

RESEARCH

Open Access



Karst vegetation coverage detection using UAV multispectral vegetation indices and machine learning algorithm

Wen Pan^{1,2}, Xiaoyu Wang³, Yan Sun^{1,2}, Jia Wang¹, Yanjie Li^{1*} and Sheng Li^{1*}

Abstract

Background Karst vegetation is of great significance for ecological restoration in karst areas. Vegetation Indices (VIs) are mainly related to plant yield which is helpful to understand the status of ecological restoration in karst areas. Recently, karst vegetation surveys have gradually shifted from field surveys to remote sensing-based methods. Coupled with the machine learning methods, the Unmanned Aerial Vehicle (UAV) multispectral remote sensing data can effectively improve the detection accuracy of vegetation and extract the important spectrum features.

Results In this study, UAV multispectral image data at flight altitudes of 100 m, 200 m, and 400 m were collected to be applied for vegetation detection in a karst area. The resulting ground resolutions of the 100 m, 200 m, and 400 m data are 5.29, 10.58, and 21.16 cm/pixel, respectively. Four machine learning models, including Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Deep Learning (DL), were compared to test the performance of vegetation coverage detection. 5 spectral values (Red, Green, Blue, NIR, Red edge) and 16 VIs were selected to perform variable importance analysis on the best detection models. The results show that the best model for each flight altitude has the highest accuracy in detecting its training data (over 90%), and the GBM model constructed based on all data at all flight altitudes yields the best detection performance covering all data, with an overall accuracy of 95.66%. The variables that were significantly correlated and not correlated with the best model were the Modified Soil Adjusted Vegetation Index (MSAVI) and the Modified Anthocyanin Content Index (MACI), respectively. Finally, the best model was used to invert the complete UAV images at different flight altitudes.

Conclusions In general, the GBM_all model constructed based on UAV imaging with all flight altitudes was feasible to accurately detect karst vegetation coverage. The prediction models constructed based on data from different flight altitudes had a certain similarity in the distribution of vegetation index importance. Combined with the method of visual interpretation, the karst green vegetation predicted by the best model was in good agreement with the ground truth, and other land types including hay, rock, and soil were well predicted. This study provided a methodological reference for the detection of karst vegetation coverage in eastern China.

Keywords UAV, Machine learning, Classification, Karst, Vegetation indices

*Correspondence:

Yanjie Li
aj7105@gmail.com
Sheng Li
Shengli@caf.ac.cn

¹ Research Institute of Subtropical Forestry, Chinese Academy of Forestry,
No. 73, Daqiao Road, Fuyang, Hangzhou 311400, Zhejiang, China

² College of Forestry, Nanjing Forestry University, Nanjing, China

³ Chun'an County Forestry Administration, Hangzhou, Zhejiang, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Karst environments are areas where slightly dissolved rock outcrops and efficient acid hydrolysis create spectacular dissolved landforms [1]. Karst areas have been distributed around the world and cover about 15% of Earth's surface [2]. The carbonate rocks in Southeast Asia, which is the largest karst area in the world, are continuously exposed and its ecological environment is extremely vulnerable to human activities [3]. Karst areas have high complexity and strong spatial and temporal heterogeneity. It is a comprehensive reflection of the staggered distribution of various ground objects such as bedrock, vegetation, and soil cover [4].

As an important part of global vegetation, karst vegetation not only provides a great carbon sink function but also provides a series of ecological services, which has always been a research hotspot in the field of global change [5]. The vegetation in karst areas is critical for maintaining fragile local ecosystems [6]. In addition, vegetation is a significant sensitive factor that reflects changes in the ecological environment of karst areas [7]. The coverage of dry vegetation such as litter and bare surface soil also plays an important role in the characterization and evaluation of the degree of land degradation [8]. Therefore, it is particularly important to figure out the classification and distribution of vegetation populations in karst areas.

In the early years, many domestic and foreign studies on vegetation coverage detection in karst areas introduced many methods. Early research mainly used fieldwork methods for vegetation coverage detection. For instance, Blasi et al. [9] conducted a field investigation and multivariate analysis of the vegetation communities in the karst tectonic basin of the Majella Massif called plant sociology methods. Bátori et al. [10] using the Moving Segmentation Window (MSW) technique, nested analysis, and Principal Coordinate Analysis (PCoA) conducted a field survey in southern and northern Hungary between 2005 and 2012 and revealed vegetation patterns in ocean trough. Although field surveys hold the advantages of high accuracy but time- and cost-consuming, and which are not suitable for the detection of vegetation in large-scale karst areas.

Recently, with the continuous improvement of the spatiotemporal and spectral resolution of remote sensing technology, massive remote sensing image datasets have emerged as the times require [11]. It is worth noting that the extraction of ground object information, especially vegetation information, employing remote sensing images has been largely studied in remote sensing research. In terms of multispectral, the overall accuracy of vegetation coverage detection was improved by 5.57% compared with traditional supervised classification

using the combination of the Back Propagation Neural Network (BPNN) model and Landsat-8 Operational Land Imager (OLI) multispectral remote sensing images [12]. Regarding hyper-spectrum, through linear spectral separation and pixel separation methods, hyperspectral images can be used to extract ecological indicators such as karst vegetation fraction and vegetation abundance, which can characterize vegetation coverage to a certain extent for vegetation inversion [13, 14]. However, most remote sensing satellite images involve visual interpretation and computer-aided digital processing of aerial photography and satellite images, which are highly subjective and inefficient, and limit the ability to distinguish and identify ground objects in karst areas [15, 16].

Karst areas are characterized by thin soil layers and exposed rocks. Since the ground cover in karst areas is often a mixture of several types (vegetated and non-vegetated), it is difficult to accurately extract the main features of vegetation cover [17]. Vegetation, soil, and rocks have high reflectance to different wavelengths of visible light, which makes the difference between karst vegetation and non-vegetation in UAV multispectral images significant. Xiao et al. [18] pointed out the method of using vegetation index to distinguish karst vegetation from non-vegetation. To solve the problems of low image resolution and inaccurate vegetation coverage detection, UAV remote sensing technology is rapidly going mainstream [19]. Among recent innovations, UAVs are suitable for tracking and assessing vegetation conditions in time, with several advantages: (1) They can fly at low flight altitudes, providing high-definition aerial imagery and high spatial resolution, fine details of vegetation can be detected, (2) flights can be flexibly scheduled according to critical moments imposed by vegetation over time, (3) they can be acquired in different ways using different sensors and the range of perception systems of vegetation spectrum (visible, infrared, thermal), (4) this technique can also generate the Digital Surface Model (DSM) with Three-Dimensional (3D) vegetation measurements by using highly overlapping images and applying an image reconstruction procedure with Structure from motion (SfM) techniques [20]. As such, UAVs are a cost-effective tool for acquiring high spatial resolution 3D data of plants and trees where satellite platforms are not feasible, filling the gap between ground-based equipment and other traditional remote sensing systems. Beyond that, UAV-based digital imagery can effectively replace data collected through laborious, subjective, and destructive manual fieldwork [21]. Due to these advantages, the UAVs are becoming quite suitable platforms for vegetation coverage detection in karst areas [22], mainly using RGB [23, 24], multispectral [25, 26], hyperspectral [27], and lidar [28] sensors.

The large amount of detailed data embedded in UAV imagery requires the development and application of powerful advanced analysis programs for extracting information related to vegetation structure and biochemical composition to better understand relevant plant traits [29, 30]. Bolin Fu et al. [31] set the flight altitude of the UAV to 105 m uniformly and used an optimized Random Forest—Decision Tree (RF-DT) model to extract vegetation communities, which explored the optimal detection variables for various types of vegetation. Zhang et al. [32] set the flight altitude to 100 m and used UAV-based hyperspectral images, combined with SVM and Edge-Preserving Filter (EPF), to automatically extract tree canopies damaged by Chinese pine caterpillars and perform more refined classification. Mäyrä et al. [33] collected data from a flight altitude of 1500 m and compared the performance of Three-Dimensional Convolutional Neural Network (3D-CNN), GBM, DL, etc. in the classification of individual tree species in hyperspectral data and used the best performing 3D-CNN as a complete tree species map was generated for the study area. These studies show that machine learning algorithms such as RF, SVM, GBM, and DL algorithms have been widely used in the detection of vegetation coverage detection. When combining different UAV flight altitudes and feature variables, the detection accuracies of various models show significant differences.

At present, the vegetation cover research of karst areas is focused on the continuous karst area in southwest China, which is the center of the karst area in Southeast Asia [34]. There are obvious differences between the karst areas in Southwest and Eastern China. The former is difficult to grow due to the sparse soil cover and

there are many tree species with strong stress resistance such as cypress. In contrast, the latter has deep soil cover, lush vegetation, and high coverage. In addition, there are few studies on vegetation cover in karst areas of eastern China. Therefore, it is necessary to understand the changes in vegetation cover in the karst areas of eastern China to achieve sustainable development of karst ecosystems.

In this study, we focus on karst ecosystems to demonstrate the use of efficient analytical methods for UAV multispectral datasets (such as RF, SVM, GBM, and DL) and aim to answer the following questions:

1. In the multispectral data of different UAV flight altitudes, how accurate are the four models of RF, SVM, GBM, and DL in identifying karst vegetation coverage?
2. In the optimal variable selection, how is the vegetation indices' importance distribution of the best models at each flight altitude?
3. Combined with the method of visual interpretation, how well do the vegetation distribution of the three flight altitudes predicted by the best model fit with the respective real vegetation distribution?

Materials

Study area

The study area is located in Wanshi, Fuyang, Hangzhou, Zhejiang, China (30°6'9"N, 119°31'46"E; Fig. 1), with a total area of 157.9 square kilometers. The region has a subtropical monsoon climate, with an average annual temperature of about 16.3 °C and an average annual

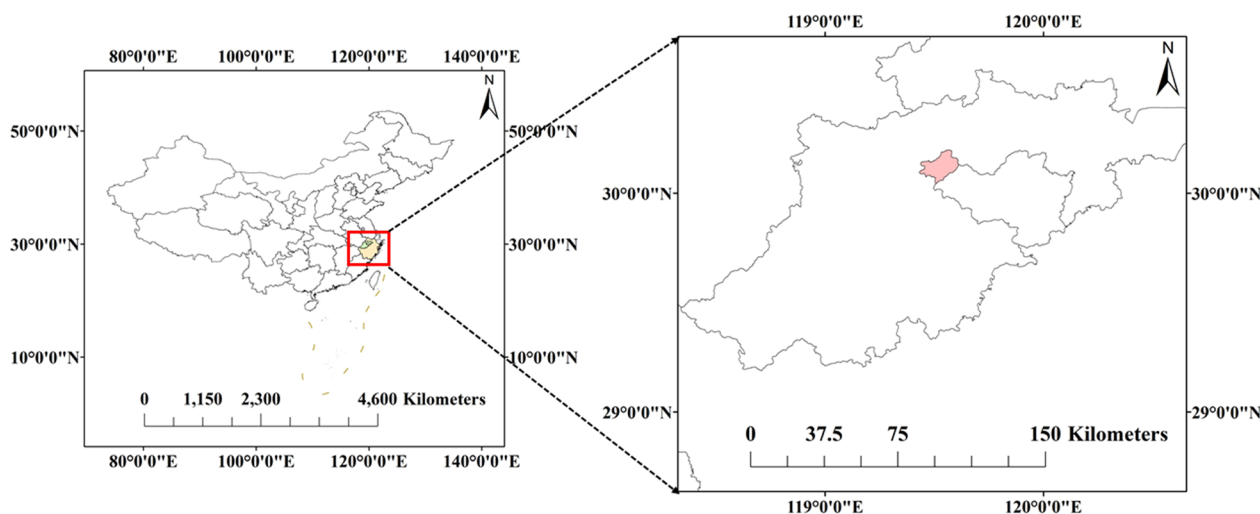


Fig. 1 The sketch map of study area

Table 1 Spectral parameters of the multispectral camera of the DJI multispectral UAV

Band name	Center wavelength (nm)	Bandwidth (nm)
Red	650	16
Green	560	16
Blue	450	16
NIR	840	26
Red edge	730	16

Table 2 Specification of DJI multispectral UAV camera

Criteria	Lens
FOV	62.7°
Focal length	5.74 mm (35 mm format equivalent: 40 mm) infinite fixed focal length
Aperture	f/2.2
ISO range for color sensors	200–800
Monochromatic sensor gain	1–8 times
Electronic global shutter	1/100–1/20000S (visible light imaging) 1/100–1/10000S (multispectral imaging)

rainfall of about 1479.3 mm. The vegetation cover in the study area is mainly cypress, miscellaneous shrubs, and *Miscanthus* (stem).

UAV multispectral imagery collection

A commercial DJI Phantom 4 multispectral UAV (DJI, Shenzhen, China), which is equipped with one Red–Green–Blue (RGB) sensor and five multispectral sensors with each of 2.08 effective megapixels, was used for capturing images of the study area. The UAV has a real-time kinematics (RTK) positioning system based on satellite navigation, which reduces the error of satellite-based position data to the centimeter level. The information about the five bands of the UAV camera is shown in Table 1. The weight, image resolution, and sensor size of the UAV multispectral camera are 1487 g, 1600 × 1300 pixels, and 4.87 × 3.96 mm, respectively. More parameters of the camera are shown in Table 2. Since the camera collects RGB and five spectral images through the sunlight sensor, it records the illuminance information of each image, which facilitates the post-calibration of multispectral images. Additionally, a control panel built into the UAV can be used to automatically adjust the radiometric reflectance calibration and directly acquire reflectance spectral data. Online access to the details of DJI Phantom 4 multispectral UAV is <https://www.dji.com/p4-multispectral/specs>.

To ensure consistent environmental conditions such as weather, light, wind direction, and wind speed, we chose

to capture UAV images at different altitudes on the same day. The UAV image data acquisition was conducted on a clear and windless day at the times from 12:00 to 2:00 pm on January 13, 2022. The higher the flight altitude, the larger the flight coverage area. The image overlaps the line by 80%. The flight altitudes were set at 100 m (flight surface: 6.01 ha), 200 m (flight surface: 14.01 ha), and 400 m (flight surface: 40.63 ha) above the ground, respectively. The total flight mission takes about two hours to complete.

UAV image data acquisition

Georeferencing of raw images is achieved through RTK systems. The camera parameters and POS parameters corresponding to each aerial image are obtained according to the calibration results of the UAV camera, the airborne Differential Global Positioning System (DGPS), and the flight controller. Parameters include 178 horizontal and vertical errors less than 0.1 m [35]. Pix4D mapper software (version 4.2.27, Pix4D SA, Switzerland) was used to stitch the complete raw images generated for georeferenced spectral reflectance and VIs mosaic calibration. Aerial triangulation was the initial step in the UAV photogrammetry workflow and could be used to determine the individual orientations of each diorama of the photogrammetry block. The SFM and Multi-View Stereo (MVS) algorithms were performed by the Pix4D mapper, performing bundle adjustment. We oriented an unlimited number of images across the block by bundling adjustments and multiple ground control points (GCPs). The GCPs for each flight has 5 points, one in each of the four corners of the flight area and one in the middle of the area. Aerial triangulation of each sensor also considered its specific lens distortion to determine the position and orientation of each sensor. Combining the above data, a dense point cloud for multi-view stereo matching could be formed, to achieve the purpose of surface reconstruction, and generate orthomosaics images [36].

Next, high-resolution GeoTIFF images including reflectance and VIs were generated, at the same location with different flight altitudes. GeoTIFF images were further processed in ENVI 5.3 (Esri Inc., Redlands, USA) to define regions of interest (ROI). The average reflection spectrum values of objects within the ROI range were taken as the reflection spectrum of the sampling point and the multispectral data corresponding to various ground objects at the sampling point were obtained.

Methods

Modeling methods

For karst vegetation coverage detection, we focused on comparing the efficiency of four models: RF, SVM, GBM, and DL.

Table 3 Details of selected vegetation indices tested in this research

Name	Formula	References
NDVI (normalized difference vegetation index)	$(\text{NIR} - \text{R})/(\text{NIR} + \text{R})$	[70]
OSAVI (optimized soil adjusted vegetation index)	$((\text{NIR} - \text{R})(1 + 0.16))/(\text{NIR} + \text{R} + 0.16)$	[71]
GNDVI (green normalized difference vegetation index)	$(\text{NIR} - \text{G})/(\text{NIR} + \text{G})$	[72]
SAVI (soil adjusted vegetation index)	$((\text{NIR} - \text{R})(1 + 0.5))/(\text{NIR} + \text{R} + 0.5)$	[73]
MSAVI (modified soil adjusted vegetation index)	$(2\text{NIR} + 1 - \sqrt{((2*\text{NIR} + 1)^2 - 8*(\text{NIR} - \text{R}))/2})$	[74]
GCI (green chlorophyll Index)	$\text{NIR}/\text{G} - 1$	[75]
RECI (red edge chlorophyll index)	$\text{NIR}/\text{E} - 1$	[75]
LCI (leaf chlorophyll index)	$(\text{NIR} - \text{E})/(\text{NIR} + \text{R})$	[76]
GRVI (green red vegetation index)	$(\text{G} - \text{R})/(\text{G} + \text{R})$	[77]
MGRVI (modified green red vegetation index)	$(\text{G}^2 - \text{R}^2)/(\text{G}^2 + \text{R}^2)$	[62]
RGBVI (red green blue vegetation index)	$(\text{G}^2 - \text{R} \times \text{B})/(\text{G}^2 + \text{R} \times \text{B})$	[78]
NDRE (normalized difference red edge index)	$(\text{NIR} - \text{E})/(\text{NIR} + \text{E})$	[79]
MACI (modified anthocyanin content index)	NIR/G	[66]
ARI (anthocyanin reflectance index)	G/NIR	[80]
MARI (modified anthocyanin reflectance index)	$(\text{G}^{(-1)} - \text{E}^{(-1)})/\text{NIR}$	[81]
VDVI (visible-band difference vegetation index)	$(2\text{G} - \text{R} - \text{B})/(2\text{G} + \text{R} + \text{B})$	[82]

NIR, E, R, G, and B represent the near-infrared, red edge, red, green, and blue bands, respectively

RF

RF is an ensemble classifier that generates multiple decision trees using randomly selected training samples and subsets of variables [37]. This classifier has become popular in the remote sensing community due to its classification accuracy [38, 39]. Not only can RF process and multicollinear high-dimensional data successfully, but it is also fast and insensitive to overfitting, and it is sensitive to sampling design [40, 41]. The variable importance measure provided by RF has been widely used in different scenarios, such as selecting the best variable to classify a specific target class [42].

SVM

SVM is a supervised learning model with associated learning algorithms for classification, regression analysis, and outlier detection of data [43]. In addition to reducing the complexity of the approximation function while ensuring the accuracy of data approximation, the SVM algorithm also has a lot of advantages in solving large samples and high-dimensional space problems [44]. At present, it has been successfully applied to spectral analysis research.

GBM

To solve the problem that it is not easy to optimize each step of the general loss function, Friedman [45] proposed the gradient boosting machine algorithm, whose idea is borrowed from the gradient descent method. The basic

principle of GBM is to train the newly added weak classifier according to the negative gradient information of the loss function of the current model and then combine the trained weak classifier into the existing model in the form of accumulation [46].

DL

A DL architecture is a multilayer stack of simple modules that attempts to learn deep features of input data hierarchically through very deep neural networks, many of which compute nonlinear input–output mappings [47]. With multiple layers of nonlinearity in DL, a system can implement extremely complex functions of its input that are sensitive to minute details and insensitive to irrelevant changes in the background and surrounding objects [48]. According to the training process, DL is first initialized hierarchically through unsupervised training and then adjusted in a supervised manner. In this scheme, high-level features can be learned from low-level features, and suitable features can eventually be used for classification [49].

Spectral vegetation indices selection

Spectral methods are considered potential methods for predicting photosynthetic pigment content, leaf spectral properties are used for reflectance spectra (e.g. VIs), and the physiology of individual trees or populations will be estimated at the stand level [50]. However, different VIs may produce different vegetation characteristic reflections and these indicators may also be affected by

Table 4 Parameters of the models at each flight altitude and total

Flight altitude (m)	Model	Classification	R (%)	F1 (%)	OA (%)
100	GBM	gv	98.71	99.03	98.61
		ro	99.67	99.51	
		so	97.12	98.34	
		wd	98.51	97.30	
	RF	gv	98.84	98.91	98.55
		ro	99.34	99.50	
		so	97.12	98.33	
		wd	98.26	97.29	
	SVM	gv	97.81	98.06	97.51
		ro	98.36	99.17	
		so	97.12	97.72	
		wd	96.52	95.10	
	DL	gv	97.68	97.87	95.83
		ro	98.36	99.17	
		so	83.95	91.07	
		wd	97.51	92.23	
200	GBM	gv	99.31	99.31	99.11
		ro	100.00	99.54	
		so	97.09	97.56	
		wd	98.82	98.82	
	RF	gv	99.31	99.03	98.70
		ro	100.00	99.77	
		so	94.17	95.10	
		wd	97.93	98.22	
	SVM	gv	98.89	98.55	97.95
		ro	97.70	98.83	
		so	93.20	92.75	
		wd	97.34	97.48	
	DL	gv	97.92	97.52	92.70
		ro	99.08	94.09	
		so	58.25	73.17	
		wd	97.04	90.36	
400	GBM	gv	98.74	98.58	98.53
		ro	99.23	99.42	
		so	90.87	93.33	
		wd	93.89	95.56	
	RF	gv	99.23	98.68	98.45
		ro	98.08	98.65	
		so	88.94	92.04	
		wd	92.36	94.84	

Table 4 (continued)

Flight altitude (m)	Model	Classification	R (%)	F1 (%)	OA (%)
all	SVM	gv	98.38	97.65	96.97
		ro	89.23	91.7	
		so	69.71	74.55	
		wd	88.65	93.55	
	DL	gv	96.53	94.13	89.67
		ro	90.77	86.61	
		so	0.48	0.96	
		wd	38.86	55.97	
	GBM	gv	98.44	98.14	97.81
		ro	96.93	97.81	
		so	90.45	93.22	
		wd	95.75	96.27	
RF	gv	98.60	98.09	97.76	
	ro	96.81	97.87		
	so	90.63	93.15		
	wd	95.58	96.06		
SVM	gv	97.01	96.35	94.68	
	ro	93.36	92.71		
	so	65.95	75.39		
	wd	91.83	92.64		
DL	gv	89.77	93.72	92.45	
	ro	93.61	95.01		
	so	81.08	72.00		
	wd	95.41	93.43		

All represents the total including 100 m, 200 m and 400 m data

Bold values represent the highest overall accuracy of the models at each flight altitude

different vegetation types. To determine the best variables for vegetation coverage detection, 5 spectral values (Red, Green, Blue, NIR, Red edge) and 16 VIs were selected, as shown in Table 3, including NDVI, OSAVI, GNDVI, SAVI, MSAVI, GCI, RECI, LCI, GRVI, MGRVI, RGBVI, NDRE, MACI, ARI, MARI, and VDVI.

The NDVI is the most commonly used indicator of vegetation greenness/vitality, showing strong correlations with Leaf Area Index (LAI) and green biomass, providing information for estimating Net Primary Production and enabling us to distinguish vegetation from non-vegetation [51]. The OSAVI can be used to reduce the effect of soil background on sparse and dry vegetation [52]. The GNDVI is highly sensitive to chlorophyll and reduces non-photosynthetic effects, which can provide valuable

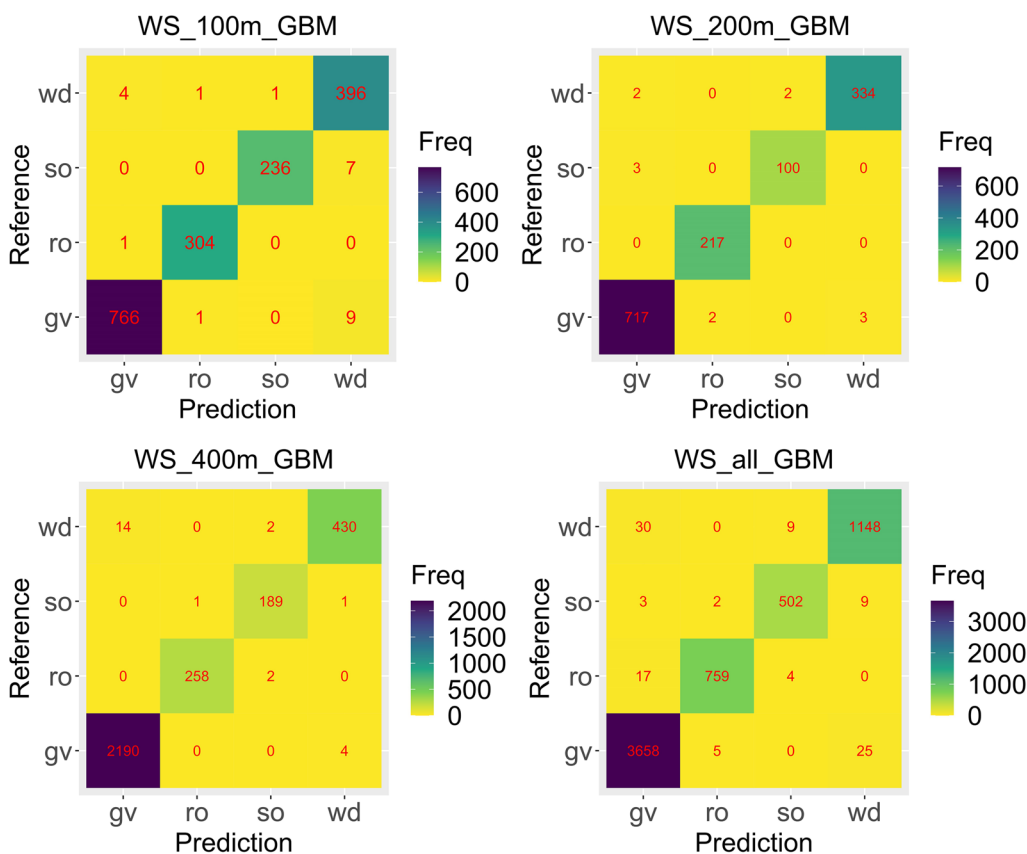


Fig. 2 Confusion matrix of best models (GBMs)

information on complex landscapes [53]. The SAVI is more sensitive to vegetation, allowing us to observe areas of potential soil degradation [54]. The MSAVI is commonly used to detect sparsely vegetated areas where soil background influences are important, minimizing external influences and enhancing vegetation signals [55]. A low MSAVI means sparse vegetation, indicating desertification [56]. The GCI is an index for estimating the chlorophyll content of various plant leaves and detecting the physiological state of vegetation, which can be used to evaluate the growth state of vegetation [57]. The RECI constructed from red edge bands is more sensitive than the traditional vegetation index in estimating vegetation biomass [58]. The LCI is a chlorophyll-sensitive vegetation index with a wide range of chlorophyll content and is hardly affected by disturbances caused by scattering [59].

The GRVI is sensitive to leaf color changes (chlorophyll and fall coloration) and can be used to differentiate green vegetation, water, and soil [60]. The GRVI uses the high reflectivity of plants in green (about 540 nm) and the absorption of the red and blue parts of the visible spectrum (400–700 nm) by plant chlorophyll to identify vegetation [61]. The squared band reflectance value helps

to amplify the difference between red, green, and blue reflectance [62]. The MGRVI is defined as the normalized difference between squared green reflectance and squared red reflectance, and thus exhibits higher sensitivity in vegetation identification [63]. The RGBVI can be used to extract vegetation cover from drone orthoimages [64]. The NDRE is an important predictor of canopy properties and is very sensitive to canopy chlorophyll content [65]. The MACI correlates with anthocyanin content in plant leaves, providing valuable information about the physiological state of plants [66]. The ARI can be used to assess vegetation health [67]. The MARI has the potential to further aid in the classification of senescent vegetation [68]. Since there are significant differences in the absorption efficiency of vegetation to different wavelengths, the VDVI can be used to detect vegetation pixels and effectively enhance vegetation information [69].

UAV data analysis methods

In this study, 100 m, 200 m, and 400 m data were randomly sampled through the *e1071* package [83], and *tidyverse* package [84] of *R* software. The collected data is

divided into the training set and validation set, of which 80% is the training set and the remaining 20% is the validation set [85]. The RF, SVM, GBM, and DL models were established on test data with flight altitudes of 100 m, 200 m, and 400 m, respectively. All modeling was performed in R software, the *randomForest* package [86] was used for RF modeling, the *caret* package [87] was used for SVM modeling and the *h2o* package [88] was used for GBM and DL modeling, respectively.

Model accuracy verification

After the models were constructed, the confusion matrix for each model was obtained through the *caret* package [87] in R software. A confusion matrix of size $n \times n$ associated with the classifier shows the predicted and actual classifications, where n is the number of distinct classes [89]. The prediction accuracy and classification error can be obtained from this matrix as follows:

$$\text{Accuracy} = (a + d)/(a + b + c + d) \quad (1)$$

$$\text{Error} = (b + c)/(a + b + c + d) \quad (2)$$

where a is the number of correct negative predictions, b is the number of incorrect positive predictions, c is the number of incorrect negative predictions, and d is the number of correct positive predictions. The fitting and predictive ability of each model were evaluated in combination with the overall accuracy. The higher the overall accuracy, the better the model fitting and prediction ability, and the higher the model accuracy.

In addition, the relationship between the overall accuracy, recall and f1 score of the machine learning model to detect karst vegetation should be considered. The parameters recall (R), F1-score (F1), and overall accuracy (OA) were used for RE, SVM, GBM, and DL model performance [30]. Overall accuracy is a widely used metric in classification, it expresses the ratio between the model and the total number of predictions on all test sets. For raw samples and predictions, recall and precision are the ratios of correct predictions to the total number of actual or predicted items in the ensemble, respectively. Generally, precision and recall are a contradictory pair of measures, when one is higher, the other tends to be lower. The F1 score is the harmonic mean of precision and recall, with 1 being the best and 0 being the worst [90].

Three quantities from the performance of a classification process in the population of all instances were used to calculate R , F1, and OA: True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) using below the equations [91]:

$$R = TP/(TP + FN) \quad (3)$$

$$F1 = 2TP/(2TP + FP + FN) \quad (4)$$

$$AA = TP + TNTP + TN + FP + FN \quad (5)$$

$$OA = (AA_1 + AA_1 + AA \dots + AA_n)/n \quad (6)$$

where P , AA, and n are the precision, the average accuracy, and the number of the classes [green vegetation (gv), rock (ro), soil (so), and weed (wd)], respectively. The accuracy of a classification process was defined as the portion of true positives and true negatives in all instances.

Important variable selection

In predictive modeling, the main concern is to identify the most important predictors included in the reduced model. This can be achieved by identifying the best predictors based on statistical characteristics such as importance or accuracy [92]. Using variable selection to develop predictive models can not only reduce the burden of data collection but also improve predictive efficiency in practice. Since many datasets have hundreds or thousands of possible predictors, variable selection is often a necessary part of predictive model development [93]. In this study, we used the *h2o* package [94] in R software to perform significant variable selection on the model with the highest overall accuracy and to determine the best predictors for the model. Two additional files show this in more detail [see Additional files 1-2].

Best model inversion and determination

To determine the best model for consistency and accuracy across all flight altitudes, we used the best models built from each flight altitude data to validate the prediction accuracy of the remaining data. After determining on which data the best model is based, it is necessary to test the prediction accuracy of the model for the remaining altitude data. The performance of the best model is tested first on the test set and then on the real set (full original images at flight altitudes of 100 m, 200 m, and 400 m). The images of the three flight altitudes predicted based on the best model are compared with the original images of the respective real data. Finally, the model with the best karst vegetation detection accuracy and karst vegetation retrieval performance was determined.

Table 5 Mutual validation parameters of optimal models at different flight altitudes

GBM model flight altitude (m)	Data flight altitude (m)	Classification	R (%)	F1 (%)	OA (%)
100	All	gv	74.95	85.03	80.41
		ro	95.53	59.22	
		so	70.63	78.87	
		wd	91.99	90.60	
	200	gv	72.58	83.71	83.19
		ro	99.54	82.92	
		so	82.52	82.52	
		wd	95.56	82.71	
	400	gv	90.35	93.69	84.96
		ro	69.23	62.61	
		so	27.40	35.85	
		wd	93.89	76.99	
200	All	gv	98.76	93.53	89.03
		ro	88.76	87.37	
		so	52.43	66.74	
		wd	75.86	82.97	
	100	gv	99.36	96.31	94.78
		ro	95.74	97.01	
		so	86.01	92.27	
		wd	90.48	91.39	
	400	gv	97.83	96.72	88.82
		ro	63.08	67.91	
		so	25.48	34.08	
		wd	88.55	80.48	
400	All	gv	87.77	92.53	80.63
		ro	91.15	74.71	
		so	64.07	47.77	
		wd	52.21	68.31	
	100	gv	98.44	98.31	80.21
		ro	97.01	69.61	
		so	39.91	49.01	
		wd	52.84	69.14	
	200	gv	78.58	87.04	66.15
		ro	87.96	53.45	
		so	30.21	25.78	
		wd	38.02	55.09	
All	100	gv	97.42	97.92	95.66
		ro	96.72	94.10	
		so	85.27	91.17	
		wd	97.26	94.90	
	200	gv	84.30	91.41	89.79
		ro	99.08	85.89	
		so	81.15	78.06	
		wd	97.39	92.82	
	400	gv	93.27	95.58	88.31
		ro	79.69	74.59	
		so	36.89	47.35	
		wd	92.29	78.24	

Table 5 (continued)

All represents the total including 100 m, 200 m and 400 m data
 Bold represents the highest overall accuracy when the optimal models at different flight altitudes were mutually validated

Results

Analysis of modeling accuracy of karst vegetation discrimination

The performance of four models using multispectral and VIs at different flight altitudes were shown in Table 4. GBM model yields the best vegetation coverage detection accuracy at all flight altitudes data, with the best accuracies from high to low of 99.11% (200 m), 98.61% (100 m), 98.53% (400 m) and 97.81% (all) respectively. Figure 2 displayed the confusion matrix of each best model (GBM) obtained at different flight altitudes. The highest accuracy of the GBM model was found at 200 m, the model accuracy gradually increases to a certain extent as the flight altitude gradually decreases.

Mutual validation analysis of optimal models at different flight altitudes

As shown in Table 5, the highest and lowest prediction accuracy rates of 95.66% and 66.15% were found when using the GBM model from all data to detect 100 m data and from 400 m data to detect 200 m data respectively. It can be found that the best model of each flight altitude has the highest accuracy to detect its own training data, but the accuracies show a downward trend to detect other flight altitude data. The above indicates that there is a certain degree of difference in the accuracy of the prediction model when training its own data and the rest of the data.

Importance analysis of VIs for karst vegetation coverage detection

It can be known that the GBM model (GBM_all) established based on the overall data works best (Table 5). Figure 3 combined with Table 5 shows that the GBM_all model has the best accuracy of 95.66% when predicting the data with a flight altitude of 100 m, while its accuracy is the worst at 88.31% when predicting the data with a flight altitude of 400 m.

In addition, the vegetation index importance order of the best GBM model at each flight altitude was shown in Fig. 4. The accuracy of GBM models based on different flight altitudes and overall data were significantly correlated with the following vegetation indices: MGRVI (100 m, 200 m), RECI (400 m), MSAVI (all). On contrary, the accuracy of GBM models based on different flight

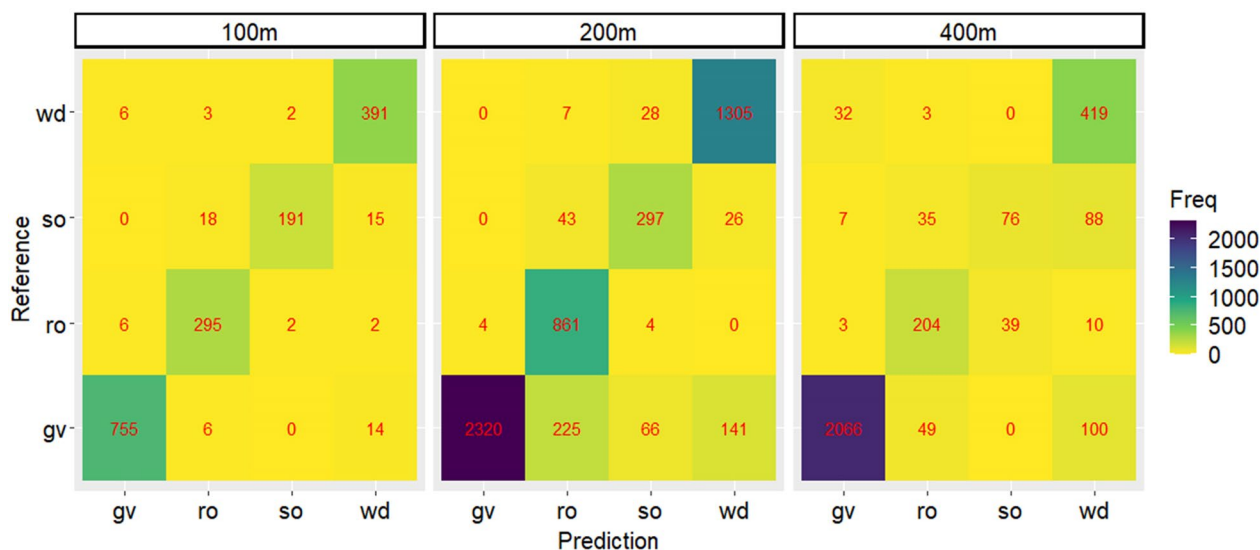


Fig. 3 Confusion matrices of GBM models constructed based on overall data for remaining data predictions

altitudes and overall data was significantly uncorrelated with the following vegetation indices: OSAVI (100 m, 400 m), SAVI (200 m), and MACI (all). The above shows that the prediction models constructed based on data from different flight altitudes have a certain similarity in the distribution of vegetation index importance.

Inversion of the best model

As shown in Figs. 5, 6 and 7 (the *x* and *y* axes in degrees represent east longitude and north latitude, respectively), the orthomosaics of the three flight altitudes predicted based on the GBM_all model were compared with the original orthomosaics of the respective real data. The orthomosaics for RGB visualization were shown on the left in Figs. 5, 6 and 7. Combined with the method of visual interpretation, the karst green vegetation predicted by the best model was in good agreement with the ground truth, and other land types including hay, rock, and soil were well predicted. It is worth noting that, unlike the southwest karst area where vegetation growth is difficult, the eastern China karst area to which the study area belongs has high vegetation and soil coverage, but low rock exposure.

Discussion

In the past, Fu et al. [95] combined four single-class SegNet models to classify karst vegetation with an overall accuracy rate of 87.34%. However, the GBM_all model proposed by our study had a higher overall accuracy of 95.66%. Flight altitude determines the final image resolution obtained as well as the effect of topography on radiance (by changing the relative angle between terrain

slopes and the UAV) [96]. Many factors affect the setting of the flight altitude, including the weather, temperature of the sampling point, and minor differences in flight altitudes. These factors may complicate the obtained image background and bring challenges to karst vegetation pixel segmentation and model accuracy detection [97]. Larriaga et al. [98] found that, contrary to expectations, the model fitting accuracy obtained at a higher UAV flight altitude was also higher. But in our research, as the flight altitude gradually decreases, the model accuracy gradually improves to a certain extent.

In our study, the vegetation indices that were significantly correlated and not correlated with the best model (GBM_all) were MSAVI and MACI, respectively, which has already been proven in previous studies. For example, Qi et al. [74] found that MSAVI could increase the dynamic range of vegetation signals, further reduce the influence of soil background, and could be effectively used for vegetation detection in karst areas. Anne et al. [99] found that excessive anthocyanins were mainly related to the juvenile or senescent state of plants, so MACI may not have a significant impact on the detection of karst vegetation coverage.

In addition, in this study, the vegetation indices that were significantly and not significantly correlated with the two flight altitude models were the MGRVI (100 m, 200 m) and the OSAVI (100 m, 400 m). This may be due to differences in lighting during UAV flights [71, 100, 101]. However, Bendig Juliane, Fern et al. [62, 102] found that MGRVI and OSAVI were mainly used for crop identification rather than forestry karst vegetation coverage detection.

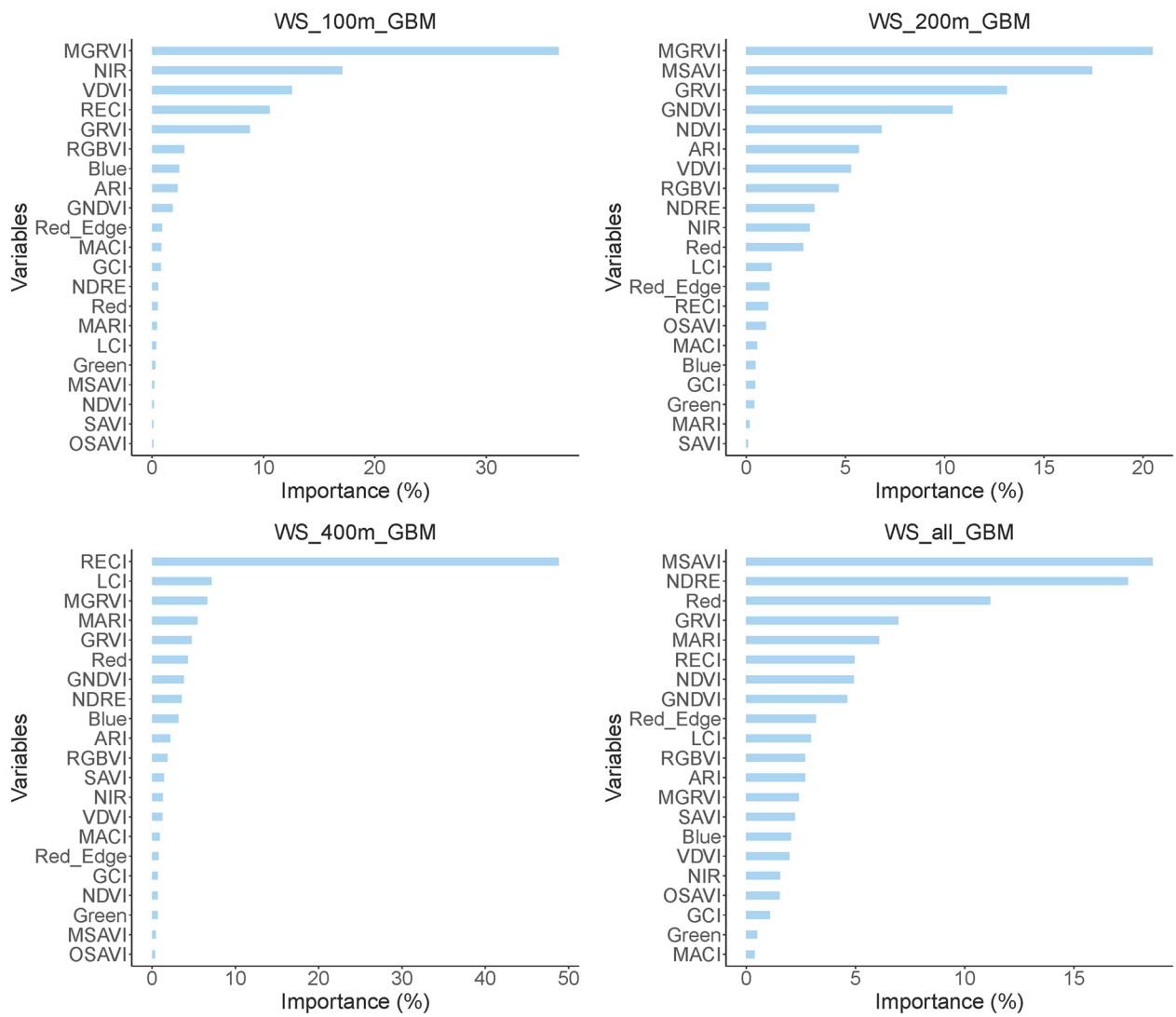


Fig. 4 The vegetation indices importance of the best GBM models for each flight altitude

Although the method proposed in this study performed well in karst vegetation coverage detection, some aspects could be further improved and explored. First, the resulting images are orthoimages taken by the drone, and vertical cropping of information may cause some loss. To reduce the mutual interference between ground features and improve the accuracy of classification, characteristic parameters can be added to the interpretation process and the layered mask method can be used to select features for each layer image [103]. Therefore, we plan to experiment with layered images in the future to reduce information that can not be displayed due to occlusion. Second, this study only compares four commonly used machine learning methods (RF, SVM, GBM, and DL). Machine learning models have complex parameters and structures [104], so

we plan to try a wider variety of machine learning methods in the future. In conclusion, UAV multispectral vegetation indices have high potential in the field of karst vegetation detection and creative results can be achieved when fused with machine learning algorithms.

Conclusion

In this work, we proposed a GBM_all model based on all flight altitudes image data from UAVs, which could accurately detect karst vegetation coverage with an overall accuracy of up to 95.66%. This study verified the prediction models constructed based on data from different flight altitudes had a certain similarity in the distribution of vegetation index importance. Combined with the

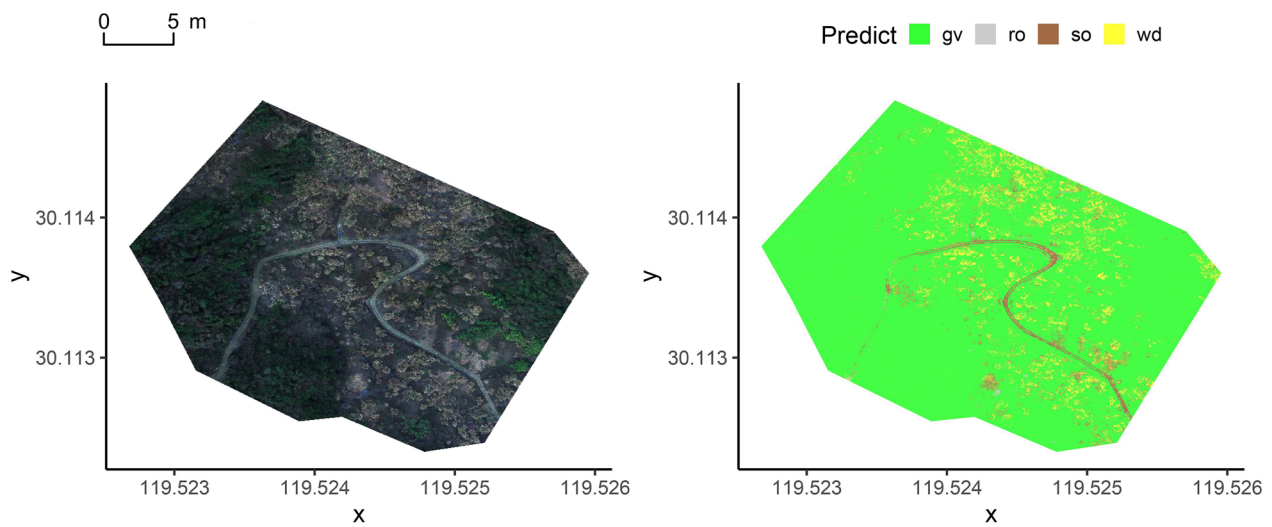


Fig. 5 The classification of the best model on the 100 m data. The gv is green vegetation, wd is hay, ro is rock, and so is soil

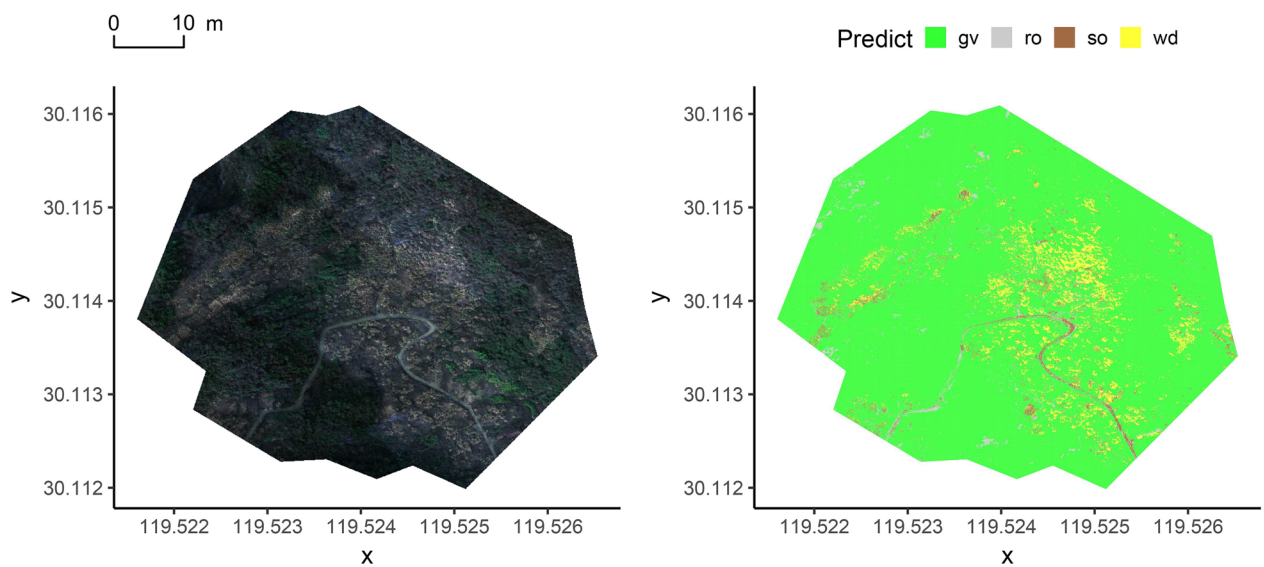


Fig. 6 The classification of the best model on the 200 m data. The gv is green vegetation, wd is hay, ro is rock, and so is soil

method of visual interpretation, we found that the karst green vegetation predicted by the best model was in good agreement with the ground truth, and other land types including hay, rock, and soil were well predicted.

UAV images were beneficial to refine the texture features of the model, improve parameter information, and were more suitable for the detection of karst continuous and complex vegetation. In this study, the combination of UAV images, multispectral vegetation index,

and machine learning algorithm has also shown good performance in karst vegetation inversion, providing a reliable and promising method for the identification of vegetation, bare rock, soil in the eastern karst area. In addition, timely and accurate detection of karst vegetation cover will provide important reference information for various forestry management departments in karst areas to reasonably determine karst vegetation restoration plans and evaluate relevant policies.

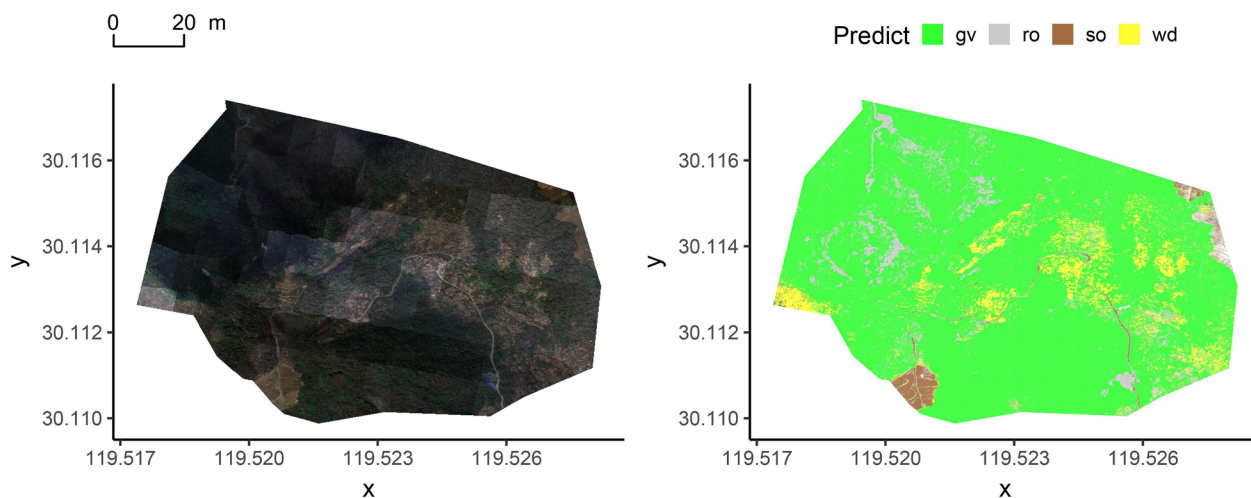


Fig. 7 The classification of the best model on the 400 m data. The gv is green vegetation, wd is hay, ro is rock, and so is soil

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-023-00982-7>.

Additional file 1. The vegetation indices importance of the RF models for each flight altitude

Additional file 2. The vegetation indices importance of the DL models for each flight altitude

Author contributions

WP experimented and wrote the manuscript. XW, JW, and YS supported the data collection and field experiments and supervised the experiments. YL and SL designed the study, supervised experiments, and performed revisions of the manuscript. All authors read and approved the final manuscript.

Funding

This work has received funding from Fundamental Research Funds of CAF No. CAFYBB2020SY017, Fundamental Research Funds of RISF: RISFZ-2021-09 and National Key R&D Program of China: 2016YFC0502605-3.

Availability of data and materials

The Data mentioned in this study are available on request from the corresponding author.

Declarations

Ethics approval and consent to participate

All authors read and approved the manuscript.

Consent for publication

All authors agreed to publish this manuscript.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Frísia S, Borsato A. Karst Develop Sedimentol. 2010;61:269–318.
- Ford D, Williams PD. Karst hydrogeology and geomorphology. New York: Wiley; 2013.
- Jiang Z, Lian Y, Qin X. Rocky desertification in Southwest China: impacts, causes, and restoration. *Earth Sci Rev*. 2014;132:1–12.
- Jiang Z, Liu H, Wang H, Peng J, Meersmans J, Green SM, Quine TA, Wu X, Song Z. Bedrock geochemistry influences vegetation growth by regulating the regolith water holding capacity. *Nat Commun*. 2020;11(1):1–9.
- Wu L, Wang S, Bai X, Tian Y, Luo G, Wang J, Li Q, Chen F, Deng Y, Yang Y. Climate change weakens the positive effect of human activities on karst vegetation productivity restoration in southern China. *Ecol Ind*. 2020;115:106392.
- Zhao S, Pereira P, Wu X, Zhou J, Cao J, Zhang W. Global karst vegetation regime and its response to climate change and human activities. *Ecol Ind*. 2020;113:106208.
- Harrington TJ, Mitchell DT. Characterization of *dryas octopetala* ectomy-corrhizas from limestone karst vegetation, western Ireland. *Can J Bot*. 2002;80(9):970–82.
- Yue Y, Wang K, Zhang B, Liu B, Chen H, Zhang M. Uncertainty of remotely sensed extraction of information of karst rocky desertification. *Adv Earth Sci*. 2011;26(3):266.
- Blasi C, Di Pietro R, Pelino G. The vegetation of alpine belt karst-tectonic basins in the central apennines (Italy). *Plant Biosys Int J Dealing Aspects Plant Biol*. 2005;139(3):357–85.
- Bátori Z, Csiky J, Farkas T, Vojtkó EA, Erdős L, Kovács D, Wirth T, Körmöczy L, Vojtkó A. The conservation value of karst dolines for vascular plants in woodland habitats of Hungary: Refugia and climate change. *Int J Speleol*. 2014;43(1):2.
- Belward AS, Skøien JO. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS J Photogramm Remote Sens*. 2015;103:115–28.
- Zhang R, Luo H, Zou Y, Liu G. Discussion on possibility of the identification of karst vegetation communities based on OLI data. In: 2014 the third international conference on agro-geoinformatics; 2014. IEEE. p. 1–7.
- Qu L, Han W, Lin H, Zhu Y, Zhang L. Estimating vegetation fraction using hyperspectral pixel unmixing method: a case study of a karst area in China. *IEEE J Sel Topics Appl Earth Observ Remote Sens*. 2014;7(11):4559–65.

Received: 24 May 2022 Accepted: 13 January 2023

Published online: 23 January 2023

14. Zhang X, Shang K, Cen Y, Shuai T, Sun Y. Estimating ecological indicators of karst rocky desertification by linear spectral unmixing method. *Int J Appl Earth Obs Geoinf*. 2014;31:86–94.
15. Song L, Yulun A, Houjiang H. Automated method based on change detection for extracting karst rock desertification information using remote sensing. *Remote Sens Technol Appl*. 2012;27(1):149–53.
16. Guimarães N, Pádua L, Marques P, Silva N, Peres E, Sousa JJ. Forestry remote sensing from unmanned aerial vehicles: a review focusing on the data, processing and potentialities. *Remote Sens*. 2020;12(6):1046.
17. Zhang Z, Ouyang Z, Xiao Y, Xiao Y, Xu W. Using principal component analysis and annual seasonal trend analysis to assess karst rocky desertification in southwestern China. *Environ Monit Assess*. 2017;189(6):1–19.
18. Xiao D, Zhou Z, Li Q, Huang D, Meng Z, Zhang Y. Construction of terrain information extraction model in the karst mountainous terrain fragmentation area based on UAV remote sensing. In: 2022 3rd international conference on geology, mapping and remote sensing (ICGMRS); 2022. IEEE. P. 716–27.
19. Pádua L, Vanko J, Hruška J, Adão T, Sousa JJ, Peres E, Morais R. UAS, sensors, and data processing in agroforestry: a review towards practical applications. *Int J Remote Sens*. 2017;38(8–10):2349–91.
20. de Castro AI, Shi Y, Maja JM, Peña JM. UAVs for vegetation monitoring: overview and recent scientific contributions. *Remote Sens*. 2021;13(11):2139.
21. Dainelli R, Toscano P, Di Gennaro SF, Matese A. Recent advances in unmanned aerial vehicle forest remote sensing—a systematic review. Part I: a general framework. *Forests*. 2021;12(3):327.
22. Riihimäki H, Luoto M, Heiskanen J. Estimating fractional cover of tundra vegetation at multiple scales using unmanned aerial systems and optical satellite data. *Remote Sens Environ*. 2019;224:119–32.
23. Moreno JL, Ortega JF, Moreno MÁ, Ballesteros R. Using an unmanned aerial vehicle (UAV) for lake management: ecological status, lake regime shift and stratification processes in a small Mediterranean karstic lake. *Limnetica*. 2022;41(2):000–000.
24. Zhou R, Yang C, Li E, Cai X, Yang J, Xia Y. Object-based wetland vegetation classification using multi-feature selection of unoccupied aerial vehicle RGB imagery. *Remote Sens*. 2021;13(23):4910.
25. Kampen M, Lederbauer S, Mund J, Immitzer M. Uav-based multi-spectral data for tree species classification and tree vitality analysis. Dreiländertagung der DGPF der OVG und der SGPF in Wien sterreich Publikationen der DGPF. 2019;28:01.
26. Tmušić G, Manfreda S, Aasen H, James MR, Gonçalves G, Ben-Dor E, Brook A, Polinova M, Arranz JJ, Mészáros J. Current practices in UAS-based environmental monitoring. *Remote Sens*. 2020;12(6):1001.
27. Dai L, Zhang G, Gong J, Zhang R. Autonomous learning interactive features for hyperspectral remotely sensed data. *Appl Sci*. 2021;11(21):10502.
28. Puliti S, Breidenbach J, Astrup R. Estimation of forest growing stock volume with UAV laser scanning data: can it be done without field data? *Remote Sens*. 2020;12(8):1245.
29. Chen G, Weng Q, Hay GJ, He Y. Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities. *Gi Sci Remote Sens*. 2018;55(2):159–82.
30. Pádua L, Adão T, Hruška J, Guimarães N, Marques P, Peres E, Sousa JJ. Vineyard classification using machine learning techniques applied to RGB-UAV imagery. In: IGARSS 2020–2020 IEEE international geoscience and remote sensing symposium; 2020. IEEE. p. 6309–12.
31. Fu B, Liu M, He H, Lan F, He X, Liu L, Huang L, Fan D, Zhao M, Jia Z. Comparison of optimized object-based rf-dt algorithm and segnet algorithm for classifying karst wetland vegetation communities using ultra-high spatial resolution uav data. *Int J Appl Earth Obs Geoinf*. 2021;104:102553.
32. Zhang N, Wang Y, Zhang X. Extraction of tree crowns damaged by *Dendrolimus tabulaeformis* Tsai et Liu via spectral-spatial classification using UAV-based hyperspectral images. *Plant Methods*. 2020;16(1):1–19.
33. Mäyrä J, Keski-Saari S, Kivinen S, Tanhuanpää T, Hurskainen P, Kullberg P, Poikolainen L, Viinikka A, Tuominen S, Kumpula T. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens Environ*. 2021;256:112322.
34. Li S-L, Liu C-Q, Chen J-A, Wang S-J. Karst ecosystem and environment: characteristics, evolution processes, and sustainable development. *Agr Ecosyst Environ*. 2021;306:107173.
35. Ma S, Zhang K. Low-altitude photogrammetry and remote sensing in UAV for improving mapping accuracy. *Mobile Inform Sys* 2022; 2022.
36. Iglhaut J, Cabo C, Puliti S, Piermattei L, O'Connor J, Rosette J. Structure from motion photogrammetry in forestry: a review. *Current Forestry Rep*. 2019;5(3):155–68.
37. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
38. Du P, Samat A, Waske B, Liu S, Li Z. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS J Photogramm Remote Sens*. 2015;105:38–53.
39. Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogramm Remote Sens*. 2012;67:93–104.
40. Dalponte M, Ørka HO, Gobakken T, Gianelle D, Næsset E. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans Geosci Remote Sens*. 2012;51(5):2632–45.
41. Millard K, Richardson M. On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. *Remote Sens*. 2015;7(7):8489–515.
42. Corcoran JM, Knight JF, Gallant AL. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in Northern Minnesota. *Remote Sens*. 2013;5(7):3212–38.
43. Mammone A, Turchi M, Cristianini N. Support vector machines. *Wiley Interdiscip Rev Comput Stat*. 2009;1(3):283–9.
44. Sluiter R, Pebesma E. Comparing techniques for vegetation classification using multi-and hyperspectral images and ancillary environmental data. *Int J Remote Sens*. 2010;31(23):6143–61.
45. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–232.
46. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38(4):367–78.
47. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.
48. Le Roux N, Bengio Y. Deep belief networks are compact universal approximators. *Neural Comput*. 2010;22(8):2192–207.
49. Chen Y, Lin Z, Zhao X, Wang G, Gu Y. Deep learning-based classification of hyperspectral data. *IEEE J Select Topics Appl Earth Observ Remote Sens*. 2014;7(6):2094–107.
50. Tao X, Li Y, Yan W, Wang M, Tan Z, Jiang J, Luan Q. Heritable variation in tree growth and needle vegetation indices of slash pine (*Pinus elliottii*) using unmanned aerial vehicles (UAVs). *Ind Crops Prod*. 2021;173:114073.
51. Castellana M, Rita A, Camarero JJ, Colangelo M, Ripullone F. Declines in canopy greenness and tree growth are caused by combined climate extremes during drought-induced dieback. *Sci Total Environ*. 2022;813:152666.
52. Leolini L, Moriondo M, Rossi R, Bellini E, Brilli L, López-Bernal Á, Santos JA, Fraga H, Bindi M, Dibari C. Use of sentinel-2 derived vegetation indices for estimating fPAR in olive groves. *Agronomy*. 2022;12(7):1540.
53. Mangewa LJ, Ndakidemi PA, Alward RD, Kija HK, Bukombe JK, Nasolwa ER, Munishi LK. Comparative assessment of UAV and sentinel-2 NDVI and GNDVI for preliminary diagnosis of habitat conditions in Burunge wildlife management area, Tanzania. *Earth*. 2022;3(3):769–87.
54. de Melo MVN, de Oliveira MEG, de Almeida GLP, Gomes NF, Morales KRM, Santana TC, Silva PC, Moraes AS, Pandorfí H, da Silva MV. Spatiotemporal characterization of land cover and degradation in the agreste region of Pernambuco, Brazil, using cloud geoprocessing on google earth engine. *Remote Sens Appl Soc Environ*. 2022;26:100756.
55. Lamaamri M, Lghabi N, Ghazi A, El Harchaoui N, Adnan MSG, Shakiul Islam M. Evaluation of desertification in the middle Moulouya basin (north-east morocco) using sentinel-2 images and spectral index techniques. *Earth Syst Environ*. 2022;1:1–20.
56. Li Q, Zhang C, Shen Y, Jia W, Li J. Quantitative assessment of the relative roles of climate change and human activities in desertification

- processes on the Qinghai-Tibet Plateau based on net primary productivity. *CATENA*. 2016;147:789–96.
57. Nadjla B, Assia S, Ahmed Z. Contribution of spectral indices of chlorophyll (RECI and GCI) in the analysis of multi-temporal mutations of cultivated land in the Mostaganem plateau. In: 2022 7th international conference on image and signal processing and their applications (ISPA); 2022. IEEE. p. 1–6.
 58. Jiang F, Sun H, Ma K, Fu L, Tang J. Improving aboveground biomass estimation of natural forests on the Tibetan Plateau using spaceborne LiDAR and machine learning algorithms. *Ecol Ind*. 2022;143:109365.
 59. Della-Silva JL, da Silva Junior CA, Lima M, da Silva RR, Shiratsuchi LS, Rossi FS, Teodoro LPR, Teodoro PE. Amazonian species evaluation using leaf-based spectroscopy data and dimensionality reduction approaches. *Remote Sens Appl Soc Environ*. 2022;26:100742.
 60. Gerardo R, de Lima IP. Monitoring duckweeds (*Lemna minor*) in small rivers using sentinel-2 satellite imagery: application of vegetation and water indices to the Lis River (Portugal). *Water*. 2022;14(15):2284.
 61. Motohka T, Nasahara KN, Oguma H, Tsuchida S. Applicability of green-red vegetation index for remote sensing of vegetation phenology. *Remote Sens*. 2010;2(10):2369–87.
 62. Bendig J, Yu K, Aasen H, Bolten A, Bennertz S, Broscheit J, Gnyp ML, Bareth G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int J Appl Earth Obs Geoinf*. 2015;39:79–87.
 63. Wang N, Guo Y, Wei X, Zhou M, Wang H, Bai Y. UAV-based remote sensing using visible and multispectral indices for the estimation of vegetation cover in an oasis of a desert. *Ecol Ind*. 2022;141:109155.
 64. Ding J, Li Z, Zhang H, Zhang P, Cao X, Feng Y. Quantifying the aboveground biomass (AGB) of Gobi Desert Shrub communities in Northwestern China based on unmanned aerial vehicle (UAV) RGB images. *Land*. 2022;11(4):543.
 65. Nasiri V, Darvishsefat AA, Arefi H, Griess VC, Sadeghi SMM, Borz SA. Modeling forest canopy cover: a synergistic use of Sentinel-2, aerial photogrammetry data, and machine learning. *Remote Sensing*. 2022;14(6):1453.
 66. Steele MR, Gitelson AA, Rundquist DC, Merzlyak MN. Nondestructive estimation of anthocyanin content in grapevine leaves. *Am J Enol Vitic*. 2009;60(1):87–92.
 67. Hati JP, Chaube NR, Hazra S, Goswami S, Pramanick N, Samanta S, Chanda A, Mitra D, Mukhopadhyay A. Mangrove monitoring in Lothian Island using airborne hyperspectral AVIRIS-NG data. *Adv Space Res*. 2022;1:1.
 68. Silva GD, Roberts DA, McFadden JP, King JY. Shifts in salt marsh vegetation landcover after debris flow deposition. *Remote Sens*. 2022;14(12):2819.
 69. Geng X, Wang X, Fang H, Ye J, Han L, Gong Y, Cai D. Vegetation coverage of desert ecosystems in the Qinghai-Tibet Plateau is underestimated. *Ecol Ind*. 2022;137:108780.
 70. Myneni RB, Hall FG, Sellers PJ, Marshak AL. The interpretation of spectral vegetation indexes. *IEEE Trans Geosci Remote Sens*. 1995;33(2):481–6.
 71. Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices. *Remote Sens Environ*. 1996;55(2):95–107.
 72. Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens Environ*. 1996;58(3):289–98.
 73. Huete AR. A soil-adjusted vegetation index (SAVI). *Remote Sens Environ*. 1988;25(3):295–309.
 74. Qi J, Chehbouni A, Huete AR, Kerr YH, Sorooshian S. A modified soil adjusted vegetation index. *Remote Sens Environ*. 1994;48(2):119–26.
 75. Gitelson AA, Gritz Y, Merzlyak MN. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J Plant Physiol*. 2003;160(3):271–82.
 76. Pu R, Gong P, Yu Q. Comparative analysis of EO-1 ALI and Hyperion, and Landsat ETM+ data for mapping forest crown closure and leaf area index. *Sensors*. 2008;8(6):3744–66.
 77. Sripada RP, Heiniger RW, White JG, Meijer AD. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agron J*. 2006;98(4):968–77.
 78. Bareth G, Bolten A, Gnyp M, Reusch S, Jasper J. Comparison of uncalibrated RGBVI with spectrometer-based NDVI derived from UAV sensing systems on field scale. *Int Arch Photogr Remote Sens Spatial Inform Sci*. 2016;41:837–43.
 79. Barnes E, Clarke T, Richards S, Colaizzi P, Haberland J, Kostrzewski M, Waller P, Choi C, Riley E, Thompson T. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In: Proceedings of the fifth international conference on precision agriculture, Bloomington, USA; 2000.
 80. van den Berg AK, Perkins TD. Nondestructive estimation of anthocyanin content in autumn sugar maple leaves. *HortScience*. 2005;40(3):685–6.
 81. Gitelson AA, Keydan GP, Merzlyak MN. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophys Res Lett*. 2006;33(11):L11402.
 82. Xiaoqin W, Miaomiao W, Shaoqiang W, Yundong W. Extraction of vegetation information from visible unmanned aerial vehicle images. *Trans Chin Soc Agricul Eng*. 2015;31(5):1.
 83. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, Lin C-C, Meyer MD. Package 'e1071'. *R J*. 2019;1:1.
 84. Wickham H, Wickham MH. Package tidyverse. Easily Install Load Tidyverse; 2017.
 85. Heermann PD, Khazenie N. Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Trans Geosci Remote Sens*. 1992;30(1):81–8.
 86. RColorBrewer S, Liaw MA. Package 'randomforest'. Berkeley: University of California, Berkeley; 2018.
 87. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Team RC. Package 'caret'. *R J*. 2020;223:7.
 88. Candel A, Parmar V, LeDell E, Arora A. Deep learning with H2O. H2O ai Inc; 2016. p. 1–21.
 89. Visa S, Ramsay B, Ralescu AL, Van Der Knaap E. Confusion matrix-based feature selection. *MAICS*. 2011;710(1):120–7.
 90. Zhang W, Liu H, Wu W, Zhan L, Wei J. Mapping rice paddy based on machine learning with Sentinel-2 multi-temporal data: model comparison and transferability. *Remote Sens*. 2020;12(10):1620.
 91. Li Y, Al-Sarayreh M, Irie K, Hackell D, Bourdot G, Reis MM, Ghamkhar K. Identification of weeds based on hyperspectral imaging and machine learning. *Front Plant Sci*. 2021;11:2324.
 92. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl*. 2019;134:93–101.
 93. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform*. 2019;20(2):492–503.
 94. Aiello S, Kraljevic T, Maj P. Package 'h2o'. *Dim*. 2015;2:12.
 95. Fu B, Liu M, He H, Fan D, Liu L, Huang L, Gao E. Comparison of multi-class and fusion of single-class SegNet model for classifying karst wetland vegetation using UAV images; 2021.
 96. Rasmussen J, Ntakos G, Nielsen J, Svendsgaard J, Poulsen RN, Christensen S. Are vegetation indices derived from consumer-grade cameras mounted on UAVs sufficiently reliable for assessing experimental plots? *Eur J Agron*. 2016;74:75–92.
 97. Mohamad N, Ahmad A, Khanan MFA, Din AHM. Surface elevation changes estimation underneath mangrove canopy using SNERL filtering algorithm and DoD technique on UAV-derived DSM data. *ISPRS Int J Geo Inf*. 2021;11(1):32.
 98. Larrinaga AR, Brotons L. Greenness indices from a low-cost UAV imagery as tools for monitoring post-fire forest recovery. *Drones*. 2019;3(1):6.
 99. Reichmuth A, Henning L, Pinnel N, Bachmann M, Rogge D. Early detection of vitality changes of multi-temporal Norway spruce laboratory needle measurements—the ring-barking experiment. *Remote Sens*. 2018;10(1):57.
 100. Zhang X, Zhang F, Qi Y, Deng L, Wang X, Yang S. New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV). *Int J Appl Earth Obs Geoinf*. 2019;78:215–26.
 101. Huete AR, Liu H, van Leeuwen WJ. The use of vegetation indices in forested regions: issues of linearity and saturation. In: IGARSS'97 1997 IEEE international geoscience and remote sensing symposium proceedings remote sensing—a scientific vision for sustainable development; 1997. IEEE. p. 1966–8.

102. Fern RR, Foxley EA, Bruno A, Morrison ML. Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland. *Ecol Ind.* 2018;94:16–21.
103. Li F, Bai J, Zhang M, Zhang R. Yield estimation of high-density cotton fields using low-altitude UAV imaging and deep learning. *Plant Methods.* 2022;18(1):1–11.
104. Castelveccchi D. Can we open the black box of AI? *Nature News.* 2016;538(7623):20.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

