

IMPROVED ARTIFICIAL IMMUNE NETWORK CROP RECOGNITION ALGORITHM BASED ON DISPERSED VEGETATION INDEX GENETIC CHAIN

Yushi Zhou¹, Xiaoping Lu^{1,*}, Xiangjun Zhang², Haikun Yu², Guosheng Cai¹

¹ Key Laboratory of Spatio-temporal Information and Ecological Restoration of Mines of Natural Resources of the People's Republic of China, Henan Polytechnic University, Jiaozuo 454003, China – lxp@hpu.edu.cn, (hpuzys, 212004020039) @home.hpu.edu.cn

² Henan Institute of Remote Sensing and Surveying and Mapping, Zhengzhou 450000, China; - (1657377018, 914309384) @ qq.com

Commission III, WG III/10

KEY WORDS: Genetics; Artificial Immunity Network; Crop Identification; Remote Sensing; Sentinel-2

ABSTRACT:

The rapid and accurate acquisition of crop planting spatial location information using remote sensing is one of the important guarantees to maintain the sustainable development of agriculture. However, the accuracy of crop identification by remote sensing is currently limited by many factors, such as the influence of other ground objects and the lack of time-series data. To overcome the above problems, this paper proposes an algorithm named improved artificial immune network crop recognition algorithm based on dispersed vegetation index genetic chain (IaiNet). This algorithm can be combined with multi-spectral data from Sentinel-2 series satellites for crop identification. As a test case, we identified and evaluated 3 different crop recognition scenarios in Henan, China. The results show that IaiNet can accurately identify the spatial distribution of crop planting. In all identification results, the accuracy is higher than 90%, and the kappa coefficient is greater than 0.9. In addition, the crop recognition results of IaiNet are significantly better than the random forest algorithm and support vector machine algorithm.

1. INTRODUCTION

Agricultural production is the basis for maintaining human survival and health and is an important strategic pillar industry for safeguarding the country's economy, people's livelihood, and political stability. In the natural resource environment, agriculture occupies a pivotal position, and sufficient agricultural resources are an important prerequisite for ensuring social stability and realizing sustainable development strategies. At present, agricultural production is continuously affected by factors such as changes in farming methods (Laporte et al., 2021; Wang, Lu, 2020), population growth (Nicolas et al., 2015), and climate change (Malhi et al., 2021), resulting in fluctuations in agricultural planting areas and grain yields. At present, the use of all advanced means to ensure the safety of agricultural production is one of the great important issues by governments around the world.

As one of the important agricultural factors, the spatial location information of crop planting distribution is the basis for formulating a series of agricultural security policies and economic plans. Various studies on agricultural production (such as yield estimation (Jiang et al., 2020), growth status monitoring (Tja et al., 2020), etc.) all rely on the accurate identification of crops. The ability to obtain the timely and accurate spatial distribution of crops and cultivated land will have a positive effect on the sustainable development of agriculture.

In recent years, with the development of satellite remote sensing technology, this technology has gradually provided strong data support for the investigation of the spatial distribution of crops. There are various methods for calculating the spatial distribution position of crops based on remote sensing observation data, including traditional supervised classification models, such as

maximum likelihood (Strahler, 1980), support vector machine (Cortes, Vapnik, 1995), artificial neural network (Gong et al., 1996), etc.; as well as emerging models and theories such as deep learning (Zhang et al., 2020). However, traditional methods are easily affected by other ground objects, resulting in lower recognition accuracy (Maxwell et al., 2018), and deep learning relies more on a large number of field investigations, professional personnel training, and stronger computer computing power (Ma et al., 2019; Vali et al., 2020). These phenomena have restricted the large-scale application of existing algorithms and models in crop recognition.

Using time-series data is a feasible idea for crop identification. Many studies have shown that multispectral remote sensing imagery based on time series is an effective means of large-scale, long-term, continuous agricultural remote sensing mapping (Ortiz et al., 2008; Potgieter et al., 2010). The vegetation index is currently the most widely used characteristic parameter to describe crop phenological changes. The vegetation index based on time series multispectral remote sensing image data can reflect the dynamic changes of different crop types over time. However, due to the physical properties of optical satellites, their sensors are often affected by factors such as clouds and fog, resulting in the inability to obtain complete time-series information, which has a great impact on crop identification based on time series data. Based on the above discussion, this paper proposes an algorithm named improved artificial immune network crop recognition algorithm based on dispersed vegetation index genetic chain (IaiNet), which can be applied to crop recognition under the premise of missing time series data. This algorithm introduces the concept of scattered vegetation index genetic chain and improves the confusion of ground objects caused by the use of a single vegetation index time series in traditional methods. In

* Corresponding author

addition, replacing the Euclidean distance with the Mahalanobis distance as a new similarity measure unit suppresses the phenomenon of being relatively sensitive to noise in the genetic chain vector. Combined with Sentinel-2 satellite imagery, the IaiNet algorithm can perform crop identification at a spatial resolution of 10 m. Finally, the method was applied to 3 different crop recognition scenarios in Henan, China in 2020. The results were evaluated using the measured samples and compared with the random forest algorithm and the support vector machine algorithm.

2. STUDY AREA

Henan Province is a major agricultural production province in China, and one of the three provinces with a national grain output exceeding 30 million tons. As a key crop monitoring area in China, it is of practical significance to select typical agricultural production counties in Henan Province as the target research area for crop identification and exploration. Therefore, this thesis uses Zhengyang County (Figure 1. a) and Huaibin County (Figure 1. b) in southern Henan as the research area to explore the capabilities of IaiNet from the perspective of different regions and different crop types. The selection of the study area is mainly based on the following factors: developed agricultural production, complex crop planting spatial distribution, many types of agricultural production crops, and the inability to obtain complete spectral time series characteristics due to the influence of climate. Based on the above factors, this paper uses winter wheat in Zhengyang County, summer peanuts in Zhengyang County, and winter wheat in Huaibin County as target crops for research to evaluate the crop recognition ability of IaiNet.

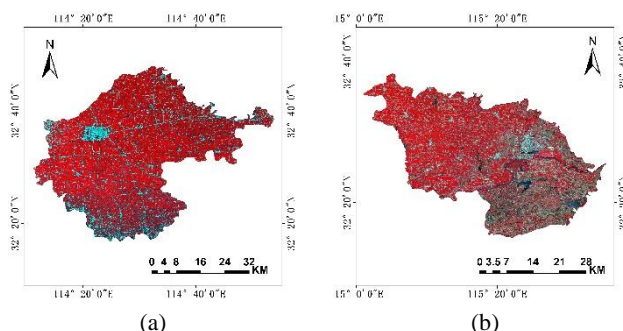


Figure 1. Schematic diagram of the location of the study area a. Zhengyang County b. Huaibin County (Note: The remote sensing image used in the figure is the average false-color image of Sentinel-2 in April 2020, and the band combination is R: near-infrared G: red B: green)

Zhengyang County is located between $114^{\circ}12'$ — $115^{\circ}53'$ east longitude and $32^{\circ}16'$ — $32^{\circ}47'$ north latitude. The county is 64.5 km long from east to west and 57 km wide from north to south.

It is under the jurisdiction of Zhumadian City, Henan Province, with a total county area. Approximately 1903 km². As of November 2020, its permanent population is approximately 625,100. Topographically, Zhengyang County is in the sloping plain area in front of the Dabie Mountains. The terrain gradually decreases from northwest to southeast, with an average elevation of 78.8 meters. The county has a suitable climate and is in the transitional zone from the north subtropical to the warm temperate zone. It has a continental monsoon humid climate. The average annual precipitation is about 960 mm and the average temperature is about 15.3 °C. This climate provides excellent conditions for the growth of a variety of crops.

Huaibin County is located between $115^{\circ}11'$ — $115^{\circ}35'$ east longitude and $32^{\circ}15'$ — $32^{\circ}38'$ north latitude. The county is 53 km long from east to west and 43 km wide from north to south. The total area of the county is approximately 1209 km². As of November 2020, the county's permanent population is approximately 549,600. Huaibin County is located between the northern and front junction depressions of the Dabie Mountains and the Huanghuai Plain. Its terrain is high in the west and low in the east, and high in the north and low in the south. Its topography is dominated by hillocks, plains, and depressions. In terms of climate, Huaibin County is in the transitional zone from the north subtropical to the warm temperate zone. It has a continental monsoon humid climate. The average annual precipitation is about 955.6 mm, the average temperature is about 15.6 °C, and the annual sunshine rate is 42%. In agriculture, Huaibin County has about 1.18 million mu of arable land, and the main crop types are wheat, rice, and corn.

In the two study areas, the growth period of winter wheat is between October and May of the following year, and the growth period of summer peanuts is between June and September.

3. MATERIALS AND PREPROCESSING

3.1 Sentinel-2 Satellite Data

Sentinel-2 was launched by ESA (European Space Agency) for land monitoring. It can provide images of vegetation, soil and water coverage, inland waterways, and coastal areas, and can also be used for emergency rescue services. It consists of two satellites, Sentinel-2A and Sentinel-2B, equipped with a multispectral imager (MSI). Sentinel-2 satellite data can provide data in 13 spectral bands ranging from visible light, near-infrared to short-wave infrared, with a ground resolution of up to 10 m (Table 1). In this paper, the Sentinel-2 L2A satellite image was used, which is the product that has undergone radiometric calibration and atmospheric correction. All Sentinel-2 images (except for cloud and monthly mean value synthesis) were obtained from Google Earth Engine, and the image acquisition time refers to the phenological period of the crops in the study area.

Sentinel-2 Bands	S2A		S2B		Spatial resolution (m)
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	
Band 1 - Coastal aerosol	442.7	21	442.2	21	60
Band 2- Blue	492.4	66	492.1	66	10
Band 3 - Green	559.8	36	559.0	36	10
Band 4 - Red	664.6	31	664.9	31	10
Band 5- Vegetation Red Edge	704.1	15	703.8	16	20
Band 6 - Vegetation Red Edge	740.5	15	739.1	15	20
Band 7 - Vegetation Red Edge	782.8	20	779.7	20	20
Band 8 - NIR ¹	832.8	106	832.9	106	10
Band 8A - Vegetation Red Edge	864.7	21	864.0	22	20
Band 9 - Water Vapor	945.1	20	943.2	21	60
Band 10 - SWIR - Cirrus	1373.5	31	1376.9	30	60
Band 11 - SWIR ²	1613.7	91	1610.4	94	20
Band 12 - SWIR	2202.4	175	2185.7	185	20

¹ NIR refers to the near-infrared band. ² SWIR refers to the shortwave infrared band.

Table 1. Spectral bands for the Sentinel-2 sensors (Sentinel-2A & Sentinel-2B)

3.2 Calculation of Vegetation Index

In this article, a total of 23 cropping indexes are selected, as shown in Table 2. In all vegetation indices, all those whose value

range is not [0,1] were normalized to ensure the unity of subsequent calculations.

Vegetation index	References	Vegetation index	References
Soil Adjusted Vegetation Index (SAVI)	(Huete, 1988)	Atmospherically Resistant Vegetation Index (ARVI)	(Kaufman, Tanre, 1992)
Transformed Soil Adjusted Vegetation Index (TSAVI)	(Baret et al., 1989)	Normalized Difference Index (NDI 45)	(Delegido et al., 2011)
Modified Soil Adjusted Vegetation Index (MSAVI)	(Qi et al., 1994)	Meris Terrestrial Chlorophyll Index (MTCI)	(Dash, Curran, 2007)
Second Modified Soil Adjusted Vegetation Index (MSAVI 2)	(Qi et al., 1994)	Modified Chlorophyll Absorption Ratio Index (MCARI)	(Daughtry et al., 2000)
Difference Vegetation Index (DVI)	(Tucker, 1979)	Sentinel-2 Red-Edge Position Index (S2REP)	(Frampton et al., 2013)
Ratio Vegetation Index (RVI)	(Major et al., 1990)	Inverted Red-Edge Chlorophyll Index (IECI)	(Castillo et al., 2017)
Perpendicular Vegetation Index (PVI)	(Richardson, Wiegand, 1977)	Pigment Specific Simple Ratio (PSSR)	(Blackburn, 1998)
Infrared Percentage Vegetation Index (IPVI)	(Crippen, 1990)	Normalized Difference Vegetation Index (NDVI)	(Rouse et al., 1974)
Weighted Difference Vegetation Index (WDVI)	(Clevers, 1988)	Modified Red Edge Normalized Difference Vegetation Index (NDVI 705)	(Gitelson, Merzlyak, 1994)
Transformed Normalized Difference Vegetation Index (TNDVI)	(Tucker, 1979)	Enhanced Vegetation Index (EVI)	(Liu, 1995)
Green Normalized Difference Vegetation Index (GNDVI)	(Gitelson et al., 1996)	2-Band Enhanced Vegetation Index (EVI 2)	(Xun et al., 2019)
Global Environmental Monitoring Index (GEMI)	(Pinty, Verstraete, 1991)		

Table 2. Selected vegetation indices

3.3 Selection of Training Samples and Verification Samples

This article mainly constructs a training sample set and verification sample set with pixels as the unit to reduce redundancy and spatial autocorrelation (Gong & Howarth, 1990). To obtain sample information, this article visually compares the 2020 site survey data (including site sampling points, drone aerial images, and site record photos) with very high resolution (Very-High-Resolution, VHR) images. All samples were selected in the form of pixels, and mixed pixels are eliminated. The above work was completed by an experienced person to ensure the accuracy and consistency of the sample. The detailed information of the sample data is shown in Table 3.

Crop Recognition Scenario	Type of Samples	Crop	Non-crop	Total
Huaibin-Winter Wheat	Training Samples	1271	1267	2538
	Validation Samples	3349	2779	6128
Zhengyang-Winter Wheat	Training Samples	2186	365	2551
	Validation Samples	5760	374	6134
Zhengyang-Summer Peanuts	Training Samples	453	468	921
	Validation Samples	1201	664	1865

Table 3. Sample data details (number of pixels)

4. METHODS

4.1 Methodological Overview

This short note mainly explains the process of improved artificial immune network crop recognition algorithm based on dispersed vegetation index genetic chain (IaiNet). First, based on Google Earth Engine (GEE), the L2A-level average monthly image data of the Sentinel-2 satellite in the study area is obtained, and then the corresponding vegetation index is calculated to construct the multi-dimensional vegetation index feature time series. Then, the importance of each vegetation index unit is calculated using the feature importance of the Gini coefficient embedded in the random forest model, and it is sorted from high to low. Immediately afterward, by determining the minimum feature set that can make the sum of importance reach 0.8, construct the genetic chain of scattered vegetation index, and use this as the antigen to input the artificial immune network algorithm. Next, through the reference vegetation index genetic chain generated by the artificial immune network algorithm, the Mahalanobis distance is used to evaluate the similarity of potential crop pixels to generate the spatial distribution layer of crop planting positions. Finally, use the verification sample to evaluate the accuracy of the results and compare them with other algorithms for final statistics and analysis. Except for Google Earth Engine, other calculations in this chapter are implemented using Python language.

4.2 IaiNet Algorithm

The artificial immune network model (aiNet) is a kind of artificial immune system model, which was originally proposed jointly by De Castro L. N. and Von Zuben F. J. (Castro & Zuben, 2001; Castro & Zuben, 2002). Referring to the human immune system, the main goal of the artificial immune network algorithm is to use

the input genetic chain of scattered vegetation index as an "antigen", and then generate "antibody", and finally use the "antibody" as a new "antigen" to identify crops. The basic steps of the IaiNet algorithm include:

Step 1: Calculate the vegetation indices for each month during the growing period of the crop.

Step 2: Import all the vegetation indices into the random forest model, calculate the feature importance scores, and arrange all the vegetation indices in descending order of the scores, and take the first several vegetation indices whose sum is greater than 0.8 to construct the genetic chain of scattered vegetation indices.

Step 3: All classes of scattered vegetation index genetic chains were regarded as "antigens" and randomly selected from them.

Step 4: The randomly screened "antigen" is cloned, and then mutation named the calculated result as "antibody".

Step 5: Use the "antibody" set to recognize the "antigen", if the "antibody" can recognize the "antigen", output the "antibody" to the "recognition antibody" set, and remove the recognized "antigen" from the "antigen" set

Step 6: When the "antigen" set is empty, output the "recognition antibody" set.

Step 7: Calculate the Mahalanobis distance between the vegetation index genetic chain of each pixel and the "recognition antibody" set, then assign the feature category attribute of the "recognition antibody" with the smallest distance to the pixel until the algorithm terminates when the entire area is traversed.

In terms of similarity measurement, the artificial immune network model used for similarity research in the past mostly uses Euclidean distance to measure the similarity relationship between vectors, but it is relatively sensitive to noise in the genetic chain vector, and lead to the occurrence of "misclassification" (Hao et al., 2018). Therefore, in this study, Mahalanobis distance was used as a new similarity measure unit. Mahalanobis distance, proposed by Indian statistician P. C. Mahalanobis, is an effective method for calculating the similarity between two unknown sample sets. Compared with Euclidean distance, it has the characteristics of "scale independent" and can exclude the interference of correlation between variables. The formula for calculating Mahalanobis distance is as follows:

$$D(\vec{x}, \vec{y}) = \sqrt{\sum_{m=1}^M \frac{(x_m - y_m)^2}{\sigma_m^2}} \quad (1)$$

In the formula, x_m and y_m are the m -th value of the vector \vec{x} and the vector \vec{y} respectively, M is the number of values in the vector, and σ_m is the standard deviation of x_m .

The IaiNet algorithm was implemented using Python programming.

4.3 Contrast Algorithm

To evaluate the ability of the IaiNet algorithm in crop identification, this paper introduces a random forest algorithm and support vector machine algorithm for comparison. Considering that the difference of input samples will bring about changes in the mapping results, the random forest algorithm and support vector machine algorithm in this paper used the same input samples and feature data as IaiNet. According to previous experience, the parameters of the random forest algorithm and the support vector machine algorithm adopt the default values recommended by the developers (Le et al., 2014). The two algorithms use the same variables as IaiNet. In the random forest algorithm, the number of trees is set to 500, the Kernel Type in the support vector machine algorithm is Radial Basis Function, the Gamma in Kernel Function is 0.25, the Penalty Parameter is 100.00, and the Pyramid Levels is 0. The above algorithms were implemented using Python programming.

4.4 Accuracy Evaluation

In this paper, Producer Accuracy (PA), User Accuracy (UA), Overall Accuracy (OA), Kappa Coefficient (kappa), and Misclassified Pixels calculated based on confusion matrix are used to measure the crop identification of IaiNet algorithm ability.

5. RESULTS AND DISCUSSION

5.1 Result of Genetic Chain Construction of Scattered Vegetation Index

The research area of this paper is located in the southern part of Henan Province, near the geographical boundary between the Qinling Mountains and the Huaihe River in the north and south, so the complete monthly average image data of Sentinel-2 cannot be obtained. For different crops, the time distribution of available data in the phenological period is shown in Table 4. Based on the above data, this paper obtained the time series data of each vegetation index and calculated its feature importance score.

Crop Recognition Scenario	Time distribution of available data
Huaibin-Winter Wheat	2019 (Nov. Dec.), 2020 (Jan. Feb. Mar.Apr.May.)
Zhengyang-Winter Wheat	2019 (Oct.Nov. Dec.), 2020 (Jan. Feb. Mar.Apr.May.)
Zhengyang-Summer Peanuts	2020 (Jun. Aug. Sept.)

Table 4. Time distribution of available data during phenology

An important evaluation criterion for obtaining the construction elements of the genetic chain of the scattered vegetation index is the feature importance score. In this paper, the feature importance scores of all vegetation indices are sorted in descending order, and the sum of the top N items is calculated. When the sum of the current N items is greater than 0.8, the vegetation index data of the first N items are the construction elements of the scattered vegetation index genetic chain.

Table 5, Table 6, and Table 7 list the genetic chain construction elements of scattered vegetation index for winter wheat in Huaibin County, winter wheat in Zhengyang County, and summer peanut in Zhengyang County, respectively. It can be found that for the same crop in different regions, the number of building elements of the genetic chain of the scattered vegetation index is not the same. For different crop identification scenarios, there are differences in the construction elements of the genetic chain of the scattered vegetation index. For the identification of winter wheat in Huaibin County, the data in May dominated; for the identification of summer peanuts in Zhengyang County, the data in August dominated; for the identification of winter wheat in Zhengyang County, each month in the phenological period played a major role effect.

Ranking	Variables	Importance Score
1	ndvi705_Mar	0.3359
2	gndvi_Mar	0.2089
3	tndvi_Mar	0.1517
4	pssr_Mar	0.1004

5	arvi_Mar	0.0433
---	----------	--------

Table 5. Construction elements of the genetic chain of dispersed vegetation index of winter wheat in Huaibin County

Ranking	Variables	Importance Score
1	ndi45_Mar	0.268
2	gndvi_Oct	0.1695
3	dvi_Jan	0.1105
4	dvi_Feb	0.0991
5	ndi45_Apr	0.0454
6	gndvi_Apr	0.0429
7	ndi45_May	0.0269
8	mcari_Jan	0.0259

Table 6. Construction elements of the genetic chain of dispersed vegetation index of winter wheat in Zhengyang County

Ranking	Variables	Importance Score
1	gemi_Jun	0.5359
2	ieci_Aug	0.0807
3	mcari_Aug	0.0743
4	msavi2_Jun	0.037
5	gndvi_Jun	0.0251
6	mtci_Aug	0.0247
7	ndvi705_Aug	0.0221
8	ieci_Jun	0.0204

Table 7. Construction elements of the genetic chain of dispersed vegetation index of summer peanuts in Zhengyang County

5.2 Crop Identification Results

The IaiNet algorithm shows good recognition potential in different crop recognition scenarios. As shown in Table 8, the crop identification calculation using the IaiNet algorithm has obtained good accuracy performance in three scenarios: Huaibin County winter wheat, Zhengyang County winter wheat, and Zhengyang County summer peanuts (accuracy > 90%, Kappa coefficient > 0.9). For the vertical comparison of three different scenarios, the recognition accuracy of winter wheat in Zhengyang County is better than the other two scenarios.

Crop Recognition Scenario	OA (%)	UA (%)	PA (%)	Kappa
Huaibin-Winter Wheat	97.14	96.27	98.47	0.9425
Zhengyang-Winter Wheat	99.27	99.24	99.98	0.9392
Zhengyang-Summer Peanuts	96.25	97.5	96.7	0.9179

Table 8. Accuracy evaluation of IaiNet crop identification results

5.3 Compared with other classification and identification methods

In this paper, the crop identification results obtained by the IaiNet algorithm were compared with the random forest algorithm (RF) and the support vector machine (SVM) algorithm to comprehensively evaluate the crop identification ability of the IaiNet algorithm.

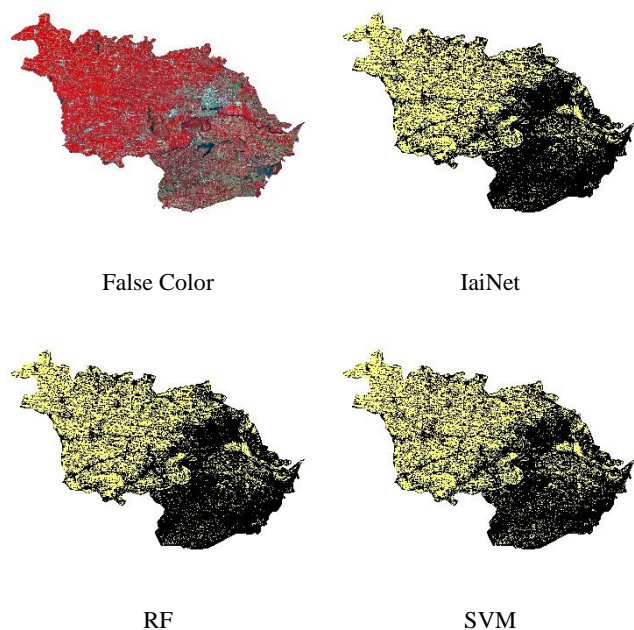


Figure 2. Comparison of crop recognition results under the recognition scene of winter wheat in Huaibin (Note: the false-color image was acquired in April, and the band combination was R: near-infrared G: red B: green)

Figure 2 shows the comparison results of crop identification for winter wheat in Huaibin. From the intuitive comparison of the results, it is found that the IaiNet algorithm has the best recognition effect of winter wheat in Huaibin. Compared with RF and SVM, the recognition results of IaiNet algorithm have the least noise and clear boundaries of cultivated land, which can effectively filter non-agricultural elements such as roads and residential areas in villages. For the area south of the Huaihe River, the actual agricultural planting situation on the surface is complex, and the wheat planting plots are fragmented and scattered, which causes certain difficulties in the identification of winter wheat in this scenario. The phenomenon in Huaibin shows that the IaiNet algorithm has certain robustness in the face of the mixed pixel influence generated by the complex surface.

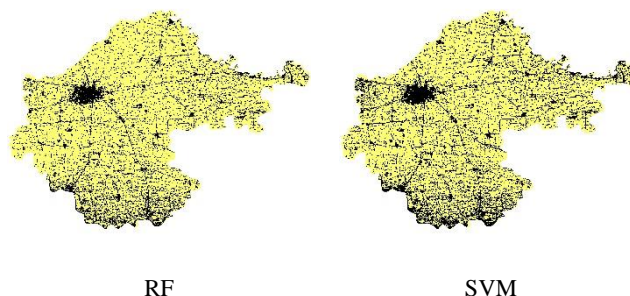
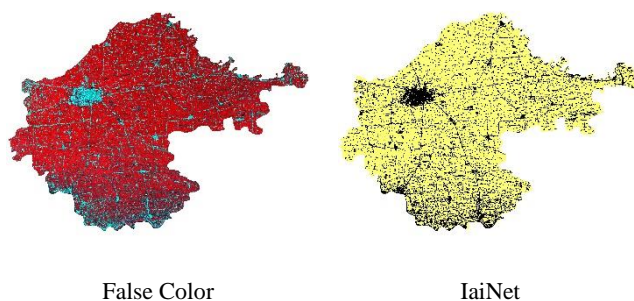


Figure 3. Comparison of crop recognition results under the recognition scene of winter wheat in Zhengyang (Note: the false-color image was acquired in April)

Figure 3 shows the comparison results of crop identification for winter wheat in Zhengyang County. It can be seen from the resulting graph that the three algorithms have little difference in the identification results of winter wheat in Zhengyang County, and all of them can better identify the spatial distribution pixels of winter wheat in Zhengyang County. However, in terms of details, RF generated more noise in the urban area of Zhengyang, and eliminated too many rural road pixels; SVM was affected by broken land in the southern river beach area of Zhengyang, which resulted in the missed classification of some wheat planting pixels, and relatively more noise pixels were generated in this area; Compared with the other two algorithms, IaiNet has obtained better results in identifying winter wheat in Zhengyang. It can filter the interference of a large number of non-wheat pixels while retaining the wheat pixels, and its details such as plot boundaries are best preserved.

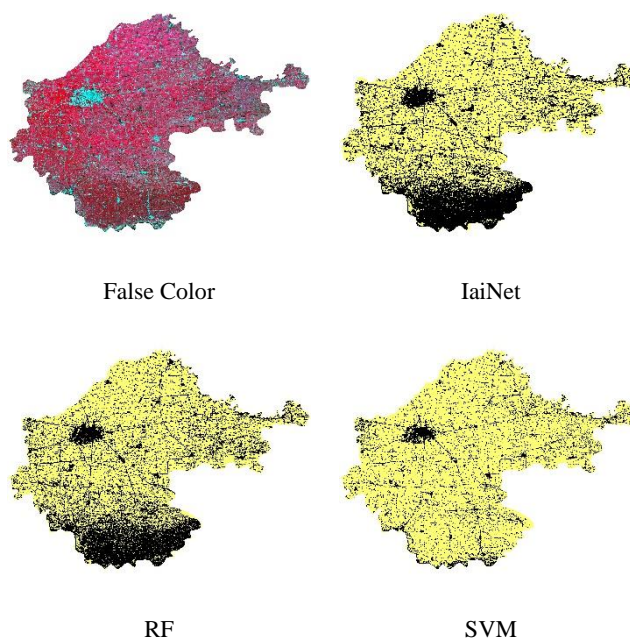


Figure 4. Comparison of crop recognition results under the recognition scene of summer peanut in Zhengyang (Note: the false-color image was acquired in August)

As shown in Figure 4, the three algorithms have produced great differences in the identification results of summer peanuts in Zhengyang. The area with the largest difference is located in the rice planting area in the south of Zhengyang County (the color is darker on the false-color image), and the IaiNet algorithm can

well identify the difference between rice and peanut pixels in this area, with less noise. Affected by the phenomenon of the same spectrum of foreign objects, RF generates more noise pixels in this area. SVM completely misclassifies rice pixels into peanuts in this area, resulting in a large loss of accuracy. Meanwhile, the IaiNet algorithm also shows better potential in the ability to strip non-agricultural elements such as rural roads and residential areas and retains the most details. This further illustrates the robustness of the IaiNet algorithm in the face of different surface conditions.

Crop Recognition Scenario	Algorithm	OA (%)	UA (%)	PA (%)	Kappa
Huaibin-Winter Wheat	IaiNet	97.14	96.27	98.47	0.9425
	RF	95.89	95.34	97.08	0.9172
	SVM	89.34	92.68	88.38	0.7839
Zhengyang-Winter Wheat	IaiNet	99.27	99.24	99.98	0.9392
	RF	99.07	99.11	99.9	0.9232
	SVM	97.2	97.43	99.57	0.788
Zhengyang-Summer Peanuts	IaiNet	96.25	97.5	96.7	0.9179
	RF	94.32	96.09	95.14	0.8755
	SVM	73.46	93.17	73.04	0.5511

Table 9. Comparison of recognition accuracy of different algorithms in different recognition scenarios

The accuracy comparison of the recognition results of different algorithms in the three recognition scenarios is shown in Table 9. Overall, the accuracy of the IaiNet algorithm is better than the other two algorithms, and the result of the SVM algorithm is the worst. In the recognition scene, the recognition accuracy of winter wheat in Zhengyang County is the best overall, and the overall recognition accuracy of winter wheat in Huaibin County and summer peanut in Zhengyang County are similar. Analysis of the reasons may lie in the following points: 1. The time series of Zhengyang County's winter wheat recognition scene is relatively complete, while the time series data are missing in both Huaibin County's winter wheat and Zhengyang County's summer peanut recognition scenes; 2. In terms of spatial distribution, the distribution of winter wheat in Zhengyang County is relatively complete, with fewer broken patterns than the other two scenarios; 3. In terms of interfering ground object type factors, the distribution of other crop types in the identification scene of winter wheat in Zhengyang County is relatively small, while the two identification scenes of winter wheat in Huaibin County and summer peanut in Zhengyang County are affected by geographical factors or plant phenological factors (such as rice identification for Zhengyang County summer peanut).

In this paper, the number of misclassified pixels of each algorithm in different recognition scenarios is counted. In Table 10, "Correct number of pixels" represents the number of pixels that were correctly identified as crops, and "Error number of pixels" was the sum of the number of pixels misclassifying crops as non-crops and non-crops as crops. The comparison shows that in the three recognition scenarios, the IaiNet algorithm has the least number of misclassified pixels, and the SVM has the worst recognition result, which produces far more misclassified pixels than the other two algorithms. From the results, the IaiNet algorithm can obtain more stable accuracy performance, but due

to various factors, some pixels will still be misclassified, resulting in a loss of accuracy.

Crop Recognition Scenario	Algorithm	Correct number of pixels	Wrong number of pixels
Huaibin-Winter Wheat	IaiNet	3224	175
	RF	3193	252
	SVM	3104	653
Zhengyang-Winter Wheat	IaiNet	5716	45
	RF	5709	57
	SVM	5612	172
Zhengyang-Summer Peanuts	IaiNet	1171	70
	RF	1154	106
	SVM	1119	495

Table 10. Statistical comparison of error segmentation pixels of different algorithms in different recognition scenarios

6. CONCLUSION

Taking the artificial immune network model as a reference, this paper proposes an improved artificial immune network crop recognition algorithm based on dispersed vegetation index genetic chain (IaiNet). In the new algorithm, the phenological period of the crop to be identified is used as the time unit to construct the monthly vegetation index data set, and then the genetic chain of the scattered vegetation index is constructed. By introducing the Mahalanobis distance, the method of measuring the similarity between "antibody" and "antigen" in the artificial immune network model is improved to realize the identification of crops. The IaiNet algorithm constructed based on the artificial immune network model can generate several "antibodies" in each category, which can effectively resist the interference caused by the phenomenon of different spectrum of same ground objects. By analyzing the results of the IaiNet algorithm in three recognition scenarios, and comparing with the results of the random forest algorithm and the support vector machine algorithm, the conclusions are as follows:

1. According to the change of the identification scene, the identification potential of vegetation index and phenological period for different crops vary. Traditional vegetation indices such as NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index) have not shown potential in the evaluation of the importance of comprehensive multi-phenological and multi-dimensional vegetation indices.
2. In the three recognition scenarios, the recognition accuracy of the IaiNet algorithm has achieved good results (accuracy > 90%, Kappa > 0.9). Compared with the random forest algorithm and the support vector machine algorithm, the advantage of crop recognition based on the IaiNet algorithm is that it can overcome the error loss caused by the phenomenon of different spectrum of same ground objects and eliminate the impact of the lack of timing. As well, the IaiNet algorithm improves the robustness of crop identification in complex surface conditions.
3. The IaiNet algorithm is still affected by different factors during crop identification, resulting in a loss of accuracy. Including timing integrity, ground fragmentation, and complexity of ground objects.

The IaiNet algorithm proposed in this paper provides a new method for crop identification cartographic production work. In the next work, it is one of the research directions to improve the IaiNet algorithm by combining the hybrid pixel decomposition technology and the object-oriented image segmentation technology.

ACKNOWLEDGEMENTS

This paper was supported by the 2016 National Key Research and Development Plan of China, grant number 2016YFC0803103, and the "Research on Key Technologies of Agricultural Remote Sensing Monitoring" project was funded by the Henan Institute of Remote Sensing and Surveying and Mapping, Data Sharing Infrastructure of Earth System Science - Data Center of Lower Yellow River Regions(<http://henu.geodata.cn>).

REFERENCES

- Baret, F., & Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote sensing of environment*, 35(2-3), 161-173.
- Baret, F., Guyot, G., & Major, D. J., 1989. Crop biomass evaluation using radiometric measurements. *Photogrammetria*, 43(5), 241-256.
- Blackburn, G. A., 1998. Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote sensing of environment*, 66(3), 273-285.
- Castillo, J. A. A., Apan, A. A., Maraseni, T. N., & Salmo III, S. G., 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 70-85.
- Clevers, J. G. P. W., 1988, July. The application of the weighted near-infrared-red vegetation index for estimating LAI at the vegetative and generative stage of cereals. In *Proc. 16th ISPRS-Congress*, Kyoto, Japan.
- Cortes, C., & Vapnik, V., 1995. Support-vector networks. *Machine learning*, 20(3), 273-297.
- Crippen, R. E., 1990. Calculating the vegetation index faster. *Remote sensing of Environment*, 34(1), 71-73.
- Dash, J., & Curran, P. J., 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Advances in Space Research*, 39(1), 100-104.
- Daughtry, C. S., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey Iii, J. E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment*, 74(2), 229-239.
- De Castro, L. N., & Von Zuben, F. J., 2001. Immune and neural network models: theoretical and empirical comparisons. *International Journal of Computational Intelligence and Applications*, 1(03), 239-257.
- De Castro, L. N., & Von Zuben, F. J., 2002. aiNet: an artificial immune network for data analysis. In *Data mining: a heuristic approach* (pp. 231-260). IGI Global.
- Delegido, J., Verrelst, J., Alonso, L., & Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors*, 11(7), 7063-7081.
- Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS journal of photogrammetry and remote sensing*, 82, 83-92.
- Gitelson, A. A., Merzlyak, M. N., & Grits, Y., 1996, May. Novel algorithms for remote sensing of chlorophyll content in higher plant leaves. In *IGARSS'96. 1996 International Geoscience and Remote Sensing Symposium* (Vol. 4, pp. 2355-2357). IEEE.
- Gitelson, A., & Merzlyak, M. N., 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of plant physiology*, 143(3), 286-292.
- Gong, P., & Howarth, P. H. I. L. P. J., 1990. An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric engineering and remote sensing*, 56(5), 597-603.
- Gong, P., Pu, R., & Chen, J., 1996. Elevation and forest-cover data using neural networks. *Photogramm. Eng. Remote Sens.*, 62, 1249-1260.
- Hao, P., Löw, F., & Biradar, C., 2018. Annual cropland mapping using reference Landsat time series—a case study in Central Asia. *Remote Sensing*, 10(12), 2057.
- Huete, A. R., 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295-309.
- Jiang, H., Hu, H., Zhong, R., Xu, J., & Lin, T., 2020. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the us corn belt at the county level. *Global Change Biology*, 26(1).
- Kaufman, Y. J., & Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE transactions on Geoscience and Remote Sensing*, 30(2), 261-270.
- Laporte, A. D., Banger, K., Weersink, A., Wagner-Riddle, C., & Smith, W., 2021. Economic and environmental consequences of nitrogen application rates, timing and methods on corn in ontario. *Agricultural Systems*, 188(4), 103018.
- Li, C., Wang, J., Wang, L., Hu, L., & Gong, P., 2014. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote sensing*, 6(2), 964-983.
- Liu, H. Q., & Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE transactions on geoscience and remote sensing*, 33(2), 457-465.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166-177.
- Major, D. J., Baret, F., & Guyot, G., 1990. A ratio vegetation index adjusted for soil brightness. *International journal of remote sensing*, 11(5), 727-740.

- Malhi, G. S. , Kaur, M. , & Kaushik, P., 2021. Impact of climate change on agriculture and its mitigation strategies: a review. *Sustainability*, 13.
- Maxwell, A. E., Warner, T. A., & Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784-2817.
- Nicolas, M. , Guadalupe, C. , Fran?ois, W. , Silvia, V. , David, M. , & Jordi, I. , et al., 2015. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sensing*, 7(10), 13208-13232.
- Ortiz, R., Sayre, K. D., Govaerts, B., Gupta, R., Subbarao, G. V., Ban, T., ... & Reynolds, M., 2008. Climate change: can wheat beat the heat?. *Agriculture, Ecosystems & Environment*, 126(1-2), 46-58.
- Pinty, B., & Verstraete, M. M., 1991. GEMI: A Non-Linear Index to Monitoring Global Vegetation Index (MSAVI). *Remote Sensing of Environment*, 48, 119-126.
- Potgieter, A. B., Apan, A., Hammer, G., & Dunn, P., 2010. Early-season crop area estimates for winter crops in NE Australia using MODIS satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(4), 380-387.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S., 1994. A modified soil adjusted vegetation index. *Remote sensing of environment*, 48(2), 119-126.
- Richardson, A. J., & Wiegand, C. L., 1977. Distinguishing vegetation from soil background information. *Photogrammetric engineering and remote sensing*, 43(12), 1541-1552.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351(1974), 309.
- Strahler, A. H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*, 10(2), 135-163.
- Tja, B. , Yi, L. , Sr, C. , Feng, L. A. , Qh, A. , & FB Hao, et al., 2020. Performance and relationship of four different agricultural drought indices for drought monitoring in china's mainland using remote sensing data - sciencedirect. *Science of The Total Environment*.
- Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.
- Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.
- Vali, A., Comai, S., & Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), 2495.
- Wang, Y. , & Lu, Y., 2020. Evaluating the potential health and economic effects of nitrogen fertilizer application in grain production systems of china. *Journal of Cleaner Production*, 264(2), 121635.
- Xun, L., Zhang, J., Cao, D., Zhang, S., & Yao, F., 2019. Crop Area Identification Based on Time Series EVI2 and Sparse Representation Approach: A Case Study in Shandong Province, China. *IEEE Access*, 7, 157513-157523.
- Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., & Chen, Q., 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sensing of Environment*, 247, 111912.