

# An Ensemble Learning Approach for Land Use/Land Cover Classification of Arid Regions for Climate Simulation: A Case Study of Xinjiang, Northwest China

Haoyang Du , Manchun Li, Yunyun Xu, and Chen Zhou 

**Abstract**—Accurate classifications of land use/land cover (LULC) in arid regions are vital for analyzing changes in climate. We propose an ensemble learning approach for improving LULC classification accuracy in Xinjiang, northwest China. First, multisource geographical datasets were applied, and the study area was divided into Northern Xinjiang, Tianshan, and Southern Xinjiang. Second, five machine learning algorithms—k-nearest neighbor, support vector machine (SVM), random forest (RF), artificial neural network (ANN), and C4.5—were chosen to develop different ensemble learning strategies according to the climatic and topographic characteristics of each subregion. Third, stratified random sampling was used to obtain training samples and optimal parameters for each machine learning algorithm. Lastly, each derived approach was applied across Xinjiang, and subregion performance was evaluated. The results showed that the LULC classification accuracy achieved across Xinjiang via the proposed ensemble learning approach was improved by  $\geq 6.85\%$  compared with individual machine learning algorithms. By specific subregion, the accuracies for Northern Xinjiang, Tianshan, and Southern Xinjiang increased by  $\geq 6.70\%$ ,  $5.87\%$ , and  $6.86\%$ , respectively. Moreover, the ensemble learning strategy combining four machine learning algorithms (i.e., SVM, RF, ANN, and C4.5) was superior across Xinjiang and Tianshan; whereas, the three-algorithm (i.e., SVM, RF, and ANN) strategy worked best for the Northern and Southern Xinjiang. The innovation of this study is to develop a novel ensemble learning approach to divide Xinjiang into different subregions, accurately classify land cover, and generate a new land cover product for simulating climate change in Xinjiang.

Manuscript received 1 July 2022; revised 9 September 2022 and 28 December 2022; accepted 16 February 2023. Date of publication 23 February 2023; date of current version 9 March 2023. This work was supported in part by the National Key Research & Development Program of China under Grant 2022YFC3800802 and in part by the National Natural Science Foundation of China under Grant 42271414 and Grant 41901318. (Corresponding author: Chen Zhou.)

Haoyang Du, Yunyun Xu, and Chen Zhou are with the Department of Geographic Information Science, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China (e-mail: duhaoyang15@mailsucas.ac.cn; yyxu.nju@gmail.com; chenzhou@nju.edu.cn).

Manchun Li is with the Jiangsu Provincial Key Laboratory of Geographic Information Technology, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China, also with the Department of Geographic Information Science, Nanjing University, Nanjing 210023, China, also with the Collaborative Innovation Center for the South Sea Studies, Nanjing University, Nanjing 210093, China, and also with the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, Nanjing 210023, China (e-mail: limanchun@nju.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSTARS.2023.3247624>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2023.3247624

**Index Terms**—Arid areas, China, ensemble learning, land use/land cover classification, machine learning.

## I. INTRODUCTION

Monitoring land use/land cover (LULC) is indispensable for investigating earth system processes, as it can provide thematic information of the earth's surface while capturing biotic and abiotic characteristics that closely correlate with the ecological conditions on the ground [1]. Moreover, accurate mapping of LULC can greatly improve analyses of environmental change and is more capable of reflecting the interactions between human activities and the geographical environment [2], [3].

Since the International Geosphere Biosphere Program (IGBP) proposed the LULC change project in 1995, numerous LULC products have been successively developed by major geoscience research institutions around the world [4], including

- 1) IGBP data and information system cover (IGBP DISCover) [5], [6], [7];
- 2) University of Maryland Land Cover dataset [8];
- 3) Global Land Cover 2000 (GLC2000) [9];
- 4) Ecosystem Classification and Land Surface Parameters database [10];
- 5) moderate resolution imaging spectroradiometer (MODIS) land cover dataset (MOD12Q1 and MCD12Q1) [11];
- 6) Climate change initiative land cover dataset (CCI-LC) [12];
- 7) the finer resolution observation and monitoring of global land cover (FROM-GLC) [13];
- 8) global LULC data (GlobeLand30; 30 m resolution) [14];
- 9) China's land-use/cover datasets (CLUDs) create by the Chinese Academy of Sciences (CAS) [15].

Although these LULC products are diverse and readily available, they are plagued by large inconsistencies in their accuracies [16]. For example, the overall accuracies of FROM-GLC, MCD12Q1, GLC2000, and GlobeLand30 are 64.9%, 78.8%, 77.9%, and 80.3%, respectively [11], [13], [14], [17]. Accordingly, it remains difficult to compare and combine these products to extract more accurate LULC information related to ecological, hydrological, and climatological studies [18], [19], especially the study of climate change. Previous studies have shown that the accuracy of land-cover data below 80% has a

great impact on the results of precipitation research, and, as the accuracy continues to decrease, the results may become worse [20]. Unfortunately, both the overall and class-specific accuracies of most datasets do not meet the common requirements for regional climate modeling [21]. Furthermore, research has shown that the accuracy of these products over Xinjiang are significantly lower than that in other regions [22], [23], thus failing to meet the high requirements for optimal use in this area and restricting further improvements in research related to regional-climate simulations, desertification monitoring, and ecosystem service assessments. Therefore, it is necessary to generate high-precision land-cover datasets for climate simulation based on existing land-use, land-cover, and some auxiliary datasets.

To overcome these limitations, the objective of this study is to present an ensemble learning method for LULC classification in Xinjiang with complex topographic areas. Furthermore, this study aims to generate a set of high-precision land-cover data via the application of multisource geographical datasets, including Landsat 8 OLI images, FROM-GLC, MCD12Q1, ESA-LC, a digital elevation model (DEM), the enhanced vegetation index (EVI), net primary production (NPP), and leaf area index (LAI). The Xinjiang Province in northwest China was chosen as a typical case, and it was divided into three subregions according to the complex topography and climate. Subsequently, different ensemble learning strategies were developed for each sub-region. Lastly, training samples were obtained, and the parameters of machine learning algorithms were calculated to assess the accuracy of each ensemble learning-enabled LULC classification. The main contribution of our study lies in taking topographic conditions into account in land-cover classification and developing a new ensemble learning method to accurately classify land cover in Xinjiang. Furthermore, we developed a new land-cover product using a two-level classification system, which solves the problem of low accuracy of land-cover classification in Xinjiang climate simulation and can be used in other research areas, such as territorial space planning and hydrological simulation. The outcomes of this study will help gain a deep understanding of the interactions between land and atmosphere in arid areas.

The article is organized as follows: Section II summarizes the theoretical basis of machine and ensemble learning in LULC classification. Section III introduces the study area and data used. Section IV describes the basic framework of the proposed method. Section V analyzes the experimental results. Section VI discusses the broader application and future works. Finally, Section VII summarizes the findings and conclusions of this study.

## II. RELATED WORK

### A. LULC Classification

LULC classification accuracy is closely related to the remote sensing data used and the classification method chosen [24]. With the advancement of computational abilities and remote sensing technologies, machine learning has gradually become one of the most effective methods for LULC classification [25],

[26]. Among the available methods, the most widely employed classification algorithms for remote sensing imagery include K-nearest neighbor (KNN), decision trees (DTs), random forest (RF), support vector machine (SVM), artificial neural network (ANN), and extreme learning machine [27].

Previous studies have shown that owing to the complexity of multisource remote sensing imagery data, distinct machine learning algorithms have different advantages when classifying certain LULC types; thus, any single classifier is limited in its ability to significantly improve the accuracy of all LULC types [28]. Accordingly, identification of the optimal classification algorithms remains challenging. For example, in one LULC classification within an individual area, SVM accuracy for irrigated herbs was 92%, while that for oak forests was only 72% [24]. Recently, classifier ensembles (i.e., multiple classifier systems) have received considerable attention in remote sensing image analyses owing to their high classification accuracy by exploiting the advantages of different classifiers while minimizing their limitations [29], [30], [31]. Various combinatorial strategies have been developed and widely used to integrate different classifications. For example, Chen et al. employed SVM, C4.5, DTs, and ANNs to construct ensemble learning classification, resulting in an overall accuracy (OA; 88.12%) and Kappa (0.87) value superior to those of any basic classifiers [32]. Hu et al. used two ensemble methods based on ANNs to classify LULC in the Zoige wetlands of China, showing that the ensemble technology improved the classification ability and stability of any single ANN [26].

In addition, the successful implementation of classification methods largely depends on the characteristics of the study area and nature of relevant data [33]. Therefore, to improve the effects of multiclassifier ensemble learning, both the machine learning algorithm and feature data characteristic must be considered [32]. Ultimately, compared to traditional classification methods, machine learning algorithms are efficient and effective because they do not rely on normal assumptions or statistical parameters [34], yet appropriate classification-algorithm selection for a given area remains essential.

### B. Machine Learning

The theoretical basis of the ensemble learning classification methods used in the present study is summarized in this section. For a deeper understanding of the theoretical background of a particular algorithm, refer to the references provided.

- 1) *KNN*: KNN is a nonparametric machine learning algorithm and one of the simplest instance-based regression and supervised classification techniques [35]. It determines the sample category by evaluating the distance relationship between an unknown category and K adjacent training samples, ultimately calculating the similarity degree between the two [36]. For further details on the KNN algorithm, see [35].
- 2) *SVM*: Support vector machine (SVM) is a machine learning method proposed by Vapnik in the 1990s and was initially applied on the recognition of hand-written digits (i.e., pattern recognition) [37], [38], [39]. It is

characterized by a small number of training samples, high noise resistance, support for high-dimensional data, and strong stability [32]. Moreover, it can generate high accuracy for modeling complex nonlinear decision boundaries and is not easy to be over fitting. Therefore, it is a supervised learning technique commonly used in a series of remote sensing applications [25]. A detailed description of the SVM algorithm can be found in [40].

- 3) *RF*: RF is a modeling classification algorithm that integrates multiple unrelated classification and regression (decision) trees using a bagging strategy. It is characterized by eliminating generalization errors to achieve nondeviation classification, thereby enhancing classification accuracy, particularly for multisource remote sensing classification [41]. Accordingly, RF has been widely employed in LULC classification due to its high accuracy [42]. For detailed introduction to RF, refer to paper [43].
- 4) *ANN*: ANN is a machine learning algorithm developed to simulate the ability of the human brain to resolve problems related to pattern recognition [34]. It is advantageous owing to its nonlinearity, strong anti-interference, high adaptability, parallel processing, and self-organization of learning-process characteristics; therefore, ANN has been applied to ever-increasing remote sensing image classification in recent years [44]. Refer to the following resources for further ANN algorithm details [45], [46].
- 5) *C4.5*: C4.5 is a decision tree algorithm modified on the basis of the previously developed Iterative Dichotomiser 3 algorithm [47], and is characterized by its advantageous strong logic, simple rule set, and effective suppression of image noise for suitable multisource remote sensing image data classification [48]. Refer to the following resources for further C4.5 details [47].

### C. Ensemble Learning

Ensemble learning is a method of training multiple machine learning algorithms to improve the predictive performance and classification accuracy based on the additive effect of the advantageous characteristics [49]. In addition to fully utilizing the respective advantages of different classifiers, ensemble learning can address the problem of over-fitting any single classifier in instances with small amounts of data [50], [51]. Refer to the following resource for further details on ensemble learning [52].

## III. STUDY AREA AND DATA

### A. Study Area

Located in the northwest China (34.40–48.10° N, 73.60–96.3° E), Xinjiang is far from oceans and maintains a typical continental, arid, or semi-arid climate. The mean temperatures in the coldest (January) and hottest (July) months are  $-17.1$  and  $27.1$  °C, respectively [53]. The average annual precipitation in the Taklimakan Desert is  $< 50$  mm·yr<sup>-1</sup>, whereas that in Tianshan is  $\sim 800$  mm·yr<sup>-1</sup> [54]. Xinjiang has a unique spatial shape for

a mountain basin [55], with the general topographical characteristics of three mountains and two basins: the Altai Mountains to the north, Tianshan in the middle, and Kunlun Mountains to the south. Between the Altai Mountains and Tianshan, there lies the Gurbantungut Desert in the Junggar Basin, while the Taklamakan Desert in the Tarim Basin lies between Tianshan and the Kunlun Mountains [56]. The unique geographical location and topographical characteristics in Xinjiang have created a unique ecosystem in the region.

### B. Data

Landsat satellite-image data are often used to classify regional-level LULC owing to their free availability and global coverage [57]. Here, because spring and summer images contain most of the phenological changes [24], [58], Landsat 8 L2 images were downloaded from the United States Geological Survey website for Xinjiang. The images from April 1, 2015 to August 31, 2015 were used in LULC classification. The EVI, normalized difference building index (NDBI), and improved normalized water index (MNDWI) were calculated via Landsat bands 2, 3, 4, 5, 6, and the composite spectral index of Landsat 8. Additionally, four sources of LULC data in Xinjiang—1) CLUDs, 2) FROM-GLC, 3) CCI-LC, and 4) MCD12Q1—were used in the present study. To ensure highly accurate LULC classification, auxiliary data including the following:

- 1) vegetation type maps of 11 populations and 54 types;
- 2) physical geographical data including elevation (DEM) and slope information;
- 3) NPP;
- 4) LAI were also employed here.

Table I lists all data and sources used in the present study.

## IV. METHODOLOGY

### A. Basic Framework

Previous studies have shown that geographical characteristics (e.g., complex landforms) and diverse climates affect LULC classification accuracy [24]. Considering the complex topography and diverse climate of the study area, Xinjiang was divided into three sub-regions, and different ensemble learning strategies were developed for each to optimize overall classification accuracy. The basic framework was as follows (see Fig. 1).

First, the subregional division according to climate and topographic features is discussed in Section IV-B. Second, the spatial consistency of different products was analyzed, and, on this basis, a hierarchical random sampling method was adopted to obtain machine learning training samples (see Section IV-C). Third, machine learning algorithms suitable for the study area were selected according to different data characteristics of Landsat 8, NPP, EVI, DEM, LAI, etc. (see Section IV-D). Lastly, a confusion matrix was used to evaluate the accuracy of different ensemble strategies of LULC classification across each subregion, and the most suitable technique was selected to draw the spatial patterns of LULC classification in Xinjiang.



TABLE I  
MULTISOURCE GEOGRAPHICAL DATASETS USED IN THE PRESENT LAND USE/LAND COVER STUDY OF XINJIANG, CHINA

	Data	Sources	Resolution (m)
1	Landsat 8	http://eros.usgs.gov/	30
2	EVI	http://eros.usgs.gov/	250
3	MNDWI	http://eros.usgs.gov/	30
4	NDBI	http://eros.usgs.gov/	30
5	LAI	https://ladsweb.modaps.eosdis.nasa.gov/	250
6	CLUDs	https://www.resdc.cn/	30
7	FROM-GLC	http://data.ess.tsinghua.edu.cn/fromglc_0_1.html	30
8	CCI-LC	http://www.esa-landcover-cci.org/?q=node/158	300
9	MCD12Q1	https://ladsweb.modaps.eosdis.nasa.gov/search/	250
10	NPP	https://ladsweb.modaps.eosdis.nasa.gov/	250
11	Vegetation type	https://www.resdc.cn/	500
12	Elevation	https://www.resdc.cn/	30
13	Slope	https://www.resdc.cn/	30

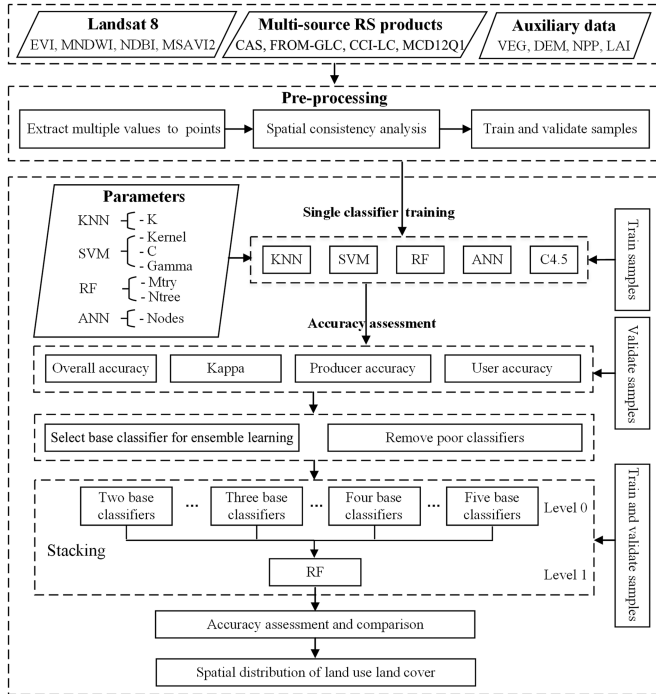


Fig. 1. Flowchart of land use/land cover (LULC) classification for Xinjiang, China.

## B. Regionalization

The spatial distribution of vegetation is affected by both climate and terrain variability. For example, changes in topographic characteristics drive the gradual shift in surface vegetation from low mountain deserts, arid grasslands, irrigated farmlands, and broadleaf forests to mid-mountain grasslands, evergreen needle-leaf, and mixed forests, and further to alpine meadows/dwarf shrubs [59]. Accordingly, during LULC classifications, the introduction of geographic divisions by climate and topography can reduce the probability of errors. Based on meteorological station data and the DEM, KNN was used to further divide the research region into three sub-regions: Northern Xinjiang, Tianshan, and Southern Xinjiang (see Fig. 2). Refer to [55] and [60] the partitioning method for details.

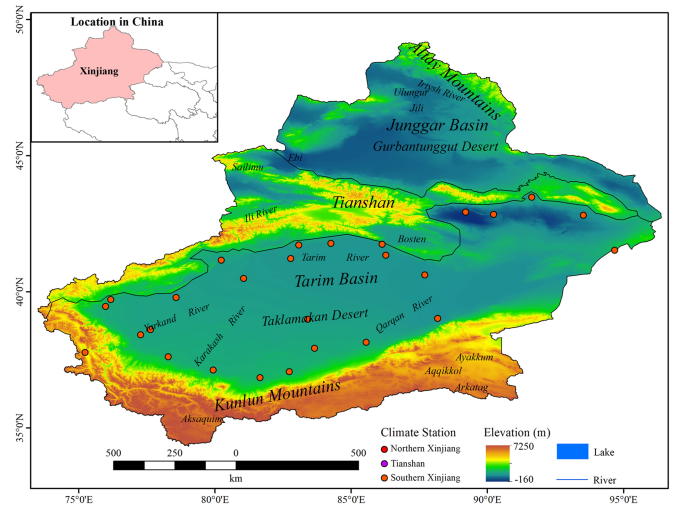


Fig. 2. Pertinent locations in Xinjiang, China.

## C. Spatial Consistency Analysis

In the present study, an inference rule for spatial data mining based on grid consistency was proposed to identify LULC types where the primary methods consisted of spatial feature extraction and consistency analyses. First, for the zoning shape to facilitate LULC classification, fishnet was created using ArcGIS v.10.2, (resolution, 0.25 km<sup>2</sup>). Second, four LULC products, vegetation type, and other auxiliary data were extracted from the fishnet feature points by extract multiple values to points. Then, the net feature points and net data were spatially aggregated to obtain the net data containing 13 types of attribute information. Lastly, the spatial consistency levels of CLUDs, FROM-GLC, CCI-LC, MCD12Q1 and vegetation type maps were analyzed using the grid comparison method under the IGBP classification systems.

## D. Machine Learning Algorithm Selection

During data mining, the characteristics of remote sensing data, as well as the advantages and disadvantages of various machine learning methods must be considered before classifying LULC [34], [41]. Combining different characteristics of Landsat 8,



NPP, EVI, DEM, LAI, etc., the most suitable machine learning algorithm for LULC classification in Xinjiang was screened according to the following process:

First, the EVI and LAI of different vegetation types can be similar; therefore, KNN classification was adopted here because it is based on the similarity of training sample data. Secondly, because remote sensing data have nonlinear properties (e.g., Landsat 8 spectral information data, DEM, etc.), the radial basis kernel function (RBF) of SVM can map nonlinear data in high-dimensional space, thereby generating classification hyperplane decision boundaries and making it linearly separable. Therefore, SVM was employed in the present study. Moreover, we observed that Landsat 8 contained “salt and pepper” noise. When compared with other machine learning techniques, RF showed the greatest insensitivity to noise, training stability, and efficiency; therefore, it was employed here. Because the spectral mixing degree of Landsat 8 data is high, especially in montane areas with complex topography, ANN was also employed here because it adapts well to the characteristics of rich texture and high spectral confusion of remote sensing. Furthermore, numerous missing values in the remote sensing data throughout alpine areas (especially in Tianshan) were observed here owing to cloud cover. Because C4.5 has a superior processing effect on samples with missing values, it was also selected here. For different machine learning combinations, model training was carried out across Xinjiang, as well as the three sub-regions to identify the most suitable parameters (see Section V-B for specific parameters).

### E. Ensemble Learning Strategy

Knowledge of the differences between base classifiers is key to constructing ensemble learning [61]. Considering the differences, advantages, and disadvantages of different machine learning techniques, different stacks of ensemble learning strategies across Xinjiang and its three subregions are proposed here. First, the random sample data were divided into two at a 3:1 ratio of training to validation data. Second,  $\geq 2$  of the base classifiers (KNN, SVM, RF, ANN, and C4.5) were selected as level 0 of the stacking ensemble strategy model, and RF was selected as level 1. Lastly, across Xinjiang and the three subregions, the ensemble learning models under different stacking strategies were trained, and the OA, producer’s accuracy (PA), user’s accuracy (UA), and Kappa were compared to obtain the optimal LULC classification ensemble strategy.

### F. Evaluation Indicators

The classification effect of machine learning algorithms represents the most critical evaluation index. A confusion matrix was used to evaluate the classification abilities of different algorithms and ensemble learning strategies. Accordingly, the OA, PA, UA, and Kappa were calculated from the confusion matrix according to the following equations:

$$OA = \frac{1}{N} \sum_{i=1}^r X_{ii} \quad (1)$$

$$UA_i = \frac{X_{ii}}{X_{+i}} \quad (2)$$

$$PA_i = \frac{X_{ii}}{X_{i+}} \quad (3)$$

$$Kappa = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^r (X_{i+} \times X_{+i})} \quad (4)$$

where  $N$  represents the total number of training samples;  $r$  is the number of rows in the confusion matrix;  $X_{ii}$  is the number of samples in row  $i$  and column  $i$  of the confusion matrix (i.e., on the diagonal at the intersection); and  $X_{i+}$  and  $X_{+i}$  are the marginal totals of row  $r$  and column  $i$ , respectively.

Although the OA and Kappa are the two most popular metrics for assessing classification accuracy, the samples used to calculate Kappa cannot always be independent in all cases because the same test set is used when evaluating the accuracy of each map [34]. Hence, the pairwise Z-score test was used to evaluate whether the differences in classification accuracy among different ensemble learning strategies were statistically significant [62]. A Z-score greater than 1.96 was considered statistically significant at the 5% level.

Field surveys and visual inspection are the ideal methods used to select samples to obtain high land-cover classification accuracy [13], [63]. Therefore, in order to compensate for the shortcomings of the confusion-matrix schedule evaluation, 42 sampling points were selected across Xinjiang, which overall covered all land cover types described in this study. For each sampling point, we identified real land cover type of the sampling point through field photography and Google Earth.

## V. RESULTS

### A. Spatial Consistency

On comparing the five remote sensing products, the inconsistency among the LULC classification was readily apparent, particularly for the forest classes. The CAS divided the forest into forested land, open forest, shrubland, and other forests, whereas other products split forest cover into evergreen and deciduous on the basis of leaf fall, and broadleaf and needleleaf on the basis of leaf size. Accordingly, to best facilitate comparison, different classification schemes were converted into IGBP schemes (Table S1, see the Supporting Information); 0 represents complete disagreement between the five products, and 5 indicates perfect consistency among them. The highest consistency of forest was 4 because the forest in the CAS classification cannot be converted into IGBP schemes, and, for other land types it was 5. Here, it was assumed that when the consistency of certain forest attributes was  $\geq 3$ , it was defined as the corresponding forest type, whereas all other land types were defined as the corresponding land type only when the consistency was  $\geq 4$ . Fig. 3 shows the spatial consistency level of the five LULC products as well as the sampling points under the IGBP classification system. Notably, complete or level 4 consistency was achieved for water areas, deserts, and the Gobi belt. In the transition zone between oasis and desert or mountains, the observed data consistency level was

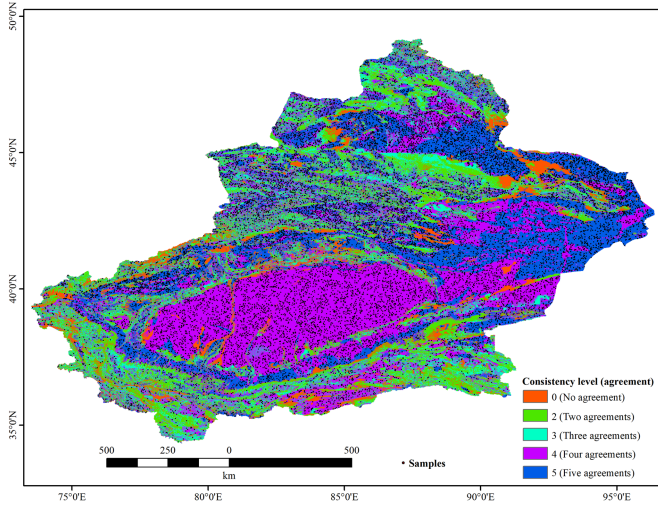


Fig. 3. Spatial consistency level between the five land cover products under the IGBP classification system.

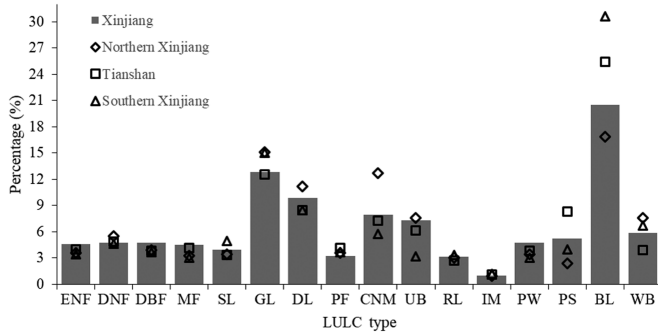


Fig. 4. Proportion of machine learning training samples in different regions: ENF, evergreen needleleaf forests; DNF, deciduous needleleaf forest; DBF, deciduous broadleaf forest; MF, mixed forest; SL, shrublands; GL, grasslands; DL, dry land; CNM, cropland/natural vegetation mosaics; UB, urban and built-up lands; RL, rural land; IM, industrial and mining; PW, permanent wetlands; PS, permanent snow and ice; BL, bare lands; WB, water body.

generally  $< 3$ . Fig. 4 shows the proportion of machine learning training samples in different regions.

### B. Machine Learning Parameters

The key to KNN classification lies in selecting an appropriate  $K$  value; thus, KNN classification accuracy was tested here using  $K$  ranging from 1 to 30. The RBF of SVM is primarily controlled by two parameters: the cost ( $C$ ) and weight of the RBF kernel ( $\gamma$ ). Here, the optimal combination of  $C$  ( $2^{-2}$ ,  $2^{-1}$ ,  $2^0$ ,  $2^1$ ,  $2^2$ ,  $2^3$ ,  $2^4$ ,  $2^5$ ,  $2^6$ ,  $2^7$ ) and  $\gamma$  ( $2^{-5}$ ,  $2^{-4}$ ,  $2^{-3}$ ,  $2^{-2}$ ,  $2^{-1}$ ,  $2^0$ ,  $2^1$ ,  $2^2$ ,  $2^3$ ,  $2^4$ ) was tested on training and validation samples. RF classification relies on the number of random samplings as a candidate variable ( $Mtry$ ) and the number of random trees ( $Ntree$ ) [34], [64]. It is recommended that  $Mtry$  be set as the square root of the input variable, and  $Ntree$  be set as a multiple of 500 to achieve the optimal classification effect [42]. Therefore, different  $Mtry$  values from 1 to 16 and  $Ntree$  values from 500 to 5000 (500 intervals) were examined here. ANN accuracy

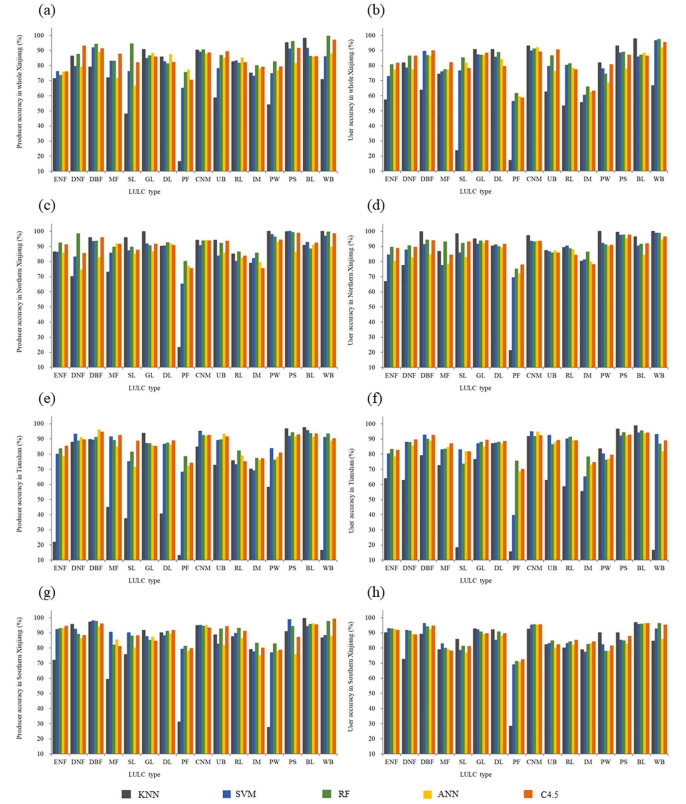


Fig. 5. (a) PA and (b) UA of different LULC types obtained using various machine learning techniques for the whole Xinjiang. (c) PA and (d) UA for Northern Xinjiang. (e) PA and (f) UA for Tianshan. (g) PA and (h) UA for Southern Xinjiang.

depends largely on the number of hidden layer nodes, which ranged from 1 to 50 in the present study [34]. Furthermore, the neurons that receive input information are then transformed using an activation function to ensure nonlinear prediction [26]. This study used the sigmoid function as the activation function [65]. Alternatively, C4.5 does not require any parameter setting due to its inherent simplicity. The machine learning parameters are shown in Table II.

### C. Classification Effect of Single Classifier

Table III shows the effects of different classifiers across Xinjiang and the three subregions, revealing that KNN had the best classification effect on all assessed regions (except for Tianshan), of which northern Xinjiang has the best performance ( $OA = 94.67\%$ ); however, according to the PA and UA (see Fig. 5), KNN maintained poor classification effects on evergreen needleleaf forests, shrublands, paddy fields, and industrial and mining across Xinjiang. Specifically, the classification accuracy for evergreen needleleaf forests, mixed forests, shrublands, dry lands, paddy fields, and industrial and mining in Tianshan was poor, whereas the classification of evergreen needleleaf forest, mixed forest, and paddy fields was poor in Northern and Southern Xinjiang. This might be due to the imbalance of KNN [66]; some of the land LULC types with high PA or UA increased the OA [25]. Therefore, OA and Kappa, in addition to PA and

TABLE II  
ADJUSTED PARAMETER VALUES OF EACH NONPARAMETRIC CLASSIFIER USED IN THE LULC CLASSIFICATION

Machine Learning	Whole Xinjiang	Northern Xinjiang	Tianshan	Southern Xinjiang
KNN	$K = 8$	$K = 5$	$K = 7$	$K = 8$
SVM	$C = 2^6, \gamma = 2^{-4}$	$C = 2^4, \gamma = 2^{-4}$	$C = 2^5, \gamma = 2^{-5}$	$C = 2^5, \gamma = 2^{-3}$
RF	$Mtry = 7$	$Mtry = 4$	$Mtry = 6$	$Mtry = 5$
ANN	$Ntree = 1000$	$Ntree = 500$	$Ntree = 1000$	$Ntree = 1000$
C4.5	$nodes = 16$	$nodes = 8$	$nodes = 11$	$nodes = 11$
	--	--	--	--

TABLE III  
OA AND KAPPA COEFFICIENTS OF DIFFERENT MACHINE LEARNING CLASSIFIERS ACROSS XINJIANG AND THE THREE SUBREGIONS (%)

Machine Learning	Whole Xinjiang		Northern Xinjiang		Tianshan		Southern Xinjiang	
	OA	Kappa	OA	Kappa	OA	Kappa	OA	Kappa
KNN	87.64	85.11	94.67	93.89	78.37	74.54	92.59	90.69
SVM	86.02	83.09	91.11	89.81	90.73	89.05	91.91	89.91
RF	86.30	84.23	92.87	91.44	90.60	88.4	92.54	90.57
ANN	85.81	82.87	92.53	87.78	89.03	87.07	92.45	88.82
C4.5	86.91	83.40	89.34	91.83	90.15	88.92	91.05	90.19

TABLE IV  
OA AND KAPPA OF CLASSIFICATION (%) UNDER EACH ENSEMBLE STRATEGY

Ensemble strategies	Whole Xinjiang		Northern Xinjiang		Tianshan		Southern Xinjiang	
	OA	Kappa	OA	Kappa	OA	Kappa	OA	Kappa
SVM, RF, ANN	90.29	87.07	96.31	95.77	94.34	93.80	96.41	95.51
SVM, RF, C4.5	88.35	85.86	95.45	95.32	94.22	93.17	96.21	96.77
SVM, C4.5, ANN	89.35	88.22	95.35	95.67	94.03	92.97	96.21	95.27
RF, C4.5, ANN	90.17	88.08	95.39	95.32	94.00	92.93	96.12	95.15
SVM, RF, C4.5, ANN	92.87	91.34	96.35	95.80	96.33	95.68	96.92	96.16

UA, must be considered when employing machine learning to classify LULC [32], [67], [68].

Except for KNN, C4.5 had the highest classification accuracy across Xinjiang (OA = 86.91%; Table III), whereas SVM had the highest classification accuracy in Tianshan (OA = 90.73%), and RF had the highest classification accuracy in Northern and Southern Xinjiang (OA = 92.87% and 92.54%, respectively). When comparing the classification effects for the three subregions, the average OAs for Southern, Northern Xinjiang, and Tianshan were 91.98%, 91.46%, and 90.13%, respectively, notably higher than the average OA across Xinjiang (86.26%). Thus, geographically subdividing the study area based on topographic conditions further improved the LULC classification accuracy by machine learning. Additionally, Tianshan maintained the lowest accuracy among the three subregions, indicating that terrain may be the reason for the low classification accuracy and the importance of considering topographic factors during LULC classification [69], [70].

#### D. Ensemble Learning Classification

Accordingly, this study adopted SVM, RF, ANN, and C4.5 to construct various ensemble learning models for LULC classification in Xinjiang, and its three subregions. The OA and Kappa of each ensemble strategy were compared to select the optimal classification method, and the accuracy results are shown in Table IV.

The greatest level of improvement in LULC classification accuracy was achieved across Xinjiang, with the OA and Kappa increasing by  $\geq 6.85\%$  and  $8.25\%$ , respectively. By subregion,

Tianshan and Northern Xinjiang were the next most improved, with OAs increasing by  $\geq 5.60\%$  and  $5.24\%$ , and Kappa values increasing by  $\geq 6.63\%$  and  $5.99\%$ , respectively. The lowest increase was observed in Southern Xinjiang, where the OA and Kappa increased by  $\geq 5.01\%$  and  $6.25\%$ , respectively. Thus, it was concluded here that ensemble learning classification based on different single classifier combinations can significantly improve LULC classification accuracy across Xinjiang and its three subregions, especially in montane areas.

Moreover, by comparing the different ensemble learning classifications in the three subregions and across Xinjiang, the OA and Kappa of the four-machine learning algorithm—SVM, RF, ANN, and C4.5—ensemble strategy in Tianshan and across Xinjiang was found to be significantly higher than those of any three-machine learning ensemble, where the OA increased by  $1.99\%$ – $2.33\%$  and  $1.88\%$ – $2.75\%$ , and the Kappa increased by  $2.58\%$ – $4.52\%$ , and  $3.12\%$ – $5.48\%$ , respectively. However, in Northern and Southern Xinjiang, the OA values of any three-machine learning ensemble strategies were similar, with an increase of  $\leq 1.00\%$  and  $0.80\%$ , and Kappa increased by  $\leq 0.48\%$  and  $1.01\%$ , respectively (see Table IV). Thus, the number and type of classifiers in ensemble learning have a strong influence, with a theoretical ideal number and type necessary for achieving an optimal classification effect [71], [72].

The significance-test results (see Table V) show that the Z-scores of pairs different ensemble learning combination strategies across Xinjiang and three subregions. There were significant differences between the combination of four different machine learning algorithms (i.e., SVM, RF, C4.5, ANN) and each of the combinations of three different machine learning algorithms (Z



TABLE V  
Z-SCORE TEST WITH ASSOCIATED PROBABILITY VALUE (P) FOR MODEL PAIRS

Comparative pairs	Z score				P value			
	Whole Xinjiang	Northern Xinjiang	Tianshan	Southern Xinjiang	Whole Xinjiang	Northern Xinjiang	Tianshan	Southern Xinjiang
(SVM, RF, ANN) versus (SVM, RF, C4.5)	1.571	0.894	0.874	0.912	0.116	0.371	0.382	0.362
(SVM, RF, ANN) versus (SVM, C4.5, ANN)	0.698	1.023	1.021	0.951	0.485	0.306	0.307	0.342
(SVM, RF, ANN) versus (RF, C4.5, ANN)	0.531	1.121	1.222	0.946	0.595	0.262	0.222	0.344
(SVM, RF, ANN) versus (SVM, RF, C4.5, ANN)	<b>2.281</b>	<b>1.293</b>	<b>2.135</b>	<b>1.234</b>	<b>0.023</b>	<b>0.196</b>	<b>0.033</b>	<b>0.217</b>
(SVM, RF, C4.5) versus (SVM, C4.5, ANN)	1.154	0.789	0.899	0.221	0.249	0.430	0.369	0.825
(SVM, RF, C4.5) versus (RF, C4.5, ANN)	1.311	0.714	0.985	0.356	0.190	0.475	0.325	0.722
(SVM, RF, C4.5) versus (SVM, RF, C4.5, ANN)	<b>2.703</b>	<b>1.112</b>	<b>2.311</b>	<b>1.321</b>	<b>0.007</b>	<b>0.266</b>	<b>0.021</b>	<b>0.187</b>
(SVM, C4.5, ANN) versus (RF, C4.5, ANN)	0.864	0.861	0.564	0.431	0.388	0.389	0.573	0.666
(SVM, C4.5, ANN) versus (SVM, RF, C4.5, ANN)	<b>2.698</b>	<b>1.432</b>	<b>2.451</b>	<b>1.403</b>	<b>0.007</b>	<b>0.152</b>	<b>0.014</b>	<b>0.161</b>
(RF, C4.5, ANN) versus (SVM, RF, C4.5, ANN)	<b>2.342</b>	<b>0.999</b>	<b>2.557</b>	<b>1.039</b>	<b>0.019</b>	<b>0.318</b>	<b>0.011</b>	<b>0.299</b>

The bold values represent the difference in Z-score and P value of each model pair is significant.

score  $\geq 1.96$ ,  $p \leq 5\%$ ) across Xinjiang and Tianshan, while there was no significant difference in northern and southern Xinjiang. This proved that the ensemble learning strategy combining four machine learning algorithms (i.e., SVM, RF, ANN, and C4.5) was superior across Xinjiang and Tianshan; whereas, the strategy combining three algorithms (i.e., SVM, RF, and ANN) worked best for Northern and Southern Xinjiang.

Additionally, compared with the ensemble learning classification across Xinjiang, the ensemble learning in the three subregions can significantly improve the accuracy of LULC classification, with the OAs of Southern Xinjiang, Northern Xinjiang, and Tianshan increasing by 6.86%, 6.70%, and 5.87%, respectively. Thus, geographically subdividing Xinjiang based on topographic conditions effectively reduced the misclassification of vegetation types caused by terrain differences, or highly heterogeneous surface vegetation in the LULC classification across the complex conditions of Xinjiang.

E Ensemble Learning Effects on Classification

Fig. 6 shows the PA and UA of different LULC types throughout Xinjiang and its subregions under different ensemble learning strategies. Compared with machine learning classification (see Fig. 5), ensemble learning significantly improved the PA and UA of different land classes. Among them, in the entire region of Xinjiang, forests (evergreen needleleaf, deciduous needleleaf, deciduous broadleaf, mixed, and shrubs), urban, industrial and mining as well as snow and ice, exhibited the most obvious improvement effects, with the PA and UA of forests increasing by 3.55%–20.22% and 4.81%–14.98%, 11.52% and 12.82%, 14.81% and 11.85%, and 12.44% and 19.68%, respectively. By subregion, the PA and UA improvement effects of Northern Xinjiang were similar to the entire region of Xinjiang

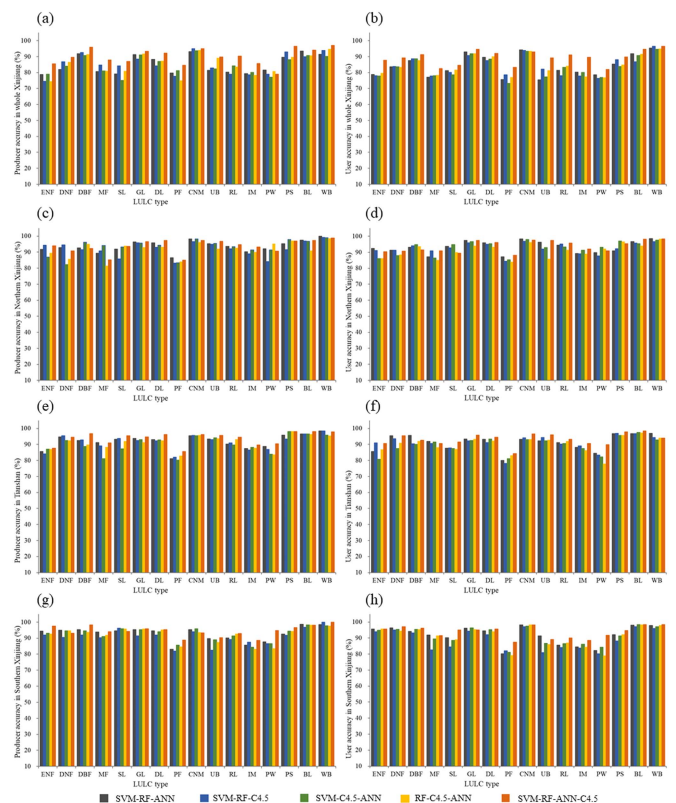


Fig. 6. (a) PA and (b) UA of different LULC types obtained by ensemble learning for the whole Xinjiang. (c) PA and (d) UA for Northern Xinjiang. (e) PA and (f) UA for Tianshan. (g) PA and (h) UA for Southern Xinjiang.

for forests (8.42%–19.93% and 8.54%–13.42%, respectively), urban (15.29% and 11.29%), paddy fields (21.41% and 18.81%),

TABLE VI  
IMPROVEMENT OF CLASSIFICATION PA AND UA (%) UNDER ENSEMBLE LEARNING OF DIFFERENT SUBREGIONS COMPARED WITH WHOLE XINJIANG

LULC type	LULC type	Northern Xinjiang		Tianshan		Southern Xinjiang	
		PA	UA	PA	UA	PA	UA
Forest	<i>Evergreen Needleleaf Forests</i>	6.16	4.58	2.00	2.87	8.78	7.87
	<i>Deciduous Needleleaf Forests</i>	3.17	2.05	4.85	6.11	5.35	7.17
	<i>Deciduous Broadleaf Forests</i>	0.77	1.79	4.97	1.43	3.53	2.98
	<i>Mixed Forests</i>	1.33	4.56	3.06	8.12	5.77	9.35
	<i>Shrublands</i>	4.98	9.04	8.53	6.96	7.62	5.62
Grasslands	<i>Grasslands</i>	2.93	2.88	1.45	1.21	1.95	1.73
	<i>Dry land</i>	4.87	4.60	7.33	6.00	3.68	5.53
Croplands	<i>Paddy Fields</i>	6.41	7.30	7.03	7.24	7.64	7.01
	<i>Cropland/Natural Vegetation Mosaics</i>	2.88	5.41	0.87	4.59	0.35	5.26
Urban and Built-up Lands	<i>Urban</i>	6.38	9.37	7.87	9.37	2.12	3.30
	<i>Rural</i>	8.33	6.35	8.21	7.97	2.32	4.31
	<i>Industrial and Mining</i>	7.60	7.21	7.53	7.69	3.61	3.69
Permanent Wetlands	<i>Permanent Wetlands</i>	13.16	7.84	11.48	7.95	8.82	0.38
Permanent Snow and Ice	<i>Permanent Snow and Ice</i>	-1.08	1.04	1.67	8.11	-3.94	2.24
Bare Lands	<i>Bare Lands</i>	3.42	1.96	3.94	3.84	4.64	3.40
Water Body	<i>Water Body</i>	2.96	1.96	0.85	-2.45	1.46	1.30

industrial and mining (17.85% and 13.98%), as well as snow and ice (11.59% and 13.68%).

The PA and UA in Tianshan increased the most dramatically for forests (6.15%–24.02% and 6.70%–18.01%, respectively), paddy fields (17.36% and 24.56%), industrial and mining (20.67% and 25.54%), and permanent wetlands (14.32% and 13.68%). In southern Xinjiang, the most significant improvement in PA and UA were observed for forests (5.07%–16.12% and 5.73%–24.64%, respectively), paddy fields (10.66% and 15.13%), industrial and mining (13.48% and 10.98%), permanent wetlands (17.84% and 13.85%), as well as snow and ice (20.64% and 12.08%). Overall, the correct classification of forests, permanent wetlands, paddy fields, urban, industrial and mining, as well as snow and ice in ensemble learning, were the keys to its higher classification performance compared to any single classifier.

Table VI shows the comparison of PAs and UAs for Xinjiang and its sub-regions. Compared with the entire region of Xinjiang, subregion ensemble learning showed the following three improvement trends for different land categories: First, shrublands, permanent wetlands, dry land, and paddy fields, showed significant improvement effects, with their PAs and UAs increasing by an average of 7.04% and 7.21%; 11.15% and 6.82%; 5.29% and 5.38%; and 7.03% and 7.18%, respectively. This may be due to the spectral and phenological characteristics of the same vegetation types across different geographical regions being too dissimilar, as it is difficult to identify the same vegetation types using ensemble learning across Xinjiang, whereas subregion level ensemble learning reduced any difference in surface spectra and vegetation phenology within a given vegetation type, thereby improving the classification. Second, the improvement effects of mixed forests, evergreen needleleaf,

deciduous needleleaf, and deciduous broadleaf forests, in addition to grasslands, croplands/natural vegetation mosaics, urban, rural, and industrial and mining were clear in one or two zones. For example, grassland classification improved significantly in northern Xinjiang but not in either of the other two sub-regions. Lastly, snow, ice, and water showed insignificant improvement effect and even negative correlation in the three sub-regions. For example, the PA of snow and ice in Northern Xinjiang and Southern Xinjiang decreased by -1.08% and -3.94%, respectively; whereas the UA of water in Tianshan decreased by -2.45%, possibly because these LULC types are unique, and do not require ensemble learning methods to identify them. The spectral signature for these land classes may be an overfit, thus resulting in the mild decrease in accuracy observed. Furthermore, the ensemble learning under complex topographic conditions here may have been limited by too few training samples.

#### F. LULC Products Based on Ensemble Learning

Fig. 7 shows the spatial distribution of LULC in Xinjiang in 2015 based on ensemble learning classification, in which the three-machine learning (SVM, RF, and ANN) ensemble strategy was used in Northern and Southern Xinjiang, and four-machine learning (including C4.5) ensemble strategy was used in Tianshan. Evergreen needleleaf forests were mainly distributed in the Altai Mountains of Northern Xinjiang. Alternatively, deciduous needleleaf and mixed forests were mainly distributed in Tianshan, whereas deciduous broadleaf and shrub forests were mainly distributed on the oasis edge of Southern Xinjiang. Grassland and sparse vegetation were mainly distributed in the Altai Mountains, Tianshan, Kunlun Mountains, and desert transition zone of northern Xinjiang; furthermore, grasslands

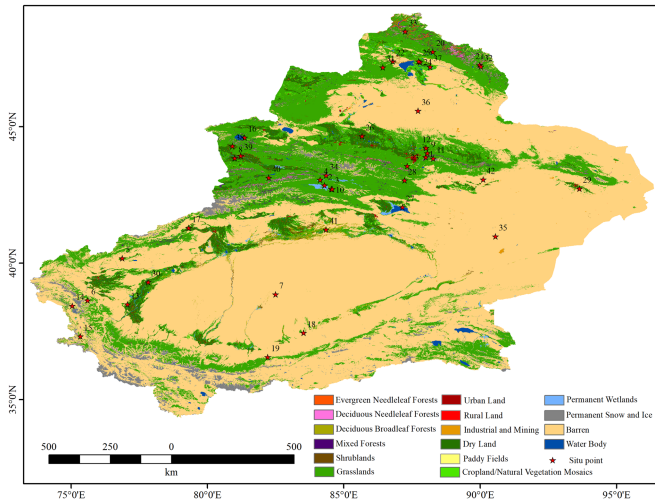


Fig. 7. Spatial distribution of LULC in Xinjiang.

at the edge of the oasis were also abundant. Dry land, paddy fields, urban, and rural were mainly distributed near river oases; wetlands around rivers and lakes; and snow/ice was mainly distributed in the Tianshan and Kunlun Mountains.

Based on the proposed method of LULC classification in Xinjiang, a map of LULC data (resolution 500 m), including the entire of Xinjiang, and its three sub-regions under IGBP classification scheme was generated. The coordinate system is WGS84 coordinate system using horizontal Mercator projection. Data can be opened, viewed, and processed using Esri ArcGIS software. Our findings can be used for climate change simulations, as well as for assessing ecological services and the hydrological cycle across Xinjiang.

## VI. DISCUSSION

This section further discusses potential improvements, comparison with other land-cover products, broader application, and future scope of this study.

### A. Limitations

In this study, owing to limited data sources, various remote sensing data with different spatial resolutions were resampled into coarse resolutions to produce the final land cover product at 500 m. However, raster data model is a representation using regular grids and the discrete cells in a raster structure cannot reflect the accurate boundary of spatial objects. Therefore, the process of resampling high-resolution data to coarse resolutions inevitably suffers from a loss of spatial information. The properties of a spatial object such as the area, shape, and location can be lost during the process [73]. The consistency of topological relations between different spatial objects may also be lost [74]. This introduces different types of errors and uncertainties when further employing the resampled product for other research.

For classified land-use data, effective approaches to reduce the loss in spatial properties have been proposed. Specifically, many studies have focused on the vector-to-raster conversion and have

put forward different methods to preserve an individual spatial property of area, topology, or shape [75], [76], and to achieve balanced preservation of various properties [77]. In addition, some researchers have paid attention to the resampling process and presented an area-preserving method for maintaining the area property after resampling categorical raster data [78]. In this study, we have chosen the majority sampling method to upscale the remote sensing data to reduce the loss in spatial properties. However, the errors and uncertainties introduced by the resampling process still remain unclear. In future, we would like to evaluate such type of errors and investigate effective approaches to further preserve important spatial properties for land cover classification.

### B. Comparison With Other Land Cover Products

The difference between classification schemes is one of the main factors leading to inconsistent classification results [16]. This also makes rigorous comparison and collaborative use of different maps challenging [13], [23]. Moreover, to compare land cover datasets that use different schemes, it is generally necessary to convert land cover products with more detailed classes to land cover classification schemes with fewer classes. This may result in loss of the ability to describe detailed land cover features and the conversion is not possible, such as forest land in CAS classification scheme cannot be converted to IGBP classification scheme [79].

Considering that main purpose of this study was to generate a new land cover data product that can be suitable for simulating climate change across Xinjiang, the internationally used IGBP land cover classification system is optimal for famous climate models such as WRF model and MM5 model for land cover classification [80]. Because the classification system of MCD12Q1 [81] and CCI-LC [12] land-cover products is an IGBP classification system or can be easily converted to an IGBP classification system, we selected the resolution of MCD12Q1 and CCI-LC land cover products to be 500 m by field sampling method, and compared the accuracy of these products with our product. As can be seen from Table S2 (see the Supporting Information), the classification accuracy of MCD12Q1 and CCI-LC land cover products in Xinjiang was only 59.52% (25 out of 42 sampling points are correct) and 66.67% (28 out of 42 sampling points are correct), which is consistent with the results of previous studies [16], [23], [79]. In comparison, the sampling accuracy of our land cover products reached 88.09% (37 out of 42 sampling points are correct).

### C. Ensemble Learning Strategy of This Study

In this study, Xinjiang, with its highly heterogenous LULC under complex topographical conditions, was chosen as a representative study region and was divided into three sub-regions according to the climatic and topographic characteristics. Then different ensemble learning strategies were formulated, and the accuracy of land cover classification was compared for each sub-region. Furthermore, significance test results (see Table V) show the Z-scores of pairs different ensemble learning combinations strategies across Xinjiang and its three subregions.



The results of this study show that in the complex topographic area of Tianshan compared with the three different ensemble learning enabled land cover classifications, the four machine learning algorithm ensemble strategies in Tianshan can improve the OA by at least 1.99%–2.33% (see Table IV), and the difference of classification accuracy passed the significance test of  $Z$  score ( $Z \geq 1.96$  and  $p \leq 0.05$ ) (see Table V). In Tianshan, with its complex topographical conditions and extremely uneven spatial distribution of LULC, the integration of more machine learning algorithms may be necessary for greater accuracy; using fewer algorithms may be insufficient for constructing an ensemble, which is stronger than the single classifier in LULC classification of Tianshan [71]. However, the OA for the northern Xinjiang and southern Xinjiang could only be improved by at most 0.99% and 0.80%, and there was no significant difference. As previous studies have indicated, an optimal number of component classifiers for an ensemble can obtain the most accurate results with these classifiers. Increasing or decreasing the number of classifiers from this ideal point may worsen the prediction or not add any benefit to the overall performance [72], [82]. Some studies suggest the number of class labels in a dataset as the ideal number of component classifiers [72], [83]. However, the real-world data are very complex and it is still challenging to determine the ideal number of classifiers in ensemble learning. More machine learning in ensemble learning classification might not always yield better classification effect [72]. For example, compared with the combination of three machine learning algorithms (i.e., SVM, RF, ANN), the OA of the combination of three machine learning algorithms (i.e., SVM, RF, C4.5, ANN) only improved by 0.04% and 0.51% in Northern and Southern Xinjiang.

#### D. Broader Application and Future Work

Previous studies have shown that the complexity of topographic conditions is the main factor leading to the high heterogeneity of LULC estimates, which greatly limits the improvement of LULC classification accuracy [24]. Accordingly, the influence of topographic conditions must be considered when classifying LULC for large areas [69]. To reduce the impact of topographic factors on ensemble learning classification, it is necessary to subdivide complex topographic areas according to climate and topographic characteristics. The method used here has been proved to be effective for improving accuracy. According to this logic, geographically subdividing the study area based on topographic conditions can potentially be an effective method for further improving the classification accuracy of typical arid regions [34], [84]. Therefore, the subregions used here could potentially be subdivided further, especially the Tianshan region, for additional improvements in accuracy [60], [85], [86].

Future efforts should focus on following factors: First, five primary types of machine learning algorithms (KNN, SVM, RF, ANN, and C4.5) were applied in the present study. A larger number or more diverse assortment of machine learning algorithms could be employed for more efficient ensemble learning. In addition, the stacking strategy is primarily used in ensemble learning for classification, and other diversity strategies should be tested in LULC classification; for example, Bagging and

Boosting, could also be assessed for their capacity to improve accuracy. Moreover, Xinjiang is only divided into three regions: Northern Xinjiang, Tianshan Mountain, and Southern Xinjiang. Future research needs to further subdivide Xinjiang to improve the classification accuracy of land cover. Such as, the Tianshan can be subdivided into the North Tianshan, the East Tianshan, and the West Tianshan [34]. Furthermore, the presented ensemble learning approach, which considers the difference in various geographical regions, can also be applied in other global locations with similar geographical characteristics, including montane areas with complex topographic conditions and arid desert-oasis mosaic landscapes with highly heterogeneous surface vegetation types. Lastly, the resolution of the generated LULC products in this study was 500 m, and higher resolution remote sensing imagery could be combined for achieving an improved resolution of the results.

## VII. CONCLUSION

Here, an ensemble learning approach was proposed for performing LULC classification in Xinjiang, northwest China. The study area was divided into three subregions according to climate and topographic characteristics, and five machine learning algorithms—KNN, SVM, RF, ANN, and C4.5—were integrated to develop different ensemble learning strategies for LULC classification. The accuracy and efficiency of each proposed ensemble learning approach were evaluated and analyzed, and the following three primary conclusions were drawn.

First, compared with individual machine learning algorithms, the ensemble learning strategy proposed here significantly improved LULC classification accuracy, with the approach having the greatest effect on Tianshan (OA and Kappa values were increased by 5.60% and 6.63%, respectively), followed by Northern Xinjiang (5.24% and 5.99%), and Southern Xinjiang (5.01% and 6.25%). In addition, the correct classification of different forest types via ensemble learning was the main contributor driving the higher classification accuracy.

Second, optimal combinations of machine learning algorithms were revealed for use in ensemble learning so that LULC classification was optimized for each subregion. Specifically, adopting SVM, RF, and ANN in the ensemble learning strategy was the most efficient for Northern and Southern Xinjiang (OA values of 96.35% and 96.92%, respectively), whereas a strategy employing SVM, RF, ANN, and C4.5 for ensemble learning was the most efficient for Tianshan (OA = 96.33%).

Lastly, the impacts of the proposed ensemble learning approach on classifying different land types could be summarized into three scenarios: First, the proposed approach was efficient for all three sub-regions, particularly for land types including shrublands, permanent wetlands, dry land, and paddy fields. Second, the approach was more successful for one or two subregions with mixed forests, evergreen needleleaf, deciduous needleleaf, and deciduous broadleaf forests, in addition to grasslands, croplands/natural vegetation mosaics, urban, rural, and industrial and mining. Third, the approach was not successful in any of the three subregions with land types of snow, ice, and water.

In summary, we introduced complex terrain and climate conditions of Xinjiang into land-cover classification, and developed a new ensemble learning method to achieve accurate land-cover classification. Furthermore, we produced a new land-cover product with a two-level classification system, which solves the problem of low accuracy of land-cover classification in Xinjiang climate simulation and can be used in other research areas, such as territorial space planning and hydrological simulation.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. L. Han and Dr. Z. Wang of the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, for their selfless assistance in collecting the field sampling sites. The authors also sincerely thank the editors and anonymous reviewers for their constructive comments, which have helped improve this paper. The datasets generated in this study have been deposited in Figshare and can be downloaded from the DOI: <https://doi.org/10.6084/m9.figshare.21006883.v1>.

#### REFERENCES

- [1] A. Di Gregorio, *Land Cover Classification System: Classification Concepts and User Manual: LCCS*. Rome, Italy: Food Agriculture Org., 2005.
- [2] A. Letourneau, P. H. Verburg, and E. Stehfest, "A land-use systems approach to represent land-use dynamics at continental and global scales," *Environ. Modelling Softw.*, vol. 33, pp. 61–79, 2012.
- [3] Z. Zhang et al., "A 2010 update of National Land Use/Cover Database of China at 1: 100000 scale using medium spatial resolution satellite images," *Remote Sens. Environ.*, vol. 149, pp. 142–154, 2014.
- [4] K. Liu and E. Xu, "Fusion and correction of multi-source land cover products based on spatial detection and uncertainty reasoning methods in central Asia," *Remote Sens.*, vol. 13, no. 2, Jan. 2021.
- [5] T. R. Loveland and A. Belward, "The IGBP-DIS global 1km land cover data set, DISCover: First results," *Int. J. Remote Sens.*, vol. 18, no. 15, pp. 3289–3295, 1997.
- [6] T. R. Loveland et al., "Development of a global land cover characteristics database and IGBP discover from 1 km AVHRR data," *Int. J. Remote Sens.*, vol. 21, no. 6/7, pp. 1303–1330, Apr. 2000.
- [7] S. P. Seitzinger et al., "International Geosphere-Biosphere Programme and Earth system science: Three decades of co-evolution," *Anthropocene*, vol. 12, pp. 3–16, 2015.
- [8] M. C. Hansen and B. Reed, "A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products," *Int. J. Remote Sens.*, vol. 21, no. 6/7, pp. 1365–1373, 2000.
- [9] E. Bartholome and A. S. Belward, "GLC2000: A new approach to global land cover mapping from Earth observation data," *Int. J. Remote Sens.*, vol. 26, no. 9, pp. 1959–1977, 2005.
- [10] S. Faroux, A. Kaptué Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. L. Moigne, "ECOLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models," *Geoscientific Model Develop.*, vol. 6, no. 2, pp. 563–582, 2013.
- [11] M. A. Friedl et al., "Global land cover mapping from MODIS: Algorithms and early results," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 287–302, Nov. 2002.
- [12] P. Defourny et al., "Land Cover CCI," Product User Guide Version, 2012, vol. 2, p. 325.
- [13] P. Gong et al., "Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.
- [14] J. Chen et al., "Global land cover mapping at 30 m resolution: A POK-based operational approach," *ISPRS J. Photogrammetry Remote Sens.*, vol. 103, pp. 7–27, 2015.
- [15] J. Liu et al., "Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s," *J. Geographical Sci.*, vol. 24, no. 2, pp. 195–210, 2014.
- [16] Y. Yang, P. Xiao, X. Feng, and H. Li, "Accuracy assessment of seven global land cover datasets over China," *ISPRS J. Photogrammetry Remote Sens.*, vol. 125, pp. 156–173, Mar. 2017.
- [17] R. Tateishi et al., "Production of global land cover data - GLCNMO," *Int. J. Digit. Earth*, vol. 4, no. 1, pp. 22–49, 2011.
- [18] J. J. Feddema et al., "The importance of land-cover change in simulating future climates," *Science*, vol. 310, no. 5754, pp. 1674–1678, 2005.
- [19] C. McAlpine et al., "A continent under stress: Interactions, feedbacks and risks associated with impact of modified land cover on Australia's climate," *Glob. Change Biol.*, vol. 15, no. 9, pp. 2206–2223, 2009.
- [20] J. Ge, J. Qi, B. M. Lofgren, N. Moore, N. Torbick, and J. M. Olson, "Impacts of land use/cover classification accuracy on regional climate simulations," *J. Geophysical Res.-Atmospheres*, vol. 112, no. D5, pp. 1–12, Mar. 2007.
- [21] F. Wu, J. Zhan, H. Yan, C. Shi, and J. Huang, "Land cover mapping based on multisource spatial data mining approach for climate simulation: A case study in the farming-pastoral ecotone of North China," *Adv. Meteorol.*, vol. 2013, pp. 1–12, 2013.
- [22] C. Giri, Z. Zhu, and B. Reed, "A comparative analysis of the global land cover 2000 and MODIS land cover data sets," *Remote Sens. Environ.*, vol. 94, no. 1, pp. 123–132, 2005.
- [23] M. Herold, P. Mayaux, C. E. Woodcock, A. Baccini, and C. Schmullius, "Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2538–2556, May 2008.
- [24] V. F. Rodriguez-Galiano and M. Chica-Rivas, "Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and digital terrain models," *Int. J. Digit. Earth*, vol. 7, no. 6, pp. 492–509, 2014.
- [25] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience Remote Sens.*, vol. 57, no. 1, pp. 1–20, 2020.
- [26] X. Hu, P. Zhang, Q. Zhang, and J. Wang, "Improving wetland cover classification using artificial neural networks with ensemble techniques," *GIScience Remote Sens.*, vol. 58, no. 4, pp. 603–623, 2021.
- [27] D. Phiri and J. Morgenroth, "Developments in Landsat land cover classification methods: A review," *Remote Sens.*, vol. 9, no. 9, 2017, Art. no. 967.
- [28] G. Lei et al., "OIC-MCE: A practical land cover mapping approach for limited samples based on multiple classifier ensemble and iterative classification," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 987.
- [29] G. Foody, D. Boyd, and C. Sanchez-Hernandez, "Mapping a specific class with an ensemble of classifiers," *Int. J. Remote Sens.*, vol. 28, no. 8, pp. 1733–1746, 2007.
- [30] F. Li, L. Xu, P. Siva, A. Wong, and D. A. Clausi, "Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2427–2438, Jun. 2015.
- [31] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.
- [32] Y. Chen, P. Dou, and X. Yang, "Improving land use/cover classification with a multiple classifier system using Adaboost integration technique," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 1055.
- [33] C. A. Di Vittorio and A. P. Georgakakos, "Land cover classification and wetland inundation mapping using MODIS," *Remote Sens. Environ.*, vol. 204, pp. 1–17, 2018.
- [34] G. Ge, Z. Shi, Y. Zhu, X. Yang, and Y. Hao, "Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms," *Glob. Ecol. Conservation*, vol. 22, 2020, Art. no. e00971.
- [35] B. Martin, "Instance-based learning: Nearest neighbour with generalisation," Univ. Waikato, Hamilton, New Zealand, Working Paper, 1995.
- [36] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," in *Proc. Int. Conf. Data Softw. Eng.*, 2015, pp. 170–174.
- [37] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [38] V. Cherkassky, "The nature of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 8, no. 6, pp. 1564–1564, Nov. 1997.
- [39] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 1997, pp. 281–287.
- [40] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

- [41] W. Su et al., "Landsat-based long-term lucc mapping in Xinlicheng Reservoir Basin using object-based classification," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 64, 2017, Art. no. 012024.
- [42] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.
- [43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [44] P. Coppin, E. Lambin, I. Jonckheere, and B. Muys, "Digital change detection methods in natural ecosystem monitoring: A review," in *Analysis of Multi-Temporal Remote Sensing Images*. Singapore: World Scientific, 2002, pp. 3–36.
- [45] S. Wang, "Artificial neural network," in *Interdisciplinary Computing in Java Programming*. Berlin, Germany: Springer, 2003, pp. 81–100.
- [46] T. Kavzoglu and P. M. Mather, "The use of backpropagating artificial neural networks in land cover classification," *Int. J. Remote Sens.*, vol. 24, no. 23, pp. 4907–4938, 2003.
- [47] J. R. Quinlan, *C4. 5: Programs for Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2014.
- [48] M. Yu and T. Ai, "Study of RS data classification based on rough sets and C4. 5 algorithm," *Proc. SPIE*, vol. 7492, 2009, Art. no. 74920B.
- [49] K. Tan et al., "Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning," *J. Hazardous Mater.*, vol. 401, 2021, Art. no. 123288.
- [50] M. Graczyk, T. Lasota, B. Trawiński, and K. Trawiński, "Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal," in *Proc. Intell. Inf. Database Syst.*, 2010, pp. 340–350.
- [51] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007.
- [52] Z. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC, 2019.
- [53] T. Wang, C. Yan, X. Song, and J. Xie, "Monitoring recent trends in the area of aeolian desertified land using Landsat images in China's Xinjiang region," *ISPRS J. Photogrammetry Remote Sens.*, vol. 68, pp. 184–190, 2012.
- [54] B. Wang et al., "Changes in extreme events of temperature and precipitation over Xinjiang, northwest China, during 1960–2009," *Quaternary Int.*, vol. 298, pp. 141–151, 2013.
- [55] H. Du et al., "Simulation and estimation of future precipitation changes in arid regions: A case study of Xinjiang, Northwest China," *Climatic Change*, vol. 167, no. 3/4, 2021, Art. no. 43.
- [56] L. Lu, C. Kuenzer, H. Guo, Q. Li, T. Long, and X. Li, "A novel land cover classification map based on a MODIS time-series in Xinjiang, China," *Remote Sens.*, vol. 6, no. 4, pp. 3387–3408, Apr. 2014.
- [57] T. Kassawmar, S. Eckert, K. Hurni, G. Zeleke, and H. Hurni, "Reducing landscape heterogeneity for improved land use and land cover (LULC) classification across the large and complex Ethiopian highlands," *Geocarto Int.*, vol. 33, no. 1, pp. 53–69, 2018.
- [58] J. R. Schriever and R. G. Congalton, "Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the Northeast," *Photogrammetric Eng. Remote Sens.*, vol. 61, no. 3, pp. 321–327, 1995.
- [59] C. Zhang et al., "Evaluation and intercomparison of high-resolution satellite precipitation estimates—GPM, TRMM, and CMORPH in the Tianshan Mountain Area," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1543.
- [60] S. Chen et al., "Added value of a dynamical downscaling approach for simulating precipitation and temperature over Tianshan mountains area, Central Asia," *J. Geophysical Res., Atmospheres*, vol. 124, no. 21, pp. 11051–11069, 2019.
- [61] T. Fan, Y. Zhu, and J. Chen, "A new measure of classifier diversity in multiple classifier system," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2008, pp. 18–21.
- [62] R. G. Congalton, R. G. Oderwald, and R. A. Mead, "Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques," *Photogrammetric Eng. Remote Sens.*, vol. 49, no. 12, pp. 1671–1678, 1983.
- [63] X. Huang, C. Weng, Q. Lu, T. Feng, and L. Zhang, "Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas," *Remote Sens.*, vol. 7, no. 12, pp. 16024–16044, 2015.
- [64] W. Shang, H. Huang, H. Zhu, Y. Lin, Z. Wang, and Y. Qu, "An improved kNN algorithm—fuzzy kNN," in *Proc. Int. Conf. Comput. Inf. Sci.*, 2005, pp. 741–746.
- [65] P. H. Akwensi, E. T. Brantson, J. N. Niipele, and Y. Y. Ziggah, "Performance evaluation of artificial neural networks for natural terrain classification," *Appl. Geomatics*, vol. 13, no. 3, pp. 453–465, 2021.
- [66] S. S. Mullick, S. Datta, and S. Das, "Adaptive learning-based  $k$ -nearest neighbor classifiers with resilience to class imbalance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5713–5725, Nov. 2018.
- [67] B. B. Damodaran and R. R. Nidamanuri, "Dynamic linear classifier system for hyperspectral image classification for land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2080–2093, Jun. 2014.
- [68] D. C. Duro, S. E. Franklin, and M. G. Dubé, "Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4502–4526, 2012.
- [69] J. Buchner et al., "Land-cover change in the Caucasus Mountains since 1987 based on the topographic correction of multi-temporal landsat composites," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111967.
- [70] H. Keshkar, W. Voigt, and E. Alizadeh, "Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery," *Arabian J. Geosci.*, vol. 10, no. 6, 2017, Art. no. 154.
- [71] H. R. Bonab and F. Can, "A theoretical framework on the ideal number of classifiers for online ensembles in data streams," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, 2016, pp. 2053–2056.
- [72] H. Bonab and F. Can, "Less is more: A comprehensive framework for the number of components of ensemble classifiers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2735–2745, Sep. 2019.
- [73] A. Galton, "Granularity-sensitive spatial attributes," *Spatial Cogn. Comput.*, vol. 3, no. 2/3, pp. 97–118, 2003.
- [74] P. Bettinger, G. A. Bradshaw, and G. W. Weaver, "Effects of geographic information system vector-raster-vector data conversion on landscape indices," *Can. J. Forest Res.*, vol. 26, no. 8, pp. 1416–1425, 1996.
- [75] C. Zhou, Y. Ou, L. Yang, and B. Qin, "An equal area conversion model for rasterization of vector polygons," *Sci. China Ser. D, Earth Sci.*, vol. 50, no. 1, pp. 169–175, 2007.
- [76] C. Zhou, D. Li, N. Xiao, Z. Chen, X. Li, and M. Li, "A topology-preserving polygon rasterization algorithm," *Cartography Geographic Inf. Sci.*, vol. 45, no. 6, pp. 495–509, 2018.
- [77] X. Huo, C. Zhou, Y. Xu, and M. Li, "A methodology for balancing the preservation of area, shape, and topological properties in polygon-to-raster conversion," *Cartography Geographic Inf. Sci.*, vol. 49, no. 2, pp. 115–133, 2022.
- [78] J. M. Johnson and K. C. Clarke, "An area preserving method for improved categorical raster resampling," *Cartography Geographic Inf. Sci.*, vol. 48, no. 4, pp. 292–304, 2021.
- [79] M. Zhang, M. Ma, P. De Maeyer, and A. Kurban, "Uncertainties in classification system conversion and an analysis of inconsistencies in global land cover products," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 4, 2017, Art. no. 112.
- [80] W. C. Skamarock et al., "A description of the advanced research WRF version 3," NCAR, Boulder, CO, USA, NCAR Tech. Note TN-4681STR, 2008.
- [81] M. A. Friedl et al., "MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 168–182, 2010.
- [82] P. Latinne, O. Debeir, and C. Decaestecker, "Limiting the number of trees in random forests," in *Proc. Mult. Classifier Syst., 2nd Int. Workshop*, 2001, pp. 178–187.
- [83] D. Hernández-Lobato, G. Martínez-Muñoz, and A. Suárez, "How large should ensembles of classifiers be?," *Pattern Recognit.*, vol. 46, no. 5, pp. 1323–1336, 2013.
- [84] H. Wang et al., "Impacts of topography on the land cover classification in the Qilian Mountains, Northwest China," *Can. J. Remote Sens.*, vol. 46, no. 3, pp. 344–359, 2020.
- [85] Y. Liu, L. Li, X. Chen, R. Zhang, and J. Yang, "Temporal-spatial variations and influencing factors of vegetation cover in Xinjiang from 1982 to 2013 based on GIMMS-NDVI3g," *Glob. Planet. Change*, vol. 169, pp. 145–155, 2018.
- [86] Q. Zhang, P. Sun, J. Li, V. P. Singh, and J. Liu, "Spatiotemporal properties of droughts and related impacts on agriculture in Xinjiang, China," *Int. J. Climatol.*, vol. 35, no. 7, pp. 1254–1266, 2015.





**Haoyang Du** is currently working toward the Ph.D. degree in geography with the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

His research interests include land cover mapping, geographic spatiotemporal data mining, and the application of geographic information systems.



**Yunyun Xu** is currently working toward the M.S. degree in human geography with the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

Her research interests include spatial Big Data analysis and mining and land use mapping.



**Manchun Li** received the Ph.D. degree in geography from Nanjing University, Nanjing, China, in 1997.

He is currently a Professor with the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University. His research interests include land cover mapping, GIS and remote sensing applications.



**Chen Zhou** received the Ph.D. degree in geography from Nanjing University, Nanjing, China, in 2018.

He is currently a Researcher with the School of Geography and Ocean Sciences, Nanjing University. His research interests include GIS spatial analysis, land cover mapping, remote sensing applications, and data integration.