边信息分类结果较有红边信息时总体精度下降0.33%。水稻实地验证点较少且均在稻田旁的道路上获取, 因此仅能通过目视解译对分类结果进行判读, 由于有无红边信息分类结果为像元级差异, 故实地验证精度未发生变

化。三叶期和返青期影像的无红边信息分类结果精度未降低的原因在于, 在重要性排序中红边波段和红边指数的排名均靠后, 对土地覆盖分类贡献度较小; 移栽期影像的特征重要性排序中, 红边波段B6位列第3, 故无红边信息分类结果精度有所下降; 而孕穗期和抽穗期影像的特征重要性排序中, 红边指数NDRE1均位居首位且重要性较高, 对土地覆盖分类贡献度较大, 故精度下降较移栽期影像更为明显。时序影像中红边指数特征重要性排名靠前, 但无红边信息分类结果精度有所降低却不明显, t 检验结果表明两种分类结果在统计学上无显著性差异($P=0.58>0.05$)。分析其原因在于, 水稻在全生育期内与其他地物类型相比有着独特的光谱特点, 即生育期前期体现水体特征, 后期体现植被特征, 故时序影像中仅利用光谱特征、植被指数与水体指数对水稻种植分布进行提取, 也可获得较好效果, 这也表明时序信息可解决因缺少红边信息而分类精度下降的问题。

4.5 紫边、黄边波段对分类精度的影响

为反映GF-6 WFV新增紫边(B7)和黄边(B8)波段对提取作物种植分布的影响, 将B7、B8波段从20个特征变量中剔除, 使用剩余18个特征并利用单时相影像及时序影像, 基于特征优选的随机森林模型提取2020年盘锦市水稻种植分布并进行精度验证(表4)。

表3 无红边信息且基于特征优选随机森林模型的2020年
盘锦市水稻分类结果精度评价

Table 3 Paddy rice classification accuracy without
red-edge information based on feature-optimized
random forest in Panjin City in 2020

遥感影像	影像对应物候期						
	三叶期	移栽期	返青期	孕穗期	抽穗期	时序	
建筑用地	95.51	95.95	95.14	97.21	97.78	97.80	
天然湿地	94.44	95.08	96.59	92.90	92.31	98.32	
F1值/%	旱地	93.33	94.74	86.36	76.19	75.00	95.56
	水稻	97.70	98.51	97.38	96.69	96.68	100.00
	水体	91.67	94.52	89.94	97.33	98.65	99.33
总体精度/%		95.83	96.83	95.00	94.83	94.83	99.00
Kappa系数		0.94	0.95	0.93	0.92	0.92	0.99
实地验证点精度/%		91.67	97.22	97.22	94.44	94.44	97.22
特征个数		7	5	7	7	9	13

表4 无B7、B8波段且基于特征优选随机森林模型的2020年
盘锦市水稻分类结果精度评价

Table 4 Paddy rice classification accuracy without bands
B7 and B8 based on feature-optimized random forest
in Panjin City in 2020

遥感影像	影像对应物候期						
	三叶期	移栽期	返青期	孕穗期	抽穗期	时序	
建筑用地	96.09	96.05	93.33	97.78	97.78	98.32	
天然湿地	92.66	95.03	96.05	96.13	96.67	98.89	
F1值/%	旱地	93.33	95.74	82.61	76.40	78.26	96.70
	水稻	97.38	98.67	97.36	96.84	96.32	1.00
	水体	31.67	94.59	88.28	97.99	97.33	1.00
总体精度/%		95.50	97.00	94.33	95.50	95.33	99.33
Kappa系数		0.94	0.96	0.92	0.93	0.93	0.99
实地验证点精度/%		91.67	97.22	97.22	94.44	94.44	97.22
特征个数		7	8	7	8	10	15

较全特征且基于特征优选随机森林模型的水稻分类精度(表2),三叶期、移栽期、孕穗期与抽穗期影像分类总体精度分别下降0.17%、0.67%、1.00%和1.34%,其中移栽期、孕穗期与抽穗期影像F1值(水稻)分别下降0.17%、0.32%和1.18%,分类精度有所降低的原因是这4期影像的紫边、黄边波段在重要性排序中排名较为靠前,对土地覆盖分类贡献度较大。较无红边信息水稻分类结果精度(表3),移栽期、孕穗期与抽穗期影像分类总体精度分别提高0.17%、0.67%和0.50%,Kappa系数均提高0.01。结果表明紫边、黄边波段能够

提高水稻分类精度,但红边信息较紫边、黄边波段分类结果精度提高效果更佳,原因在于红边波段或红边指数特征重要性排名较紫边和黄边波段更为靠前。时序影像分类精度与基于特征优选随机森林模型的水稻分类精度(表2)相一致,略高于无红边信息水稻分类结果精度(表3)且差异无统计学意义($P=0.47>0.05$),结果表明,基于无紫边、黄边波段的时序影像可获得较高分类精度,时序信息可解决缺少紫边、黄边波段而分类精度下降的问题。此外,因分类结果为像元级差异且水稻实地验证点位于稻田旁的道路上,故经目视解译判读后,水稻实地验证点精度未发生变化。

5 结论

本研究采用不同物候期的GF-6 WFV影像构建光谱特征、植被指数、水体指数和红边指数等特征变量,基于特征优选的随机森林模型对2020年盘锦市水稻种植分布进行提取并得到以下结论:

(1)不同物候期的单时相影像水稻分类结果表明,利用单时相GF-6 WFV影像并基于特征优选的随机森林模型能够较高精度地获取水稻种植分布信息,总体精度、F1值(水稻)和实地验证点精度均在90%以上。综合目视解译、验证样本精度评价与实地验证点精度评价,处于水稻移栽期的影像分类结果最佳,总体精度、F1值(水稻)、Kappa系数与实地验证点精度分别为97.67%、98.84%、0.97和97.22%。

(2)与单时相影像分类结果相比,利用时序影像提取水稻种植分布,分类精度显著提高,总体精度、F1值(水稻)与Kappa系数分别为99.33%、100.00%和0.99,较移栽期影像分类结果分别提高1.66%、1.16%和0.02。利用时序影像基于随机森林方法进行分类时,较多的特征变量造成模型复杂度较高,分类速度较慢,因此,实际应用中,有红边信息时利用水稻移栽期的单时相影像进行水稻种植分布提取即可满足精度要求且分类效率较高;无红边信息时可利用时序影像进行水稻种植分布提取,从而解决无红边信息时分类精度较低的问题。

(3)红边波段和红边指数在特征重要性排名较高的影像中,特别是在孕穗期和抽穗期等植被覆盖度较高的时期,引入红边波段和红边指数能够提高分类精度,减少错分、漏分现象,优于相

同分类方法下未引入红边波段和红边指数的分类结果。研究表明,红边波段和红边指数的引入能够增强作物识别能力,在作物精细提取方面具有重要作用和广阔前景,但可能由于盘锦市地物类型简单、种植结构单一,引入红边信息后分类精度提高效果不显著。

(4) 根据重要性排序可看出紫边和黄边波段在土地覆盖分类中有较好表现,但没有红边信息贡献度大,因此引入紫边和黄边波段的分类结果略优于相同分类方法下未引入紫边和黄边波段的分类结果,但较于红边信息精度提高效果不显著。

水稻因独特的种植方式,与其他作物有着明显的光谱差异,且GF-6 WFV在传统波段的基础上新增红边(B5、B6)、紫边(B7)和黄边(B8)波段,能够提高水稻信息提取精度,故在种植结构简单的地区(如地势平坦且耕地连片的中国东北地区),本研究方法分类精度较高且适用性较好。而中国南方作物种植结构复杂(如地势起伏较大、地形较为复杂的四川丘陵地区、湖南省等),水稻田破碎化严重,且因气候条件,南方多种植双季稻,故本研究方法及结论需进行下一步验证与分析。此外,研究区内不同土地覆盖类型面积差异较大,本研究中样本点的选取存在不同类别样本数不平衡问题,下一步可对数据集进行预处理,从而提高非平衡数据集的分类精度。

参考文献(References)

- Breiman L. 1996. Bagging predictors. *Machine Learning*, 24(2): 123-140 [DOI: 10.1023/A:1018054314350]
- Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5-32 [DOI: 10.1023/A:1010933404324]
- Chen A X and Li Y C. 2020. Rice recognition of different growth stages based on Sentinel-2 images in mountainous areas of Southwest China. *Transactions of the Chinese Society of Agricultural Engineering*, 36(7): 192-199 (陈安旭, 李月臣. 2020. 基于 Sentinel-2 影像对西南山区的不同生长期水稻识别. *农业工程学报*, 36(7): 192-199) [DOI: 10.11975/j.issn.1002-6819.2020.07.022]
- Chen J, Chen J, Liao A P, Cao X, Chen L J, Chen X H, He C Y, Han G, Peng S, Lu M, Zhang W W, Tong X H and Mills J. 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103: 7-27 [DOI: 10.1016/j.isprsjprs.2014.09.002]
- Clevers J G P W and Gitelson A A. 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation*, 23: 344-351 [DOI: 10.1016/j.jag.2012.10.008]
- Dash J and Curran P J. 2004. The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, 25(23): 5403-5413 [DOI: 10.1080/0143116042000274015]
- Elvidge C D and Chen Z K. 1995. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, 54(1): 38-48 [DOI: 10.1016/0034-4257(95)00132-K]
- Fang F P and Cheng S H. 2018. The development of rice science, technology and industry in China. *Journal of Agriculture*, 8(1): 100-106 (方福平, 程式华. 2018. 水稻科技与产业发展. *农学学报*, 8(1): 100-106) [DOI: 10.11923/j.issn.2095-4050.Cjas2018-1-100]
- Feng R, Yu W Y, Ji R P, Zhang Y S, Wu J W and Chen P S. 2017. Comparison and analysis of several extraction methods for lakes and reservoirs based on FY3B/MERSI. *Science of Surveying and Mapping*, 42(7): 147-152 (冯锐, 于文颖, 纪瑞鹏, 张玉书, 武晋雯, 陈鹏狮. 2017. FY3B/MERSI 数据的湖泊湿地面积提取对比分析. *测绘科学*, 42(7): 147-152) [DOI: 10.16251/j.cnki.1009-2307.2017.07.024]
- He Y, Huang C, Li H, Liu Q S, Liu G H, Zhou Z C and Zhang C C. 2019. Land-cover classification of random forest based on Sentinel-2A image feature optimization. *Resources Science*, 41(5): 992-1001 (何云, 黄翀, 李贺, 刘庆生, 刘高焕, 周振超, 张晨晨. 2019. 基于 Sentinel-2A 影像特征优选的随机森林土地覆盖分类. *资源科学*, 41(5): 992-1001) [DOI: 10.18402/resci.2019.05.15]
- Horler D N H, Dockray M and Barber J. 1983. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2): 273-288 [DOI: 10.1080/01431168308948546]
- Hou M J, Yin J P, Ge J, Li Y C, Feng Q S and Liang T G. 2020. Land cover remote sensing classification method of alpine wetland region based on random forest algorithm. *Transactions of the Chinese Society for Agricultural Machinery*, 51(7): 220-227 (侯蒙京, 殷建鹏, 葛静, 李元春, 冯琦胜, 梁天刚. 2020. 基于随机森林的高寒湿地地区土地覆盖遥感分类方法. *农业机械学报*, 51(7): 220-227) [DOI: 10.6041/j.issn.1000-1298.2020.07.025]
- Huang J W, Li Z Y, Chen E X, Zhao L and Mo B P. 2021. Classification of plantation types based on WFV multispectral imagery of the GF-6 satellite. *Journal of Remote Sensing*, 25(2): 539-548 (黄建文, 李增元, 陈尔学, 赵磊, 莫冰萍. 2021. 高分六号宽幅多光谱数据人工林类型分类. *遥感学报*, 25(2): 539-548) [DOI: 10.11834/jrs.20219090]
- Huang J X, Hou Y Z, Su W, Liu J M and Zhu D H. 2017. Mapping corn and soybean cropped area with GF-1 WFV data. *Transactions of the Chinese Society of Agricultural Engineering*, 33(7): 164-170 (黄健熙, 侯裔焯, 苏伟, 刘峻明, 朱德海. 2017. 基于 GF-1 WFV 数据的玉米与大豆种植面积提取方法. *农业工程学报*, 33(7): 164-170) [DOI: 10.11975/j.issn.1002-6819.2017.07.021]
- Immitzer M, Vuolo F and Atzberger C. 2016. First experience with

- Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 8(3): 166 [DOI: 10.3390/rs8030166]
- Jia K, Li Q Z, Tian Y C and Wu B F. 2011. A review of classification methods of remote sensing imagery. *Spectroscopy and Spectral Analysis*, 31(10): 2618-2623 (贾坤, 李强子, 田亦陈, 吴炳方. 2011. 遥感影像分类方法研究进展. *光谱学与光谱分析*, 31(10): 2618-2623) [DOI: 10.3964/j.issn.1000-0593(2011)10-2618-06]
- Kim H O and Yeom J M. 2014. Effect of red-edge and texture features for object-based paddy rice crop classification using RapidEye multi-spectral satellite image data. *International Journal of Remote Sensing*, 35(19): 7046-7068 [DOI: 10.1080/01431161.2014.965285]
- Li Q J, Liu J, Mi X F, Yang J and Yu T. 2021. Object-oriented crop classification for GF-6 WFV remote sensing images based on Convolutional Neural Network. *National Remote Sensing Bulletin*, 25(2): 549-558 (李前景, 刘珺, 米晓飞, 杨健, 余涛. 2021. 面向对象与卷积神经网络模型的GF-6 WFV影像作物分类. *遥感学报*, 25(2): 549-558) [DOI: 10.11834/jrs.20219347]
- Li W J, Guo X L, Yang L B, Yan M, Zou C X, Fang Y H, Sun H and Huang J F. 2020. Accurate recognition of wine grapes using multi-feature optimization based on GF-6 satellite images. *Transactions of the Chinese Society of Agricultural Engineering*, 36(18): 165-173 (李文杰, 郭晓雷, 杨玲波, 闫鸣, 邹晨曦, 方亚华, 孙涵, 黄敬峰. 2020. 基于GF-6卫星影像多特征优选的酿酒葡萄精准识别. *农业工程学报*, 36(18): 165-173) [DOI: 10.11975/j.issn.1002-6819.2020.18.020]
- Li Z P, Liu Z H, Li Z G, Tang P Q, Tan J Y and Yang P. 2014. Progress and prospect of application of remote sensing to rice spatial distribution. *Chinese Journal of Agricultural Resources and Regional Planning*, 35(6): 9-18 (李志鹏, 刘珍环, 李正国, 唐鹏钦, 谭杰扬, 杨鹏. 2014. 水稻空间分布遥感提取研究进展与展望?. *中国农业资源与区划*, 35(6): 9-18) [DOI: 10.7621/cjarrp.1005-9121.2014602]
- Liang J, Zheng Z W, Xia S T, Zhang X T and Tang Y Y. 2020. Crop recognition and evaluation using red edge features of GF-6 satellite. *Journal of Remote Sensing (Chinese)*, 24(10): 1168-1179 (梁继, 郑镇炜, 夏诗婷, 张晓彤, 唐媛媛. 2020. 高分六号红边特征的农作物识别与评估. *遥感学报*, 24(10): 1168-1179) [DOI: 10.11834/jrs.20209289]
- Lu C L, Bai Z G, Li Y C, Wu B, Di G D and Dou Y F. 2021. Technical characteristic and new mode applications of GF-6 satellite. *Spacecraft Engineering*, 30(1): 7-14 (陆春玲, 白照广, 李永昌, 武斌, 邸国栋, 窦毅芳. 2021. 高分六号卫星技术特点与新模式应用. *航天器工程*, 30(1): 7-14) [DOI: 10.3969/j.issn.1673-8748.2021.01.002]
- Lu D, Fu X J and Yue M J. 2020. Development status and policy suggestions of rice industry in Liaoning Province. *Liaoning Agricultural Sciences*, (6): 57-59 (陆娣, 付雪娇, 岳铭鉴. 2020. 辽宁省水稻产业发展现状及政策建议. *辽宁农业科学*, (6): 57-59) [DOI: 10.3969/j.issn.1002-1728.2020.06.014]
- Mehdaoui R and Anane M. 2020. Exploitation of the red-edge bands of Sentinel 2 to improve the estimation of durum wheat yield in Grombalia region (Northeastern Tunisia). *International Journal of Remote Sensing*, 41(23): 8986-9008 [DOI: 10.1080/01431161.2020.1797217]
- Su W, Zhao X F, Sun Z P, Zhang M Z, Zou Z C, Wang W and Shi Y L. 2019. Estimating the corn canopy chlorophyll content using the Sentinel-2A image. *Spectroscopy and Spectral Analysis*, 39(5): 1535-1542 (苏伟, 赵晓凤, 孙中平, 张明政, 邹再超, 王伟, 史园莉. 2019. 基于Sentinel-2A影像的玉米冠层叶绿素含量估算. *光谱学与光谱分析*, 39(5): 1535-1542) [DOI: 10.3964/j.issn.1000-0593(2019)05-1535-08]
- Sun M X, Liu M, Sun Q Q, Zhang P, Jiao X, Sun D F and Shi Y Y. 2020. Response of new bands in GF-6 to land use/cover based on linear spectral mixture analysis model. *Transactions of the Chinese Society of Agricultural Engineering*, 36(3): 244-253 (孙敏轩, 刘明, 孙强强, 张平, 焦心, 孙丹峰, 史云扬. 2020. 利用光谱混合分解模型分析GF-6新增波段对土地利用/覆被的响应. *农业工程学报*, 36(3): 244-253) [DOI: 10.11975/j.issn.1002-6819.2020.03.030]
- Wang L J, Kong Y R, Yang X D, Xu Y, Liang L and Wang S G. 2020. Classification of land use in farming areas based on feature optimization random forest algorithm. *Transactions of the Chinese Society of Agricultural Engineering*, 36(4): 244-250 (王李娟, 孔钰如, 杨小冬, 徐艺, 梁亮, 王树果. 2020. 基于特征优选随机森林算法的农耕地土地利用分类. *农业工程学报*, 36(4): 244-250) [DOI: 10.11975/j.issn.1002-6819.2020.04.029]
- Yan L and Jiang W W. 2016. Progress in the study of vegetation cover classification of multispectral remote sensing imagery. *Remote Sensing for Land and Resources*, 28(2): 8-13 (闫利, 江维薇. 2016. 多光谱遥感影像植被覆盖分类研究进展. *国土资源遥感*, 28(2): 8-13) [DOI: 10.6046/gtzyyg.2016.02.02]
- Yang J Y, Zhou Z X, Du Z R, Xu Q Q, Yin H and Liu R. 2019. Rural construction land extraction from high spatial resolution remote sensing image based on SegNet semantic segmentation model. *Transactions of the Chinese Society of Agricultural Engineering*, 35(5): 251-258 (杨建宇, 周振旭, 杜贞容, 许全全, 尹航, 刘瑞. 2019. 基于SegNet语义模型的高分辨率遥感影像农村建设用地提取. *农业工程学报*, 35(5): 251-258) [DOI: 10.11975/j.issn.1002-6819.2019.05.031]
- Zhang P and Hu S G. 2019. Fine crop classification by remote sensing in complex planting areas based on field parcel. *Transactions of the Chinese Society of Agricultural Engineering*, 35(20): 125-134 (张鹏, 胡守庚. 2019. 地块尺度的复杂种植区作物遥感精细分类. *农业工程学报*, 35(20): 125-134) [DOI: 10.11975/j.issn.1002-6819.2019.20.016]
- Zhang Q Y, Li Z, Xia C Z, Chen J and Peng D L. 2019. Tree species classification based on the new bands of GF-6 remote sensing satellite. *Journal of Geo-Information Science*, 21(10): 1619-1628 (张沁雨, 李哲, 夏朝宗, 陈健, 彭道黎. 2019. 高分六号遥感卫星新增波段下的树种分类精度分析. *地球信息科学学报*, 21(10): 1619-1628) [DOI: 10.12082/dqxxkx.2019.190116]

Remote sensing extraction of paddy rice in Northeast China from GF-6 images by combining feature optimization and random forest

ZHANG Yueqi, REN Hongrui

College of Mining Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Abstract: Searching for an efficient, high-precision method for mapping paddy rice planting distribution in Northeast China has important implications for accurate paddy rice yield estimation and agricultural policy making.

In this paper, paddy rice planting distribution was mapped by feature optimization random forest method in Panjin City, Liaoning Province. Based on the land coverage types, 2000 samples of 1000 paddy rice samples, 250 water samples, 300 wetland samples, 150 dry land samples, and 300 construction land samples were acquired. Training samples and testing samples accounted for 70% and 30%, respectively. In addition, 36 paddy rice field validation points were obtained through field surveys. The spectrum features, vegetation indexes, water index, and red edge indexes were constructed by using the GF-6 WFV images taken in the periods of May 11, May 25, June 1, June 6, July 20, and August 22 in 2020, and these images corresponded to the trefoil stage, transplanting stage, returning green stage, booting stage, and heading stage according to the phenological phase of paddy rice in Panjin City, respectively. The returning greening stage image was covered by June 1 and June 6. The feature importances of single temporal images and time series images were calculated, and out-of-bag (OOB) estimations on different feature combination models were performed based on OOB data. The optimal input features were selected after comprehensively considering the accuracy and complexity. Then, the feature optimization random forest model was established to extract the paddy rice planting area and spatial distribution information in Panjin City in 2020.

According to the testing samples and the paddy rice field validation points, the accuracy evaluation of classification results showed the following: (1) Based on the single temporal images with different phenological phases, all the classification accuracies were 94% and above. The classification result of the image in the paddy rice transplanting stage was the best that the overall accuracy, F1 score (paddy rice), Kappa coefficient, and field validation point accuracy were 97.67%, 98.84%, 0.97, and 97.22%, respectively. (2) On the basis of comparison with the classification results of single temporal images, using time-series images for land coverage classification and paddy rice information extraction effectively improved the classification accuracy and reduced misclassification and omission, and the paddy rice classification map polygons were more regular. The overall accuracy, F1 score (paddy rice), Kappa coefficient, and field validation points accuracy with time series images were 99.33%, 100%, 0.99, and 97.22%, respectively. (3) Through analyzing of the paddy rice extraction results with or without red edge bands and red edge indexes, the classification accuracy was improved by the introduction of red edge information. This paper proved that based on the feature optimization random forest model, the paddy rice information was accurately extracted by using the single temporal image of paddy rice transplanting stage. Compared with single temporal image, using time-series images improved the classification accuracy. Considering the complexity and running speed of the model, the single temporal image of paddy rice transplanting stage was used to extract paddy rice planting area to meet the accuracy requirement in practical applications. (4) Through analyzing the results of paddy rice extraction without purple band and the yellow band, this paper proved the introduction of purple and yellow bands can improve the classification accuracy, but the effect of improving the accuracy of the classification result was inferior to the red edge information.

Improving the classification accuracy of paddy rice and enhancing crop recognition capabilities by red edge information, purple band, and yellow band, showed the GF-6 satellite had broad application prospects in crop precise identification and area extraction.

Key words: remote sensing, random forest, red edge band, feature optimization, GF-6, paddy rice, purple band, yellow band

Supported by Research Project Supported by Shanxi Scholarship Council of China (2022-055)