Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data

Siwen Feng, Jianjun Zhao, Tingting Liu, Hongyan Zhang, Zhengxiang Zhang, and Xiaoyi Guo

***Abstract*—In this paper, the random forests method and the support vector machine in machine learning are explored and compared to the traditional statistical-based maximum likelihood method with 126 features from Sentinel-2A images. The spectral reflectance of 12 bands, 96 texture parameters, 7 vegetation indices, and 11 phenological parameters are successfully extracted from Sentinel-2A images in 2017. The classification result shows that the optimal combination of 13 features yields overall accuracies of traditional classification and machine learning classification of 88.96% and 98%, respectively. Short-wave infrared information shows a significant effect on distinguishing rice, corn, and soybean. The water vapor band plays a significant role in distinguishing between corn and rice. In the multiclassification problem, the machine learning methods have robustness with the identification accuracy of greater than 95% for each crop type, whereas the traditional classification result shows imbalanced accuracies for different crops.**

***Index Terms*—Crop type identification, machine learning, random forest (RF), Sentinel-2A, support vector machine (SVM).**

I. Introduction

REMOTE-SENSING data play an important role in crop classification [1]–[3]. Remote-sensing data with high spatial and temporal resolutions improve the accuracy and real-time performance of large-scale agricultural remote-sensing monitoring and are also used for various agricultural monitoring types and government decision-making, such as land lease use, commodity supply, crop area estimation, crop water demand calculation, and risk prediction [4]–[9]. Simultaneously, the effective use of existing low-cost data sources to obtain timely and accurate crop classification information is a scientific challenge.

Currently, there are two main methods of crop classification using satellite remote-sensing technology: One method is to use the spectral features of high spatial resolution data combined with multitemporal characteristics and some auxiliary features (such as elevation information and textural information) to improve classification accuracy [10]–[12]. The other way is to use high temporal resolution data by analyzing crop phenological sequence patterns or time series vegetation indices, combined with some spectral features [13]–[17]. The first method is mainly based on the different spectral characteristics of different crops, but the phenomenon of foreign bodies with the same spectrum has a certain impact on classification accuracy. Although increasing elevation and textural and multitemporal characteristics can improve crop classification accuracy according to crop characteristics, there is still much room for improvement. The other crop classification method involves studying the growth pattern and phenological characteristics of crops. The time series data of high temporal resolution images are used for identification. This method effectively improves the classification accuracy but also has higher requirements for data sources. The acquisition of images with the same quality over successive time periods also limits widespread implementation of the method.

In previous studies, crop identification was usually performed using single-source, high spatial, or high temporal resolution remote-sensing data or multisource fusional data. Landsat data are widely used in crop classification studies because of their characteristics of being available at no cost and having high spatial and high temporal resolutions [18], [19]. However, the limited features from single date images cannot meet the requirement of fine crop type classification accuracies. Therefore, Landsat data are often used in conjunction with other data, such as Moderate Resolution Imaging Spectroradiometer (MODIS) and Sentinel, to improve temporal resolution and to obtain the phenological features [9], [11], [20]. Multisource data fusion improves classification accuracy but also introduces different sensor calibration errors and the tedious work of processing multisource data [9]. MODIS data are also widely used in the paper of crop time series data due to their high temporal resolution and

Manuscript received May 4, 2019; accepted June 5, 2019. Date of publication July 1, 2019; date of current version September 29, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 41771450, 41871330, 41630749, 41571078, 41571489, and 41601438, in part by the Fundamental Research Funds for the Central Universities under Grant 2412019BJ001, in part by the Foundation of the Education Department of Jilin Province in the 13th Five-Year Project under Grant JJKH20190282KJ, and in part by the National Key Research and Development Project under Grant 2016YFA0602301. *(Corresponding author: Jianjun Zhao.)*

S. Feng is with the School of Geographical Sciences, Northeast Normal University, Changchun 130024, China, and also with the College of Urban and Environmental Sciences, Peking University, Beijing 100871, China (e-mail: 1801214097@pku.edu.cn).

J. Zhao, H. Zhang, Z. Zhang, and X. Guo are with the School of Geographical Sciences, Northeast Normal University, Changchun 130024, China (e-mail: zhaojj662@nenu.edu.cn; Zhy@nenu.edu.cn; zhangzx040@nenu.edu.cn; guoxy914@nenu.edu.cn).

T. Liu is with the National Drought Mitigation Center, University of Nebraska-Lincoln, Lincoln, NE 68583-0988 USA (e-mail: gnuliutingting@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2019.2922469

availability [16], [21], [22]. However, due to the spatial resolution of the research area, it is only suitable for large-scale crop identification. In the past two years, Sentinel data have been used in agricultural applications and data fusion studies, such as crop water demand calculations, crop phenology monitoring and disaster monitoring at no cost and with high spatial and temporal resolutions [9], [13], [23]. Also, the finer temporal resolution and unique vegetation red edge band information in Sentinel data provide richer features for crop type fine classification.

Machine learning is a way to study the computer's ability to simulate new human learning behaviors, acquire new human skills, and reorganize existing structures to continuously improve a computers' performance [24]. The application of machine learning has spread to various branches of artificial intelligence, such as expert systems, pattern recognition, intelligent robots, and many other fields. Machine learning for remote-sensing inversion includes artificial neural networks, support vector machines (SVMs), random forests and other collection methods, as well as case-based reasoning and neuro-fuzzy and genetic algorithms [25]. Compared with traditional classification methods, machine learning effectively utilizes more features and has the advantages of simple operation, short time consumption, and robustness in different data volumes and different classification types. Machine learning based on the SVMs, random forest, and deep neural network methods has many important applications in crop identification based on remote-sensing images and has achieved a high level of precision compared with the methods used in previous studies [11], [12], [17]. Among these methods, the SVMs is proven to be weakly sensitive to feature space dimensions and realizes the robustness of the Hughes effect, which is applied to multifeature identification problems [26]. The random forest method is weakly affected by data size sensitivity and noise in the data [27] and has a strong generalization ability in multiclassification problems. Therefore, SVMs and random forests have become important research methods for crop multiclassification problems.

Considering the current low-cost, high-resolution data source requirements for crop identification and the extraction and use of richer interpretation characters and identification features [28], high temporal and high spatial resolution Sentinel-2A data are used as the data source in this paper. The city of Yushu in the Jilin Province, China, is selected as the research area and a characteristic system combining phenological parameters, vegetation indices, spectral reflectance, and texture features is established. The SVMs and random forest method are compared with traditional classification methods for multifeature crop identification. The random forest method is used for feature selection in this paper, and the best model for multifeature crop classification is constructed.

The overall goal of this paper is to explore the application performance of random forest and SVMs methods in the identification of multifeature crop types, such as texture, spectroscopy, vegetation index, and phenological parameters. The specific objectives are to:

1) explore the method for extraction of phenological parameters, texture parameters, vegetation indices, and spectral features based on Sentinel-2A high temporal and high spatial resolution data;
2) explore the best feature combination scheme based on the evaluation of feature importance using the random forest method;
3) to compare the classification performance between the traditional classification maximum likelihood method and SVMs and random forest in machine learning for multifeature crop classification and to identify crop types with high precision.

II. Materials

A. Study Area

The Jilin Province has fertile soil and is an important grain commodity base in China. For many years, the grain commodity rate, commodity volume, per capita possession, and volume of transfers have ranked first in the country. The city of Yushu is located in the hinterland of the Songliao Plain, with a total area of 4712.49 km^2. The geographical position of Yushu is located at 44°30'57"–45°15'02"N, 126°01'44"–127°05'09"E, and the city is part of the temperate continental monsoon climate zone. The annual average temperature is 4.1–5.6 °C, and the average annual rainfall is 500–900 mm. The annual average sunshine amount is 2200–3000 h, and the rain and hot seasons are synchronized. The agricultural production conditions in Yushu are superior. Yushu is the country-level city with the largest planting area in the Jilin Province. Yushu has the title of "the first granary in the world," with a total planting area of 3911.26 km^2. The main crop types are corn, which is followed by rice and soybean.

B. Data Source

1) Sentinel-2A Dataset: The Sentinel-2A satellite is the first satellite in the multispectral imaging mission under the Copernicus program jointly implemented by the European Commission and the European Space Agency. Sentinel-2A is mainly used for land observations with global high-resolution and high revisiting capabilities, biophysical change mapping, monitoring of coastal and inland waters, and risk and disaster mapping.

The data source for this paper is Sentinel-2 L1C data (https://scihub.copernicus.eu/dhus/#/home), which includes a total of 33 views from the beginning of March 2017 to the end of September 2017. The average cloud cover of the images is shown in Table I. The L1C product is the upper surface apparent reflectance after radiation correction and geometric correction. Each L1C product consists of a 100 km^2 orthophoto (UTM/WGS84), and the map coordinates of the images are corrected with the digital elevation model.

Since the Sentinel-2A L1C product has been radiometrically scaled, orthographically projected, and geometrically corrected, only atmospheric correction of the data is required in the preprocessing. Among the methods for atmospheric correction of Sentinel-2 images, the Sen2cor method is proven to be more accurate than the SMAC and 6S models. Therefore, the Sen2cor method is used for atmospheric correction of images, and the

TABLE I
SENTINEL-2A DATA INFORMATION

Date	Number of frames	Cloud cover
2017-3-27	4	15.72
2017-4-16	4	0.01
2017-5-26	4	0.59
2017-6-15	5	3.97
2017-7-5	4	0.40
2017-7-25	4	0.13
2017-9-3	4	0.34
2017-9-23	4	8.14
Total	33	-

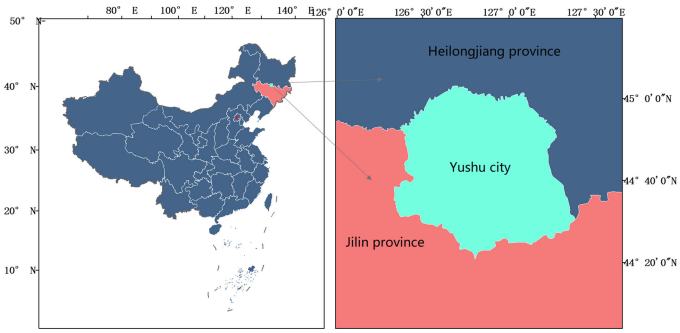

Fig. 1. Location map of Yushu.

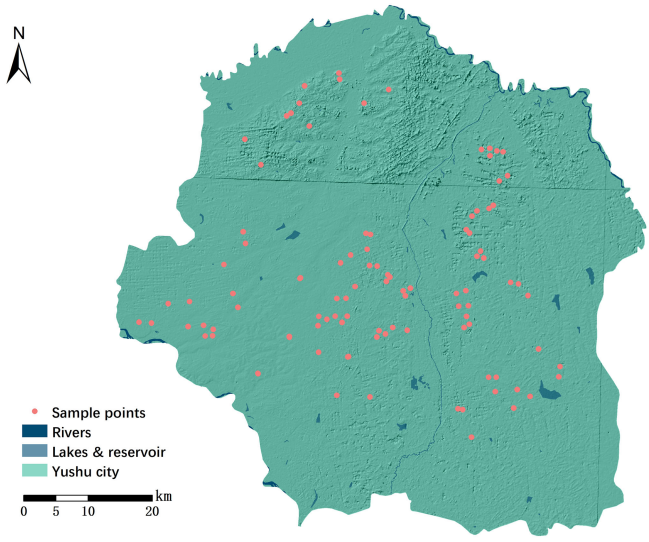

Fig. 2. Crop classification sample distribution map in Yushu.

correction is obtained for the atmospheric bottom reflectance data in 12 effective bands, except for the 10th band.

2) Field Sampling Data: This paper used field sampling data from March 29th to 30th, 2018 as verification data (see Fig. 2). Since the time of planting had not yet been reached, the type of crops planted in 2017 can be identified. Combined with Google Earth imagery, corn, soybean, rice, and other samples were selected. The single pixel is used as a single sample to reduce the correlation between samples and improve the reliability of

TABLE II
NUMBER OF SAMPLE PIXELS

Sample category	Number of pixels
Rice	11691
Corn	13628
Soybean	9362
Others	3939

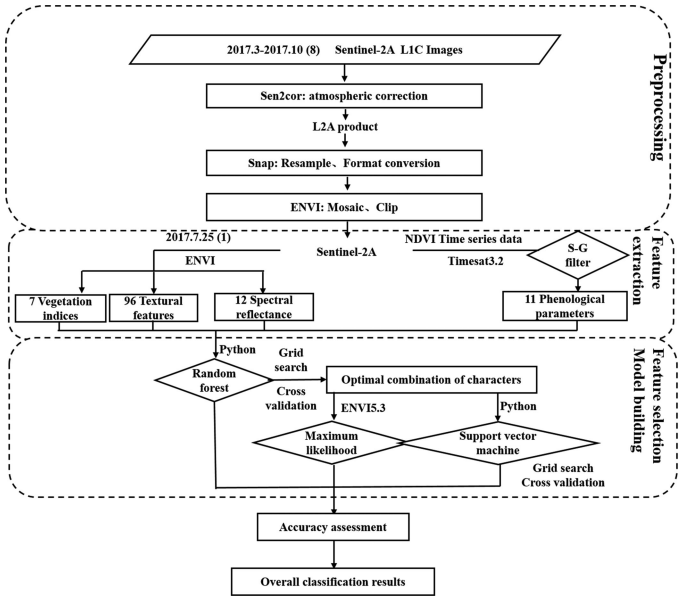

Fig. 3. Technical route map.

the classification results. The number of sample pixels is shown in Table II.

C. Technical Route

The SVMs method was used to extract cultivated land as the research area and spectral and textural features based on normalized difference vegetation index (NDVI) time series data. Timesat software was used to extract phenological parameters and 126 feature vectors of samples were used to select features with the random forest method in the Scikit-learn toolkit (http://scikit-learn.org/stable/). Based on the best feature combination, the classification accuracy of the random forest method, SVMs method, and maximum likelihood method are compared. The specific technical route is shown in Fig. 3.

D. Feature Extraction

1) Cultivated Land Extraction: Based on the sampled farmland data, the cultivated land, forestland, water bodies, building land, roads, and other land uses were selected as interest areas according to the Google Earth image. The preprocessed image for July 25, 2017 was divided into cultivated land, forestland, water, and others using the SVMs method. In this paper, all the farmland band information was extracted using the extracted cultivated land as a mask.

TABLE III
VEGETATION INDEX CALCULATION FORMULA

Vegetation index	Calculation formula
NDVI	$\frac{(\rho_{NIR}-\rho_{RED})}{(\rho_{NIR}+\rho_{RED})}$
EVI	$2.5*\frac{(\rho_{NIR}-\rho_{RED})}{(\rho_{NIR}+6*\rho_{RED}-7.5*\rho_{BLUE}+1)}$
$NDVI_{705}$	$\frac{(\rho_{750}-\rho_{705})}{(\rho_{750}+\rho_{705})}$
$mNDVI_{705}$	$\frac{(\rho_{750}-\rho_{705})}{(\rho_{750}+\rho_{705}-2*\rho_{BLUE})}$
mSR_{705}	$\frac{(\rho_{750}-\rho_{BLUE})}{(\rho_{750}+\rho_{BLUE})}$
LSWI	$\frac{(\rho_{NIR}-\rho_{SWIR1})}{(\rho_{NIR}+\rho_{SWIR1})}$
GCVI	$\frac{\rho_{NIR}}{\rho_{GREEN}}-1$

2) Spectral and Vegetation Index Feature Extraction: The NDVI [29] and enhanced vegetation index (EVI) [30] are selected as the broadband greenness indexes in this paper. The NDVI reflects the difference between the scattering of green leaves in the near-infrared range and the chlorophyll absorption in the red band. As an enhanced NDVI, EVI solves for the influence of soil background and atmospheric aerosol on dense vegetation, and saturation is not easy to reach. Due to the red edge band of vegetation unique to Sentinel-2A, the red edge normalized vegetation index (NDVI705) [31], [32], modified red edge normalized difference vegetation index (mNDVI705) [32], [33], and modified red edge simple ratio index (mSR705) [32], [33] in the narrow band greenness index were selected to construct the model. The land surface water index (LSWI) [34], which is sensitive to leaf moisture and soil moisture, was selected to distinguish between corn and rice. The green chlorophyll vegetation index (GCVI) [35] was chosen because of its strong linear relationship with the leaf area indices of soybean and corn. The vegetation index calculation formula is shown in Table III, where ρ_705 and ρ_750 represent the surface reflectance of the 5th and 6th bands, respectively.

3) Texture Feature Extraction: The gray level co-occurrence matrix (GLCM) is a classical analysis method for extracting texture by calculating the conditional probability density between gray levels of images [36], [37]. According to the basic principle of the GLCM, the gray value of the image pixel presents a certain correlation in space, and the spatial correlation is statistically analyzed. The gray value of the relevant pixel is rewritten using various texture descriptors; thus, various texture features can be presented [38]. The texture parameters calculated in this paper include mean, variance, homogeneity, contrast, dissimilarity, entropy, and second moment correlation. There are eight types of texture parameters, totaling $12 \times 8 = 96$ texture indicators.

4) Phenological Feature Extraction: Based on the eight surface reflectance data obtained from the beginning of March 2017 to the end of September in the research area, the NDVI of each image is calculated separately, and Timesat software is used to denoise the NDVI time series data and fit the NDVI curve. Finally, all phenological parameters of all pixels for the Yushu farmland in 2017 are extracted.

In this paper, S-G filtering is used in Timesat software [39] for fitting. Crop phenological parameters were extracted from the smoothed, denoised NDVI time series curve. Start of season is defined as the time when NDVI amplitude increases to 50% in the left half of the fitting function. The end of season is defined as the time when NDVI amplitude decreases to 50% in the right half of the fitting function. When the input farmland NDVI value is in the range of 0–1, the setting of the start of season and end of season values is 0.43, and the setting of the S-G window size is 3×3 pixels. The left derivative and right derivative are both set as the time when NDVI increases (decreases) to 80% of the fitting function amplitude. The length of the season is the difference between the start of the season and the end of the season. The position of the middle of the season is the mean when the left half fitting function increases to 80% and the right half fitting function reduces to 80%. The maximum fitted data is the maximum value in the fit function during the growth season, which may occur at a different position than in the middle of season. The base value is the mean of the minimum value of the left half fitting function and the right half fitting function. The amplitude is the difference value between the peak value of the NDVI time series curve and the function reference value. The large integral describes the growth season function integral from the beginning to the end of the growth season. The small integral describes the difference function integral from the beginning of the growth season to the end of the growth season and the reference level.

III. METHODS

The paper is divided into five stages: farmland extraction; feature extraction; feature optimization; optimal parameter calculation; and final classification model application. A total of 14 404 pixel samples were obtained through sampling. During the process of estimating the optimal parameters, to avoid overfitting in machine learning, the classification of random forest and SVMs involves randomly extracting 2/3 of the sample set as training data A (9602 pixels), and the remaining 1/3 of the sample data (4802 pixels) is used as test data B. The training data are divided into training sample set Aa and verification sample set Ab. The three-fold cross-validation combined with the grid search is used to verify the data to obtain the optimal parameters of the model. After the completion of the model construction, the remaining 1/3 of the test data, which is not involved in the model construction, is used to evaluate the model accuracy. The classification based on the maximum likelihood involves the manual use of the region of interest for 2/3 of the samples as the training samples (the region of interest used for classification), and the remaining 1/3 are used as the verification samples.

A. Support Vector Machines

SVM is a machine learning classification method based on the statistical Vapnik–Chervonenkis dimension theory and

structural risk minimization principle proposed by Cortes and Vapnik in 1995 [39], [40]. SVM is considered a robust classifier for the Hughes effect due to its weak sensitivity to feature space dimensions [41].

Previous research results show that when there is a large sample data volume, the radial basis function is more convergent than other kernel functions and it has better classification performance [42], [43]. In this paper, we use the radial basis function (RBF) as the kernel function. When using the classifier based on the SVMs in the Scikit toolkit, we need to determine two more important parameters, C and gamma, in the RBF function. Supposing that a dataset $\{(x_i, y_i)\}_{i=1}^{l}$ of labeled patterns is given, where $y_i \in \{-1, 1\}$, SVM is to construct a hyperplane $w^T x + b = 0$ to classify the patterns into two classes. SVM classification [40] adopts kernel function $k(x, y)$ to replace inner product operation in high-dimensional space, and introduces slack variables ξ to get the trade-off between margin w and misclassification errors, and the primal problem is optimized as follows:

$$\min J(w, b, \xi) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}\xi_i$$
$$\begin{cases} \text{s.t. } y_i\left(w^T x_i + b\right) \geq 1 - \xi_i \\ \xi_i \geq 0,\ i = 1, \ldots, l. \end{cases} \tag{1}$$

C is the penalty coefficient (cost), that is, the tolerance of the error. The higher the value of *C*, the lower the tolerance, and the more likely overfitting will occur. The smaller the value of *C*, the easier it is to fit. Values of *C* that are too large or too small will reduce the generalizability of the model. Gamma is a parameter that comes with RBF as a kernel function. *A* Gaussian RBF kernel is formulated as

$$k(x, y) = \exp\left(-\frac{||x - y||^2}{2\sigma^2}\right). \tag{2}$$

Centered by the pattern itself, each training pattern shapes a covering sphere with the same radius that is subject to the width σ. And the relationship between σ and gamma is followed as

$$\text{gamma} = \frac{1}{2\sigma^2}. \tag{3}$$

Thus, the formula (2) can be rewritten as follows:

$$k(x, y) = \exp\left(-\text{gamma} * ||x - y||^2\right). \tag{4}$$

Gamma implicitly determines the distribution of data after mapping to a new feature space. The larger the gamma, the lower the number of support vectors, and the smaller the gamma value, the greater the number of support vectors. The number of support vectors affects the speed of training and prediction. In this paper, a grid search is used in python combined with three-fold cross-validation to determine the optimal parameters *C* and gamma of the SVMs with RBF as the kernel function.

B. Random Forest Classification

Since the manual intervention required in random forest classification is small, the data classification of multifeatures can be processed, and the running time is fast, which is better for robust performance and generalization ability than a single decision tree. The random forest method was selected in this paper for classification and feature analysis.

Random forest classification is a new multidecision tree classification method proposed by Leo Breiman in 2001 to integrate multiple trees through the idea of integrated learning [44]. Information gain and Gini impurity can be selected as the decision tree portioning attribute segmentation node standard. Information gain is used in the algorithm for generating ID3, C4.5, and C5.0 decision trees. Since the entropy expresses the content of information, the smaller the entropy value, the more ordered the subset, and the larger the entropy gain, the better the subset homogeneity. If $p_k(k = 1, 2 \ldots, |y|)$represents the proportion of the *k*th sample in set D, the calculation formula of the information gain is as follows:

$$\text{Ent}(D) = -\sum_{k=1}^{|y|} p_k \log_2 p_k. \tag{5}$$

The Gini coefficient refers to the expected error rate that randomly applies a certain result from a set to a certain item of data in the set. Assuming that the samples are divided into m classes, the Gini index of the binary tree node A is calculated as (5). In the formula, p_i represents the probability of belonging to class *i*. When Gini(*A*) = 0, all samples belong to one class, and the classification effect is the best.

$$\text{Gini}(A) = 1 - \sum_{i=1}^{m} p_i^2. \tag{6}$$

C. Maximum Likelihood Method

The most commonly used supervised learning methods in remote-sensing image processing are statistical classification methods, including the maximum likelihood method, minimum distance method, and parallelepiped method. Among these methods, the classification accuracy and stability of the maximum likelihood method are the best [45]. The maximum likelihood classifier (MLC) is a supervised classifier based on the Bayes criterion and statistical analysis for evaluating similarities between other pixels and training categories. The discriminant rule is established based on the criterion of misclassification probability or minimum risk. The mean and variance of the training samples are used. In the MLC construction classification function, it is assumed that the multifeature sample data have a multidimensionally normal distribution; that is, each dimension of each type of data forms a normal distribution on its own number axis, and the data vectors of unknown categories are calculated and compared. The attribution category of the data is determined by calculating and comparing the probability that the data vector of the unknown category belongs to each category. In this paper, 2/3 of the sample data were selected as the classification interest area, and the maximum likelihood method was used for the classification.

IV. Result

A. Feature Selection Based on Random Forest

1) Feature Importance Analysis: The feature importance evaluation based on random forest includes two indicators: feature importance degree based on impurity and average accurate reduction rate. Since the feature importance degree based on impurity is biased to select features with more categorical variables to give higher importance, and when there are related features, the feature importance associated with the features is low, so the average accurate reduction rate is chosen in this paper to analyze the feature importance.

Using the average accurate reduction rate to measure the feature importance is to directly measure the influence of each feature on the prediction accuracy of the model. The change in the accuracy of the model is observed by scrambling the order of the values of a certain feature. For the unimportant features, the impact on the accuracy of the model is small, and the important features will greatly reduce the classification accuracy. Using the entire sample dataset as training data, in which the optimal parameters are determined by three-fold cross-validation and grid search, the average accurate reduction rate scores of each feature based on random forest for crop classification are shown in Table IV. The higher the score, the greater the impact on the accuracy of the model.

According to the average accurate reduction rate scores of 126 features, the first band spectral reflectivity has the most important influence on the accuracy of the model compared with the other features. A single feature can affect 0.57% of the overall classification accuracy for the 126 features. The spectral reflectance and the mean of the 1st, 9th, and 11th bands have a higher impact on the accuracy of the models, which can affect the overall accuracy by more than 0.3 percentage points. Among the most important dozen bands, the influence of the mean value of the band is more obvious than the influence of the band itself.

Among the spectral features, the impact of the spectral reflectance of the 1st band (coast/aerosol), the 11th band (shortwave infrared), and the 9th band (water vapor) on the accuracy is higher than the impact of the green band and the red-edged vegetation band, which is higher than that of the visible light spectrum. Among the texture features, the influence of the mean value is particularly prominent. The degree of influence of GCVI and mSR705 on the overall accuracy is higher than that of other vegetation indices in this experiment. The base value of the function and the beginning of the growing season have a great influence on the overall accuracy of the model that is higher than the influence of the other phenological parameters on the overall accuracy.

According to research by Yaping Cai [11], crop phenological information and uniform sample distribution can effectively improve classification performance, and shortwave infrared information provides reliable information for soybean and corn identification. In this paper, both the shortwave infrared reflectance and start of season have a high impact on the model, which can also confirm the research idea. To further explore the influence of the most important feature on the overall accuracy and the other top-level features on the classification effect, the spatial distribution of different crop features is observed through the three-dimensional (3-D) feature space scatter plot. Four representative feature spaces are selected for analysis, as shown in Fig. 4.

According to the spatial distribution of the 3-D features of different crops, in the feature space composed of features with high feature importance, the distributions of the same crops are concentrated, and the differences among crop species are significant; the lower the feature importance, the more dispersed the same crops in the feature space. As shown in Fig. 4(a)–(d), the mean value of the 9th band (steam band) is significant for distinguishing between corn and rice. Rice has lower mean values of the water vapor band and the short-wave infrared band than the other studied crops. As shown in Fig. 4(a)–(c), the mean value of the short-wave infrared band also plays a significant role in distinguishing between corn and soybean. Soybean has a significantly higher average short-wave infrared radiance than corn and a lower GCVI value than corn. Other crops have higher mean values of short-wave infrared radiance and a lower average radiance of the water vapor bands, with distribution features similar to those of soybean. The spectral reflectance of the 1st band can be used to distinguish soybean and other crops. Among the features with high feature importance, the differentiation of the distribution of various crops in the base value of the function dimension is not obvious, but the degree of aggregation within the class is obvious. In addition, the base values of corn and rice are lower than those of soybean and other crops.

2) Optimal Feature Combination: Taking the average accurate reduction rate of the features as the standard, the features are sorted in descending order of importance, and the number of features is gradually increased. Then, using the above optimal model constructed by random forest, the overall sample is subjected to three-fold cross-validation to obtain the overall classification accuracy of three overall precisions. The obtained results are shown in Fig. 5.

As seen in the figure, the overall accuracy of the random forest model in multifeature classification increases as the number of features increases, while as the growth rate gradually decreases. In this process, when the number of features of the classification increases from one to two, the overall accuracy increases by 19.78%, and an overall accuracy of 94.11% is obtained with the first four features. Then the amplitude decreases as the number of features increases. The increase in the classification accuracy of the first four features is more obvious. The overall accuracy of the model is 98.3% when the number of features of the classification reaches 13. Subsequently, the change in overall accuracy is not obvious as the number of features increases.

The increasing trend of overall accuracy with an increasing number of features is unstable. When the 126 features are used for classification, the overall accuracy can reach 98.8%. When the number of features is less than 13, the overall accuracy increases as the number of features increases, and the rising trend is obvious. When the number of features used for classification is more than 13, the overall accuracy increases slowly, and in some parts of the line, the accuracy does not increase, increases minimally, or decreases. Additionally, this unstable growth will become increasingly obvious as the number of features increases.

TABLE IV
FEATURE IMPORTANCE MEASUREMENT

No.	Feature name	Average accurate reduction rate (%)	No.	Feature name	Average accurate reduction rate (%)
1	band1	0.57	64	secondmoment_b2	0.02
2	mean_b9	0.40	65	secondmoment_b12	0.02
3	mean_b1	0.39	66	homogeneity_b9	0.02
4	mean_b11	0.35	67	homogeneity_b8a	0.02
5	band11	0.35	68	homogeneity_b8	0.02
6	band9	0.26	69	homogeneity_b3	0.02
7	basevalue_s1	0.24	70	homogeneity_b2	0.02
8	mean_b5	0.20	71	homogeneity_b1	0.02
9	band3	0.20	72	entropy_b3	0.02
10	variance_b5	0.17	73	entropy_b2	0.02
11	band5	0.17	74	entropy_b12	0.02
12	GCVI	0.17	75	entropy_b11	0.02
13	mean_b4	0.16	76	entropy_b1	0.02
14	mSR705	0.16	77	dissimilarity_b9	0.02
15	startseason_s1	0.15	78	dissimilarity_b8	0.02
16	band6	0.15	79	dissimilarity_b3	0.02
17	band2	0.14	80	dissimilarity_b2	0.02
18	mean_b8a	0.13	81	correlation_b9	0.02
19	mean_b6	0.12	82	correlation_b5	0.02
20	contrast_b5	0.12	83	correlation_b3	0.02
21	band12	0.12	84	correlation_b2	0.02
22	mNDVI705	0.11	85	correlation_b12	0.02
23	mean_b8	0.10	86	correlation_b11	0.02
24	band4	0.10	87	contrast_b9	0.02
25	mean_b7	0.09	88	contrast_b8a	0.02
26	mean_b3	0.09	89	contrast_b7	0.02
27	NDVI705	0.09	90	contrast_b3	0.02
28	mean_b12	0.07	91	LSWI	0.02
29	dissimilarity_b5	0.07	92	EVI	0.02
30	band8a	0.07	93	variance_b12	0.01
31	variance_b6	0.06	94	secondmoment_b7	0.01
32	mean_b2	0.06	95	secondmoment_b4	0.01
33	leftderivative_s1	0.06	96	secondmoment_b1	0.01
34	largeintegral_s1	0.06	97	homogeneity_b7	0.01
35	band7	0.06	98	homogeneity_b12	0.01
36	variance_b3	0.05	99	homogeneity_b11	0.01
37	smallintegral_s1	0.05	100	entropy_b8a	0.01
38	rightderivative_s1	0.05	101	entropy_b7	0.01
39	lengthseason_s1	0.05	102	entropy_b6	0.01
40	dissimilarity_b6	0.05	103	entropy_b4	0.01
41	contrast_b6	0.05	104	dissimilarity_b8a	0.01
42	amplitude_s1	0.05	105	dissimilarity_b7	0.01
43	variance_b9	0.04	106	dissimilarity_b4	0.01
44	variance_b4	0.04	107	dissimilarity_b1	0.01
45	positionmseason_s1	0.04	108	correlation_b8a	0.01
46	homogeneity_b5	0.04	109	correlation_b8	0.01
47	endseason_s1	0.04	110	correlation_b7	0.01
48	band8	0.04	111	correlation_b6	0.01
49	NDVI	0.04	112	correlation_b4	0.01
50	maxfitdata_s1	0.03	113	correlation_b1	0.01
51	homogeneity_b6	0.03	114	contrast_b8	0.01
52	entropy_b9	0.03	115	contrast_b4	0.01
53	entropy_b5	0.03	116	contrast_b11	0.01
54	contrast_b2	0.03	117	contrast_b1	0.00
55	variance_b8a	0.02	118	variance_b8	0.00
56	variance_b7	0.02	119	secondmoment_b8a	0.00
57	variance_b2	0.02	120	secondmoment_b8	0.00
58	variance_b11	0.02	121	secondmoment_b11	0.00
59	variance_b1	0.02	122	entropy_b8	0.00
60	secondmoment_b9	0.02	123	dissimilarity_b12	0.00
61	secondmoment_b6	0.02	124	dissimilarity_b11	0.00
62	secondmoment_b5	0.02	125	contrast_b12	0.00
63	secondmoment_b3	0.02	126	homogeneity_b4	0.00

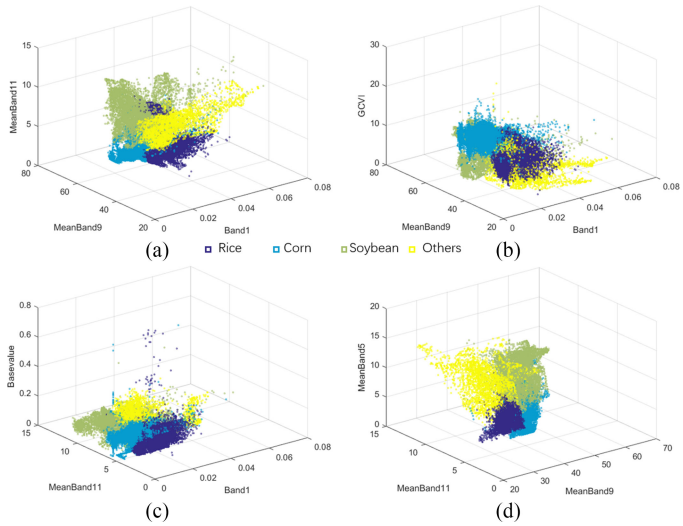

Fig. 4. Spatial distribution of 3-D features of crop identification.

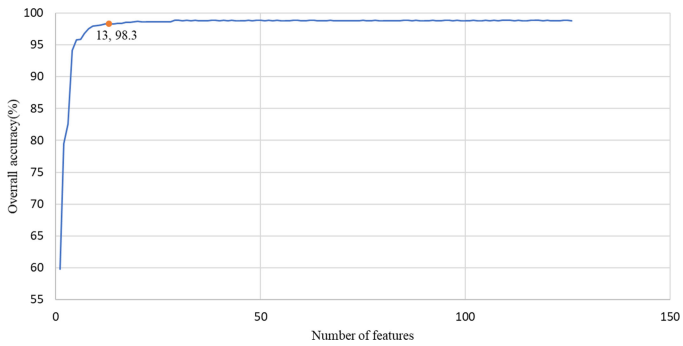

Fig. 5. Relationship between the number of features and the accuracy of random forest.

There are two reasons for this phenomenon. On the one hand, in the calculation of the overall accuracy, the model disturbs the sample for training fitting and verification. Each fitting result is not necessarily consistent with the verification data, and there will be differences in accuracy. The mean value of the three-fold cross-validation was used in this paper as the last accuracy to minimize the inconsistency caused by the training and fitting of randomly selected samples. On the other hand, the influence of the number of features on the overall accuracy of the model is sorted according to the descending order of feature importance. The importance of the added features is decreasing, and an increasing amount of feature noise content is added to the latter; thus, the accuracy of the model appears to grow unsteadily as the number of features grows.

The final optimized combination of features comprises 13 features: the spectral reflectance of the 1st, 3rd, 5th, 9th, and 11th bands, GCVI, the base value of the NDVI fit function, the mean of the 1st, 4th, 5th, 9th, and 11th bands, and the variance of the 5th band. This combination is used for the subsequent classification methods.

B. Classification and Accuracy Assessment

Based on the identified optimal combination of 13 features, the optimal parameter combination of the SVMs is obtained by grid search combined with three-fold cross-validation: the kernel function is the RBF, C is 100.0, and gamma is 0.01. The optimal parameters of the random forest are also obtained: the information gain is the decision tree segmentation criterion, and the number of decision trees is set to 200.

Classification accuracy is an important indicator used to measure the advantages and disadvantages of classification methods. Common metrics include user accuracy, mapping accuracy, overall classification accuracy, and Kappa coefficient.

In this paper, user accuracy, mapping accuracy, overall classification accuracy, and Kappa coefficient are selected as accuracy evaluation indicators based on the confusion matrix calculation. The accuracy of the classification method is comprehensively evaluated as follows:

Overall classification accuracy (p_0)

$$p_0 = \sum_{i=1}^{n} p_{ii}/p. \tag{7}$$

User accuracy of the *i*th class of data (p_{u_i})

$$\mathbf{p}_{u_i} = p_{ii}/p_{i+}. \tag{8}$$

Mapping accuracy of the *j*th class of data (p_{A_J})

$$\mathbf{p}_{A_J} = p_{jj}/p_{+j}. \tag{9}$$

Kappa coefficient

$$Kappa = \frac{p\sum_{i=1}^{n} p_{ii} - \sum_{i=1}^{n}(p_{i+}p_{+j})}{p^2 - \sum_{i=1}^{n}(p_{i+}p_{+j})}. \tag{10}$$

p_{ij} represents the number of cells occupied by the *i*th class in the classification result and the *j*th class of the measured data. $p_{i+} = \sum_{i=1}^{n} p_{ij}$ is the sum of the number of *i*th class pixels in the classification result. $p_{+j} = \sum_{i=1}^{n} p_{ij}$ is the sum of the *j*th class pixels in the measured data. *P* is the total number of samples, and *n* is the number of classes.

To avoid the overfitting problem in multidimensional feature data classification with machine learning, the Scikit toolkit is used to perform the three-fold cross-validation of the SVMs and random forest in the accuracy verification, and the extracted sample data are randomly scrambled and then 2/3 of the data are extracted as the training sample. The cross-validation results are stable at a given value, indicating that the classification accuracy is credible, and then, the confusion matrix is generated. The remaining 1/3 of the sample area of interest is used for accuracy verification, and the accuracy verification index is shown in Table V.

According to the accuracy verification of the three classification methods in Table V, the overall accuracy of the traditional classification method is 88.96%, which is the lowest among the research methods. The traditional classification result shows that the optimal combination of features determined by random forests can effectively identify rice, corn, and soybeans, but the identification accuracy is very low for other crops

TABLE V
CLASSIFICATION METHOD ACCURACY COMPARISON

Classification methods	User accuracy (%)				Mapping accuracy (%)				Kappa	Overall accuracy (%)
	Rice	Corn	Soybean	Other	Rice	Corn	Soybean	Other		
SVM	99.64	98.58	98.78	98.87	99.64	98.89	98.40	98.72	0.986	98.98
RF	99.67	98.82	98.06	98.33	99.62	98.82	98.40	97.66	0.984	98.84
MCL	94.82	95.86	89.19	39.43	94.28	91.70	81.83	72.47	0.842	88.96

and somewhat low for soybean. This proves that the result of the traditional classification method for dry farmland classification is not good. The identification based on multifeatured sample training is not sensitive and cannot identify other crops from the main identified crops well, although the overall accuracy of the traditional method is not low. Compared with the traditional classification, the Kappa coefficient and the overall accuracy of the best model based on the SVM and random forest method in machine learning reach more than 0.98, and the user accuracies of all crops are above 90%. These results show that based on the multifeatured classification of this batch of samples, the classification performance of machine learning is significantly better than the traditional classification and can appropriately distinguish corn, rice, soybean, and other crops.

According to the comparison of the classification accuracy of each category, the classification accuracy of rice and corn in the traditional classification method is above 90%, indicating that the traditional classification method can better distinguish between dry and paddy farmlands with the optimal feature combination, although precise category identification cannot be carried out. In the machine learning results, rice has the highest identification accuracy, mainly because the differences between the features of rice farmlands and those of corn, soybeans, and other dry crops are obvious and easy to detect. The identification accuracies for corn, soybeans, and other crops are also higher than 98%. Which are higher than the overall accuracies of multilayer perceptron for 92.7% and ensemble of 1-D and 2-D convolutional neural networks for 93.5% and 94.6% [46]. There is no problem with large differences in accuracy due to the uneven distribution of sample numbers. This result shows that machine learning is robust to multiclassification problems with multifeatured and uneven samples. In this unbalanced multiclassification study, SVM performs as well as random forest.

Overall, the traditional classification method has a significant effect on identifying paddy farmlands and dry farmlands, but it is not effective for finer crop identification in dry farmlands. Compared to machine learning, the traditional classification method cannot identify other crops using the existing optimal combination of available features. In the traditional classification result, the classification accuracy of different crops is obviously different, and the accuracy is obviously affected by the distribution of the number of samples. By contrast, the machine learning method can effectively identify the crop category by training the multifeatured samples and can also be combined with cross-validation and the grid search in the training process to construct the best model to avoid overfitting problems. The methods are not affected by the imbalance of the sample distribution, and the classification accuracies of different crops are all high.

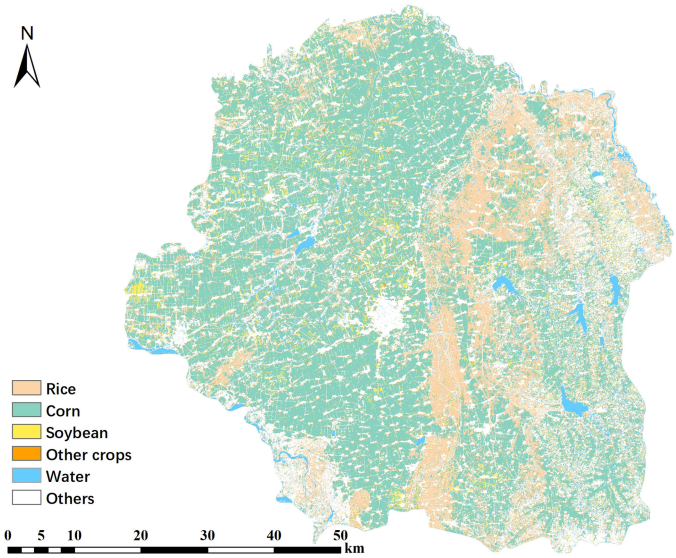

Fig. 6. Crop classification results in Yushu.

C. Classification Result

According to the optimal combination of features and optimal random forest model, the crops of Yushu were classified. The model is called to train and fit all the sample data, and the classification results are combined with the mask data of the extracted farmland to obtain the farmland crop type identification result (see Fig. 6).

V. DISCUSSION

In this paper, the phenological features, vegetation index features, spectral features, and textural features required for crop classification were obtained from the same data source of high temporal and spatial resolution images. This successfully avoided the sensor correction error in the fusion of the multisource data and effectively utilized the resolution advantages of high temporal and spatial resolution images and the characteristics of the red edge vegetation band.

Based on the ranking of feature importance and the analysis of feature number, the findings are as follows.

Among the optimal combination of 13 features identified from 126 features, the mean of the band spectral reflectance and spectral reflectance have the greatest impact on the overall accuracy. According to the spatial distribution of crop features, the shortwave infrared band is the most obvious for distinguishing rice

from other crops. Rice has the lowest mean values of short-wavelength infrared reflectance and water vapor band spectral reflectance compared to corn, soybean, and other crops. Therefore, the identification accuracy for rice is the highest among the classification results obtained by the optimal combination of features. The mean value of reflectance in the short-wave infrared band is also significant for distinguishing between corn and soybean. In the feature space, the aggregation of two types of crops is more obvious. Therefore, the average value of the short-wave infrared band is significant for distinguishing rice, corn, and soybean. As shown by the spatial distribution of features map, the mean value of the water vapor band is significant for distinguishing between corn and rice. Soybean has a significantly higher short-wave infrared reflectance than corn and a lower GCVI value than corn, and thus the feature distribution of soybean and corn crops is obvious, and the identification accuracies of these two crops are also high in all of the research methods, ranking second only to the identification accuracy of rice. Paddy farmlands can be identified most easily from spatial distribution of features map. In addition, the differences in crop distribution between the mean value of the short-wave infrared reflectance and GCVI information can be used to identify soybeans and corn in dry farmlands. Other crops are the most difficult to identify from dry farmlands. According to the spatial distribution of features, the coastal/aerosol band is most effective in identifying crops. However, the maximum likelihood method in the traditional classification method has poor identification accuracy for other crops when utilizing the existing optimal feature combination and cannot be used for finer identification. By contrast, machine learning methods obtain a more balanced and high-precision identification result for each type of crop based on the optimal combination of available features.

According to the order of feature importance, among the features with high feature importance, such as the short-wave infrared, water vapor, and red edge vegetation bands, the mean value of the spectral reflectance has a higher degree of importance than the spectral reflectance itself. The influence of the mean value on the classification accuracy is generally higher than that of the band reflectance itself, indicating that the mean of the reflectance reflects the differences in different crops more significantly than the reflectance itself.

The spatial distribution of the 3-D features of other typical high-importance features reveals that the higher the average accurate reduction rate, the more concentrated the distribution of samples within the crop, and the more obvious the difference between different types of crops. In this research, we also find that if the feature composition of the 3-D feature space is changed according to the order of decreasing feature importance, the distribution within the crop classes is gradually dispersed. Therefore, the high feature importance score is the result of a more concentrated intra-class aggregation distribution and a larger inter-class difference distribution in the spatial distribution of features of the various crops.

When the features are sorted in descending order of feature importance or when the number of features increases, the average accurate reduction rate decreases, indicating that the influence on the overall accuracy becomes smaller and smaller. Accordingly, the rate of the overall accuracy increase will become slower and slower, and there may even be unstable growth with a downward trend within a certain range due to the random selection of samples when constructing the model.

The optimal combination of 13 features determined with the random forest method is taken as the classification features, combined with the optimal parameters determined by the combination of grid search and cross-validation, and the classification accuracy of different classification methods is compared. The findings are as follows.

As shown by the spatial distribution of features, the difficulty of identifying rice is lower than that of corn and soybean, and the difficulty of identifying corn and soybean is lower than that of other crops. Therefore, the classification results of the traditional classification method show that the identification accuracy of rice is higher than that of corn, soybean, and other crops. The identification accuracy of other crops is 39%, indicating a not-ideal result. By contrast, the machine learning methods effectively utilize the existing combination of features to successfully obtain user accuracies and mapping accuracies of other crops of greater than 90%. Because the sample size distribution of different crops in this classification is not balanced, in the machine learning results, the classification accuracy of different crops does not have a notable effect, indicating balanced high precision. These results show that the machine learning method is not affected by the imbalance of sample size distribution in the classification, representing another advantage of machine learning methods over the traditional classification method. The comparison of the classification accuracies of random forest and SVM reveals that the accuracies of the two results are both finely balanced. These results indicate that in the multiclassification problem of an unbalanced sample distribution, SVM and random forest have similar classification ability.

According to the optimal combination of features determined using random forest, the overall accuracies of the maximum likelihood, SVM, and random forest methods are all above 90%, while as the accuracies of the machine learning methods are above 99%. In the classification of maximum likelihood, all other crops in the test sample were identified as soybean and corn, and the problem of overfitting appeared. Since other crop samples cannot be obtained one at a time, all the features of other crops cannot be utilized in the process of training samples using different classification methods. The other crops, as a category in the process of feature selection and training, also have a certain complexity. Therefore, the classification effect of other crops under each classification method is the worst relatively. According to the 2016 Jilin Province Statistical Yearbook, the planting area of other crops except corn, soybean, and rice in the study area accounted for 6.5% of the total, and the highest area accounted for 1.7%, which was small enough to be negligible.

The accuracy of various types of crops using SVM and random forest is different, but both show relatively low precision in the recognition of other crops. The SVM method is constructed by building the high-dimensional feature space optimal classification hyperplane. The distribution of sample data is obvious in the spatial distribution of the 3-D feature composed of the optimal combination of features. The recognition accuracy of rice,

corn, and soybean is relatively high. The random forest method employs random sampling to use the information entropy gain as the standard to train decision trees. The method of fitting the sample features and crop category to construct the forest has the highest recognition accuracy for corn and soybean. In the multiclassification problem of unbalanced sample size distribution, the SVM has a classification ability equal to that of the random forest.

In this paper, the combination of the optimal parameters and optimal combination of features determined by machine learning places the overall accuracy of sample classification at over 98%. However, the overall accuracy of crop classification in Yushu is also affected by the accuracy of farmland extraction and the quality of sample selection. In the use of machine learning methods for the fine identification of multifeature crops, high-quality sample data in each category to be classified is required to ensure classification accuracy. It is difficult to obtain corresponding classification accuracy for the sample, which has a small proportion and is difficult to obtain.

In this paper, 11 phenological parameters were extracted. However, data quality indicators such as cloud cover and noise directly affected the availability of data information, and the number of data periods directly affected the extraction of phenological features from the time series data. Spectra, textures, vegetation indices, and phenological features can be simultaneously acquired with Sentinel-2A data. The problem of decloudding and denoising with Sentinel-2A needs to be resolved in order to take advantage of the high temporal and spatial resolution and the added red edge information of vegetation.

VI. Conclusion

In this paper, 126 features of spectroscopy, vegetation index, texture, and phenological parameters were extracted from preprocessed Sentinel-2A images, and the optimal combination of 13 features was determined based on random forest feature selection. The sample training and accuracy verification were carried out using the SVM, random forest, and maximum likelihood methods, and the results are significant for the application of high spatial temporal resolution data in crop type identification.

The main conclusions of this paper are as follows.

1) High temporal and spatial resolution Sentinel-2A images can be used successfully to extract rich spectral, textural, vegetation index, and crop phenological features. The optimal combination of features consists of the spectral reflectance of the 1st, 3rd, 5th, 9th, and 11th bands, GCVI, the base value of the function, the mean of the 1st, 4th, 5th, 9th, and 11th bands, and the variance of the 5th band and results in overall accuracies of traditional classification and machine learning classification of 88.96% and 98%, respectively. The optimal parameters for SVM in machine learning are $C = 100.0$, gamma $= 0.01$, and a radial basis function as the kernel function. The best number of decision trees for random forest is 200.
2) In descending order of feature importance, the overall accuracy of random forest classification increases as the number of features increases. After 13 features, the increasing trend of overall accuracy becomes unstable, and the accuracy tends to be stable. In the spatial distribution of features with high feature importance, the aggregation within the intra-class crop distribution is obvious, and the inter-class difference is significant.
3) The short-wave infrared information has a significant effect on distinguishing rice, corn, and soybean; the water vapor band plays a significant role in distinguishing between corn and rice; GCVI helps to distinguish between corn and soybean, and the coastal/aerosol band is more effective in extracting other crops from dry fields. The mean value of the spectral reflectance is more reflective of crop characteristics than the spectral reflectance itself.
4) The difficulty of identifying rice is lower than that of corn and soybean, and the difficulty of identifying corn and soybean is lower than that of other crops. The performance of the machine learning methods for the fine identification of multifeature crops is significantly better than the traditional classification method. In the multiclassification problem, the accuracies of the machine learning methods are not affected by the uneven distribution of the sample size, and an identification accuracy of greater than 95% can be obtained for each type of crop. The SVM method has similar classification ability as the random forest method.

References

[1] C. M. Biradar *et al.*, "A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 11, no. 2, pp. 114–129, 2009.
[2] Q. Hu *et al.*, "Exploring the use of Google Earth imagery and object-based methods in land use/cover mapping," *Remote Sens.*, vol. 5, no. 11, pp. 6026–6042, 2013.
[3] G. Zhang *et al.*, "Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 157–171, 2015.
[4] D. K. Bolton and M. A. Friedl, "Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics," *Agricultural Forest Meteorol.*, vol. 173, no. 2, pp. 74–84, 2013.
[5] D. B. Lobell *et al.*, "A scalable satellite-based crop yield mapper," *Remote Sens. Environ.*, vol. 164, pp. 324–333, 2015.
[6] S. Vanino *et al.*, "Capability of Sentinel-2 data for estimating maximum evapotranspiration and irrigation requirements for tomato crop in central Italy," *Remote Sens. Environ.*, vol. 215, pp. 452–470, 2018.
[7] B. D. Wardlow, S. L. Egbert, and J. H. Kastens, "Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains," *Remote Sens. Environ.*, vol. 108, no. 3, pp. 290–310, 2007.
[8] W. U. Wen-Bin *et al.*, "How could agricultural land systems contribute to raise food production under global change?" *J. Integr. Agriculture*, vol. 13, no. 7, pp. 1432–1442, 2014.
[9] M. Wu *et al.*, "Monitoring cotton root rot by synthetic Sentinel-2 NDVI time series using improved spatial and temporal data fusion," *Scientific Rep.*, vol. 8, no. 1, pp. 2045–2322, 2018.
[10] A. Bannari *et al.*, "Estimating and mapping crop residues cover on agricultural lands using hyperspectral and IKONOS data," *Remote Sens. Environ.*, vol. 104, no. 4, pp. 447–459, 2006.
[11] Y. Cai *et al.*, "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach," *Remote Sens. Environ.*, vol. 210, pp. 35–47, 2018.
[12] P. Kumar *et al.*, "Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data," *Int. J. Remote Sens.*, vol. 36, no. 6, pp. 1604–1617, 2015.
[13] D. Bargiel, "A new method for crop classification combining time series of radar images and crop phenology information," *Remote Sens. Environ.*, vol. 198, pp. 369–383, 2017.
[14] Y. Shao *et al.*, "An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data," *Remote Sens. Environ.*, vol. 174, pp. 258–265, 2016.

[15] A. Veloso *et al.*, "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote Sens. Environ.*, vol. 199, pp. 415–426, 2017.
[16] N. Vorobiova and A. Chernov, "Curve fitting of MODIS NDVI time series in the task of early crops identification by satellite images," *Procedia Eng.*, vol. 201, pp. 184–195, 2017.
[17] Y. Zhan *et al.*, "The effect of EVI time series density on crop classification accuracy," *Optik*, vol. 157, pp. 1065–1072, 2018.
[18] M. C. Hansen and T. R. Loveland, "A review of large area monitoring of land cover change using Landsat data," *Remote Sens. Environ.*, vol. 122, no. 1, pp. 66–74, 2012.
[19] J. Townshend *et al.*, "Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges," *Int. J. Digit. Earth*, vol. 5, no. 5, pp. 373–397, 2012.
[20] J. K. Gilbertson, J. Kemp, and A. V. Niekerk, "Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques," *Comput. Electron. Agriculture*, vol. 134, pp. 151–159, 2017.
[21] Y. Chen *et al.*, "Mapping croplands, cropping patterns, and crop types using MODIS time-series data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 69, pp. 133–147, 2018.
[22] Q. Hu *et al.*, "How do temporal and spectral features matter in crop classification in Heilongjiang Province, China?" *J. Integr. Agriculture*, vol. 16, no. 2, pp. 324–336, 2017.
[23] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.
[24] C. Huyck, E. Verrucci, and J. Bevington, "Remote sensing for disaster response: A rapid, image-based perspective," *Earthq. Hazard, Risk, Disasters*, vol. 7, pp. 1–24, 2014.
[25] S. Liang *et al.*, "Recent progress in land surface quantitative remote sensing," *J. Remote Sens.*, vol. 20, no. 5, pp. 875–898, 2016.
[26] B. Bauer-Marschallinger, D. Sabel, and W. Wagner, "Optimisation of global grids for high-resolution remote sensing data," *Comput. Geosci.*, vol. 72, pp. 84–93, 2014.
[27] V. F. Rodriguez-Galiano *et al.*, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, no. 1, pp. 93–104, 2012.
[28] W. Na *et al.*, "Identification of main crops based on the univariate feature selection in Subei," *J. Remote Sens.*, vol. 21, pp. 519–530, 2017.
[29] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.
[30] A. R. Huete *et al.*, "A comparison of vegetation indices over a global set of TM images for EOS-MODIS," *Remote Sens. Environ.*, vol. 59, no. 3, pp. 440–451, 1997.
[31] A. Gitelson and M. N. Merzlyak, "Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation," *J. Plant Physiol.*, vol. 143, no. 3, pp. 286–292, 1994.
[32] D. A. Sims and J. A. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote Sens. Environ.*, vol. 81, no. 2, pp. 337–354, 2002.
[33] B. Datt, "A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using eucalyptus leaves," *J. Plant Physiol.*, vol. 154, no. 1, pp. 30–36, 1999.
[34] X. Xiao *et al.*, "Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data," *Remote Sens. Environ.*, vol. 82, no. 2, pp. 335–348, 2002.
[35] A. A. Gitelson *et al.*, "Remote estimation of leaf area index and green leaf biomass in maize canopies," *Geophys. Res. Lett.*, vol. 30, no. 5, p. 1248, 2003.
[36] J. Hirsch, "The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery," *Int. J. Remote Sens.*, vol. 26, no. 4, pp. 733–745, 2005.
[37] O. Regniers, L. Bombrun, V. Lafon, and C. Germain, "Supervised classification of very high resolution optical images using wavelet-based textural features," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3722–3735, Jun. 2016.
[38] R. M. Haraclick, "Texture features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[39] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[40] O. Bousquet, "New approaches to statistical learning theory," *Ann. Inst. Statistical Math.*, vol. 55, no. 2, pp. 371–389, 2003.
[41] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sep. 2006.
[42] Z. Huang *et al.*, "Credit rating analysis with support vector machines and neural networks: A market comparative study," *Decis. Support Syst.*, vol. 37, no. 4, pp. 543–558, 2004.
[43] K. J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing,* vol. 55, no. 1, pp. 307–319, 2003.
[44] L. Breiman, "Random forests, machine learning 45," *J. Clin. Microbiol.*, vol. 2, pp. 199–228, 2001.
[45] C. Lee and D. Landgrebe, "Feature extraction and classification algorithms for high-dimensional data," Purdue University, West Lafayette, IN, USA, ECE Tech. Rep. 212, Jan. 1993.
[46] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 99, pp. 1–5, May 2017.

Siwen Feng was born in Shijiazhuang, China, in 1996. She received the Bachelor of Science degree in geographical information science from Northeast Normal University, Changchun, China, in 2018. She is currently working toward the master's degree with the College of Urban and Environmental Sciences, Peking University, Beijing, China.

Her current research interests include soil erosion and modeling.

Jianjun Zhao received the B.S. degree in geographical information system and the doctor's degree in cartography and geographical information system from Northeast Normal University, Changchun, China, in 2007 and 2013, respectively.

From 2013 to 2015, he was a Postdoctoral Fellow majoring in environmental science and engineering in the School of Environment, Northeast Normal University. From 2011 to 2012, he was trained as a Joint Doctor with the University of Rhode Island, Kingston, RI, USA. He is currently an Associate Professor with the School of Geographical Sciences, Northeast Normal University. His research interests include global climate change, land surface phenology, and remote sensing.

Dr. Zhao was the recipient of the third prize of Natural Science Academic Achievement award in Jilin province, China.

Tingting Liu received the BS degree in information management and information systems from Shanxi University, Taiyuan, China, the MA degree in national economics from Guangxi Normal University, Guilin, China, and the Ph.D. degree in environmental and natural resources economics from the University of Rhode Island, Kingston, RI, USA in 2003, 2006 and 2014, respectively.

She is currently a Drought Policy and Impact Analyst with National Drought Mitigation Center, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA, following a 3-year appointment as a Postdoc Fellow with the Oak Ridge Institute for Science and Education, U.S. Environmental Protection Agency's Cincinnati office. She was an Assistant Professor with the School of Economics and Management, Guangxi Normal University, Guilin, China, prior to coming to the USA in 2009.

Hongyan Zhang received the B.S. degree in geography from Northeast Normal University, Changchun, China, in 1985, and the Ph.D. degree in cartography and geographical information system from the Institute of Geographic Science and Resources, Chinese Academy of Sciences, Beijing, China, in 2005.

He is currently a Professor with the School of Geographical Sciences, Northeast Normal University.

Zhengxiang Zhang received the B.S. degree in geography and the doctor's degree in cartography and geographical information system from Northeast Normal University, Changchun, China, in 1997 and 2010, respectively.

He is currently a Professor with the School of Geographical Sciences, Northeast Normal University.

Xiaoyi Guo received the B.S. degree in surveying engineering from the Jilin Institute of Architecture and Civil Engineering, Changchun, China, in 2009, and the Ph.D. degree in cartography and geography information system from Northeast Normal University, Changchun, China, in 2015.

He is currently an Associate Professor with Northeast Normal University. His research interests include vegetation dynamics and mapping of ecological disturbance using optical remote sensing.