



OPEN

Machine learning algorithm for estimating karst rocky desertification in a peak-cluster depression basin in southwest Guangxi, China

Yali Zhang¹, Yichao Tian^{1,2✉}, Ying Li³, Donghua Wang⁴, Jin Tao¹, Yongwei Yang¹, Junliang Lin¹, Qiang Zhang¹ & Luhua Wu⁵

Karst rocky desertification (KRD) has become one of the most serious ecological and environmental problems in karst areas. At present, mapping KRD with a high accuracy and on a large scale is still a difficult problem in the control of KRD. In this study, a random forest (RF) based on maximum information coefficient and correlation coefficient feature selection is proposed to predict KRD. Nine predictors stood out as feature factors to estimate KRD. Rock exposure was the most important predictor, followed by fractional vegetation cover for the prediction of KRD processes. The kappa and classification accuracy indexes were to evaluate the performance of the model. We recorded overall accuracy rate and kappa index values of 94.7% and 0.92 for the testing datasets respectively. The RF model was then used to predict the KRD in 2001, 2011, 2016, and 2020, and it was found that the KRD in the study area has exhibited a positive trend of improvement. Therefore, the use of multisource remote sensing data combined with the RF model can obtain better prediction results of KRD, thereby providing a new idea for large-scale estimation of the KRD in peak-cluster depression.

Karst rocky desertification, similar to desertification, is a landscape characterized by a large area of exposed bedrock due to vegetation destruction and soil erosion in karst areas^{1,2}. It is not only a process of land degradation but also the result of land degradation. Typically, karst rocky desertification areas are characterized by thin surface soil, weak land productivity, and a poor anti-disturbance ability. Its ecological and environmental problems have been successively identified³, including an extremely fragile ecological environment, loss of biodiversity⁴, and ecosystem degradation^{5,6}. The ecological environmental security problems caused by karst rocky desertification have seriously affected people's living environment and sustainable development⁷, and thus, KRD has drawn intensive interest in the field of global environmental change⁸.

Remote sensing technology has been widely used in military, agricultural, medical, and geographical mapping research due to its advantages of fast acquisition, high resolution, low cost, and good security. With the advancement of remote sensing techniques, karst rocky desertification assessment based on remote sensing data has also been rapidly developed⁹. This is exemplified in the work undertaken by Liu et al.¹⁰, which involved using a multispectral remote sensing Landsat 8 image to calculate the brightness temperature and determining the degree of karst rocky desertification by setting the brightness temperature threshold in Pingguo County, Guangxi. Another study was conducted by Zhang et al.¹¹, in which based on a hyperspectral Hyperion image, the abundances of vegetation and exposed rock were extracted to monitor and evaluate the karst rocky desertification using the retrieved annual vegetation coverage from medium-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) data. In a similar case, Zhang et al.¹² assessed karst rocky desertification (KRD) in southwestern China. The information on karst rocky desertification was extracted using high spatial resolution GF-1 wide field of view (WFV) satellite data for the Nandong underground river basin. Overall, the remote

¹School of Resources and Environment, Beibu Gulf University, Qinzhou 535011, China. ²Key Laboratory of Marine Geographic Information Resources Development and Utilization in the Beibu Gulf, Beibu Gulf University, Qinzhou 535011, China. ³College of International Studies, Beibu Gulf University, Qinzhou 535011, China. ⁴College of Environmental Science and Engineering, Guilin University of Technology, Guilin 541004, China. ⁵School of Economics and Management, Tongren University, Tongren 554300, China. ✉email: tianyichao1314@hotmail.com

sensing data used to monitor the degree of karst rocky desertification have transitioned from multispectral¹³ to hyperspectral images^{14,15}, and low resolution data have been replaced by high resolution data^{16,17}.

Based on remote sensing data, considerable research has been conducted on karst rocky desertification monitoring techniques and methods. Traditionally, satellite imagery for karst rocky desertification mapping has relied on visual interpretation or human–computer interactive interpretation¹⁸. This is evident in the case of Huang and Cai¹⁹, in which the human–computer interaction interpretation method was used to interpret remote sensing images acquired in 1974, 1993, and 2001, and then, the spatial pattern and changes in the karst rocky desertification in the middle of Guizhou Province were analyzed. Similarly, based on Thematic Mapper (TM) remote sensing data, Wang et al.²⁰ mapped the karst rocky desertification in northern Guangdong through visual interpretation, and the accuracy of their karst rocky desertification map interpretation reached 93.6%. Although human visual interpretation can more accurately classify karst rocky desertification from remote sensing images, it demands considerable professional knowledge and is always time-consuming, which strongly hinders its efficiency. Therefore, it can only be used for the assessment of the degree of karst rocky desertification on a small scale, making it difficult to conduct such research over a large area⁴. To overcome these limitations, scholars have used unsupervised classification, spectral hybrid analysis, and machine learning algorithms to extract karst rocky desertification information. For example, Li and Wu¹³ used the decision tree and fuzzy maximum likelihood methods to classify KRD. When the vegetation fraction, bedrock exposure, and slope factor were added to the classifier, the classification accuracy improved from 84.23 to 91.71%. Chen et al.²¹ utilized the Classification And Regression Tree (CART) method, which increased the unsupervised classification, and normalized difference vegetation index (NDVI) data participation decision classification to classify the KRD, which effectively avoided the problem of artificial subjectivity. After this, based on Advanced Land Observing Satellite (ALOS) imagery, Qi et al.²² assessed the feasibility of using the dimidiate pixel model (DPM) and spectral mixture analysis (SMA) approaches for KRD monitoring. Alternatively, a combination of spectral analysis and the vegetation index can be used to extract karst rocky desertification information with a high accuracy, but this method has difficulty identifying the degree of karst rocky desertification in similar shaded areas²³. Supervised classification and spectral hybrid analysis can speed up the annotation process. However, high-quality expert-annotated samples are still a prerequisite for achieving accurate results using intelligent approaches. Unsupervised classification, to a certain extent, avoids the subjectivity of the artificial selection of samples, but has difficulty distinguishing between types of ground objects with small spectral characteristic differences. In the background of the big data era, machine learning methods such as support vector machines, random forest models, and neural networks have been extensively used in the fields of hydrology, meteorology, and ecology, and they have also provided a new direction in the extraction of karst rocky desertification information²⁴. Machine learning combined with the factors influencing karst rocky desertification can not only overcome human subjectivity, but also efficiently identify the degree of karst rocky desertification in large areas⁴.

A recent case study reported by Zhang et al.²⁵ argues that the optimal factor influencing karst rocky desertification is of great significance to the evaluation of the degree of karst rocky desertification in karst areas. The vegetation coverage, rock exposure rate, and slope are usually used as grading factors for the degree of karst rocky desertification^{26,27}. However, the factors influencing karst rocky desertification are complex and diverse. In support of this, Huang et al.⁸ utilized artificial neural networks (ANNs) to identify the importance of different environmental factors to karst rocky desertification. Zhang et al.²⁸ used correlation analysis to study the relationships between karst rocky desertification, temperature, and rainfall and pointed out that karst rocky desertification is not sensitive to responses to climate change and there is a certain lag. Bai et al.²⁹ explored the influence of lithology on karst rocky desertification using a combination of mathematical modeling and spatial analysis. Zhang et al.²⁵ concluded that the karst rocky desertification index (KRDI) is a good indicator of karst desertification, and the higher the KRDI value, the higher the degree of desertification. In addition, the fragile ecological environment and unreasonable human interactions have promoted the aggravation of karst rocky desertification. For example, Li and Xiong³⁰ qualitatively analyzed the impact of human activities on the degree of karst rocky desertification and pointed out that traditional agricultural activities have less impact on karst rocky desertification, while sudden short-term destructive economic activities are the humanistic motivations for initiating large-scale karst rocky desertification. Yao et al.³¹ studied the relationships between the degree of karst rocky desertification and the gross domestic product (GDP) and population density using superposition analysis and concluded that areas with higher population densities and higher GDPs also had higher degrees of karst rocky desertification. Shi et al.³² used night light remote sensing data to verify the impact of human activities on karst rocky desertification, and their study showed that the total night light (TL) associated with severe karst rocky desertification was concentrated in Guizhou and Yunnan. In addition, a large and growing body of literature has investigated the temporal and spatial distributions and change characteristics of karst rocky desertification, and these studies have achieved some results³³. However, the research areas have been concentrated in Guizhou and Yunnan, and the research scale has usually been the county scale^{34,35}. Few studies have been conducted on the karst rocky desertification in peak-cluster depression basins³⁶.

The karst peak-cluster depression basin in southwest Guangxi is a typical area where tropical karst and non-karst landforms intersect in the world, and it is also one of the hotspots of global biodiversity and ecosystem services. In addition, strong climate change, geological movements, and unreasonable human activities have caused karst rocky desertification to become the most serious environmental problem in this area, threatening the ecological security and economic and social development in the karst region in the peak-cluster depression basin in southwest Guangxi.

In response to these issues, in this study, a typical karst peak-cluster depression watershed in southwest Guangxi was selected as the study area. The main objectives of this study were as follows: (1) to analyze the relevant factors that may affect the development and evolution of karst rocky desertification; (2) to identify the optimal karst rocky desertification characteristic factors; and (3) to invert the spatial and temporal patterns of

	RE	FVC	S	LAI	FPAR	ET	P	LST	DEM	SP	ST	POP	L	DI
MIC	0.89	0.58	0.36	0.54	0.62	0.54	0.45	0.29	0.27	0.15	0.05	0.23	0.05	0.33
R	0.87	-0.77	0.48	-0.69	-0.76	-0.66	-0.19	0.42	0.41	-0.08	-0.05	0.24	0.03	0.19

Table 1. MIC and correlation coefficients on the factors influencing of KR D. *RE* rock exposure, *FVC* fractional vegetation cover, *S* slope, *LAI* leaf area index, *FPAR* fraction of photosynthetically active radiation, *ET* evapotranspiration, *LST* land surface temperature, *DEM* elevation, *P* precipitation, *SP* slope aspect, *ST* soil type, *POP* population, *L* lithology, *DI* drought index.

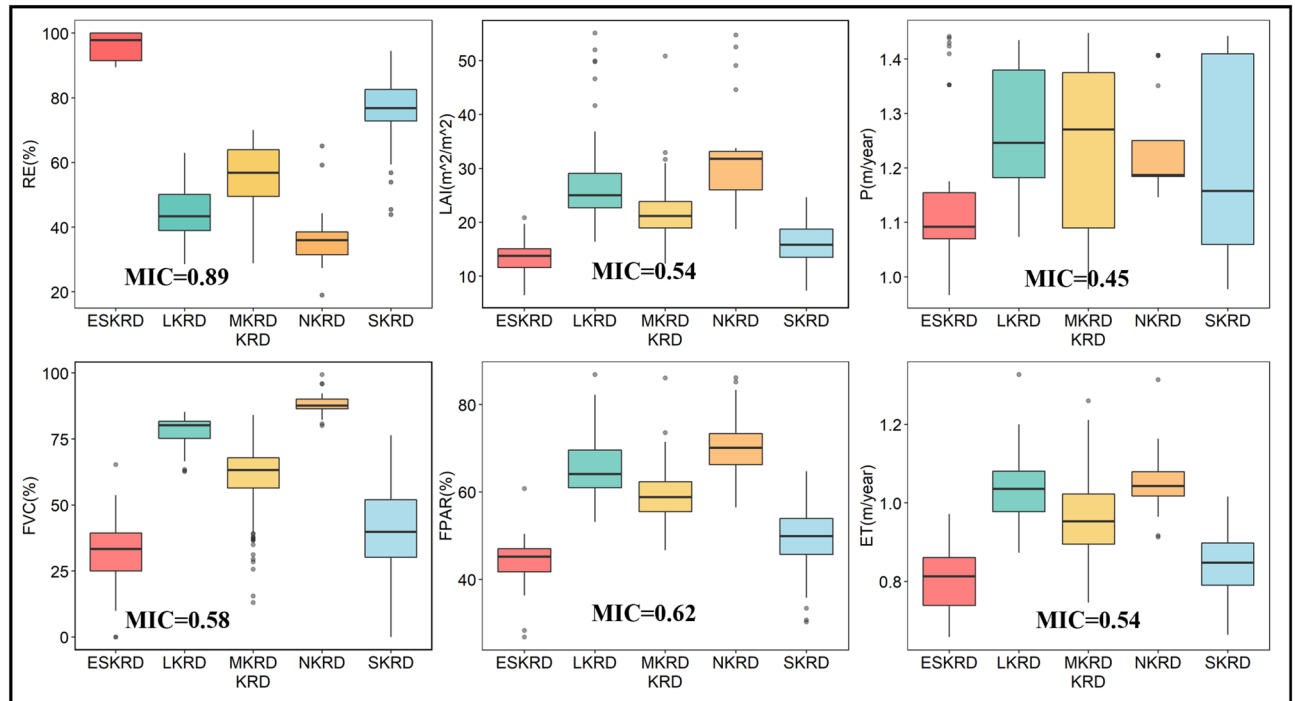


Figure 1. Feature factors of KR D in 2020 selected based on MIC. Abbreviations: NKRD—No karst rocky desertification. LKRD—Light karst rocky desertification. MKRD—Medium karst rocky desertification. SKRD—Severe karst rocky desertification. ESKRD—Extremely severe rocky desertification.

karst rocky desertification from 2001 to 2020. The results of this study not only provide ideas for karst rocky desertification monitoring in peak-cluster depressions but also provide a data reference for government decision-makers and environmental managers to make macroscopic decisions.

Results

Feature factors. The R and MIC scores were used as measurements of linear and nonlinear correlations, and the results shown in Table 1.

The MIC, which is a nonlinear variable discovery method, revealed that 6 features of the 14 variables were relevant to the KR D in the peak-cluster depression basin in southwest Guangxi (Fig. 1). As is shown in Fig. 1, these six features were the RE, FVC, LAI, FPAR, ET, and P (MIC > 0.4). Among them, the RE and FVC had the strongest impact on the karst rocky desertification.

As can be seen from Table 1, the RE, FVC, S, FPAR, ET, LAI, DEM and LST factors exhibited strong correlations with the degree of karst rocky desertification (Fig. 2), while the correlations between the drought index, lithology, soil type, population density, and slope direction and the degree of karst rocky desertification were low and only revealed the correlation degree between each factor and the karst rocky desertification linearly. Therefore, the RE, FVC, S, LST, P, ET, LAI, DEM and FPAR were selected as the feature factors for inverting the spatial distribution of the karst rocky desertification. Figure 3 shows the selected feature factors via MIC values and correlation coefficients.

Model accuracy and variable importance. A summary of the parameters characterizing the accuracy of the RF models is presented in Table 2, which indicates that the overall accuracy rate of the random forest model is 94.7% and the kappa coefficient is 0.92. So, the random forest model is more reliable in mapping the karst rocky desertification. The importance of the input features obtained from the RF model can be used to measure

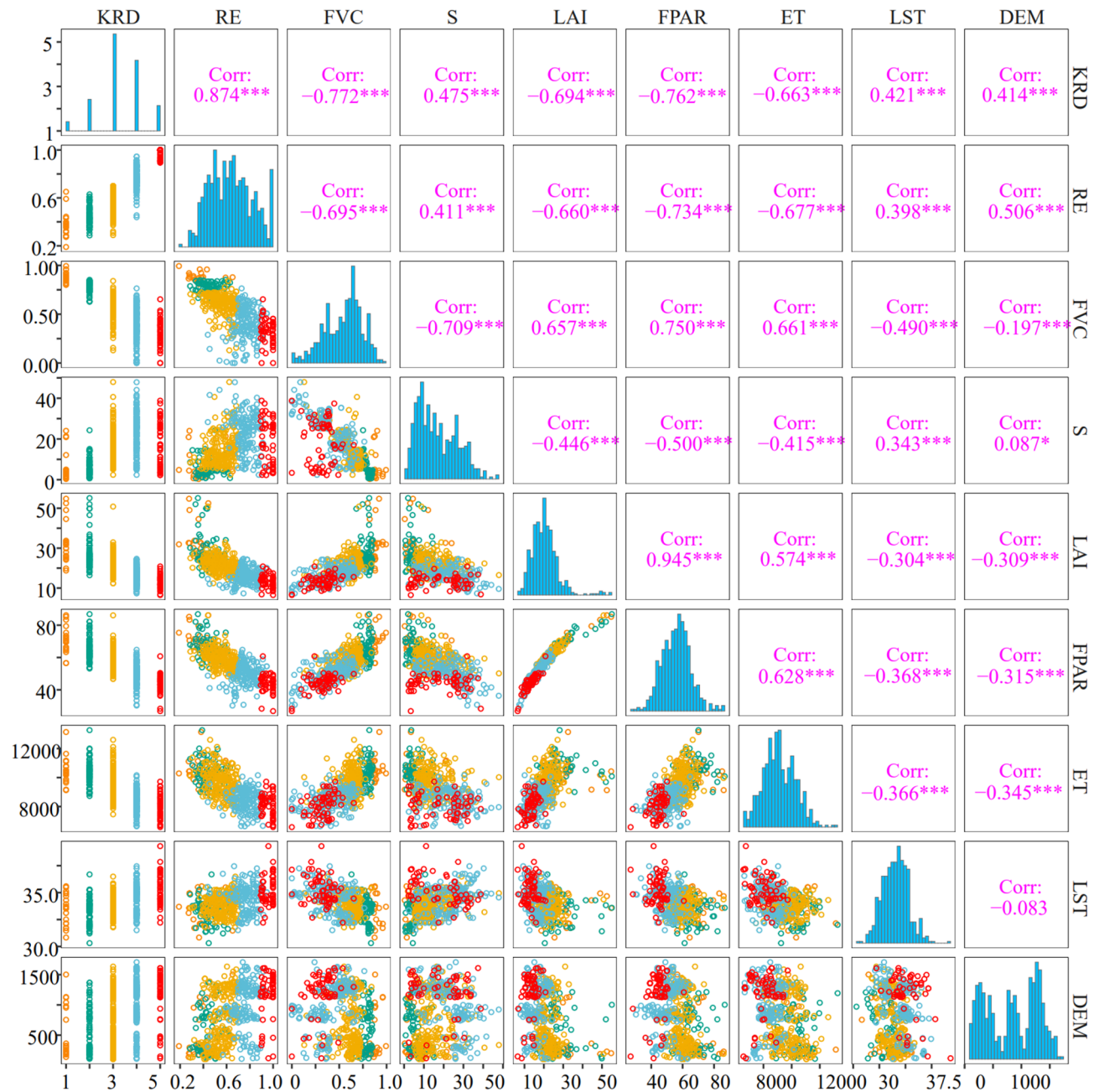


Figure 2. Feature factors of KRD selected based on correlation coefficient. *Significant correlation at the 0.05 level (both sides); **significant correlation at the 0.01 level (both sides); ***significant correlation at the 0.001 level (both sides).

their contributions to the classification accuracy (Fig. 4). Specifically, the variable importance scores are as follows: RE > FVC > FPAR > S > LAI > LST > P > ET > DEM. Notably, as the predominant features, RE and FVC score about twice compared with the following feature FPAR.

Spatial and temporal distributions of karst rocky desertification. Figure 5 shows the distribution of the karst rocky desertification in the peak-cluster depression basin in southwest Guangxi. The evolution of the study area was characterized by KRD in space, indicating that the KRD generally improved, and the areas of SKRD and ESKRD continuously decreased. Specifically, in the early part of the study period (from 2001 to 2006), the SKRD and ESKRD accounted for large areas. The macroscopic pattern of the spatial distribution of the karst rocky desertification shows that the upper reaches of the basin were dominated by SKRD and ESKRD, the central region was dominated by MKRD, and the lower region was characterized by LKRD and MKRD. However, from 2011 to 2020, the degree of karst rocky desertification was contained, and the severe and extremely severe karst rocky desertification were scattered in small areas. The macroscopic pattern of the spatial distribution of

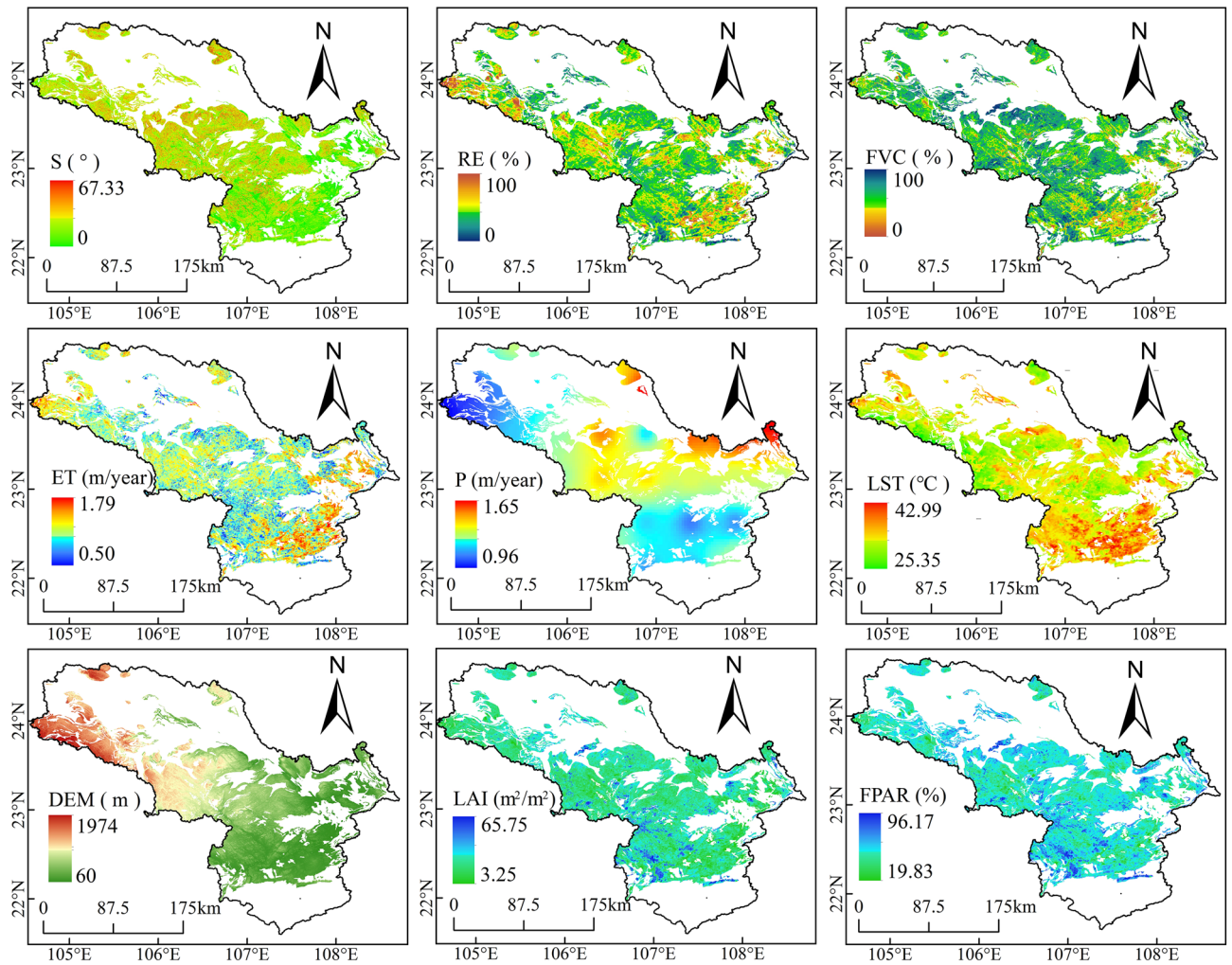


Figure 3. Feature factors of KRD in 2020 selected based on MIC and R in the peak-cluster depression basin in southwest Guangxi, China. Maps were generated using QGIS 3.26.2 (<https://www.qgis.org/en/site/>).

Overall accuracy (OA)	Kappa	User's accuracy (UA)					Producer's accuracy (PA)					Run time
		NKRD	LKRD	MKRD	SKRD	ESKRD	NKRD	LKRD	MKRD	SKRD	ESKRD	
94.7%	0.92	100%	88.9%	96.8%	92.7%	92.9%	100%	80%	93.8%	97.4%	100%	4.5 s

Table 2. Accuracy of the RF, including the kappa, OA, UA, and PA.

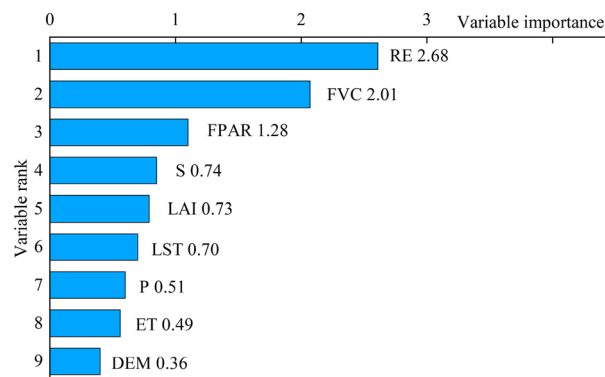


Figure 4. The importance of feature factors of KRD in the peak-cluster depression basin in southwest Guangxi, China.

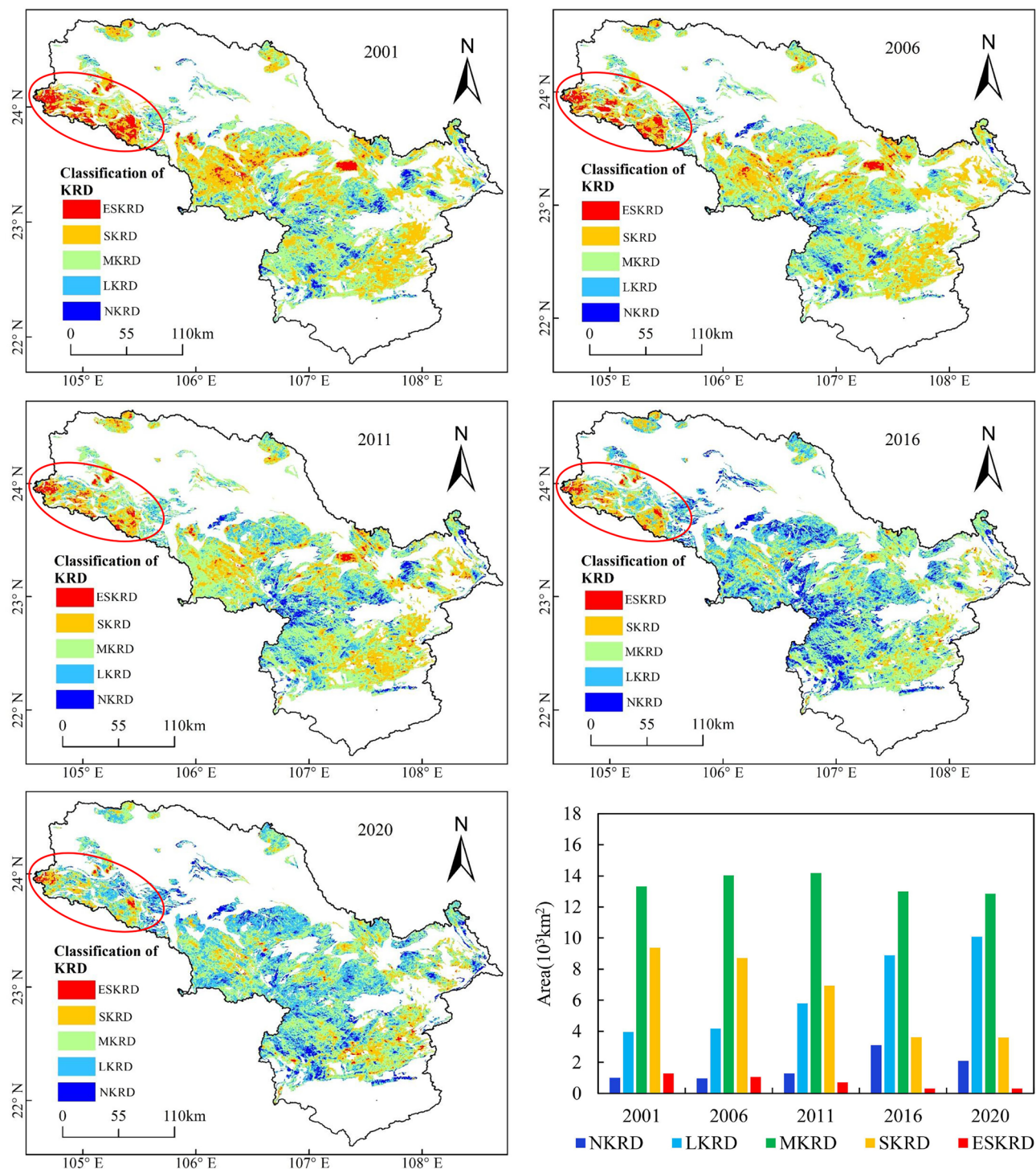


Figure 5. Spatial distribution of the karst rocky desertification. Maps were generated using QGIS 3.26.2 (<https://www.qgis.org/en/site/>).

the karst rocky desertification shows that only a few SKRD and ESKRD areas were distributed in Guangan County in the upper reaches of the basin.

In general, since 2010, there has been an obvious decrease in the total area of karst rocky desertification in the peak-cluster depression basin in southwest Guangxi, indicating that the karst rocky desertification problem generally exhibits a reverse trend, primarily manifested as a decrease in the level of karst rocky desertification. However, there are still a large number of moderate rocky desertification. In view of this, we suggest: (1) The primary task of karst rocky desertification control should follow the laws of nature, increase protection, reduce human disturbance, and strengthen the protection of potential rocky desertification land. For example, relevant government departments appropriately promote the pace of local urbanization, increase the intensity

Year	LKRD		MKRD		SKRD		ESKRD		Total karst area (10 ³ km ²)
	Area (10 ³ km ²)	Proportion (%)	Area (10 ³ km ²)	Proportion (%)	Area (10 ³ km ²)	Proportion (%)	Area (10 ³ km ²)	Proportion (%)	
2001	3.95	13.65	13.31	46.03	9.37	32.39	1.29	4.47	27.92
2006	4.17	14.39	14.04	48.55	8.71	30.11	1.05	3.63	27.96
2011	5.79	20.03	14.19	49.06	6.95	24.02	0.70	2.43	27.63
2016	8.89	30.73	13.00	44.94	3.61	12.48	0.31	1.08	25.81
2020	10.10	34.91	12.85	44.42	3.58	12.39	0.30	1.05	26.83

Table 3. Change in the karst rocky desertification areas of different levels.

of ecological migration and rural population transfer, and effectively alleviate the population pressure in karst areas. (2) The ideal policies for the prevention and treatment of the rocky desertification peak-cluster depression basin in southwest Guangxi should properly handle the relationship between economic development and ecological protection so that local residents can gradually change their dependence on their original farming livelihoods. (3) Different methods should be effectively utilised according to the specific karst environments. Ecological restoration work should be carried out scientifically and rationally, particularly in areas with severe rocky desertification.

As can be seen from Table 3, the total area of KRD changed from 27,920 km² to 26,830 km² in the 20 years from 2001 to 2020; and the net area changed to 1090 km² with a reduction rate of 54.5 km² a⁻¹. The MKRD, SKRD, and ESKRD areas decreased from 2001 to 2020 in the peak-cluster depression basin in southwest Guangxi. In particular, the proportions of the SKRD and ESKRD areas decreased from 32.39%, and 4.47% to 12.39%, and 1.05%, respectively. Generally, the comprehensive control effect of the karst rocky desertification was remarkable, and the overall karst rocky desertification exhibited a trend of improvement, but LKRD and MKRD were still widely distributed.

In the period from 2001 to 2020, the rocky desertification levels in the study area tended to decrease. The reason for the change of karst rocky desertification is mainly due to the rapid development of urbanization as well as social economy, and a large number of rural laborers have shifted from traditional agriculture to other industries, which has slowed down the pressure on land³⁷. The reduction in farmland area improved farmland management and increased regional gross industrial product, which together with the continuously rising gross domestic product of the tertiary industry caused a positive rocky desertification development³⁸. In addition, the Program of Conversion from Cropland to Forest and Grassland had been applied to restore the vegetation ecosystem since 2005, which has contributed to the gradual development of the ecological and environmental conditions of karst rocky desertification in a benign direction³⁹.

Discussion

The occurrence of karst rocky desertification is a dynamic evolutionary process in time and space, and it is the result of the joint influences of the natural environment and human activities^{40–42}. The formation factors of karst rocky desertification are complex and diverse, and an inversion model of karst rocky desertification needs to combine reasonable karst rocky desertification feature factors to produce reliable results. Feature extraction or generation is a critical step in the recognition process since the designated attributes strongly influence the recognition results⁴³. Too many of the variables available may introduce noise or may not provide information to identify KRD⁴⁴. When the feature factors used in the machine learning model are insufficient, the model will be under-fitted, which will lead to a certain deviation in the predicted results. However, adopting too many features will increase the search space of the model and the run time of the model, and concurrently, the corresponding model's construction process will be more complicated⁴⁵. In addition, irrelevant factors will interfere with the model. Therefore, it is necessary to strategically identify the variables related to KRD, which can produce the best effect from the karst rocky desertification prediction model. Related studies have pointed out that the rock exposure rate and vegetation cover contribute the most to karst rocky desertification extraction¹⁷, which is consistent with the findings of this study.

As is shown in Table 1, the MIC and correlation coefficient between the KRD and RE are 0.87 and 0.89, respectively. In addition, the results of the variable importance assessment provided in Fig. 4 indicate that the RE and FVC were the two most important variables affecting karst rocky desertification in the peak-cluster depression in southwest Guangxi. This was also proved by Xi et al.⁴⁶ and Gu et al.³¹. From these two studies, it is evident that the NDVI had the best correlation with the karst rocky desertification, and the RE had a significant positive correlation with the occurrence intensity of the karst rocky desertification. Specially, based on geographical detector technology, Xi et al. obtained the contribution rates of the RE and FVC to karst rocky desertification, which were 44% and 42%, respectively. Gu et al. measured the importance of various factors to karst rocky desertification using the partial least squares regression model (PLS). However, this study attempted to input only the rock exposure rate and the vegetation coverage into the machine learning model, and the classification effect of the model was not good.

Moreover, data from several studies suggest that the leaf area index, LST, slope, rainfall, and ET can also be used as important factors and indicators for analyzing and evaluating the degree of rock desertification. For example, using Landsat 8 data, Deng et al.⁴⁷ summarized the spatial distribution of the LST in karst areas, revealing that land surface temperature can describe the characteristics of karst rocky desertification in karst areas to a certain extent. Likewise, Li et al.⁴⁸ found that biophysical parameters such as the surface vegetation coverage and

leaf area index can better reflect the distribution of karst rocky desertification. Overall, these results demonstrate that the LST and LAI can better reflect the distribution of karst rocky desertification, which is consistent with the results of this study. As is shown in detail in Table 1, the LAI and FPAR also exhibit good correlations with the degree of karst rocky desertification. When all of the factors were incorporated into the machine learning model, the accuracy of the model classification did not improve (kappa coefficient was 0.86), but the run time of the model was longer. Thus, the scientific selection of the feature factors is particularly important when inverting karst rocky desertification. In this study, the feature variables selected using the MIC and correlation coefficient were added to the random forest model, which increased the overall accuracy of the model classification to 94.7% and the kappa coefficient to 0.92.

Traditional karst rocky desertification monitoring methods mainly rely on ground surveys, which require a great deal of time and money. Owing to the terrain limitations, they can only be carried out in areas with low altitude slopes and are not suitable for investigation in peak-cluster depression areas⁴⁹. Although visual interpretation of remote sensing images for monitoring rock desertification is not limited by the topography, it also has disadvantages, including a low interpretation efficiency, easily influenced by human subjectivity, and difficulty guaranteeing the accuracy. In this study, based on machine learning, the spatial distribution of the karst rocky desertification in a karst area was mapped using remote sensing data and auxiliary data. Using a traditional machine learning algorithm has certain advantages, but the selection of the feature vectors and the determination of the model parameters all have a certain impact on the accuracy of the prediction model⁵⁰. Recently, little research has been conducted on mapping karst rocky desertification information based on machine learning algorithms. Pu et al.¹⁷ compared the accuracy of three algorithms, i.e., the random forest (RF), bagged decision tree (BDT), and extreme random tree (ERT) algorithms, and determined that their overall accuracies (OAs) were 85.21%, 80.85%, and 78.93%, respectively. Xu et al.⁴ used a support vector machine model (SVM) to evaluate the karst rocky desertification areas in Liujiang, Changshun, and Zhenyuan. The overall accuracies in these areas were 85.50%, 84.00%, and 84.86%, respectively; and the kappa coefficients reached 0.8062, 0.7917, and 0.8083, respectively. Obviously, the differences in the research areas and the setting of the model parameters have a certain influence on the accuracy of the prediction model⁴⁵. Some scholars have proposed that among many machine learning algorithms, the RF algorithm has the advantages of simple training, high computational efficiency, and high stability in the changing of parameter values in a classification model⁵¹. Although RF algorithm was faster to train and more stable, The accuracy of the random forest model depends on the settings of the internal parameters⁵². So, in our study, an iterative backward feature elimination procedure was used to reduce the number of less relevant variables until the internal accuracy (calculated on the basis of the OOB error) no longer varies. Using this approach significantly increases classification accuracy⁵³. In addition, another limitation of the random forest model is that the accuracy of the model depends on the quality of the samples. Previous studies have reported that the sizes of the training samples sets were found to influence the performance of the RF classifier⁵⁴. So, in order to reduce misclassification, the sensitivity of RF classification to the sampling design also needs to be considered⁵³. According to the results of many experiments, the model has the highest accuracy when the training set is total of 75% and the test set is 25% in this study. In this study, the RF model, which was optimized via iteration, was used to map the KRD in the peak-cluster depression basin in southwest Guangxi based on machine learning. The overall accuracy of identifying the karst rocky desertification was 94.7%, and the kappa coefficient was 0.92. Therefore, this study provides an effective method of KRD monitoring.

Although quantitative analysis of the driving factors of karst rocky desertification development and evolution was conducted in this study, the in-depth relationships and causality between the different influencing factors need to be explored further. Therefore, in future research, it is necessary to comprehensively consider more factors affecting karst rocky desertification to better reveal the development and evolution mechanisms of karst rocky desertification. In addition, hyperspectral images with a high spectral resolution and rich texture features would be very suitable for the study of karst rocky desertification areas, and they have been widely used in other fields. Thus, a machine learning model based on hyperspectral data, combined with optimized algorithms, is a new way to extend remote sensing image information extraction techniques in karst rocky desertification areas.

Conclusions

Based on RF classifiers and using multisource remote sensing imagery, the spatio-temporal patterns of karst rocky desertification in the peak-cluster depression basin in southwest Guangxi, China, were monitored. The main conclusions are as follows:

In this study, six feature factors were identified using an MIC value of >0.4 as the selection standard, and the Pearson correlation method was used to filter the variable set. Concurrently, based on the results of these two filters, nine factors (RE, FVC, SLOPE, LAI, FPAR, ET, P, DEM and LST) were selected as the optimal factors for the inversion of the karst rocky desertification.

According to the nine feature factors, the RF algorithm was optimized via iteration. The optimized RF model was then used to predict the KRD in 2001, 2011, 2016, and 2020. The overall accuracy of the RF model was 94.7%, and the kappa coefficient was 0.92.

In general, the karst rocky desertification in the study area exhibited a positive trend of improvement. Specifically, both the area and the degree of karst rocky desertification decreased. Despite the remarkable effect of the comprehensive management of karst rocky desertification, areas of light and moderate karst rocky desertification are still widely distributed in the study area.

The RF method with the feature selection would be a better method for karst rocky desertification mapping compared with the common method. So, the accuracy of the optimal monitoring scheme in the peak-cluster depression basin in southwest Guangxi, China, could also be investigated for other regions. In addition, our

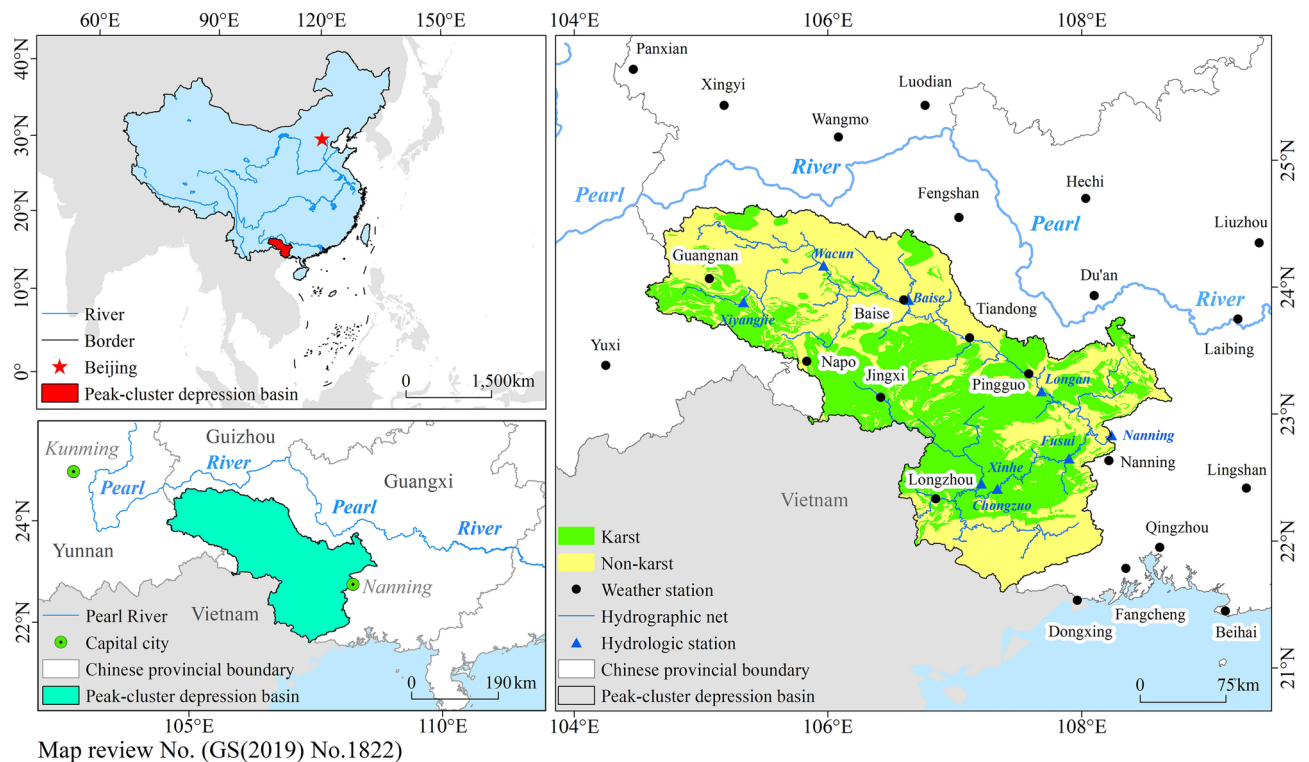


Figure 6. The location of the study area. Maps were generated using QGIS 3.26.2 (<https://www.qgis.org/en/site/>).

research provides technical support and data sources for the implementation of projects such as returning farmland to forests, soil and water conservation, and rocky desertification prevention and control.

Materials and methods

Study area. The peak-cluster depression basin in southwest Guangxi is located in the slope zone from the Guizhou Plateau to Guangxi Basin, in which karst landforms are widely developed. This region is characterized by a fragile ecosystem, large area, wide distribution, and diverse geomorphological types of carbonate rocks. As a typical peak-cluster depression karst area in southwestern China⁵⁵, the geographical position of the study area is $104^{\circ} 33' - 108^{\circ} 43' E$ and $21^{\circ} 35' - 24^{\circ} 39' N$, covering an estimated area of $61,485.16 \text{ km}^2$, and its altitude ranges from 500 to 1700 m. In addition, it possesses a typical tropical and subtropical humid and hot monsoon climate, with a mean annual temperature of $13 - 14^{\circ} C$ and a mean annual precipitation of $900 - 1600 \text{ mm}$ ⁵⁶.

Furthermore, this area is not only an important ecological barrier in the Pearl River Basin, but also an important water conservation area and biodiversity priority protection area in China, as well as one of the areas where ethnic minorities mostly live in the Guangxi Zhuang Autonomous Region³. In addition, as a border area, the bilateral political relationships and the situations in neighboring countries have a significant impact on the sustainable development of the ecological environment in the border area. Moreover, the study area is the most convenient sea and land route from China to Vietnam and even Association of Southeast Asian Nations (ASEAN) countries and is an important hub of the “One Belt, One Road” initiative (Fig. 6).

Data sources and preprocessing. Measured data, as well as a great deal of MODIS remote sensing data and auxiliary data, were employed in this study. Specifically, five MODIS products (MOD09A1, MOD13Q1, MOD11A2, MOD16A3, and MOD15A2) were used (Table 4), and the auxiliary data included elevation, slope, aspect, precipitation, lithology, soil type, drought index, and population data (Table 5).

MODIS data. MODIS data products are widely used in the fields of land use/cover research, natural disaster monitoring and analysis, and marine ecological environment, and they play an important role in ecological environment research and applications at global and regional scales. In this study, the MODIS datasets (tiles h27v06 and h27v062) for 2001, 2006, 2011, 2016, and 2020 were used to calculate the karst rocky desertification. They were all downloaded from the United States Geological Survey (<https://earthdata.nasa.gov/>). To begin this process, the MODIS Reprojection Tool (MRT) was used for the data extraction, mosaicking, and reprojection. Specifically, the fraction of photosynthetically active radiation (FPAR) and the leaf area index (LAI) products were acquired from MOD15A2. In the same way, the land surface temperature (LST) and evapotranspiration (ET) were retrieved from the MOD11A2. In the follow-up phase, of the seven bands of the MOD09A1 product, only bands 2 and 7 were used to generate the corresponding rock exposure rate (Table 6). After this, the fractional

Data Sources	Data type	Variables	Temporal resolution	Spatial resolution	Coverage
MODIS	MOD09A1	Surface reflectance B2, B7	8 days	500 m	Global
	MOD13Q1	NDVI	16 days	250 m	Global
	MOD11A2	LST	Monthly	1 km	Global
	MOD16A3	ET	Yearly	500 m	Global
	MOD15A2	LAI/ FPAR	8 days	500 m	Global

Table 4. Detailed specifications of the remote sensing datasets.

Data	Temporal Resolution	Spatial Resolution	Sources
Soil type (ST)		1 km	https://www.resdc.cn/
DEM		90 m	http://www.gscloud.cn/#page1/2
Slope (S)		90 m	http://www.gscloud.cn/#page1/2
Slope aspect (SP)		90 m	http://www.gscloud.cn/#page1/2
Drought index (DI)	Yearly	0.05	http://climexp.knmi.nl/selectfield_obs2.cgi?id=someone@somehere
Lithology (L)		Shp	http://www.karstdata.cn/
Population (POP)		1 km	https://www.worldpop.org/geodata/listing?id=64
Precipitation (P)	Monthly	1 km	http://data.cma.cn/

Table 5. Relevant parameters and sources of the auxiliary data.

Factor	Formula	References
RE	$NDRI = \frac{B7 - B2}{B7 + B2}$ $RE = \frac{NDRI - NDRI_r}{NDRI_r - NDRI_o}$	Zhu ⁵⁷
FVC	$FVC = \frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}}$	Mishra and Chaudhuri ⁵⁸

Table 6. FVC and RE were extracted using QGIS3.26.2 (<https://www.qgis.org/en/site/>). NDRI normalized difference rock index.

vegetation cover (FVC) was calculated using the NDVI data from MOD13Q1, which provides a 16-day composite with 250 m spatial resolution data, including NDVI products (Table 6). Finally, before further analysis, the above product datasets were uploaded to QGIS3.26.2 (<https://www.qgis.org/en/site/>), where they were projected to the World Geodetic System (WGS) 1984, Universal Transverse Mercator (UTM) zone 48 N projected coordinate system. They were further resampled to a spatial resolution of 250 m for uniformity. The boundary of the study area was used as a mask for cutting to ensure the same processing extent.

Auxiliary data processing. The sources and details of the auxiliary data are shown in Table 5. These data were processed using the following methods. First, using the QGIS3.26.2 (<https://www.qgis.org/en/site/>), the soil type, population, precipitation, digital elevation model (DEM), slope, and slope direction data were resampled to the same spatial resolution as the MODIS product data using a nearest neighbor algorithm and by replicating the pixels. Second, the drought index datasets, which were derived from the product data of the Climate Research Unit self-calibrated Palmer Drought Severity Index (CRUscPDSI) with a 1 km spatial resolution, were converted to a point layer. Then, the datasets were interpolated to raster files using the Kriging technique to be consistent with the spatial resolution of the other data.

In particular, the vector data of the lithology data were converted to a raster layer with a 250 m × 250 m spatial resolution for this study. In addition, for uniformity, all of the auxiliary data were further transformed to the World Geodetic System (WGS) 1984, Universal Transverse Mercator (UTM) zone 48 N projected coordinate system. Finally, the auxiliary data were extracted using the study area boundary as a mask to generate the rocky desertification impact factor data.

Karst rocky desertification survey data. The field data were collected during the spring and summer of 2020 (March–August). A total of 527 sampling plots, 30 × 30 m each¹⁷, were established. They were located randomly along the road so that they would be easy to reach. Within each plot, the longitude, latitude, and elevation of the sample centroids were recorded using a high accuracy global navigation satellite system (GNSS). Then, the area of bare rock and vegetation coverage were measured using a high-precision handheld global positioning system (GPS) measuring instrument. In addition, the vegetation type, landscape type, and surrounding environment were also recorded according to visual observations. Ultimately, the indexes of the vegetation coverage and the rock exposure rate were calculated to determine the classification of the karst rocky desertification (Fig. 7).

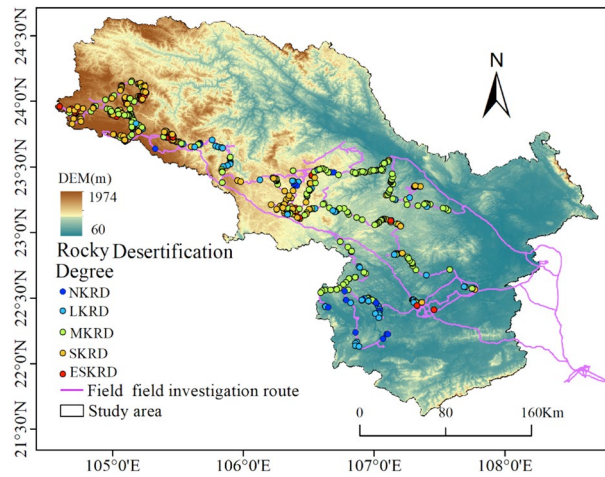


Figure 7. Field data survey map of karst rocky desertification. Map was generated using QGIS 3.26.2 (<https://www.qgis.org/en/site/>).

Classification of KRD	Exposure rate of bedrock (%)	Vegetation and soil coverage (%)	Distribution character of the exposed rock
NKR	< 30	70	None
LKR	30–50	50–70	Point
MKR	50–70	30–50	Point, line
SKR	70–90	20–30	Line, patch
ESKR	> 90	20	Patch

Table 7. Field classification standard for karst rocky desertification.

Based on the vegetation coverage, rock exposure rate, and rock distribution obtained from the survey, based on previous studies⁵⁹, and combined with the landforms in the study area, the karst areas in this study were classified into five types: non-rocky desertification, light karst rocky desertification, moderate karst rocky desertification, severe karst rocky desertification, and extreme karst rocky desertification. The classification standard is shown in Table 7.

Feature selection method. The maximal information coefficient (MIC), which was introduced by Reshef et al. in 2011⁶⁰, is a powerful approach for detecting various relationships between variables, and it was developed on based on mutual information. The larger the MIC value between the two variables is, the stronger the correlation is, and vice versa. And, the Pearson’s correlation coefficient, a statistical method that captures the dependence of two variable correlations, is frequently used for the decorrelation of variables and for feature extraction⁶¹.

Random forest classifier. The random forest (RF) classifier is a non-parametric ensemble classification method based on a large number of regression trees, especially with two important parameters as the number of decision trees and the number of split nodes⁶². The disadvantage of RF was that the split rules for classification are unknown⁶³. However, because of its high stability and its ability to perform efficient processing of large-scale data⁶⁴, the random forest classifier is a more practical integrated learning method, and it can effectively reduce the error of a single classifier and improve the classification accuracy using multiple classifiers for voting classification. Random forest algorithms have randomness in sample and feature selection, which makes it difficult for random forest to fall into overfitting and gives it a good antinoise ability^{65,66}. In this study, a bagging integrated random forest classification algorithm was used to predict the degree of karst rocky desertification.

The specific steps were as follows.

First, the degree of karst rocky desertification was defined as the dependent variable, and the vegetation coverage, the rock exposure, and other factors were used as explanatory variables to select the optimal number of leaf nodes by setting RFL_{leaf} = 5, 10, 20, ..., 500.

Second, ① the entire dataset was randomly split into calibration (70%) and validation (30%) datasets to model the KR. The calibration dataset was used to train the models with all of the relevant variables identified by the MIC. The independent validation sets were used to evaluate the predictive performance of the RF model. ② The kappa and classification accuracy indexes were calculated for the validation datasets. ③ If the kappa coefficient was greater than 0.95, the trained model was saved. Otherwise, steps ① and ② were repeated.

Finally, based on the karst rocky desertification characteristic data, the trained RF model was employed to monitor the karst rocky desertification dynamics during 1990–2020 in the peak-cluster depression basin in southwest Guangxi.

Accuracy assessment. The stability and reliability of the model algorithm are the basis of the subsequent research, so it is very important to measure the accuracy of the model. In this study, to test the effectiveness of the RF algorithm, performance measurement metrics, including the overall accuracy, users' accuracy, producers' accuracy, and kappa coefficient, were adopted.

Technical route and workflow. A flowchart of the entire process used in this study is shown in Fig. 8, which can be divided into three parts.

- (1) **Data preprocessing and the extraction of desertification indicators**
Based on the MODIS data and auxiliary data, the karst rocky desertification factor data were extracted, including the vegetation coverage, bedrock exposure rate, surface temperature, leaf area index, photosynthetic utilization efficiency, elevation, slope, slope direction, lithology, soil type, evapotranspiration, population density, and annual precipitation data. The optimum characteristics of the karst rocky desertification factors were selected via the MIC and Pearson's correlation coefficient.
- (2) **RF model setting**
The purpose of this step was to adjust the model parameters. The optimal numbers of trees and leaves for the RF classifier were determined by plotting the Out of Bag (OOB) error versus the number of trees and by determining the threshold number of trees for which the error was stable. The number of trees to be used in the RF classifier was chosen as 250, which is not too computationally expensive but is large enough to stabilize the model error among the ensemble of the decision trees. Similarly, the number of leaves used in the RF classifier was 5, and the OOB error was the smallest.
- (3) **Karst rocky desertification mapping.**
Finally, based on karst rocky desertification feature sets, the trained RF model was used to monitor the karst rocky desertification dynamics during 2001–2020 in the peak-cluster depression basin in southwest Guangxi.

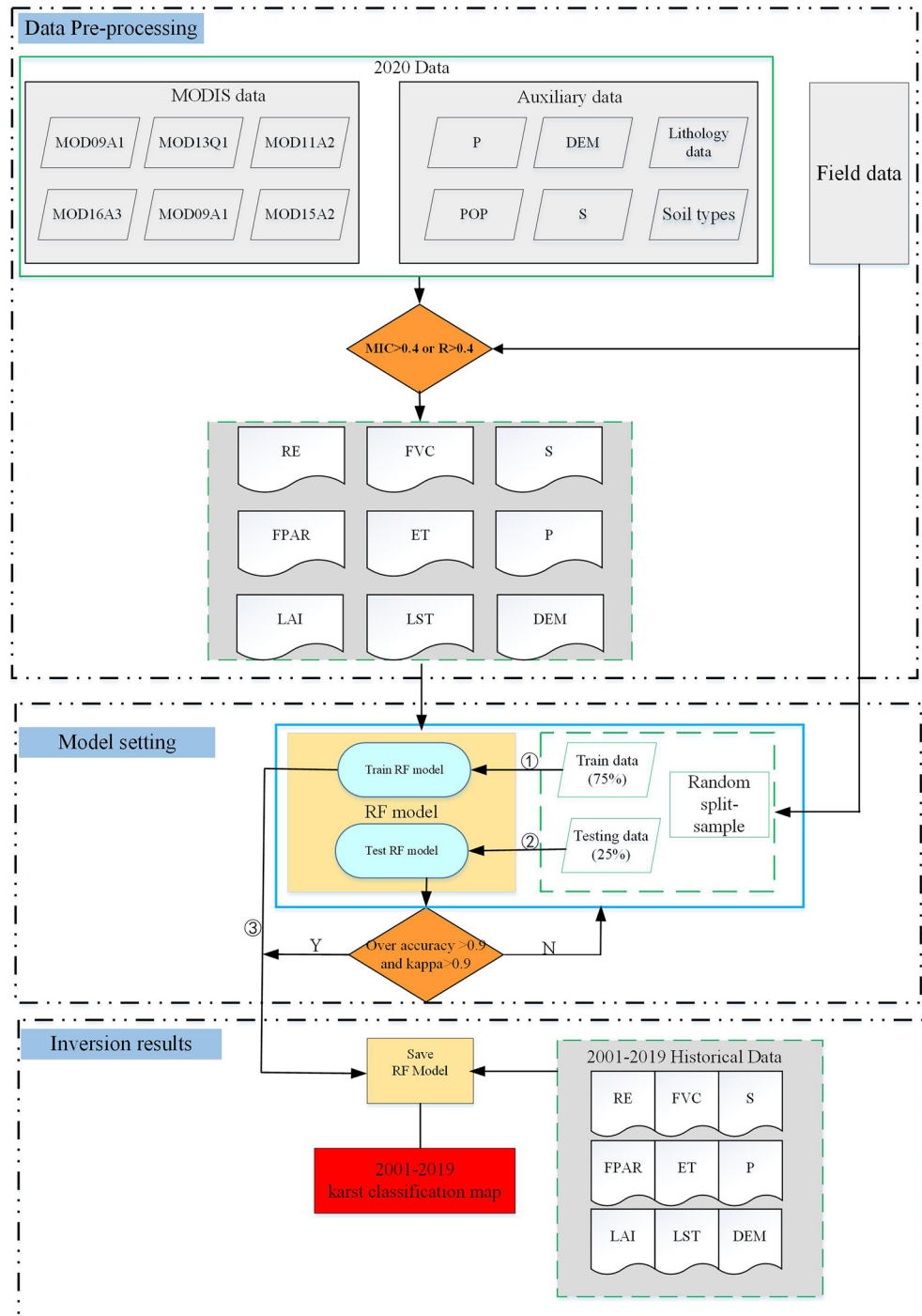


Figure 8. Flow chart of karst rocky desertification mapping in peak cluster depression in Southwest Guangxi, China.

Data availability

The data that support the findings of this manuscript are available from the corresponding author, T.Y, upon reasonable request.

Received: 29 July 2022; Accepted: 30 September 2022

Published online: 09 November 2022

References

1. Wang, S., Liu, Q. & Zhang, D. Karst rocky desertification in southwestern China: Geomorphology, landuse, impact and rehabilitation. *Land Degrad. Dev.* **15**(2), 115–121 (2004).
2. Jiang, M. *et al.* Geologic factors leadingly drawing the macroecological pattern of rocky desertification in southwest China. *Sci. Total Environ.* **458–460**, 419–426 (2013).
3. Jiang, Z., Lian, Y. & Qin, X. Rocky desertification in Southwest China: Impacts, causes, and restoration. *Earth-Sci. Rev.* **132**, 1–12 (2014).
4. Xu, E., Zhang, H. & Li, M. Object-based mapping of karst rocky desertification using a support vector machine. *Land Degrad. Dev.* **26**(2), 158–167 (2012).
5. Li, Y., Bai, X., Wang, S. & Tian, Y. Integrating mitigation measures for karst rocky desertification land in the Southwest mountains of China. *Carbonates Evaporites* **34**, 1095–1106 (2018).
6. Lan, J. Responses of soil organic carbon components and their sensitivity to karst rocky desertification control measures in Southwest China. *J. Soil. Sediment.* **21**, 978–989 (2020).
7. Gao, J., Du, F., Zuo, L. & Jiang, Y. Integrating ecosystem services and rocky desertification into identification of karst ecological security pattern. *Landscape Ecol.* **36**, 2113–2133 (2020).
8. Huang, X. *et al.* Driving factors and prediction of rock desertification of non-tillage lands in a karst basin, Southwest China. *Pol. J. Environ. Stud.* **30**(4), 3627–3635 (2021).
9. Chen, S., Zhou, Z., Yan, L. & Li, B. Quantitative evaluation of ecosystem health in a karst area of South China. *Sustain. Basel* **8**(10), 975 (2016).
10. Liu, F., He, B. Y. & Kou, J. F. Landsat thermal remote sensing to investigate the present situation and variation characteristics of karst rocky desertification in Pingguo County of Guangxi, Southwest China. *Sci. Soil Water Conserv.* **15**(02), 125–131 (2017).
11. 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.* **31**, 86–94 (2014).
12. 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.* **189**(6), 1–19 (2017).
13. Li, S. & Wu, H. Mapping karst rocky desertification using Landsat 8 images. *Remote Sens. Lett.* **6**(9), 657–666 (2015).
14. Yang, S. X., Lin, H., Hou, F., Zhang, L. P. & Hu, Z. L. Estimating karst area vegetation coverage by pixel unmixing. *Bull. Surv. Mapp.* **5**, 23–27 (2014).
15. Xiong, Y., Yue, Y. M. & Wang, K. L. Comparative study of indicator extraction for assessment of karst rocky desertification based on hyperion and ASTER images. *Bull. Soil Water Conserv.* **33**(03), 186–190 (2013).
16. Dai, G., Sun, H., Wang, B., Huang, C., Wang, W., Yao, Y., *et al.* Assessment of karst rocky desertification from the local to regional scale based on unmanned aerial vehicle images: A case-study of Shilin County, Yunnan Province, China. *Land Degrad. Dev.* 1–14 (2021).
17. Pu, J., Zhao, X., Dong, P., Wang, Q. & Yue, Q. Extracting information on rocky desertification from satellite images: A comparative study. *Remote Sens.* **13**(13), 2497 (2021).
18. Yue, Y. M. *et al.* Remote sensing of indicators for evaluating karst rocky desertification. *Procedia Environ. Sci.* **15**(04), 722–736 (2011).
19. Huang, Q. & Cai, Y. Spatial pattern of Karst rock desertification in the middle of Guizhou Province. Southwestern China. *Environ. Geol.* **52**(7), 1325–1330 (2006).
20. Wang, J., Li, S., Li, H., Luo, H. & Wang, M. Classifying indices and remote sensing image characters of rocky desertification lands: a case of karst region in Northern Guangdong Province. *J. Desert Res.* **5**, 765–770 (2007).
21. Chen, F. *et al.* Assessing spatial-temporal evolution processes and driving forces of karst rocky desertification. *Geocarto Int.* 1–22 (2019).
22. Qi, X., Zhang, C. & Wang, K. Comparing remote sensing methods for monitoring karst rocky desertification at sub-pixel scales in a highly heterogeneous karst region. *Sci. Rep-UK* <https://doi.org/10.1038/s41598-019-49730-9> (2019).
23. Yue, Y. *et al.* Spectral indices for estimating ecological indicators of karst rocky desertification. *Int. J. Remote Sens.* **31**(8), 2115–2122 (2010).
24. Yan, Y., Hu, B. Q., Han, Q. Y. & Li, Y. L. Early warning for karst rocky desertification in agricultural land base on the 3S and ANN technique: A case study in Du'an County, Guangxi. *Carsologica Sin.* **31**(01), 52–58 (2012).
25. Zhang, J. *et al.* Spectral analysis of seasonal rock and vegetation changes for detecting karst rocky desertification in southwest China. *Int. J. Appl. Earth Obs. Geoinf.* <https://doi.org/10.1016/j.jag.2021.102337> (2021).
26. Liu, Y., Wang, J. & Deng, X. Rocky land desertification and its driving forces in the karst areas of rural Guangxi, Southwest China. *J. Mt. Sci-Engl.* **5**(4), 350–357 (2008).
27. Li, Y., Xie, J., Luo, G., Yang, H. & Wang, S. The evolution of a karst rocky desertification land ecosystem and its driving forces in the Houzhaihe Area, China. *J. Ecol.* **5**, 501–512 (2015).
28. Zhang, Y. R., Zhou, Z. F. & Ma, S. B. Rocky desertification and climate change characteristics in typical karst area of Guizhou Province over past two decades. *Environ. Sci. Technol.* **37**(09), 192–197 (2014).
29. Bai, X. Y., Wang, S. J., Chen, Q. W. & Cheng, A. Y. Constrains of lithological background of carbonate rock on spatio-temporal evolution of karst rocky desertification land. *Earth Sci.* **35**(4), 691–696 (2010).
30. Li, L. & Xiong, K. Study on peak-cluster-depression rocky desertification landscape evolution and human activity-influence in South of China. *Eur. J. Remote Sens.* 1–9 (2020).
31. Yao, Y. H., Shuo, N. D. Z., Zhang, J. Y., Hu, Y. F. & Kou, Z. X. Spatiotemporal characteristics of karst rocky desertification and the impact of human activities from 2010 to 2015 in Guanling County, Guizhou Province. *Prog. Geogr.* **38**(11), 1759–1769 (2019).
32. Shi, K., Yang, Q. & Li, Y. Are karst rocky desertification areas affected by increasing human activity in Southern China? An empirical analysis from nighttime light data. *Int. J. Environ. Res. Public Health.* **16**(21), 4175 (2019).
33. Luo, X. L. *et al.* Analysis on the spatio-temporal evolution process of rocky desertification in Southwest Karst area. *Acta Ecol. Sin.* **41**(02), 680–693 (2021).
34. Yang, Q., Jiang, Z., Yuan, D., Ma, Z. & Xie, Y. Temporal and spatial changes of karst rocky desertification in ecological reconstruction region of Southwest China. *Environ. Earth Sci.* **72**(11), 4483–4489 (2014).
35. Zhang, C., Qi, X., Wang, K., Zhang, M. & Yue, Y. The application of geospatial techniques in monitoring karst vegetation recovery in southwest China. *Prog. Phys. Geog.* **41**(4), 450–477 (2017).

36. Ying, B., Xiao, S., Xiong, K., Cheng, Q. & Luo, J. Comparative studies of the distribution characteristics of rocky desertification and land use/land cover classes in typical areas of Guizhou province, China. *Environ. Earth Sci.* **71**(2), 631–645 (2013).
37. Luo, X. *et al.* Analysis on the spatio-temporal evolution process of rocky desertification in Southwest Karst area. *Acta Ecol. Sin.* **41**(2), 680–693 (2021).
38. Chong, G. *et al.* Characteristics of changes in karst rocky desertification in southern and western China and driving mechanisms. *Chin. Geogr. Sci.* **31**, 1082–1096 (2021).
39. Guo, B. *et al.* A novel-optimal monitoring model of rocky desertification based on feature space models with typical surface parameters derived from LANDSAT_8 OLI. *Degrad. Dev.* **32**(17), 5023–5036 (2021).
40. Chen, F. *et al.* Spatio-temporal evolution and future scenario prediction of karst rocky desertification based on CA–Markov model. *Arab. J. Geosci.* **14**, 1262 (2021).
41. Wu, X., Liu, H., Huang, X. & Zhou, T. Human driving forces: Analysis of rocky desertification in karst region in Guanling County, Guizhou Province. *Chin. Geogr. Sci.* **21**(5), 600–608 (2011).
42. Chen, H. *et al.* The evolution of rocky desertification and its response to land use changes in Wanshan Karst area, Tongren City, Guizhou Province, China. *J. Agr. Resour. Environ.* **37**(01), 24–35 (2020).
43. Zerrouki, N., Dairi, A., Harrou, F., Zerrouki, Y. & Sun, Y. Efficient land desertification detection using a deep learning-driven generative adversarial network approach: A case study. *Concurr. Comp-Pract. E.* <https://doi.org/10.1002/cpe.6604> (2021).
44. Keskin, H., Grunwald, S. & Harris, W. Digital mapping of soil carbon fractions with machine learning. *Geoderma* **339**, 40–58 (2019).
45. Tian, Y. *et al.* Aboveground mangrove biomass estimation in Beibu Gulf using machine learning and UAV remote sensing. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2021.146816> (2021).
46. Xi, H. *et al.* Spatio-temporal characteristics of rocky desertification in typical Karst areas of Southwest China: A case study of Puding county, Guizhou province. *Acta Ecol. Sin.* **38**(24), 8919–8933 (2018).
47. Deng, Y. *et al.* Relationship among land surface temperature and LUCC, NDVI in typical karst area. *Sci. Rep-UK.* 296–306 (2018).
48. Li, S. M., Yu, L. W., Gan, S. & Yang, Y. M. Study on inversion relationship between vegetation Index and leaf area index of rocky desertification area in southeast Yunnan based on ETM+. *J. Kunming Univ. Sci. Technol. (Nat. Sci.)* **40**(06), 31–36 (2015).
49. Yan, X. & Cai, Y. Multi-Scale anthropogenic driving forces of karst rocky desertification in Southwest China. *Land Degrad. Dev.* **26**(2), 193–200 (2013).
50. Meyer, H., Reudenbach, C., Wollauer, S. & Nauss, T. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecol. Model.* <https://doi.org/10.1016/j.ecolmodel.2019.108815> (2019).
51. Cracknell, M. & Reading, A. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci-UK* **63**, 22–33 (2014).
52. Feng, K. *et al.* Monitoring desertification using machine-learning techniques with multiple indicators derived from MODIS images in Mu Us Sandy Land, China. *Remote Sens.* **14**, 2663. <https://doi.org/10.3390/rs14112663> (2022).
53. Belgiu, M. & Drăguț, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm* **114**, 24–31 (2016).
54. Chutia, D., Bhattacharyya, D. K., Sarma, K. K., Kalita, R. & Sudhakar, S. Hyperspectral remote sensing classifications: A perspective survey. *Trans. GIS* <https://doi.org/10.1111/tgis.12164> (2015).
55. Song, T. Q., Peng, W. X., Du, H., Wang, K. & Zeng, F. Occurrence spatial-temporal dynamics and regulation strategies of karst rocky desertification in southwest China. *Acta Ecol. Sin.* **34**(18), 5328–5341 (2014).
56. Zhu, L.F. *Study on the Spatial-Temporal Variation of Vegetation Coverage and Karst Rocky Desertification based on MODIS Data*. Ph.D. Dissertation, Southwestern University. Chongqing, China (2018).
57. Yang, Q. *et al.* Spatio-temporal evolution of rocky desertification and its driving forces in karst areas of Northwestern Guangxi, China. *Environ. Earth Sci.* **64**, 383–393 (2011).
58. Mishra, N. & Chaudhuri, G. Spatio-temporal analysis of trends in seasonal vegetation productivity across Uttarakhand, Indian Himalayas, 2000–2014. *Appl. Geogr.* **56**, 29–41 (2015).
59. 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.* **31**, 86–94 (2014).
60. Reshef, D. *et al.* Detecting novel associations in large data sets. *Science* **334**(6062), 1518–1524 (2011).
61. Li, W. *et al.* Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2020.139099> (2021).
62. Abdelhakim, A., El, H., Luis, E., Salah, E. & Abdelghani, C. Retrieving crop albedo based on radar sentinel-1 and random forest. Approach. *Remote Sens.* **13**(16), 3181 (2021).
63. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M. & Rigol-Sanchez, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm* **67**, 93–104 (2012).
64. Dharumarajan, S., Bishop, T., Hegde, R. & Singh, S. Desertification vulnerability index—an effective approach to assess desertification processes: A case study in Anantapur District, Andhra Pradesh, India. *Land Degrad. Dev.* **29**(1), 150–161 (2017).
65. Li, P. *et al.* Dynamic monitoring of desertification in ningdong based on landsat images and machine learning. *Sustainability* **14**, 7470. <https://doi.org/10.3390/su14127470> (2022).
66. Pacheco, A. D. P., Junior, J. A. D. S., Ruiz-Armenteros, A. M. & Henriques, R. F. F. Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in Central Portugal using landsat-8, sentinel-2, and terra imagery. *Remote Sens.* **13**, 1345. <https://doi.org/10.3390/rs13071345> (2021).

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 42061020, 42261052), Natural Science Foundation of Guangxi Zhuang Autonomous Region, (Grant No. 2018JJA150135), Guangxi Key Research and Development Program, (Grant No. AA18118038), Science and Technology Department of Guangxi Zhuang Autonomous (Grant No. 2019AC20088), The Program of Improving the Basic Research Ability of Young and Middle-aged Teachers in Guangxi Universities (Grant No. 2021KY0431), High level talent introduction project of Beibu Gulf University, (Grant No. 2019KYQD28).

Author contributions

Conceptualization, Y.T. and Y.Z.; methodology, Y.T. and Y.Z.; validation, Y.Z., Y.T. and D.W.; formal analysis, Y.T.; investigation, Y.Z., J.T., Q.Z., Y.L., Y.T., Y.Y. and D.W.; resources, Y.Y.; data curation, Y.Z., J.L. and Y.T.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z. and Y.L.; visualization, Y.L., L.W.; supervision, Y.T.; project administration, Y.T. All the authors have reviewed the manuscript and agreed the submission and publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.T.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 Author(s) 2022