Study on Crop Classification Using Sentinel-2 Satellite Data

Weili Zeng[1], Qiaomei Su[1], Rong Pan[1], Jinlong Fan[2], Jean Boucha[3]
1: Taiyuan University Of Technology, China, People's Republic of;
2: NSMC, China, People's Republic of;
3: Universite Catholique de Louvain, Belgium

ABSTRACT

Precise classification of crops is an important way to realise precision agriculture. In order to improve the extraction accuracy of crop information from medium and high resolution remote sensing images in farming areas, this paper uses Sentinel-2 satellite data and DEM data, constructs spectral features, remote sensing index features, texture features and terrain features, and uses the random forest model based on feature preference to finely extract corn in some mountainous areas of Shanxi Province, China, and explores the effects of different feature combinations on the extraction accuracy of corn. The results show that the multi-feature combination effectively improves the classification accuracy of crops, in which the classification result of fusing four types of feature variables is the best, and its overall accuracy reaches 94%. The preferred texture features and terrain features help to present the information of mountainous areas, and can significantly improve the extraction accuracy of the mountainous crop information. Random Forest algorithm can effectively perform data mining of feature variables while still ensuring high extraction accuracy, and can be used as a powerful tool for bulk crop extraction.

INTRODUCTION

With the rapid development of remote sensing technology, remote sensing data has been widely used in crop information extraction. The sentinel satellite data of the European Space Agency (ESA) has the characteristics of multi-band and short revisit period, which provides sufficient data guarantee for crop remote sensing identification and application research. At present, the integration of multi-feature variables into crop classification has gradually become the development trend of crop remote sensing extraction research.

OBJECTION

The study area is part of Shanxi Province. The region is located in north China, on the east bank of the middle reaches of the Yellow River, with complex and diverse landform types. It belongs to the temperate continental monsoon climate, with sufficient light throughout the year, excellent irrigation conditions in the region, and fertile land, which is suitable for the growth of crops, and the main crops are corn, wheat, sorghum, potatoes and so on.

The research objective is to propose an accurate identification method of crop planting structure based on sentinel-2 data using random forest algorithm with multi-feature combination optimisation, and explore the potential of high-resolution remote sensing data in crop identification and classification under complex spatial planting structure.

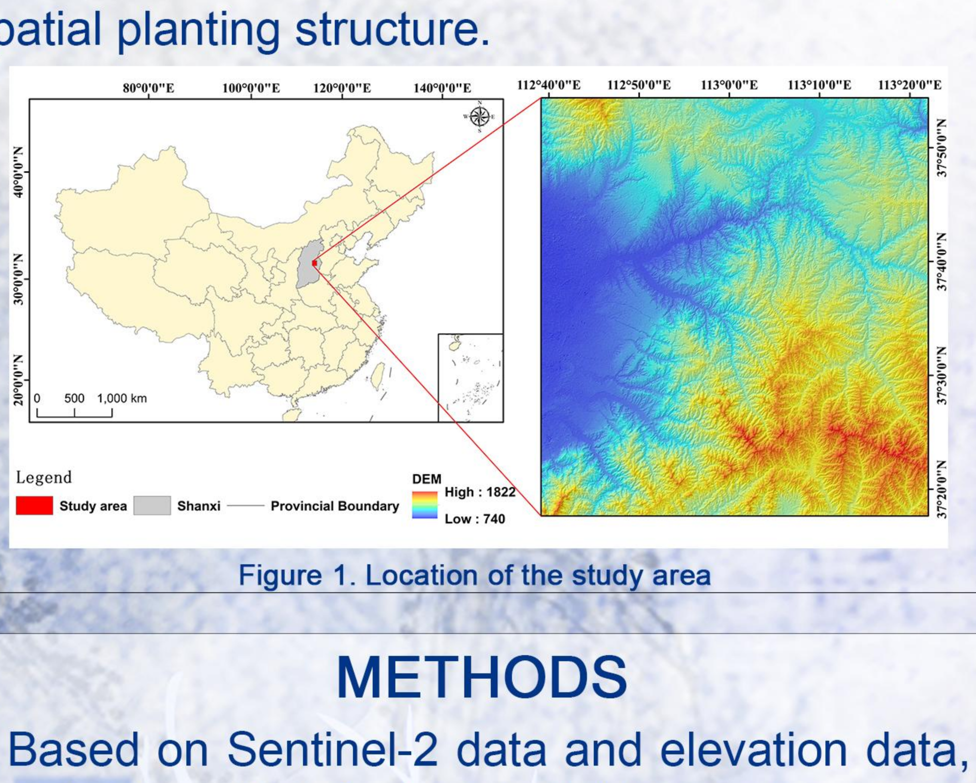

Figure 1. Location of the study area

METHODS

Based on Sentinel-2 data and elevation data, construct multiple feature variables and carry out feature preference and combination, and explore the best method to improve the extraction accuracy of corn for different feature combination schemes.

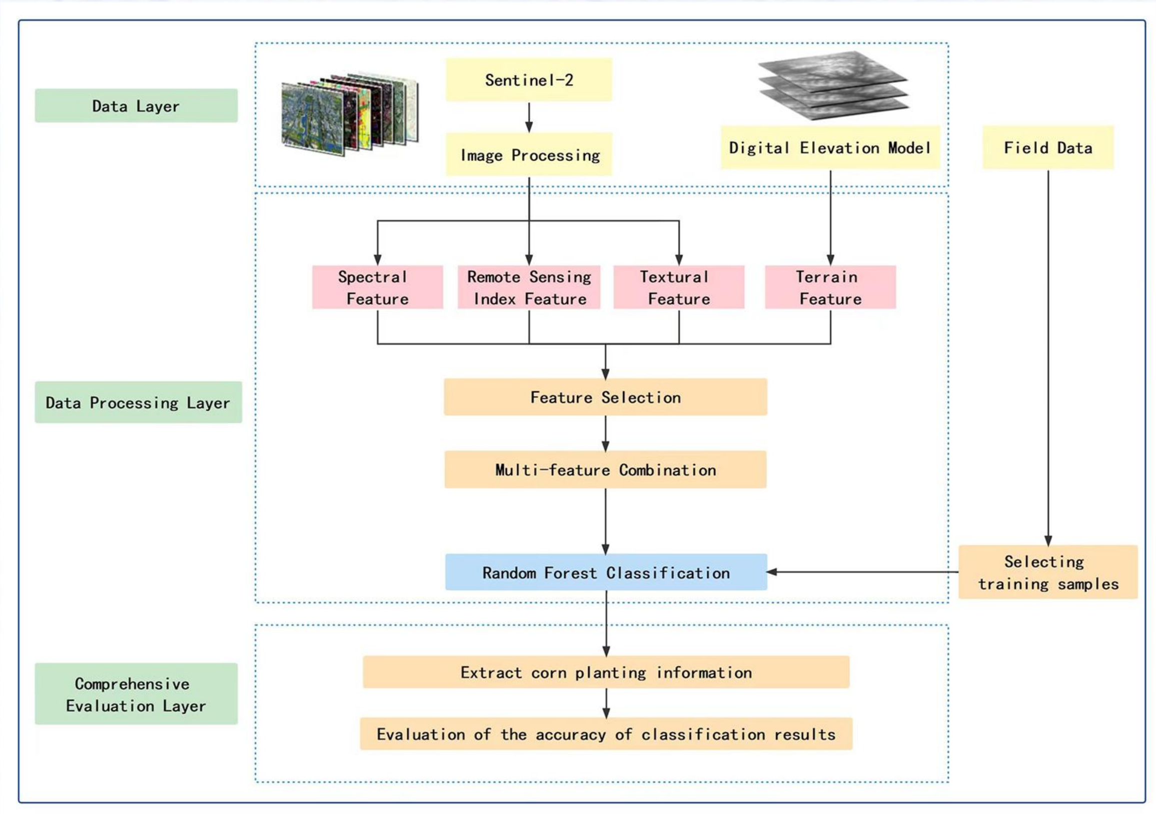

Figure 2. The technical route of this study

Construct feature variables

Spectral feature	Sentinel-2 band: B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12	
RS index feature	$\mathrm{NDVI}=\frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+\rho_{red}}$	MSAVI $=\frac{2\rho_{nir}+1-\sqrt{(2\rho_{nir}+1)^2-8(\rho_{nir}-\rho_{red})}}{2}$
	$\mathrm{NDWI}=\frac{\rho_{green}-\rho_{nir}}{\rho_{green}+\rho_{nir}}$	$\mathrm{RNDVI}=\frac{\mathrm{NDVI}-\mathrm{NDVI}_{mix}}{\mathrm{NDVI}_{max}-\mathrm{NDVI}_{min}}\times 255$
	$\mathrm{LSWI}=\frac{\rho_{nir}-\rho_{swir1}}{\rho_{nir}+\rho_{swir1}}$	$\mathrm{ARVI}=\frac{\rho_{nir}-(2\rho_{red}+\rho_{blue})}{\rho_{nir}+(2\rho_{red}+\rho_{blue})}$
	$\mathrm{RVI}=\frac{\rho_{red}}{\rho_{nir}}$	$\mathrm{EVI}=\frac{2.5(\rho_{nir}-\rho_{red})}{\rho_{nir}+6\rho_{red}-7.5\rho_{blue}+1}$
	$\mathrm{DVI}=\rho_{nir}-\rho_{red}$	$\mathrm{NDTI}=\frac{\rho_{swir1}-\rho_{swir2}}{\rho_{swir1}+\rho_{swir2}}$
Texture feature	$\mathrm{Mean}=\sum_{i,j=0}^{N-1} i\cdot P_{i,j}$	$\mathrm{Homogeneity}=\sum_{i,j=0}^{N-1}\frac{i\cdot P_{i,j}}{1+(i-j)^2}$
	$\mathrm{Entropy}=\sum_{i,j=0}^{N-1} i\cdot P_{i,j}\ln P_{i,j}$	$\mathrm{Angular\ Second\ Moment}=\sum_{i,j=0}^{N-1} i\cdot (P_{i,j})^2$
	$\mathrm{Variance}=\sum_{i,j=0}^{N-1} i\cdot P_{i,j}(i-\mathrm{Mea})^2$	$\mathrm{Dissimilarity}=\sum_{i,j=0}^{N-1} i\cdot P_{i,j}\lvert i-j\rvert$
	$\mathrm{Contrast}=\sum_{i,j=0}^{N-1}(i-j)^2\cdot P_{i,j}$	$\mathrm{Correlation}=\sum_{i,j=0}^{N-1} i\cdot P_{i,j}\frac{(i-\mathrm{Mea}_r)(j-\mathrm{Mea}_c)}{\sqrt{\mathrm{Var}_r\cdot \mathrm{Vaar}_c}}$
Terrain feature	Elevation, Slope, Aspect, Hillshade	

Feature combination schemes

Scheme	Feature combination
1	Spectral Feature
2	Spectral Feature + Remote Sensing Index Feature
3	Spectral Feature + Texture Feature
4	Spectral Feature + Topographic Feature
5	Spectral Feature + Remote Sensing Index Feature + Texture Feature
6	Spectral Feature + Remote Sensing Index Feature + Topographic Feature
7	Spectral feature + Texture feature + Topographic Feature
8	Spectral Feature + Remote Sensing Index Feature + Texture Feature + Topographic Feature

RESULTS

In the feature importance analysis, elevation has the highest importance score, indicating that elevation plays an important role in crop information extraction in mountainous areas. Then, NDTI, LSWI, Mean, indicating that remote sensing index features and texture features play a key role in crop information extraction.

To determine the number of features, OOB was used for evaluation. When the number of features is 18, the OOB accuracy reaches the peak, which is 0.943.Therefore, the feature variables with the top 18 importance were selected as the input features for the random forest model based on feature preference and feature combination.

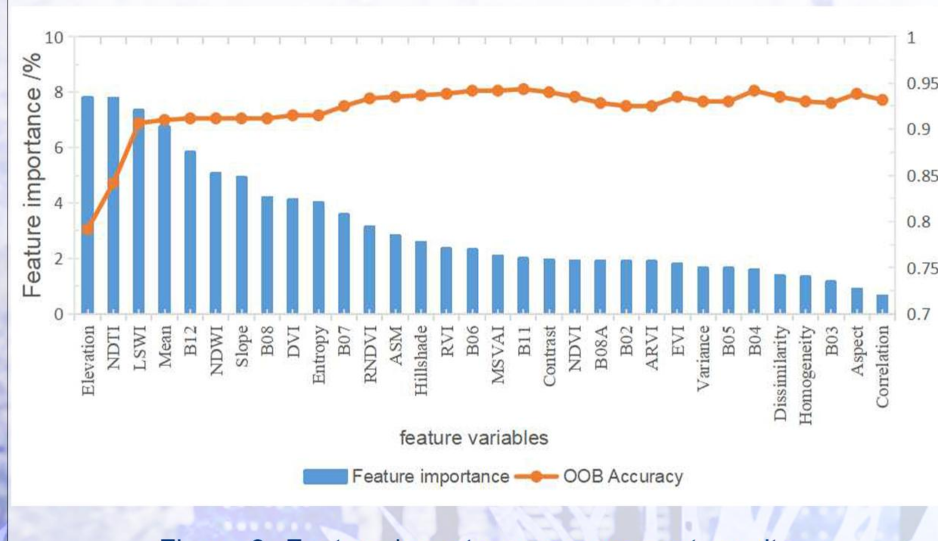

Figure 3. Feature importance assessment results and OOB accuracy of different feature number

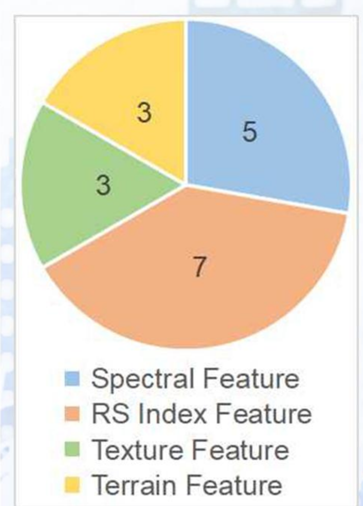

Figure 4. The number of different features in the preferred features

Compared with single feature variables, the overall accuracy and Kappa coefficient of classification based on multi-feature optimization combination for crop extraction are significantly improved. The best classification result is scheme 8, based on spectral features + remote sensing index features + texture features + terrain features, which has an overall accuracy of 94%, a Kappa coefficient of 0.92, and an F1 Score of 91.7%, which is better than the other schemes, and improves 12.8%, 0.15, and 18.31% over the results of the classification using only spectral features, respectively.

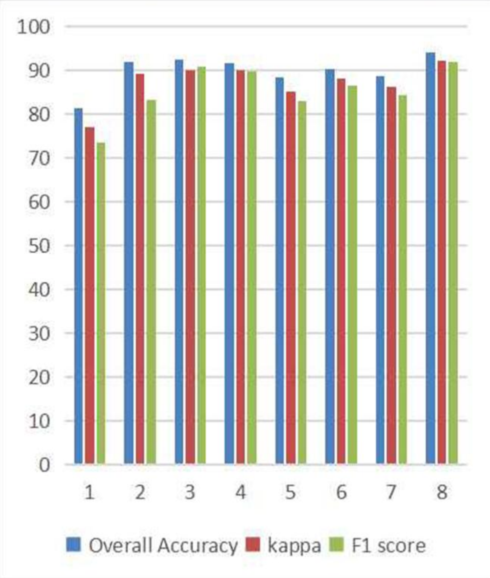

Figure 5. The accuracy of different schemes

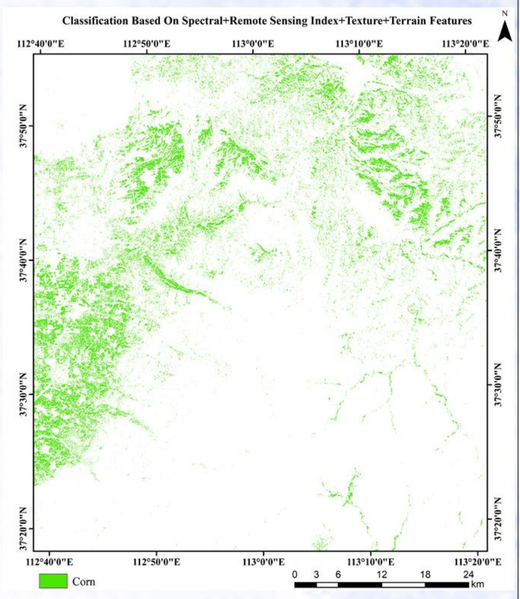

Figure 6. Corn extraction results of scheme 8

DISCUSSION

Based on Sentinel satellite data, this study adopts a feature optimization algorithm combined with a random forest classifier to realize the extraction of information on the planting structure of typical crops in the study area. The region of this study is relatively small and only focuses on one growing season, and the applicability and robustness of the conclusions of this study will be examined by conducting corn planting area extraction studies on a larger scale for multiple growing seasons. For the next study, whether the feature selection algorithm based on multi-temporal phase and multi-feature can be combined with high-resolution remote sensing imagery and generalized to other regions needs to be further explored.

CONCLUSIONS

1. Spectral features are the basis of feature classification, and most of the crop information can be captured by spectral features, but the effect of corn extraction is ordinary.

2. Adding remote sensing index features, texture features, and terrain features can effectively improve the classification accuracy.

3. The integration of texture features and topographic features contributes to the presentation of information, enriches the information content of multi-spectral images, and is beneficial to the extraction of crop information.

4. The random forest model can achieve fast and accurate crop information extraction. It can be used as a powerful tool for bulk crop extraction.

MAJOR REFERENCES

[1] Fan, H. Land-cover mapping in the Nujiang Grand Canyon: integrating spectral, textural, and topographic data in a random forest classifier. Int.J.Remote Sens. 2013, 34, 7545–7567.

[2] Fan, J.; Zhang, X.; Zhao, C.; Qin, Z.; De Vroey, M.; Defourny, P. Evaluation of Crop Type Classification with Different High Resolution Satellite Data Sources. Remote Sens. 2021, 13, 911

[3] Shafizadeh-Moghadam, H.; Khazaei, M.; Alavipanah, S.K.; Weng, Q. Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors. GIScience & Remote Sensing. 2021, 1–15.

DRAGON SYMPOSIUM 2023