ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

Land Use Classification and Driving Factor Analysis in Taihang Mountain Area Based on GEE

Jinying Tang[1, a], Haitao Jing[1, b, *], Panpan Liu[1, c], Qi Liu[1, d] and Shi He[1, e]

[1] Department of Surveying and Mapping, Henan Polytechnic University, Henan, Chinese

[a] tjy19960220@163.com, [b] jht@hpu.edu.cn, [c]212004020001@home.hpu.edu.cn, [d]1217427369@qq.com, [e]heshi@hpu.edu.cn

Abstract

Taking the land use and driving factors in Taihang Mountain Area as the research object, the purpose is to provide reference for ecological environmental protection, water and soil conservation and livelihood improvement in mountains areas. This study made full use of Sentinel optical images, ESA WorldCover 10m land use and ALOS DEM data to analyze the importance of red edge, spectrum, indices and terrain characteristics to the accuracy of land use classification in a large region based on GEE cloud platform. Then, the random forest algorithm was used to classify meticulously land use type in Taihang Mountains, and the drivers of the land use variations was analyzed combined with socio economic, meteorological and environmental data on the basis of PLUS model. The results show that, (1) Sentinel-2's red edge index NDVI_ re2, slope and elevation are more conducive to the classification of land use types, and the importance scores are 48.15, 46.79 and 45.60 respectively; (2) the overall accuracy and Kappa coefficient are 85% and 0.81 through feature optimization, respectively; (3) the significant factors affecting land use change in Taihang Mountains are elevation and rainfall.

Keywords

Taihang Mountains; Land use; Google Earth Engine; Random forest algorithm; PLUS model.

1. INTRODUCTION

Mountainous areas account for more than half of China's land area, and their land use changes have always attracted the attention of the scientific community [1-2]. The Taihang Mountain Range belongs to the important mountain range and geographical dividing line in the eastern region of China, located on the eastern edge of the second step of China's topography. [3]. The process of land use change not only affects the biocarbon cycle in Taihang Mountain, but also directly affects the sustainable development of Taihang Mountain [4-6]. At the same time, land use change in this area is a characterization signal of the direct impact of human activities on the natural state of land surface. Therefore, research on spatiotemporal changes in land use is essential to restore ecological carrying capacity, ensure food security and effective supply of agricultural products.

In recent years, scholars at home and abroad have carried out a lot of research on land use change and its driving mechanism, and obtained many research results. The concept of land use dates back to the study of land conditions in southern Germany by German scholars in the early 19th century, and then gradually emerged [7]. S. Bansal et al. [8] analyzed land use change and driving factors, and concluded that socioeconomic, demographic and other driving factors affect land use change in India's densely populated Yamuna River Basin. With the proposal of land use

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

patterns, the research on its driving forces has been widely concerned, Cao Na et al. [9] took the Hebei section of the Taihang Mountain Range as an example, combined with socio-economic factors, and qualitatively and quantitatively discussed the potential impacts of population, economy, science and technology and policy on land use change. Zhao Chenguang et al. [10] studied land use change in the artificial vegetation restoration area in the northeast of the Tengger Desert, and found that the main driving factors were artificial afforestation area and rainfall. The driving factors of land use change in many study areas are mainly elevation, temperature, population, slope, etc. [11-13], among which the driving force of land use change in mountainous areas is mainly elevation [14].

In this study, based on the GEE cloud platform, combined with Sentinel-2 Level-1C data and ALOS DEM data with 12.5 meters spatial resolution and land use data with 10 meters spatial resolution, the random forest algorithm was used to carry out the land use classification accuracy assessment research in the Taihang Mountain area, and the land use changes and driving factors were studied. PLUS is then used to analyze land-use changes and drivers. The temporal and spatial changes of land use in Taihang Mountain were studied, and the driving factors of the changes were analyzed, in order to have reference significance for environmental protection, ecological construction and rational utilization of land resources in Taihang Mountain.

2. STUDY AREA DATA

2.1. Overview of the study area

As an important geographical dividing line, the North China Plain in the east belongs to deciduous broad-leaved forests, while the Loess Plateau in the west is dry grassland and forest steppe zone [15]. Location: 113°17'E~114°23'E and 35°32'N~36°04'N, including 115 county-level administrative districts, with a total area of about 13.7×104 km^2, an average altitude of 861 m, and the terrain is high in the northwest and low in the southeast [16,17] (Figure 1). The climate of the Taihang Mountains belongs to the temperate semi-humid continental monsoon climate, with an average annual temperature of 9-11 °C, an average annual precipitation of 500-520mm, and significant intra-annual changes in precipitation and temperature [18]. The main land use types are arable land, forest land and grassland, accounting for more than 90% of the total area. With the disorderly change of land for production and living functions, the contradiction between people and land has intensified.

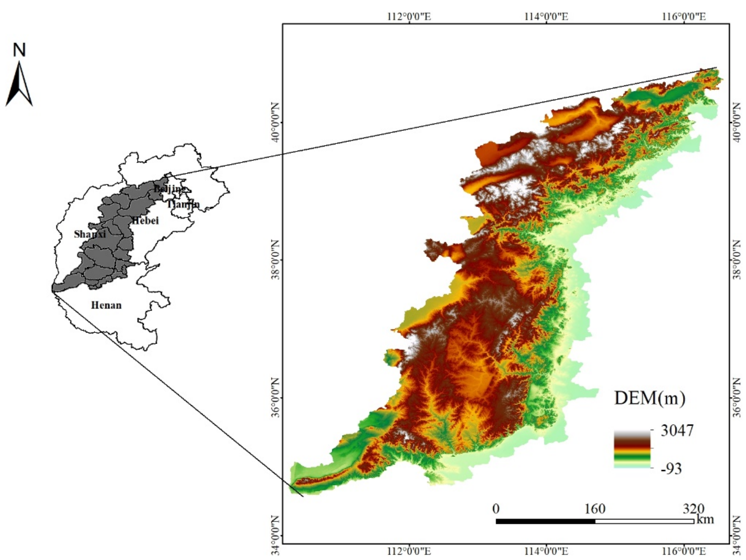

Figure 1. The location and Sentinel-2 imagery of the study region in 2021

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

2.2. Data sources and preprocessing

In this study, Sentinel-2 Level-1C data, land use data with 10 meters spatial resolution, and ALOS DEM data resampled to 10 meters were used as data sources, and the spectral index and topographic features of the Taihang Mountain area were extracted from the GEE platform, and all the calculated indices were added as new bands to the preprocessed Sentinel-2 images to make the feature layers required for classification. Based on ESA's 10-meter land use data, random sampling was carried out in the made feature layer, training data and validation data were generated, and the samples were divided into 696 training samples and 304 validation samples according to the ratio of 7:3.

Sentinel-2 remote sensing data is classified on the GEE cloud platform, and the raster data of nine land use drivers in the Taihang Mountain area are selected, and the land use change in the Taihang Mountain is analyzed by using the PLUS model [19]. The data from this study are shown in the table below (Table 1).

Table 1. The dataset used in this study

Directory	Data	Time	Spatial resolution	Data source
Data used to classify	Sentinel-2 Level-1C	2016-2021	10m	Google Earth Engine
	ESA WorldCover 10m v100	2020	10m	Google Earth Engine
	ALOS DEM	2011	12.5m	PIE-Engine Studio (https://engine.piesat.cn/)
Socioeconomic drivers	Population GDP	2015	1000m	Resource and Environment Science and Data Center (http://www.resdc.cn/DOI)
	Distance from railway Distance from road	2020	1000m	National Catalogue Service For Geographic Information (https://www.webmap.cn/)
Meteorological and environmental drivers	Annual precipitation Annual temperature	2000-2020	1000m	National Earth System Science Data Center (http://www.geodata.cn/)
	DEM Slope Aspect	2011	12.5m	PIE-Engine Studio (https://engine.piesat.cn/)

3. RESEARCH METHODS

3.1. Research ideas

Firstly, the time, region and cloud cover of Sentinel-2 data are screened, and the filtered data is preprocessed by de-clouding, mosaicking and clipping, and classification features are constructed, including spectral features, index calculation and terrain analysis, and the feature set required for classification is obtained and all feature values are normalized. Then, based on ESA's global land use data with a spatial resolution of 10 meters in 2020, the data in the Taihang Mountains were screened out and the categorical values in the data were reordered to 0-10 for subsequent accuracy assessment. The categorical values and attributes of land use data in the Taihang Mountain Area were randomly sampled as label values to generate categorical samples. The random forest algorithm [20] is used to analyze the importance of feature variables, select features, and evaluate the classification and accuracy. The specific flowchart of the research idea is shown in Figure 2.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

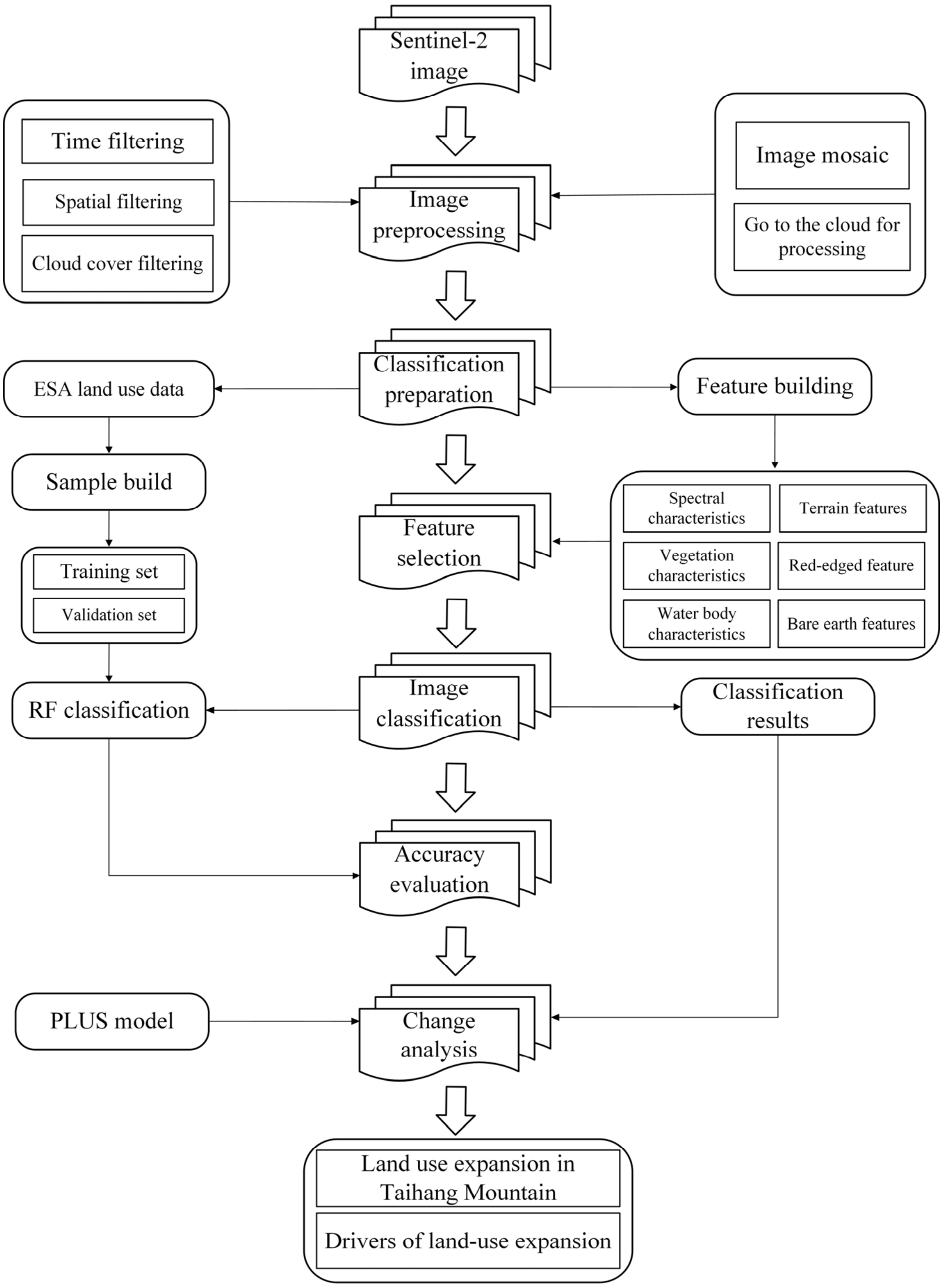

Figure 2. The flow chart of overall research ideas

3.2. Categorical feature index construction

In this study, a dataset of 34 categorical features including spectral information, remote sensing index and terrain factor was constructed. The spectral characteristics are 13 bands of Sentinel-2 images; Based on band calculation, 18 remote sensing indices were obtained, including vegetation index, water index, bare ground and building index, and red edge index [21], among which the red edge index has high value for vegetation classification [22]. Topographic features are important indicators for the classification of mountain land use types, and the ALOS DEM data resampled to 10 meters is imported into GEE and two terrain indices, slope and aspect, are calculated as the classification characteristics of this study. In summary, according to the distribution characteristics of land cover types in the Taihang Mountain area, six classification characteristics were constructed: spectrum, vegetation, water body, bare land and architecture, red edge and topography. The names and quantities of each characteristic variable are shown in Table 2.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

Table 2. Categorical feature set

Characteristic category	Feature naming	quantity
Spectral characteristics	Aerosols, Blue, Green, Red, Red Edge 1, Red Edge 2, Red Edge 3, NIR, Red Edge 4, Water vapor, Cirrus, SWIR 1, SWIR 2	13
Vegetation index	NDVI, GNDVI, EVI, SAVI, MAVI, ARVI	6
Water index	NDWI, MNDWI, RNDWI, EWI	4
Building and bare land index	NDBI, BSI	2
Red edge index	NDVI_re1, NDVI_re2, NDVI_re3, ND_re1, ND_re2, CL_re	6
Terrain index	DEM, slope, aspect	3

3.3. Random forest algorithm

The random forest algorithm, proposed by Breiman in 2001, inherits the Bagging ensemble learning theory and the nonparametric classification and regression method of random space molecules [22, 23], and can effectively run huge data sets and process thousands of input variables [24]. In the learning stage, the random forest model will first use the bootstrap sampling method, and then construct N decision trees with N subsamples put back in the input data. M features are then randomly selected among all features at each node in the tree, split tests are performed according to the Gini coefficient and the optimal features are found [25]. Finally, based on the independent prediction results of all decision trees, the most likely category of classification targets is determined by voting, and the comprehensive out-of-bag error (OOB) is obtained.

OOB error can not only reflect classification accuracy, but can also be used to calculate feature importance score (VIM) [26]. After each decision tree obtains the OOB error (Bo), for each feature participating in the decision tree operation, keep the value of other features unchanged, randomly shuffle the out-of-band data value of the feature variable, and recalculate the OOB error (Bn). The difference and percentage of OOB errors of all decision trees for both classes are disrupted by the VIM of the feature. For any feature m, the decision tree number is t, and the feature importance score V(m) can be expressed as:

$$\mathrm{V}(m) = \frac{1}{N}\sum_{t=1}^{N} B_{n_t}^{m} - B_{O_t}^{m} \tag{1}$$

$B_{n_t}^{m}$ is the OOB error of the t_{th} decision tree when any feature m value is not shuffled

$B_{O_t}^{m}$ is the OOB error of the t_{th} decision tree when the arbitrary feature m value is scrambled

4. RESULTS AND ANALYSIS

4.1. Feature importance analysis

The random forest algorithm can analyze and evaluate the importance and contribution of feature variables, which can improve the classification accuracy and reduce the redundancy of data. Figure 3 shows the importance distribution of feature variables analyzed by the random forest model, and the higher the importance score, the greater the image and contribution of the feature to the classification results. It can be seen from the figure that the importance scores of NDVI_re2, SLOPE and DEM feature variables exceeded 45, 48.15, 46.79 and 45.60, respectively. Among them, Sentinel-2's unique red edge band red edge index NDVI_re2=((B8A-B6))/(B8A+B6)) has the largest image of the classification results, and the main feature type in the study area is grassland, which shows that the reflectivity of grassland to the B8A and B6 red-edge bands in the Taihang Mountains is higher than that of other bands, and these two bands have better effects in identifying vegetation such as grassland. With the increase of elevation and slope, cultivated land and building land gradually decreased, and the cover area of forest and grassland gradually increased, so the importance of DEM and SLOPE in the classification was high.

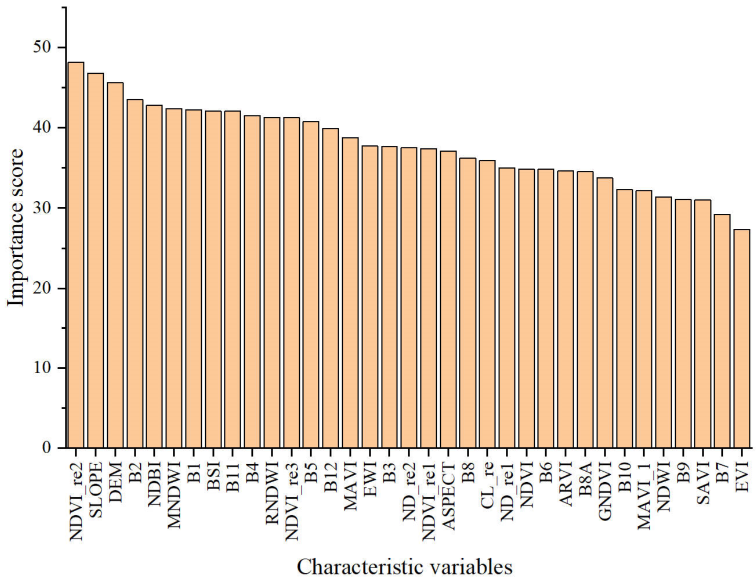

Figure 3. Importance distribution of characteristic variables

4.2. Classification results and accuracy analysis

In the GEE environment, the random forest model was constructed using the above 34 features and the specified number of samples, and the optimal adjustment of the number of classification trees was performed, and the final results showed that the number of classification trees was 71, and the classification accuracy reached the highest 0.992 (Figure 4). As can be seen from Table 3, the random forest model based on the above parameters has achieved excellent classification results in both the training and validation sets. The overall accuracy and Kappa coefficient on the training set reached 0.99, and the producer accuracy and user accuracy of specific land use categories in the training set were higher than 0.95; the overall accuracy and Kappa coefficient on the validation set were 0.85 and 0.81, respectively, and the producer accuracy and user accuracy of specific land use categories were between 0.72~0.86, but the producer accuracy and user accuracy of sparse vegetation were 0.21 and 0.67, respectively, which were lower than the classification accuracy of other categories. The classification results are shown in Figure 5. It can be seen from the figure that the area of grassland and sparse vegetation in the Taihang Mountains increased by 2429.26km^2 and 1301.14km^2 respectively between 2016 and 2021, while the remaining land types including forest land, arable land, construction land and water decreased by 122.70km^2, 3236.82km^2, 171.16km^2 and 199.72km^2, respectively. In the process of land use type change, although the forest land decreases, but the grassland and sparse vegetation show an increasing trend, and the growth area is much higher than the area of forest land decrease, but most of the areas of grassland growth are located in higher areas, and the reduced part of forest land appears in the northern part of Taihang Mountain, which belongs to the scope of Beijing, indicating that the forest land cover in Beijing is in a downward trend.

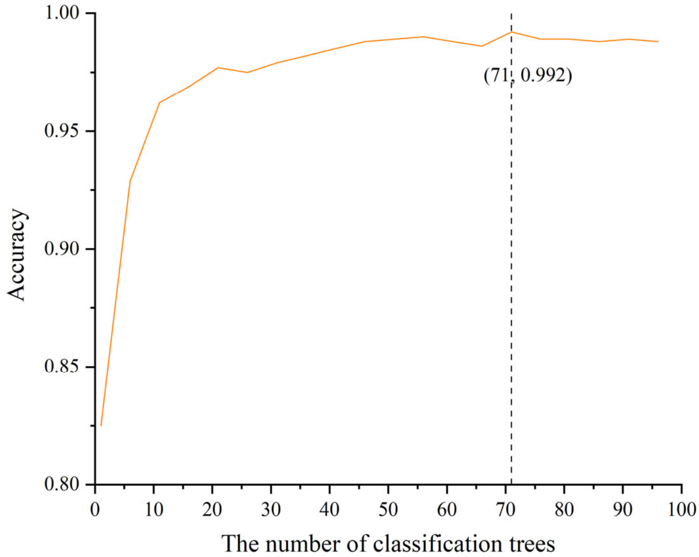

Figure 4. The optimized results of Random Forest arguments

Table 3. The accuracy evaluation of classification result

		Forest	Grassland	Cultivate land	Building land	Sparse vegetation	Waters
Train	Producer Accuracy	1	0.99	1	0.95	0.97	0.98
	Users Accuracy	0.98	1	0.99	0.99	0.96	0.97
	Over Accuracy	0.99					
	Kappa	0.99					
Validation	Producer Accuracy	0.81	0.84	0.86	0.75	0.21	1
	Users Accuracy	0.82	0.72	0.78	0.75	0.67	0.85
	Over Accuracy	0.85					
	Kappa	0.81					

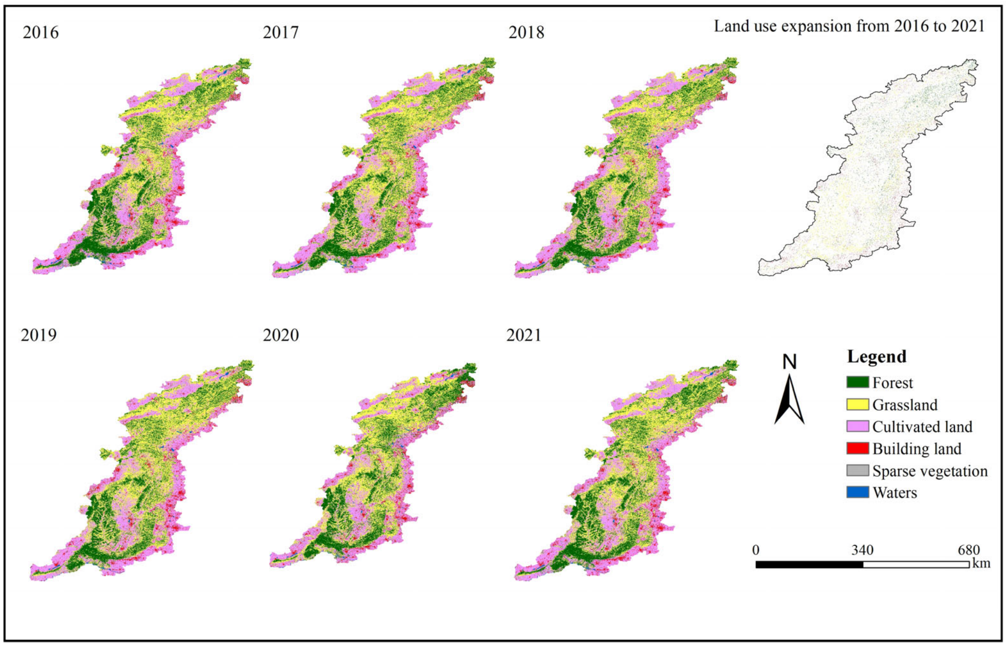

Figure 5. The result of classification and land expansion

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

4.3. Analysis of land use change and drivers

Based on the results of land use classifications in 2016 and 2021, the degree of transition between land use types during the study period can be obtained (Figure 6). It can be seen that the mutual transformation between the two land types of forest land and grassland accounted for the largest proportion, 4.03%. The Land Expansion Analysis Strategy (LEAS) module based on the PLUS model uses seven selected drivers to analyze the changes of forest land and grassland types in land use change in the Taihang Mountains. As can be seen from Figure 7, the eastern part of the Taihang Mountain region is relatively flat, so it has a developed economy and a large population. Figures 7E and 7F show that the western and northern parts of the Taihang Mountains are cooler and receive higher rainfall than other areas. From Figures 8A and 8B, the range of forest land change was mainly concentrated in the northeast of Taihang Mountain, and grassland mainly increased in the western and northern areas with higher elevation, which was also in line with the spatial distribution of socio-economic drivers and natural drivers in the Taihang Mountain region. From the contributions of the drivers in Figures 8C and 8D, it can be seen that the main drivers of land use type change are elevation, which shows that elevation still has a relatively important influence on land use type change in mountainous areas.

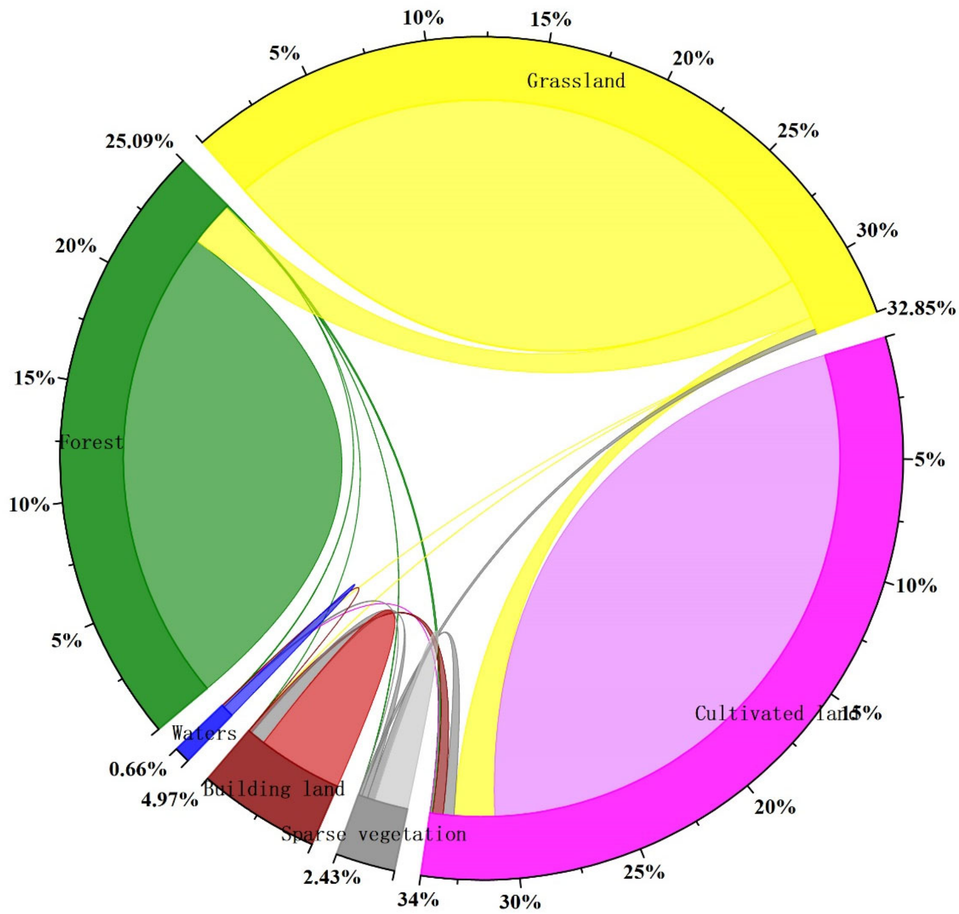

Figure 6. Land use change in Taihang Mountains from 2016 to 2021

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

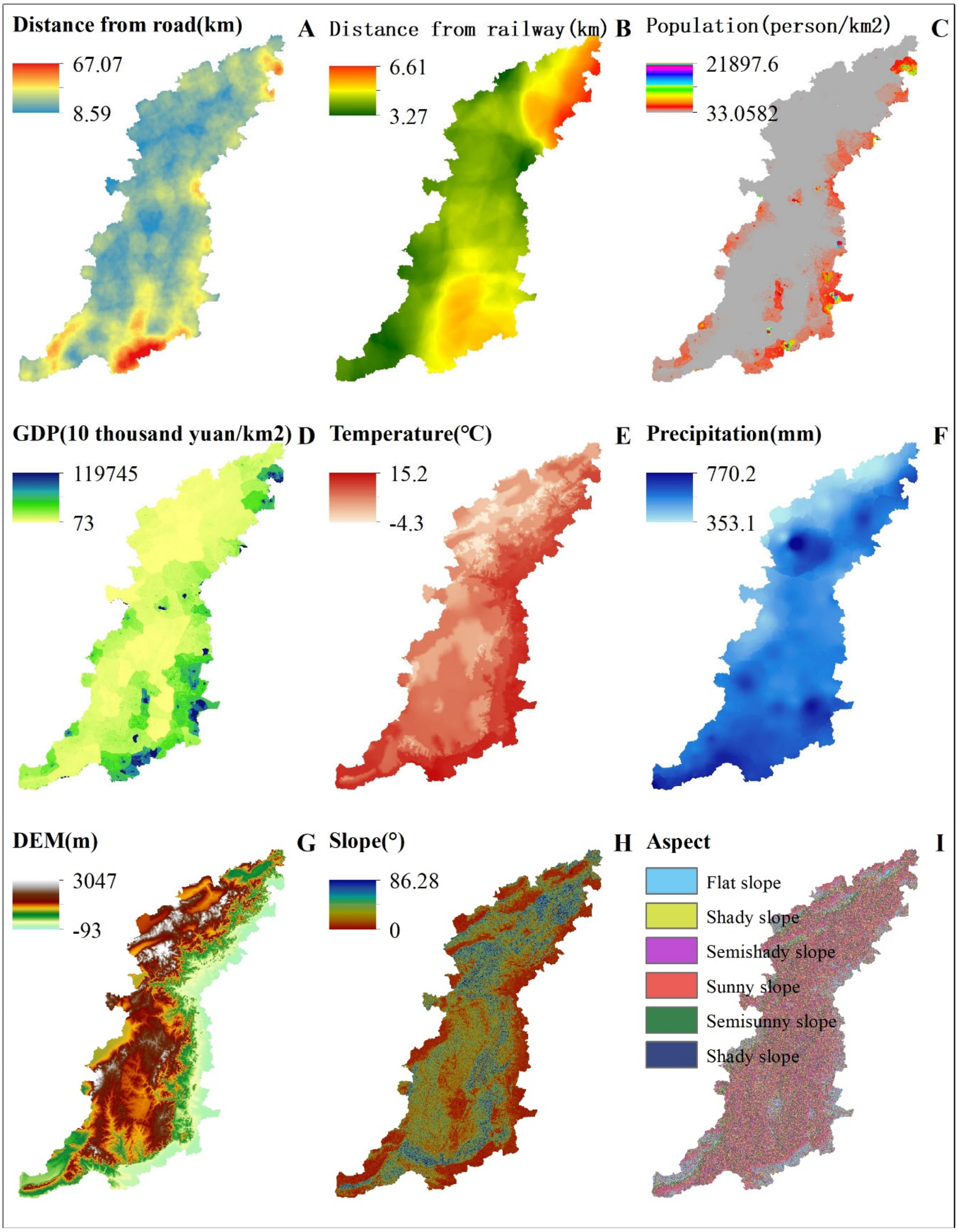

Figure 7. Drive factors of change of land use in Taihang Mountains

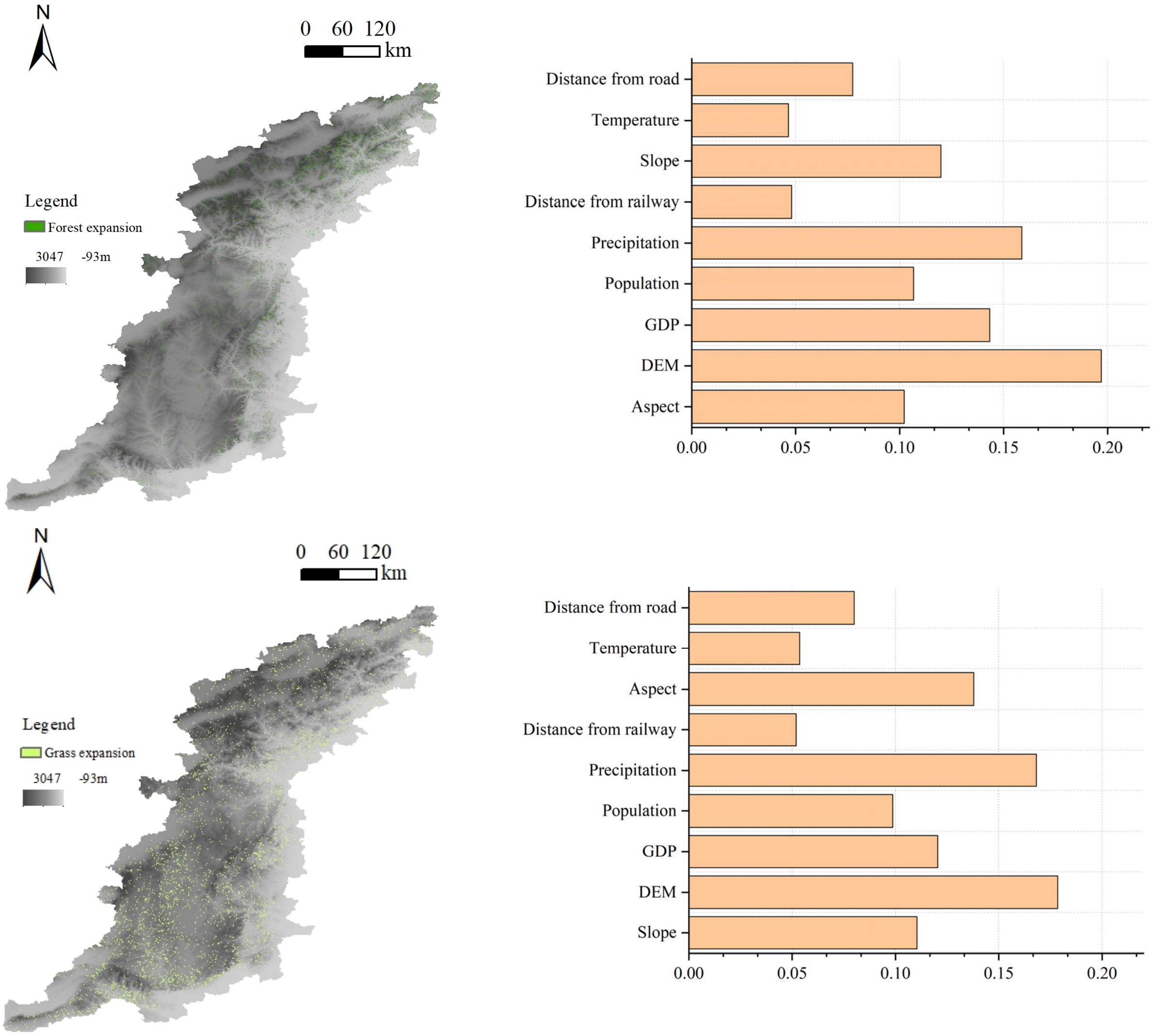

Figure 8. The importance of the contribution of each variable to the change of vegetation types

5. CONCLUSION AND DISCUSSION

5.1. Conclusion

This study jointly developed a pixel-based random forest classification algorithm based on Sentinel passive remote sensing data and GEE environment, the method runs in a short time, significantly improves the efficiency of data processing and analysis, all processing and analysis processes are realized in Google Cloud Platform, which obviously releases the pressure of local computer hardware, and adopts the mode of GEE plus machine learning algorithm to quickly realize the classification of land use types, and the overall accuracy of the final classification results is 85% and 0.81, and the classification quality is high. The spectral features, index features and topographic features can be used to better mine the information implied by remote sensing images, improve the classification accuracy of complex features in the study area, and through the results of feature importance analysis, it can be seen that the red edge index and topographic factors play an important role in the classification of land use types in mountainous areas. In the analysis of the drivers of land use change, it is found that the most important influencing factor in mountainous areas is still elevation, because elevation can not only affect the range of human activities, but also affect meteorological changes, and then affect the distribution and growth of mountain vegetation.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

5.2. Discuss

(1) Uncertainty in GEE land-use classification. Due to the limitation of accuracy and data selection, Sentinel data with high accuracy but less data volume was selected in this study, and the data were screened within the optional range of Sentinel data as much as possible, but there was still insufficient data. Another uncertainty factor is the common cloud and shadow influencing factors in remote sensing data, because of the limitation of the amount of data, so the period of images selected in this study is 12 months a year, and the four seasons are not distinguished, which causes a large number of clouds, shadows, snow and other noise in the images, and sparse vegetation mostly exists at high altitudes, which will be significantly greater than other land types. In the future, the method of combining Landsat remote sensing data with Sentinel data can be considered to make up for the lack of data caused by the lack of data.

(2) The potential impact of elevation on land-use change in mountain areas. In this study, the results of the PLUS model analysis show that elevation has the greatest impact on land use change in mountainous areas. The main reason is that changes in elevation affect many factors such as precipitation, temperature and human activities in mountain areas, which in turn affect changes in land use in mountain areas [27,28]. Combined with the land use changes in this study, it can be seen that urbanization in the Taihang Mountain will further aggravate the reduction of forest land, but this problem will be significantly improved by the Sanbei shelter forest project and the Taihang Mountain greening project. In the future development, the forest land in the city should be strictly protected, and the area of forest land can be increased by increasing the urban green area. This study only preliminarily classifies the land use types of Taihang Mountain, and analyzes the changing factors, and the process and influence mechanism between the two variables can be clarified through new methods in the future.

REFERENCES

[1] Z. J. Liu, X. X. Yu, L.Li, et al. Vulnerability assessment of Eco-environment in Yimeng Mountainous area of Shandong Province based on SRP conceptual model.Chinese Journal of Applied Ecology, 22(2011) No. 8,p.2084–2090. (In Chinese)

[2] M.W. WANG,H. XU: Research review of Karst mountain eco-environmental Geological problems and Sustainable development in Yunnan. Ecological Economy, 30(2014) No. 9,p.185–187. (In Chinese)

[3] S.L.JI : Cultural connotation and utilization of Taihang Mountain Cultural Belt. China Development Watch, 24(2020):5.(In chinese)

[4] H.Sheng,J.Li,P.Zhou,et al. Effect of Land Use Change on Labile Organic Carbon Fractions of SoilDerived from Granite. Ecology and Environmental Sciences, 2015, 24(7): 1098-1102. (In Chinese)

[5] X.T. Zheng: Thinking on the preparation of village land use planning under the background of rural revitalization strategy. Modern Agriculture, 03(2022),p.85-87. (In Chinese)

[6] W.Q. NI, X.P. ZHANG, L. YANG, et al. Study on the Scale Effect of Ecosystem Service Value Estimation Based on Land Use. Ecological Economy, 38(2022) No. 9,p.170-178. (In Chinese)

[7] S.H. TAN, S.X. NI: Progress in the Research on Land Use since 20th Century. Areal Research and Development, 25(2006) No. 5,p.84-89. (In Chinese)

[8] BANSAL S, SRIVASTAV S K, ROY P S, et al. An analysis of land use and land cover dynamics and causative drivers in a thickly populated yamuna river basin of india . Applied Ecology and Environmental Research,3(2016).

[9] N. CAO, J.F. GE :The Driving Mechanism Analysis of Land Use Change of Taihang Mountain Area in Hebei. Journal of Agricultural Mechanization Research, 10(2007) ,p.12-15+24. (In Chinese)

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

[10] C.G. ZHAO, Y.S. CHENG, H.Y. LI, et al. Land use /cover change of an artificial vegetation system in the northeastern edge of Tengger Desert. Journal of Arid Land Resources and Environment, 35(2021) No. 6,p.131-138. (In Chinese)

[11] J.DU, S.K.QIU, Z. WANG, et al. Temporal and Spatial Changes and Driving Factors of Land Use along the Yellow River in Henan Province. Henan Science, 40(2022) No. 6,p.922-930. (In Chinese)

[12] H. ZHANG, W. DENG, Y. WANG:Analyses on the spatio-temporal evolution and driving factors of land use in India during 2000—2020.World Regional Studies, 31(2011) No. 4,p.786-799.(In Chinese)

[13] X.H. JIANG, H.Q. GAO, Y.X. LEI: Changes and driving factors of land use types in Kuye Basin in recent 40 years. South-to-North Water Transfers and Water Science & Technology, 18(2020)No. 4,p.17-38. (In Chinese)

[14] D.D.JIANG, F.YANG, W.B.MA, et al. Evolution and Scenario Prediction of Ecosystem Service Value in Dalou Mountain Area. Research of Environmental Sciences, 35(2022) No. 7,p.1670-1680. (In Chinese)

[15] Taihang Mountain,Information on: http://www.people.com.cn

[16] X.R. LI, H. GAO, L.P.HAN, et al. Spatio-temporal variations in vegetation NPP and the driving factors in Taihang Mountain Area. Chinese Journal of Eco-Agriculture, 25(2017)No. 4,p.498–508. (In Chinese)

[17] Q. LIU,H.T. ,JING,P.P. LIU, et al. Research on the Relationship between Temporal and Spatial Changes of NDVI and Climate Factors in Taihang Mountain. Journal of Northwest Forestry University 25(2022),p.1-12. (In Chinese)

[18] Y.B. ZHANG, Q.X. MENG, H.QIN, et al. Flora and geographic pattern of mountain forests at community level in Taihang Mountains : Results based on plant community survey.Chinese Journal of Applied Ecology,30(2019) No. 10, p. 3395－3402. (In Chinese)

[19] BELGIU, DRAGUT:Random forest in remote sensing: A review of applications and future directions. ISPRS J PHOTOGRAMM, 114(2016)p.24-31.

[20] CLERICI N, VALBUENA C, CESAR A, et al. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia, Journal of Maps, 13(2017)No. 2,p.718-726.

[21] X.G.NING,W.T.CHANG,H.WANG, et al. Extraction of marsh wetland in Heilongjiang Basin based on GEE and multi-source remote sensing data. National Remote Sensing Bulletin, 26(2022) No. 2,p.386-396. (In Chinese)

[22] MARKUS I , FRANCESCO V , CLEMENT A:First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. Remote Sensing, 8(2016) No. 3,p.166.

[23] BREIMAN L. Random forests,Machine Learning, 45(2001) No. 1,p.5-32.

[24] G.Z.WANG, H.L. JIN, X.H.GU, et al. Remote Sensing Classification of Autumn Crops Based on Hybrid Feature Selection Model Combining with Relief F and Improved Separability and Thresholds. Transactions of the Chinese Society for Agricultural, 52(2021) No. 2,p.199-210. (In Chinese)

[25] Y.P. CHEN,W.J. YUN,X. ZHOU, et al. Classification and Extraction of Land Use Information in Hilly Area Based on MESMA and RF Classifier, Transactions of the Chinese Society for Agricultural, 48(2017) No. 7,p.136-144. (In Chinese)

[26] M.J. HOU, J.P. YIN, J. GE, et al. Land Cover Remote Sensing Classification Method of Alpine Wetland Region Based on Random Forest Algorithm,Transactions of the Chinese Society for Agricultural, 51(2020) No. 7,p.220-227. (In Chinese)

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0020

[27] M,Z,XI,Z.QZHAO,P.S.WU, et al. Changes and Predictions of Land Use in Mountain Section of the Hutuo River Basin Based on Improved CA-Markov Model,Journal of Northwest Forestry University, 36(2021) No. 4,p.150-158. (In Chinese)

[28] K.J.ZHENG,C.LI,Y.M.WU,et al. Temporal and spatial variation of landscape ecological risk and influential factors in Yunnan border mountainous area , Acta Ecologica Sinice,42(2022) No. 18,p.7458-7469.(In Chinese)