

王小飞,张方敏,任祖光,等. 基于机器学习算法的河南省冬小麦面积提取研究[J]. 江苏农业科学,2024,52(6):215-224.
doi:10.15889/j.issn.1002-1302.2024.06.028

基于机器学习算法的河南省冬小麦面积提取研究

王小飞¹, 张方敏¹, 任祖光², 张世豪³, 高歌^{1,4}

(1. 南京信息工程大学气象灾害预报预警与评估协同创新中心/江苏省农业气象重点实验室, 江苏南京 210044;

2. 中联智慧农业股份有限公司, 安徽芜湖 241080; 3. 桂林电子科技大学广西图像图形与智能处理重点实验室, 广西桂林 541004;

4. 国家气候中心中国气象局气候研究开放实验室, 北京 100081)

摘要:为了精准获取河南省冬小麦空间分布及面积数据,基于 2003—2021 年 250 m MODIS-NDVI 时间序列遥感数据集,通过设置不同的阈值条件获得高质量的样本数据,采用深度神经网络(DNN)、随机森林(RF)和支持向量机(SVM)算法,自动从 NDVI 时序数据中提取冬小麦特征,分别训练出非线性模型,在 250 m 尺度对河南省冬小麦分布和面积进行识别。结果表明,基于 DNN 算法的河南省冬小麦面积识别模型精确率为 97.26%,总体一致性为 97.97%;基于 RF、SVM 算法的精确率分别为 91.51% 和 89.31%,总体一致性均在 90% 以下。和 RF、SVM 算法相比,DNN 算法在精度上有明显的提升,能够更好地反映河南省冬小麦的时间变化趋势和空间面积分布。该研究说明,运用中等分辨率长时间序列影像结合 DNN 算法,在一定程度上可以更准确识别大区域的农作物信息。

关键词:冬小麦;深度神经网络;NDVI;遥感;时间序列

中图分类号:S127 **文献标志码:**A **文章编号:**1002-1302(2024)06-0215-10

冬小麦是河南省的主要农作物之一,其产量占全国小麦总产量的 28% (2021 年统计年鉴)。因此,获取冬小麦准确的空间分布和种植面积信息,有助于充分发挥其潜在的生产力,促进我国农业的发展和粮食安全,这一工作具有非常重要的意义^[1]。遥感技术在作物类型识别和分类中发挥着关键作用,能够有效提取出不同作物的长势特征和

种植面积,是推进农业现代化的关键一步^[2]。由于不同农作物在光谱特征上的相似度较高,仅凭借单一时间点的遥感影像进行分类容易出现“异物同谱”和“同物异谱”的问题,从而影响地物的识别分类精度^[3]。针对这个问题,学者们提出使用长时间序列的植被指数反映不同农作物的物候特征^[4],进而提高农作物分类的识别精度^[5]。

近年来,学者们利用 MODIS^[6-7]、Landsat^[8]、哨兵^[9]、GF^[10]等系列卫星数据开展了冬小麦的遥感识别研究。使用的方法主要包括机器学习和非机器学习 2 种^[11-12]。传统的非机器学习方法在特定条件下表现良好,但其难以泛化到不同的时间、地点或环境条件下,因此受到了一定的限制。比如,这些方法通常依赖于手动定义的规则和阈值来进

收稿日期:2023-04-19

基金项目:江苏省碳达峰碳中和科技创新专项资金(编号: BK20220017)。

作者简介:王小飞(1995—),男,河南周口人,硕士研究生,主要从事农业遥感研究。E-mail:1573979951@qq.com。

通信作者:张方敏,博士,教授,主要从事农业与生态气象研究, E-mail:fmin.zhang@nuist.edu.cn;高歌,博士,研究员,主要从事气候变化与评估研究, E-mail:gaoge@cma.gov.cn。

[31] 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.

[32] Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices[J]. Remote Sensing of Environment, 1996, 55(2):95-107.

[33] 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.

[34] 郭 锐. 植被指数及其研究进展[J]. 干旱气象, 2003(4):71-

75.

[35] 徐建华. 计量地理学[M]. 北京:高等教育出版社,2006.

[36] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[37] 周敏姑,邵国敏,张立元,等. 无人机多光谱遥感反演冬小麦 SPAD 值[J]. 农业工程学报,2020,36(20):125-133.

[38] 田军仓,杨振峰,冯克鹏,等. 基于无人机多光谱影像的番茄冠层 SPAD 预测研究[J]. 农业机械学报,2020,51(8):178-188.

[39] 牛庆林,冯海宽,周新国,等. 冬小麦 SPAD 值无人机可见光和多光谱植被指数结合估算[J]. 农业机械学报,2021,52(8):183-194.

行特征提取和分类,对于复杂的植被覆盖类型和变化不够灵活^[13]。Busetto 等^[14]、Skakun 等^[15]、Chu 等^[16]、Tao 等^[17]、Qiu 等^[18]和 Dong 等^[19]基于 MODIS、Landsat 时序数据分别采用 PhenoRice 算法、高斯混合法、双高斯函数方法、冬小麦冬前高峰特征法、物候特征法和时间加权的动态时间归整法进行了冬小麦特征提取。采用机器学习方法进行冬小麦遥感识别具有明显优势,机器学习方法能够自动从遥感数据中学习并提取多维特征,无需事先定义特定规则或阈值^[20]。利用大量的训练样本数据,机器学习模型能够建立复杂的特征-类别关联,从而提高冬小麦识别的准确性^[21]。

随着机器学习算法的快速发展,深度学习算法通过多层结构和大规模数据训练,能够实现高度的特征学习和模式识别,在各种应用中能取得卓越的性能^[22],明显强于传统机器学习算法^[23-26],因此,在基于遥感的地物分类应用也得到了广泛关注。张乾坤等利用 Landsat-8 数据,通过构建相对最优的卷积神经网络结构实现了新疆地物高精度分类^[27];Sun 等使用 Landsat-8 数据构建时间序列,利用深度神经网络(deep neural network, DNN)算法完成美国北达科他州地区不同作物的分类,总体精度超过了 82%^[28];姚锦一等采用 Sentinel-2 数据,结合 DNN 算法对色楞格河流域的水体信息进行了提取,总体精度达到了 97.65%^[29];Kussul 等结合 Landsat-8 和 Sentinel-1A 遥感数据,建立时间序列数据用于提取作物的物候特征,采用卷积神经网络算法实现了乌克兰基辅地区不同农作物的分类,精度可达 94.6%^[30]。

MODIS 提供了较长时间序列的时序数据,解决了中高分辨率影像数据量大、重访周期长、大范围内同一时间影像缺失等问题。因此,本研究基于 2003—2021 年 MODIS 的 250 m 归一化植被指数(normalized difference vegetation index, NDVI)时序数据,利用具有更强的学习能力和更好的泛化性能的 DNN 算法提取河南省冬小麦种植面积,并基于随机森林(random decision forests, RF)、支持向量机(support vector machine, SVM)的机器学习算法的结果进行对比,分析 3 种机器学习算法下的河南省冬小麦种植面积的提取精度。

1 研究区域与数据

1.1 研究区域概况

河南省是我国内陆腹地的一部分,地理位置介

于 110.21~116.39°E、31.23~36.22°N 之间。河南省属于温带季风气候区,四季分明,夏季雨水集中,气候炎热,冬季干冷。年平均气温为 10.5~16.7℃,年降水量 407.7~1 295.8 mm,受东南季风影响,呈现由东南向西北逐渐减少的趋势。河南省地势高低不平,西部以山地和丘陵为主,东部以平原为主。平原地区土层深厚,土质肥沃,棕壤和褐土是主要土壤类型,适合冬小麦的生长。河南省是全国冬小麦主要产区之一。

1.2 研究数据

1.2.1 遥感数据 NDVI 数据来源于 2003—2021 年 MODIS MYD13Q1 产品(<http://earthexplorer.usgs.gov/>),16 d 融合产品,1 年 23 期,空间分辨率为 250 m。使用 MODIS 数据处理工具(MODIS Reprojection Tool)将 MYD13Q1 数据从 HDF 格式转换为 Tiff 格式,进行重投影和裁剪等操作,得到河南省冬小麦每年生长季内(10 月 8 日到次年 6 月 17 日)17 期的 NDVI 数据(因闰年和平年的原因,所述日期和实际选取的日期会有 1 d 的误差)^[31]。

1.2.2 样本数据 为获取河南省主要地物类型分布,结合 Google Earth 和中国气象科学数据共享服务网(<http://data.cma.cn/>)下载的参考样本点数据,在冬小麦生长季将河南省地物类型分为 6 大类别,即冬小麦、单峰类作物、其他双峰类作物、林地、建筑用地和水体(图 1)。在研究区每年共获得 9 970 个左右的样本点(30 580 个左右的像元点),包括冬小麦 6 000 个左右(20 000 个左右的像元点)、单峰类作物 550 个左右(600 个左右的像元点)、其他双峰类作物 900 个左右(1 000 个左右的像元点)、林地 1 500 个左右(5 700 个左右的像元点)、建筑用地 1 000 个左右(3 100 个左右的像元点)、水体 20 个左右(180 个左右的像元点)。随机选取 75% 的样本点作为训练集,剩下的 25% 样本点用于测试集,用来建立冬小麦识别的训练模型。

1.2.3 验证数据 为了更好地验证本研究基于 DNN 算法的分类精度,使用 CLCD 30 m 的土地覆盖数据集(<https://zenodo.org/records/4417810>)^[32]、国家青藏高原科学数据中心提供的 2001—2018 年 250 m 华北平原农作物种植区分布数据集(<https://data.tpdc.ac.cn/>)^[33]和河南省统计年鉴 2003—2021 年的冬小麦种植面积(<https://tjj.henan.gov.cn/>)进行精度验证。

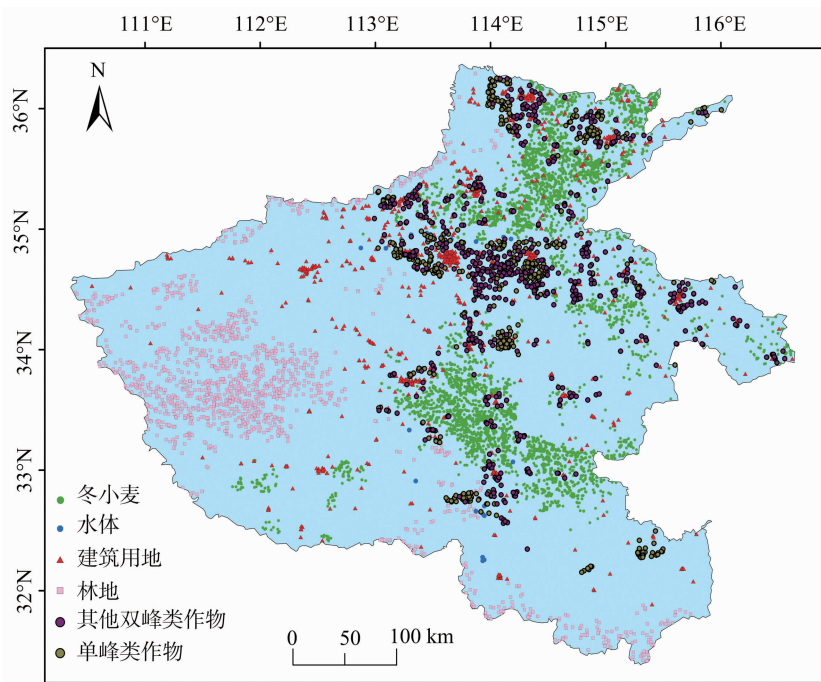


图1 河南省样本点分布示意图(2018 年)

2 研究方法

本研究技术路线如图 2 所示。首先,从样本点中获取 NDVI 时序数据,然后利用设定的阈值过滤掉低质量样本。然后,运用机器学习模型(包括 DNN、RF 和 SVM)结合样本的 NDVI 时序数据,对冬小麦进行识别模型的训练。最后将 MODIS - NDVI 时序数据输入经过训练的冬小麦识别模型,对冬小麦进行遥感识别,并且对 3 种机器学习算法的模型精度和冬小麦识别精度分别进行详细分析。

2.1 阈值设置

阈值主要是剔除冬小麦样本中出现的低质量像元和一些异像元。冬小麦的生长状况受雨水、日照等天气条件影响比较大,比如在 2003 年、2006 年、2008 年、2013 年一些地区的冬小麦 NDVI 值偏低,因此在冬小麦的样本中通过设置阈值,过滤掉低质量的样本。经过多次试验测试,认为位于冬小麦孕穗期和抽穗期 4 月 14 日、4 月 30 日、5 月 16 日的 NDVI 值能反映出小麦的长势,在 NDVI 值偏低的 4 年,任意时间节点的 NDVI 值大于 0.78 即判断为质量高的冬小麦样本,在其他 NDVI 年均值较高的年份,3 个时间节点的 NDVI 值大于 0.8 判断为质量高的冬小麦样本。

在其他单峰类作物、双峰类作物、林地、建筑用地和水体的样本中,提取时间序列的 NDVI 值同样

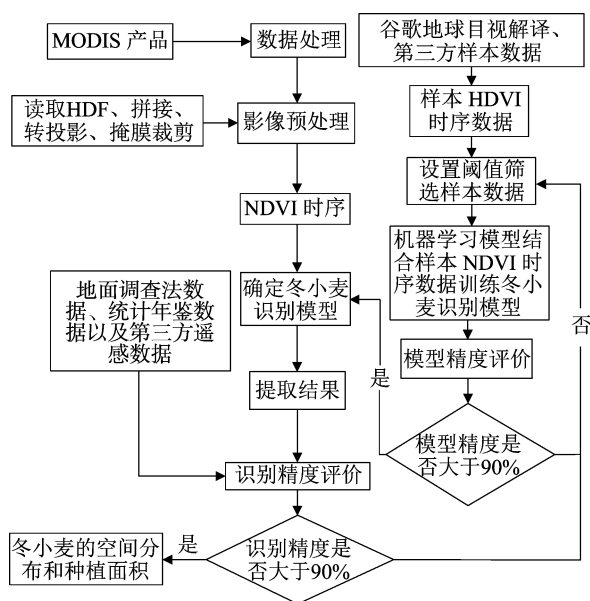


图2 河南省冬小麦面积提取技术路线

需要剔除一些异常值。经过多次试验,认为单峰类作物和双峰类作物在每个时间节点的 NDVI 值不为空值即可,建筑用地的每个时间节点的 NDVI 值不应超过 0.4。

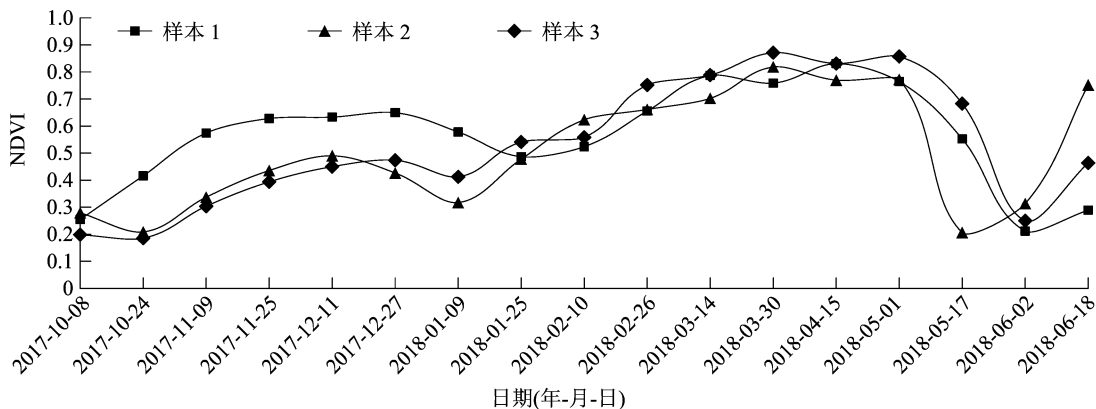
2.2 特征选择

由于河南省南北部气候差异的原因,小麦种植和收获的时间会存在一定差异,因此,需要确定识别冬小麦的起始和结束的时间,进而建立 NDVI 时间序列的特征曲线。冬小麦种植时间在秋季的

10—11 月之间,该季节夏玉米、大豆等作物秋收完成,此时的 NDVI 值介于 0.2~0.3 之间,河南省冬小麦种植时间差异不大,选择 10 月 8 日作为起始时间。

河南省冬小麦收获时间大部分集中在 5 月 20 日之后和 6 月 20 日之前,由图 3 可知,在 2018 年 6 月 2 日之后 NDVI 的曲线走势出现明显不同,这是由于 2 地的冬小麦成熟时间不一致导致的。样本 2 所在区域为南阳地区,经过调研,此地区的冬小麦

成熟时间大部分集中在 5 月下旬,收割结束之后,会种植夏粮作物,NDVI 值在 6 月 17 日会达到 0.4 以上,而此时河南省东部和北部的小麦正处于收割阶段,因此最后一个特征值选择 6 月 2 日的 NDVI 值。综上,选择从 10 月 8 日到次年的 6 月 2 日共 16 期的 NDVI 值作为特征值,可以有效区分冬小麦、单峰类植物、其他双峰类植物、林地、建筑用地以及水体(图 4)。



样本 1 位于 114.31°E、35.61°N, 样本 2 位于 113.43°E、35.57°N, 样本 3 位于 113.86°E、33.64°N

图3 河南省冬小麦 NDVI 时间序列曲线(2018 年)

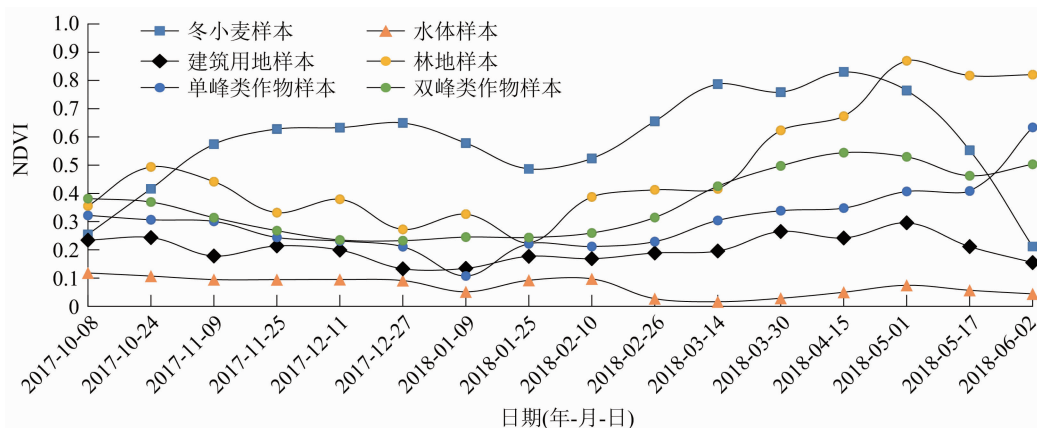


图4 河南省不同地物 NDVI 时间序列曲线(2018 年)

2.3 构建基于 DNN 的冬小麦训练模型

使用 Python 编程语言对数据进行预处理,在基于 TensorFlow 的深度学习框架下进行 DNN 模型的搭建、训练和测试。

2.3.1 DNN 网络层数和神经元个数设置 网络结构采用输入层、全连接层和输出层,其中输入层是将输入的数据压成一维的数据,再用全连接层(设置 128 个神经元)接收一维数据,进行向量处理,最后由输出层输出结果,因为输出结果有 6 个(冬小麦、单峰类作物、双峰类作物、林地、建筑用地、水体),输出层的神经元个数设置为 6。

2.3.2 DNN 激活函数设置 在全连接层中使用 relu 激活函数,可以规避 sigmoid 函数饱和区对网络收敛的限制,加快模型训练速度。在神经网络的输出层中,采用 softmax 作为激活函数对全连接层的输出进行变换,使得输出符合概率分布。这样的处理可以将每个输出值限制在 $[0, 1]$ 之间,并且所有输出的总和为 1,从而更好地表示分类概率。

$$\text{relu}(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

式中: x 为来自上一层神经网络的输入向量。

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad j=1, \dots, k. \quad (2)$$

式中: σ 是 softmax 激活函数; z_j 表示神经网络模型中第 j 个节点的输出值; K 表示模型输出的节点个数,即分类的类别个数。

2.3.3 DNN 训练方法设置 在 compile 中配置训练方法时,选择 adam 算法。该算法能够基于训练数据迭代更新神经网络的权重,通过计算梯度的一阶估计和二阶估计为不同的参数设计独立的自适应性学习率。本研究使用 adam 默认参数,其中一阶矩阵指数衰减率为 0.9,二阶矩阵指数衰减率为 0.99,稳定性常数为 10^{-8} 。

2.3.4 DNN 训练参数设置 迭代次数 (epochs) 设置为 100,这表示将对数据进行 100 次迭代训练。批量大小 (batch size) 设置 64,每次迭代从训练数据中随机选择 64 个样本进行训练。validation_split 设置为 0.25,每次会自动随机选择 75% 的数据为训练集,剩下 25% 的数据为测试集。验证频率 (validation frequency) 设置为 1,每次迭代后都进行 1 次验证,以监测模型性能。

最后,使用训练好的模型对 MODIS - NDVI 遥感影

像进行分类,提取出冬小麦的空间分布和种植面积。

2.4 构建基于 RF 和 SVM 的冬小麦训练模型

为验证 DNN 算法的优越性,将其与机器学习框架下的 RF 和 SVM 算法进行对比。基于 RF 和 SVM 算法的冬小麦训练模型使用和 DNN 算法一样的数据集。其中,RF 训练模型的参数设置:子树的数量为 100,每棵树的最大深度为 9,随机种子为 42;SVM 训练模型的参数设置:使用线性基作为 SVM 的核函数,惩罚因子 C 为 1,松弛变量 γ 为 10^{-1} ,随机种子为 42。RF 和 SVM 将数据集 75% 的样本作为训练集,剩下 25% 的样本为测试集。

3 结果与分析

3.1 识别结果

根据 2003—2021 年的样本数据,基于 DNN 训练出冬小麦识别模型,利用这个模型遍历读取了冬小麦生长季内的 16 期 MODIS - NDVI 遥感影像,依次得到河南省 2003—2021 年冬小麦的空间分布图(图 5),并统计得到冬小麦面积时间序列图(图 6)。由图 5 和图 6 可知,2003—2021 年河南省的冬小麦空间分布格局变化不明显,河南省冬小麦种植面积整体呈上升趋势。

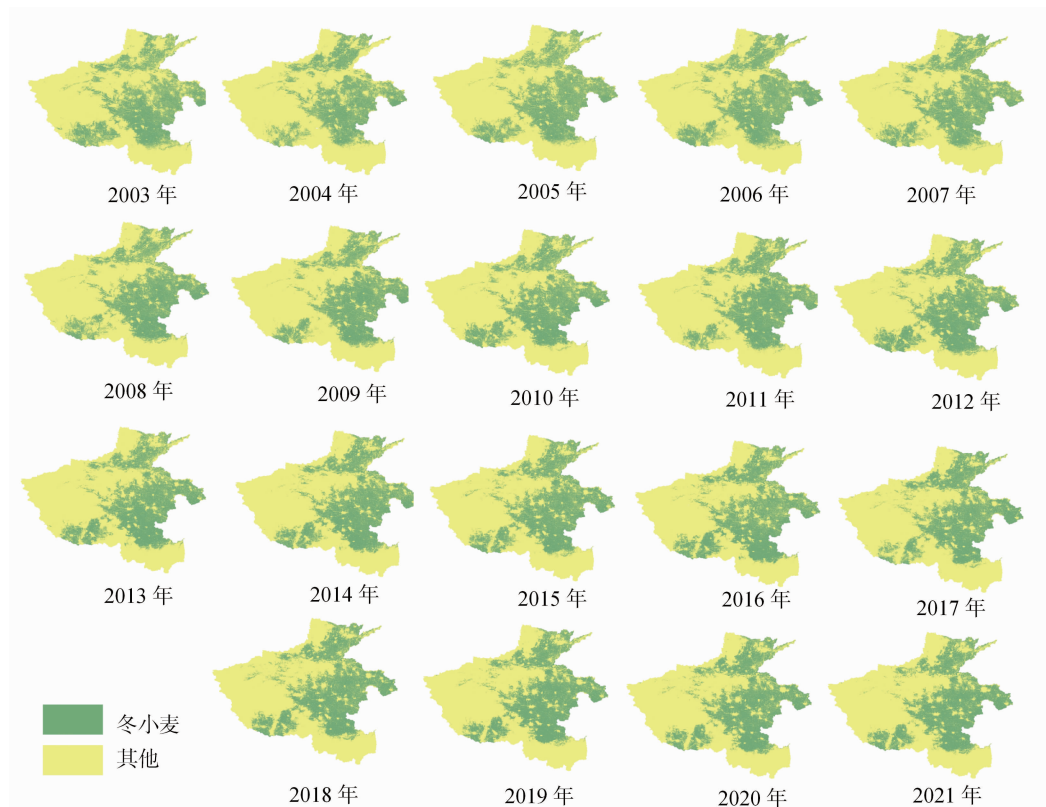


图5 基于 DNN 算法的河南省冬小麦空间分布图(2003—2021 年)

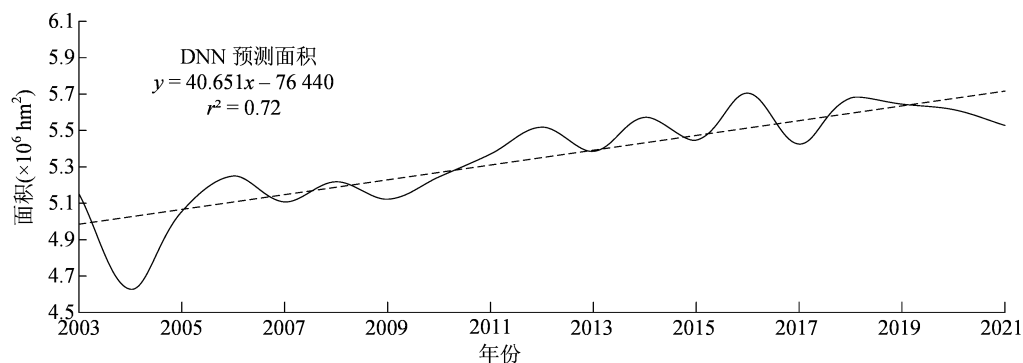


图6 基于 DNN 算法的河南省冬小麦总面积变化时间序列图(2003—2021 年)

3.2 分类精度分析

3.2.1 模型精度分析 在本研究中,通过精确率和损失率来评估模型的优劣,精确率衡量了模型预测的阳性样本中真正阳性样本的比例,而损失率则代表了模型预测错误的样本所占的比例^[34]。如表 1 所示,基于 DNN 算法的多年精确率为 97.26% (最高为 2003 年的 98.43%,最低为 2012 年的 95.14%),比 RF、SVM 算法精确率分别提升 5.75、

7.95 百分点。基于样本点的多年损失率为 7.07%,其中最高为 2012 年的 12.27%,最低为 2003 年的 4.12%,比 RF、SVM 算法损失率分别降低 19.79、22.46 百分点。该结果与张荣群等^[3]、王连喜等^[23]、张锦水等^[35]和徐青云等^[36]对比,精确率分别提升 8.31、2.10、1.77、11.54 百分点,表明该模型的精度较好,可以对河南省冬小麦的空间分布和面积进行提取。

表 1 2003—2021 年不同分类方法精确度和损失度对比

年份	DNN 精确率 (%)	DNN 损失率 (%)	RF 精确率 (%)	RF 损失率 (%)	SVM 精确率 (%)	SVM 损失率 (%)
2003	98.43	4.12	94.61	19.34	92.62	20.10
2004	98.15	5.54	94.29	20.42	92.32	22.36
2005	97.98	5.67	92.74	21.15	85.14	31.96
2006	95.80	12.00	92.27	21.23	91.40	25.81
2007	95.74	10.37	90.73	32.59	88.32	32.96
2008	96.99	8.54	91.93	27.97	87.72	33.36
2009	95.95	9.83	90.76	27.61	88.79	31.91
2010	96.69	8.01	91.94	25.58	89.81	29.14
2011	96.10	9.14	91.44	29.44	88.88	33.59
2012	95.14	12.27	89.09	30.41	84.74	40.62
2013	97.52	6.13	91.49	27.61	90.63	27.47
2014	98.26	4.38	91.94	25.89	90.97	26.79
2015	98.18	5.21	90.48	29.02	89.29	30.53
2016	98.07	4.95	94.09	29.02	89.63	30.82
2017	98.27	4.63	89.18	34.15	89.82	29.29
2018	97.56	6.19	92.38	22.01	91.00	26.57
2019	97.89	5.16	90.17	25.15	89.31	28.42
2020	97.67	5.88	89.22	31.60	87.51	29.95
2021	97.51	6.28	89.97	30.13	88.97	29.41
平均	97.26	7.07	91.51	26.86	89.31	29.53

3.2.2 识别精度分析 使用 ArcMap 软件对 2003—2021 年 CLCD 30 m 的土地覆盖数据集进行 250 m 重采样,然后分别与使用 DNN、RF 和 SVM 方

法的 250 m 冬小麦分类图进行空间对比。以 2005、2015 年为例(图 7),2005 年基于 DNN 的方法冬小麦分布较稀疏,2015 年 3 种方法的结果较一致。遍

历每个年份的 250 m 分辨率的冬小麦分类图和 CLCD 数据集,比较 2 个数据集中相同像元位置上的类别,并记录冬小麦和 CLCD 数据集中相同类别的像元数量,统计每个年份的冬小麦总像元数量,并

计算相同类别像元数量与冬小麦总像元数量的比值,得到 2003—2021 年 DNN 预测的冬小麦空间分布与 CLCD 数据集耕地的重叠率达到 98% 以上,RF 和 SVM 预测的冬小麦空间分布重叠率为 90% 左右。

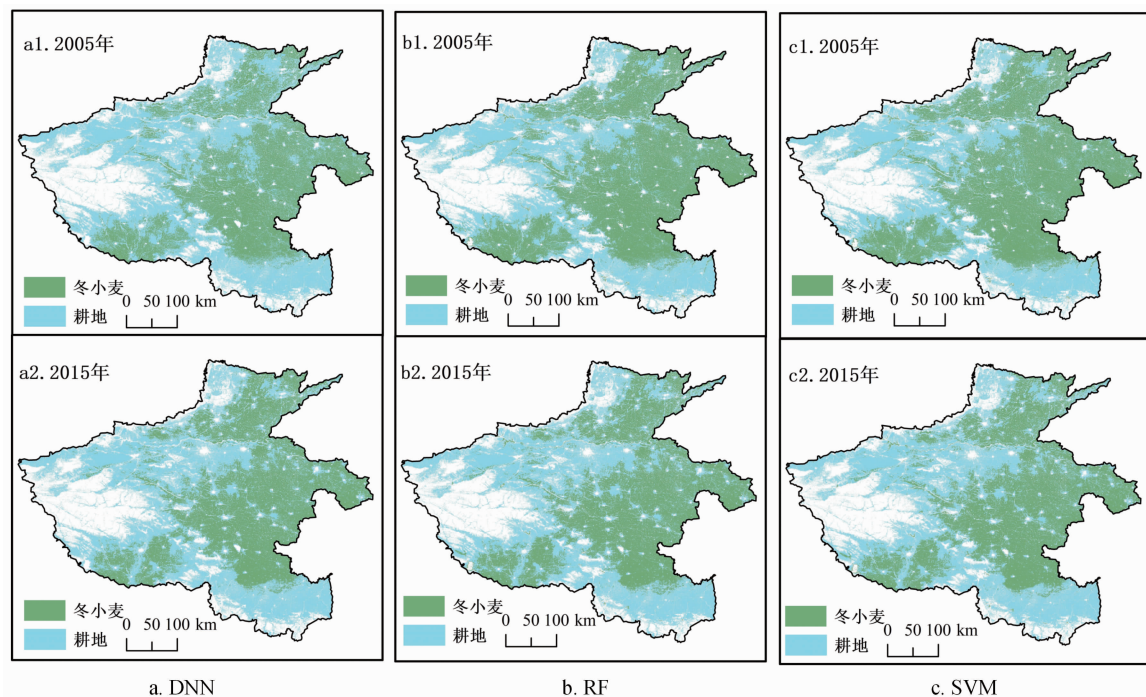


图7 2005 年和 2015 年基于 3 种算法的冬小麦空间分布

以国家青藏高原科学数据中心提供的华北平原农作物种植区分布图作为实际数据,使用 DNN、RF 和 SVM 方法得出的结果作为预测数据,比较实际冬小麦面积数据和 DNN、RF 和 SVM 方法预测的冬小麦数据。结果表明,2003—2018 年河南省冬小麦面积与 DNN 预测的面积一致性达到 95% 以上,与 RF 和 SVM 预测的面积一致性在 90% 以下。以 2018 年商丘和 2017 年周口的结果为例(图 8),一致性分别为 95.29%、98.62%,与 RF 方法的一致性分别为 92.02%、89.43%,与 SVM 的一致性分别为 92.43%、87.56%。对比结果进一步说明了使用基于 DNN 的冬小麦训练模型能够较好地获取河南省 2003—2021 年冬小麦空间分布信息。

同样采用上述方法,与河南省统计年鉴中的冬小麦播种面积数据进行对比,结果(表 2)表明,基于 DNN 方法计算 2003—2021 年的分类面积与统计面积的一致性平均为 97.97% (最高为 2016 年的 99.99%,最低为 2003 年的 93.28%),并且和基于统计面积的变化趋势一致,均为递增趋势。而 RF 和 SVM 算法预测面积在 2003—2006 年误差率高达 20%~35%,导致基于 RF 和 SVM 算法的 2003—

2021 年河南省冬小麦预测面积呈递减趋势,与统计年鉴面积趋势变化呈相反结果。结果说明了基于 DNN 的冬小麦训练模型优于 RF 和 SVM 算法,亦能很好地反映河南省冬小麦的多年面积变化情况。

4 讨论

尽管冬小麦生长周期内的 NDVI 时间序列曲线基本一致,但是由于每年气候变化和作物管理方式不同,每年的 NDVI 时间序列曲线仍有变化^[37],如图 9 所示(该冬小麦样本点位于 116.19°E、33.88°N),2019 年冬小麦的 NDVI 明显低于其他年份,2012 年 10 月到 12 月的 NDVI 值远高于其他年份,2008 年 1 月 9 日的 NDVI 值达到历年最低。因此,本研究根据当年的样本集训练出当年的冬小麦模型,会减小多年间作物生长差异对模型精度的影响。

本研究所使用的是 250 m MODIS 遥感数据,存在混合像元的情况。为了提高分类结果的准确性,本研究将混合像元中实际冬小麦占比大于 50% 的像元分类为冬小麦,实际冬小麦占比小于 50% 的像元分类为非冬小麦,由于冬小麦的时序曲线变化和其他双峰类作物的曲线类似,需要收集更多的其他

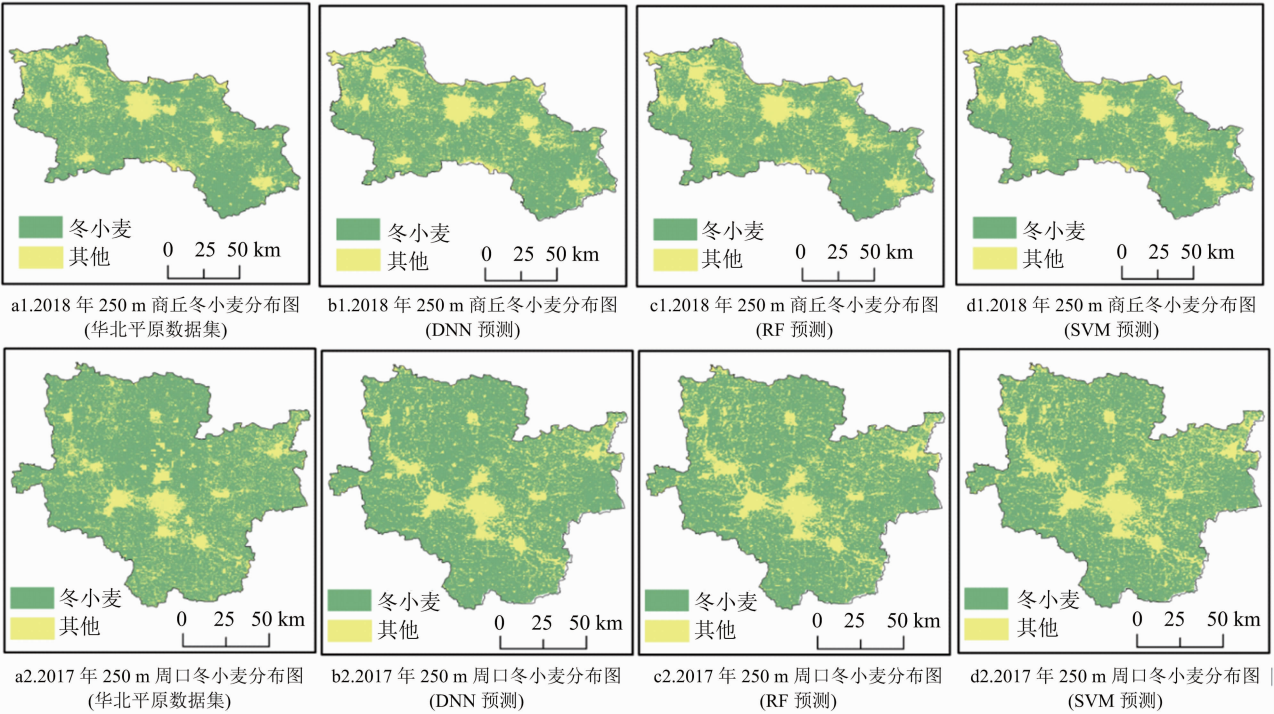


图8 3 种算法冬小麦空间精度对比图(以商丘、周口为例)

表 2 2003—2021 年河南省冬小麦预测面积和省统计年鉴面积

年份	冬小麦面积(×10 ⁶ hm ²)				DNN 一致性 (%)	RF 一致性 (%)	SVM 一致性 (%)
	DNN 预测	RF 预测	SVM 预测	统计年鉴			
2003	5.150 77	7.005 91	7.388 63	4.804 60	93.28	68.58	65.03
2004	4.627 43	7.067 03	7.218 01	4.856 00	95.29	68.71	67.28
2005	5.054 79	6.154 79	6.211 80	4.962 70	98.18	80.63	79.89
2006	5.249 69	6.617 98	6.730 36	5.208 47	99.21	78.70	77.39
2007	5.107 54	5.011 94	5.019 54	5.213 33	97.97	96.14	96.28
2008	5.218 51	6.134 51	6.351 24	5.260 00	99.21	85.74	82.82
2009	5.122 39	5.597 20	5.496 55	5.263 30	97.32	94.03	95.76
2010	5.245 00	5.612 65	5.507 59	5.280 00	99.34	94.07	95.87
2011	5.370 46	5.908 03	5.860 67	5.323 33	99.12	90.10	90.83
2012	5.518 49	5.199 41	5.252 11	5.468 80	99.10	95.07	96.04
2013	5.385 14	6.313 53	6.258 80	5.517 98	97.59	87.40	88.16
2014	5.572 21	6.049 76	6.040 68	5.581 24	99.84	92.26	92.39
2015	5.446 97	5.377 98	5.365 83	5.623 14	96.87	95.64	95.42
2016	5.705 44	6.559 66	5.337 61	5.704 91	99.99	86.97	93.56
2017	5.425 58	5.036 43	4.947 29	5.714 64	94.94	88.13	86.57
2018	5.678 14	5.455 93	5.478 37	5.739 85	98.92	95.05	95.44
2019	5.644 63	5.298 86	5.337 48	5.706 65	98.91	92.85	93.53
2020	5.614 81	5.876 86	5.429 18	5.673 67	98.96	96.54	95.69
2021	5.527 66	5.513 94	5.453 79	5.673 79	97.42	97.18	96.12
平均	—	—	—	—	97.97	88.62	88.64

双峰类作物的样本,充分发挥 DNN 对于大样本大数据处理的优势,减少混合像元带来的误差。因此,

未来的研究可以采用多种高分辨率遥感数据和多种机器学习算法相结合的方法提高识别精度。

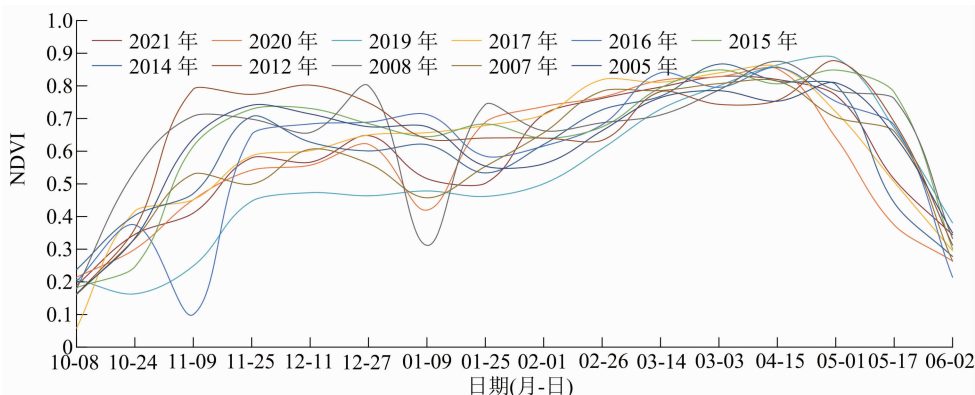


图9 不同年份的冬小麦 NDVI 曲线变化

DNN 是由多个感知器组成的集成学习模型,具有很强的非线性拟合能力,通过本研究 3 种机器学习算法的精度对比可以发现,DNN 的分类能力远远优于其他 2 种机器学习算法,但是也存在一定的不确定性。在分类的结果中,有小部分冬小麦像元会识别成其他类型,其原因可能是该地区的天气相关,如冬季的雨雪天气、春季的干旱以及管理农作物方式的不同都会导致 NDVI 的异常,也可能是由于模型的不足,会对精度的评价产生一定的影响。因此需要对 DNN 模型的分类过程进行更深入的研究,以揭示其“黑盒”性质,提高对其分类结果的信任度和可靠性^[38]。

5 结论

本研究以 2003—2021 年的 MODIS - NDVI 为数据源,根据河南省的地势和气候完成对地表混合像元和异常值的处理,利用处理后的 NDVI 时间序列作为特征值,构建基于 DNN、RF 和 SVM 的冬小麦识别机器学习算法模型,与已知的数据对比对识别的冬小麦进行精度验证。结果表明,在空间上,基于 DNN 算法的 2003—2021 年冬小麦分布与 CLCD 数据集在耕地上的重叠率达到 98% 以上,与青藏高原科学数据中心提供的华北平原农作物种植区的一致性达 95% 以上,与河南省统计年鉴中的冬小麦面积一致性平均为 97.97%,并且变化趋势保持一致。基于 RF 和 SVM 算法的冬小麦分布的识别结果与相同数据集的对比显示,在重叠率和一致性上这 2 种算法均低于 DNN 的冬小麦识别结果,并且不能很好地反映冬小麦的时间变化趋势。结果说明了基于 DNN 算法的模型能更有效地提取反映作物生长信息的时间和空间特征,提高了大范围冬小麦的空间分布和面积数据的精确度。

参考文献:

- [1] 孙九林. 中国农作物遥感动态监测与估产总论[M]. 北京:中国科学技术出版社,1996:1-8.
- [2] 何忠,谢国雪,林垚君,等. 遥感技术在我国甘蔗产业的应用综述[J]. 国土资源信息化,2020(4):22-27.
- [3] 张荣群,王盛安,高万林,等. 基于时序植被指数的县域作物遥感分类方法研究[J]. 农业机械学报,2015,46(增刊1):246-252.
- [4] 饶新宇,李红军,张圣微,等. 草原植被长势遥感监测方法适宜性研究[J]. 中国生态农业学报(中英文),2021,29(12):2084-2092.
- [5] Liang S L. Quantitative remote sensing of land surfaces[M]. New Jersey: John Wiley & Sons,2003:494-501.
- [6] Bognár P, Kern A, Pásztor S, et al. Yield estimation and forecasting for winter wheat in Hungary using time series of MODIS data[J]. International Journal of Remote Sensing, 2017, 38(11): 3394-3414.
- [7] Pan Y Z, Li L, Zhang J S, et al. Winter wheat area estimation from MODIS - EVI time series data using the Crop Proportion Phenology Index[J]. Remote Sensing of Environment, 2012, 119:232-242.
- [8] Skakun S, Vermote E, Franch B, et al. Winter wheat yield assessment from Landsat 8 and Sentinel - 2 data: incorporating surface reflectance, through phenological fitting, into regression yield models[J]. Remote Sensing, 2019, 11(15):1768.
- [9] Xu F, Li Z F, Zhang S Y, et al. Mapping winter wheat with combinations of temporally aggregated Sentinel - 2 and Landsat - 8 data in Shandong Province, China[J]. Remote Sensing, 2020, 12(12):2065.
- [10] Huawei Mou A, Huan Li B, Yuguang Zhou C, et al. Estimating winter wheat straw amount and spatial distribution in Qihe County, China, using GF - 1 satellite images[J]. Journal of Renewable and Sustainable Energy, 2021, 13(1):013102.
- [11] Zhang X W, Qiu F, Qin F. Identification and mapping of winter wheat by integrating temporal change information and Kullback - Leibler divergence[J]. International Journal of Applied Earth Observation and Geoinformation, 2019, 76:26-39.
- [12] Zhang Q X, Wang G F, Wang G J, et al. Identifying winter wheat using landsat data based on deep learning algorithms in the North

- China Plain[J]. Remote Sensing,2023,15(21):5121.
- [13] Shi W Z, Zhang M, Zhang R, et al. Change detection based on artificial intelligence: state - of - the - art and challenges [J]. Remote Sensing,2020,12(10):1688.
- [14] Busetto L, Zwart S J, Boschetti M. Analysing spatial - temporal changes in rice cultivation practices in the Senegal River Valley using MODIS time - series and the PhenoRice algorithm [J]. International Journal of Applied Earth Observation and Geoinformation,2019,75:15 - 28.
- [15] Skakun S, Franch B, Vermote E, et al. Early season large - area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model[J]. Remote Sensing of Environment,2017,195:244 - 258.
- [16] Chu L, Liu Q S, Huang C, et al. Monitoring of winter wheat distribution and phenological phases based on MODIS time - series; a case study in the Yellow River Delta, China [J]. Journal of Integrative Agriculture,2016,15(10):2403 - 2416.
- [17] Tao J B, Wu W B, Zhou Y, et al. Mapping winter wheat using phenological feature of peak before winter on the North China Plain based on time - series MODIS data [J]. Journal of Integrative Agriculture,2017,16(2):348 - 359.
- [18] Qiu B W, Hu X, Chen C C, et al. Maps of cropping patterns in China during 2015—2021 [J]. Scientific Data,2022,9:479.
- [19] Dong J, Fu Y Y, Wang J J, et al. Early - season mapping of winter wheat in China based on Landsat and Sentinel images [J]. Earth System Science Data,2020,12(4):3081 - 3095.
- [20] Amani M, Ghorbanian A, Ali Ahmadi S, et al. Google earth engine cloud computing platform for remote sensing big data applications; a comprehensive review [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2020,13:5326 - 5350.
- [21] Li Y S, Li X W, Zhang Y J, et al. Cost - efficient information extraction from massive remote sensing data; when weakly supervised deep learning meets remote sensing big data [J]. International Journal of Applied Earth Observation and Geoinformation,2023,120:103345.
- [22] Goodfellow I, Bengio Y, Courville A. Deep learning [M]. Cambridge: The Massachusetts Institute of Technology Press, 2016: 164 - 172.
- [23] 王连喜, 徐胜男, 李 琪, 等. 基于决策树和混合像元分解的江苏省冬小麦种植面积提取 [J]. 农业工程学报, 2016, 32(5): 182 - 187.
- [24] 李 龙, 李旭青, 吴 伶, 等. 基于决策树和神经网络的农作物分类研究——以廊坊市为例 [J]. 红外, 2019, 40(3): 24 - 31, 36.
- [25] 苏腾飞, 刘全明, 苏秀川. 基于多种植被指数时间序列与机器学习的作物遥感分类研究 [J]. 江苏农业科学, 2017, 45(16): 219 - 224.
- [26] 杨 野, 杨德昌, 孙 红, 等. 基于无人机多光谱遥感和机器学习算法的新疆棉花生物量估算 [J]. 江苏农业科学, 2023, 51(15): 179 - 187.
- [27] 张乾坤, 蒙继华, 任 超. 构建地块二维表征及 CNN 模型的作物遥感分类 [J]. 遥感学报, 2022, 26(7): 1437 - 1449.
- [28] Sun Z H, Di L P, Fang H, et al. Deep learning classification for crop types in North Dakota [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2200 - 2213.
- [29] 姚锦一, 王卷乐, 严欣荣, 等. 基于深度神经网络的蒙古国色楞格河流域水体信息提取 [J]. 地球信息科学学报, 2022, 24(5): 1009 - 1017.
- [30] Kussul N, Lavreniuk M, Skakun S, et al. Deep learning classification of land cover and crop types using remote sensing data [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(5): 778 - 782.
- [31] 李炳军, 张一帆. 不同生育期气候变化对河南省冬小麦产量影响的量化分析 [J]. 江苏农业科学, 2022, 50(12): 238 - 246.
- [32] Yang J, Huang X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019 [J]. Earth System Science Data, 2021, 13(8): 3907 - 3925.
- [33] Li J D, Lei H M. Tracking the spatio - temporal change of planting area of winter wheat - summer maize cropping system in the North China Plain during 2001—2018 [J]. Computers and Electronics in Agriculture, 2021, 187: 106222.
- [34] Pleşoiu A I, Stupariu M S, şandric I, et al. Individual tree - crown detection and species classification in very high - resolution remote sensing imagery using a deep learning ensemble model [J]. Remote Sensing, 2020, 12(15): 2426.
- [35] 张锦水, 赵光政, 洪友堂, 等. 基于像元物候曲线匹配的生长季内河北省冬小麦空间分布识别 [J]. 农业工程学报, 2020, 36(23): 193 - 200.
- [36] 许青云, 杨贵军, 龙慧灵, 等. 基于 MODIS NDVI 多年时序数据的农作物种植识别 [J]. 农业工程学报, 2014, 30(11): 134 - 144.
- [37] Wu C Y, Chen J M. Diverse responses of vegetation production to interannual summer drought in North America [J]. International Journal of Applied Earth Observation and Geoinformation, 2013, 21: 1 - 6.
- [38] 田富有, 吴炳方, 曾红伟, 等. 基于多层神经网络与 Sentinel - 2 数据的大豆种植区识别方法 [J]. 地球信息科学学报, 2019, 21(6): 918 - 927.