引文格式：戴晨,吴昕悦,王秀琴,等. 利用监督学习的水稻生育期识别技术研究[J]. 地理空间信息, 2025, 23(4):87-90. doi:10. 3969/j.issn. 1672-4623.2025.04.019

利用监督学习的水稻生育期识别技术研究

戴　晨[1]，吴昕悦[1]，王秀琴[2]，曹　晨[2]，乔　娜[1]，张芯瑜[1]
（1. 镇江市气象局，江苏 镇江 212003；2. 镇江市丹徒区气象局，江苏 镇江 212100）

摘　要：利用遥感技术识别水稻生育期，能够有效提高农田管理精细化水平并科学指导农事活动。研究综合利用Sentinel-2卫星影像资料、水稻生育期观测资料以及土地利用数据，构建水稻生育期样本光谱、指数与纹理特征，利用随机森林模型进行特征优选，基于K近邻、支持向量机、决策树分类的监督学习算法对水稻移栽、分蘖、拔节孕穗、抽穗扬花、灌浆成熟以及收获收割共6个关键生育期构建识别模型，结果表明：使用随机森林算法对特征进行重要性评价，其中能够反映植被长势的指数特征最具有特征代表性；支持向量机模型在水稻生育期识别中展现出较明显优势，总体分类精度达到84.56%，Kappa系数为0.813。

关键词：水稻；遥感；监督学习；生育期识别；Sentinel-2
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）04-0087-04

Research on Rice Growth Stage Identification Technology Based on Supervised Learning

DAI Chen[1], WU Xinyue[1], WANG Xiuqin[2], CAO Chen[2], QIAO Na[1], ZHANG Xinyu[1]
(1. Zhenjiang Meteorological Bureau, Zhenjiang 212003, China;
2. Zhenjiang Dantu District Meteorological Bureau, Zhenjiang 212100, China)

Abstract: Utilizing remote sensing technology to identify rice growth stage can effectively enhance the precision of agricultural field management and provide scientific guidance for farming activities. We integrated Sentinel-2 satellite images, rice growth stage observation data, and land-use information to construct rice growth stage samples with spectral, index, and texture features, employed random forest model for feature selection, and utilized supervised learning algorithms including K-nearest neighbors, support vector machines and decision trees to construct identification models for six key rice growth stages, such as transplanting, tillering, booting, heading, filling, and maturing. Results indicate that the random forest algorithm is employed to evaluate the importance of features, where index features reflecting vegetation growth status are found to be most representative. The support vector machine model demonstrates notable advantages in rice growth stage identification, achieving an overall accuracy of 84.56% and Kappa coefficient of 0.813.

Key words: rice, remote sensing, supervised learning, growth stage identification, Sentinel-2

水稻的生育周期主要包括秧苗、移栽、返青、分蘖、拔节、孕穗、抽穗、扬花和成熟期[1]。在水稻的多个生育期之中，往往需要对水稻进行多样化的田间管理措施，例如，在分蘖期，通过主动的晒田，可以有效抑制水稻的无效分蘖[2]；适宜的穗肥施肥能够提高水稻扬花期间叶片叶绿素含量以及含氮量，进而使水稻产量得到提升[3]；抽穗扬花期是水稻育种工作的关键时期，准确得到水稻的抽穗期信息有助于提升育种工作质量[4]。

传统的水稻生育期监测主要是依靠人力实地观测或者基于农业气象站点的观测[5]，存在难以通过单次观测获取大范围水稻生育信息的劣势。因遥感技术具备宏观、迅速的大范围监测能力，成为水稻种植面积监测不可或缺的技术手段[6-7]，且作物不同生育阶段往往体现出不同的光谱特征，这也让基于遥感技术的水稻生育期识别成为可能。在水稻生育期遥感识别领域，学者早期主要利用中低分辨率遥感影像，通过阈值法或者生育期光谱曲线形态法开展相关研究，例如，郑小波等[8]使用MODIS的合成NDVI序列与地面农业气象观测站数据，分析水稻生育期、长势与NDVI值的关系，结果发现归一化植被指数NDVI可以作为水稻各生育期判别的重要指标；孙华生等[9]通过分析水稻各生育期增强植被指数变化特征，通过阈值法对水稻生育期进行识别，结果表明生育期误差大多小于16 d；Sakamoto等[10]利用多时相MODIS数据计算增强植被指数曲线并检测出最大、最小以及拐点，基于Coiflet小波变换方法对水稻生长期进行判别，结果表明，生育期识别结果与统计数据相差仅10 d左右。随着高分辨

项目来源：江苏省气象局青年基金资助项目（KQ202328）；镇江市重点研发计划资助项目（SH2022019）。
第一作者简介：戴晨（1994—），硕士研究生，工程师，研究方向为生态农业遥感，E-mail：257@qq.com。

率卫星遥感、无人机遥感以及机器学习技术的日趋成熟，当前越来越多学者致力于高分辨率影像的水稻生育期识别研究，例如，杨振忠等[11]通过利用 K 近邻（KNN）算法结合可见光大气修正指数（visible atmospherically resistant index，VARI）构建水稻生育期识别模型，生育期识别准确率可达86%以上。徐建鹏等[12]则利用RAdam神经网络结合生育期观测样本构建模型，模型对生育期识别准确率达到97%以上。目前，利用卫星遥感技术结合机器学习方法针对江苏南部区域开展大面积水稻生育期识别技术研究的工作尚少，本研究旨在通过水稻生育期样本收集处理、特征筛选以及模型对比，形成适用于江苏南部水稻生育期识别的有效技术方案。

作为三大主粮作物之一的水稻，基于监督学习方法对其展开生育期识别监测，可以为农业管理部门和农民提供重要的科学信息，也能为农业保险理赔服务提供科学依据，有针对地指导农民的田间管理活动，对地方农业可持续发展具有重要意义。

1 研究区概况与数据源

1.1 研究区概况

镇江市地处江苏省西南部，以丘陵地形为主，属北亚热带季风气候。年平均气温17.1℃，年平均降水量1 222.3 mm，年累计日照时数为1 912.7 h，夏季热量丰富，雨量充沛，气候条件较优越，孕育出适宜水稻种植的环境。本研究的实验区镇江润果农业农场（133.29°E，37.59°N）正位于镇江东北部，坐落于镇江市京口区姚桥镇现代农业产业园区，占地2万6千余亩，每年5—10月种植作物主要以水稻为主，为本研究提供了大量的水稻生育期实验样本数据。

1.2 数据来源介绍

本研究所使用的遥感数据主要为2022年Sentinel-2A MSI多光谱影像，包括空间分辨率为10 m的蓝、绿、红和近红外波段，空间分辨率为20 m的4个红边波段和2个短波红外波段，并将其中20 m分辨率的光谱波段通过重采样的方式升尺度至10 m。水稻生育期样本资料主要通过外业调查和水稻生育期观测资料分析的方式获取，时间为2022年5-11月，收集到的样本信息包括水稻的种植方式（水直播、机插秧、旱直播）、栽播时间（6月上、中、下旬）、种植品种（南粳9 108、南粳46），并通过GPS定位在遥感图上进行标注生成样本矢量图。其他辅助数据包括2021版全球土地覆盖产品[13]，主要用于辅助样本选取。

2 研究方法

2.1 遥感影像筛选预处理及样本集构建方式

研究通过利用Google Earth Engine（GEE）平台中存储的经过大气校正和正射校正的Sentinel-2多光谱影像表面反射率产品，结合镇江水稻生育期观测资料以及润果水稻种植矢量数据对Sentinel-2多光谱影像进行筛选，共计得到2022年6—10月过境润果农场的影像152景，并使用Sentinel-2 Cloud Probability数据集对研究区域内Sentinel-2影像进行去云处理，去除云概率大于50%的像素，而后进一步对所得到的影像进行通过逐景分析，最终得到5景无云高质量影像并对应水稻生育期观测资料，对不同种植时间分别进行生育期提取，得到水稻移栽、分蘖、拔节孕穗、抽穗扬花、乳熟成熟以及收获期全生育期样本。因抗杂稻品种生育周期较短，属于早熟品种，10月下旬已完成收获，因此将其作为收获期样本，具体各生育期样本及说明如表1所示。

表1 水稻生育期样本数据说明

生育期	影像时间	水稻地块	移栽时间	样本数目
移栽	2022-07-02	南粳9108	6月下旬	5 316
分蘖	2022-07-12	南粳46	6月上旬	7 136
拔节孕穗	2022-08-09	南粳9108	6月中旬	5 514
抽穗扬花	2022-10-03	南粳46	6月中旬	17 973
乳熟成熟	2022-10-25	南粳9108	6月上旬	9 905
收获	2022-10-25	抗杂稻	6月中旬	5 827

2.2 分类特征选取与计算

为有效完成水稻各生育期监督学习分类，首先需要对输入特征进行预选，本研究通过研读相关文献，初步获得了29个输入特征，其中光谱特征10个，指数特征12个，纹理特征7个。

2.3 特征选择方案

在利用遥感信息进行地物分类以及作物生育期识别过程，大量的特征数量存在一定冗余特征或者噪音数据，如果将所有特征都代入模型进行训练，会严重影响模型的精度和效率，造成“维度灾难”，这就需要对特征进行优选，剔除冗余特征，提高模型的泛化能力。随机森林因其具有分类准确度高、受过拟合以及缺失值影响小、能够在预测过程中确定特征重要性等特点，被越来越多的应用于机器学习特征选择领域，本研究首先将6个不同生育期样本分别按照NDVI

数值进行排序，各去除排名前10%以及后10%的像素点疑似噪点信息，再分别对特征进行归一化处理，消除量纲影响并采取随机森林算法进行特征重要性排序来进行特征选取。

2.4　模型建立方案

研究通过利用水稻的光谱特征、指数特征以及纹理特征，分别利用 *K* 近邻（K nearest neighbors，KNN）[14]、决策树（decision trees，DT）[15]、支持向量机（support vector machines，SVM）[16]3种机器学习方法进行训练，比较不同模型在实现水稻生育期识别中的性能表现，得到最优模型。其中KNN算法使用欧式距离进行距离计算，DT使用基尼系数来获取最优分割特征，SVM的核函数为径向基。

2.5　精度评价方法

混淆矩阵[17]（confusion matrix）也称误差矩阵，是监督分类识别领域进行精度评价的一种常用方法，用 *n* 行 *n* 列的矩阵形式来表示。具体评价指标有总体精度（overall accuracy，OA）、制图精度（producer′s accuracy，PA）、用户精度（user′s accuracy，UA）及Kappa系数（kappa coefficient，KC），利用以上精度指标可以有效评价图像分类的精度。

3　结果与分析

3.1　基于随机森林重要性评价的生育期识别特征优选

随机森林算法可以对特征变量的重要程度和贡献进行分析评价，在提高模型精度的同时减少数据的冗余和处理工作量。为确保用于特征选取的随机森林算法参数尽可能接近最优参数，使用参数迭代方案对随机森林模型最优参数进行选取，最终得到最优参数分别为n_estimators=20，max_depth=7，min_samples_split=4。这样建立随机森林算法模型对29个输入特征进行重要性评分计算，得到排序结果如图1所示。从图中可以看出，对不同生育期类别最为敏感的3个特征分别为比植被指数（RVI）、增强植被指数（EVI）以及归一化植被指数（NDVI），特征重要性评分均在0.1以上，整个水稻生育期中，植被光谱的变化是各阶段最主要的差异所在，而从图中也可以看出，纹理特征对于生育期识别分类的精度提升并没有太大的参考价值，这主要是因为哨兵2号卫星的分辨率尚不足以区分水稻生长过程中所产生的纹理变化，这说明对于10 m分辨率的卫星数据而言，在识别水稻的特征选择中，能够反映植被长势的指数特征对水稻生育期识别具有十分重要的意义。本研究以重要性评分0.03为界，选用重要性排名前11的特征作为最优特征组合作为生育期识别模型算法的输入特征，其中指数特征有9个，光谱特征与纹理特征各有一个，分别为近红外波段B8以及NDVI总和平均值。

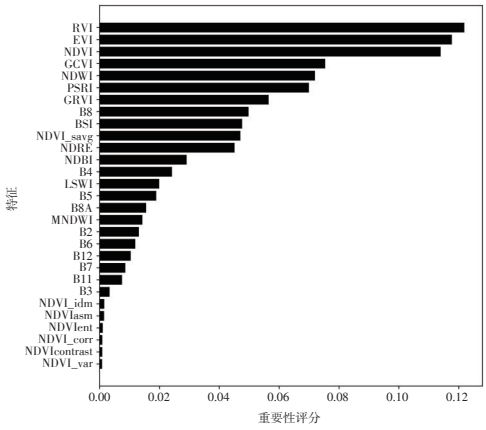


图1　基于随机森林模型的特征重要性评分

3.2　不同生育期模型识别效果对比

本研究水稻生育期模型构建是基于Python语言Scikit-learn库包中 *K* 近邻、决策树以及支持向量机开展的。每种算法均使用格点搜索方法对模型参数设置进行最优化选取，其中 *K* 近邻算法需要精选的参数为查询近邻样本的数量，动态调整范围为3~20；决策树算法需要调整其最大深度（max_depth）和能够分离出内部节点的最小样本数（min_samples_split），调整范围分别为1~10和2~20；支持向量机采用径向基核函数，惩戒参数（*C*）按照0.01、0.1、1、10、100进行动态调整，通过格点搜索的反复实验，最终获得各模型最优参数设置如表2所示。

表2　各分类算法的最优参数设置

算法	参数设置
KNN	$K=7$
DT	Max_depth=8；min_samples_split=20
SVM	Kernel=RBF；$C=100$

实验过程中首先对样本数据按照7∶3的比例进行随机分层，70%的样本集作为训练样本，剩余30%作为测试集用于模型的精度评价，对训练样本进行训练过程中对各模型均采用5折分层交叉验证的方式进行性能评估择选，即将训练样本按照4∶1的比例进行分层抽样，其中4份作为训练集，剩余1份作为验证集，以此过程反复操作10次，得出模型的性能评价参数如表3所示。可以看到，3种分类模型的总体精度均在83%以上，Kappa系数也均在0.79以上，其中支持向量机模型的Kappa系数达到0.813。

表3 不同模型的精度评价表

识别模型	总体精度/%	Kappa系数
KNN	83.04	0.797
DT	84.37	0.811
SVM	84.56	0.813

利用基于支持向量机算法构建的模型对6–11月过境研究区的共6幅Sentinel–2遥感影像开展水稻生育期识别，结果表明6月10日RGB影像图上显示田块有水体的镜面反射特征和裸土特征，模型输出结果能够较好的识别出移栽后的水田信息，同时将裸土地识别为收获期地块，同月17日部分较早种植的地块进入分蘖期，模型识别出的地块总体边界较为清晰。在7月12日、8月6日、10月3日以及11月7日，大部分田块陆续进入拔节孕穗、抽穗扬花、乳熟成熟以及收获收割阶段，模型输出结果对发育期中晚期的乳熟成熟以及收获收割明显较拔节孕穗和抽穗扬花的识别精度更高，这是由于中晚期水稻的光谱特征变化显著，而拔节孕穗和抽穗扬花2个阶段的周期相近，总体水稻的细微光谱特征差异无法在10 m分辨率遥感影像上得到凸显，因此可以看见7月12日以及8月6日的模型输出结果显示地块较破碎且边界模糊，存在一定错分漏判现象。总体来看，水稻生育期模型的输出结果能够较好体现出随时间维度的生长发育周期变化趋势。

利用基于支持向量机的水稻生育期识别模型对测试集的6 028个不同生育期样本进行模型测试，具体性能表现如表4所示，由于拔节孕穗、抽穗扬花以及乳熟成熟3个生育期的差异变化不显著，因此模型在识别三者时的分离能力较弱，出现了彼此之间的误判，其中拔节孕穗的识别精度较低，用户精度处于80%以下，除此以外，模型对于各个生育期的制图精度均在80%以上，尤其是移栽期、分蘖期和乳熟成熟期，用户精度均达到85%以上，总体来看，基于支持向量机算法构建识别模型在水稻生育期分类上有良好的的性能表现，总体分类精度达到84.56%，Kappa系数为0.813。

表4 SVM模型水稻识别性能参数混淆矩阵

水稻生育期		模型输出							制图精度/%
		移栽	分蘖	拔节孕穗	抽穗扬花	乳熟成熟	收获	合计	
真实样本	移栽	713	16	15	20	25	57	846	84.28
	分蘖	11	803	44	27	33	13	931	86.25
	拔节孕穗	14	23	708	33	41	12	831	85.20
	抽穗扬花	15	32	52	913	43	23	1 078	84.69
	乳熟成熟	26	43	83	71	1235	33	1 491	82.83
	收获	56	11	15	23	21	725	851	85.19
	合计	835	928	917	1 087	1 398	863	6 028	
用户精度/%		85.39	86.53	77.21	83.99	88.34	84.01		

4 结论与讨论

本研究以江苏省镇江市姚桥镇润果农业农场为研究区，基于多时相Sentinel–2多光谱影像和水稻生育期观测资料，提取水稻生育期光谱、指数和纹理特征并生成大量样本，利用随机森林算法进行特征筛选，从而获取最优特征组合，基于DT、KNN和SVM算法构建水稻生育期识别模型，并对模型进行性能参数对比，结论如下：

1）使用随机森林算法对分类特征进行重要性评价，结果显示三类特征重要性排序为指数特征>光谱特征>纹理特征，纹理特征对于水稻生育期分类影响最小，最为敏感的3个特征分别为比植被指数（RVI）、增强植被指数（EVI）以及归一化植被指数（NDVI），特征重要性评分均在0.1以上。

2）基于SVM算法构建的模型在3种模型中的性能表现最优，显示出较好的水稻生育期提取精度，总体分类精度达到84.56%，Kappa系数为0.813。模型对于各个生育期的制图精度均在80%以上，尤其是移栽期、分蘖期和乳熟成熟期，用户精度均达到85%以上。

尽管SVM算法相较于其他算法在本研究中体现出更强的泛化能力，但是对于大范围的水稻生育期识别，仍存在模型验证样本量小、水稻种植结构矢量数据缺失等问题，同时江南汛期雨云覆盖率较高，如何有效克服云雨天气带来的遥感光学影像资料缺失，以及水稻发育中期生育期间隔短、光谱特征差异小的问题仍然是水稻生育期识别亟需解决的难题。因此，多源卫星融合方案对比以及小样本深度学习模型算法将是后续研究的重点。

参考文献

[1] 杨振忠. 基于遥感数据的水稻生育期识别研究[D] 武汉:武汉大学,2019

（下转第113页）