Article

Remote Sensing Extraction of Agricultural Land in Shandong Province, China, from 2016 to 2020 Based on Google Earth Engine

Hui Liu [1,2], **Mi Chen** [1,2,*], **Huixuan Chen** [1,2], **Yu Li** [3], **Chou Xie** [4,5], **Bangsen Tian** [4,5], **Chu Wang** [1,2]
and Pengfei Ge [1,2]

1   College of Resources Environment and Tourism, Capital Normal University, Beijing 100048, China
2   State Key Laboratory Incubation Base of Urban Environmental Processes and Digital Simulation,
     Capital Normal University, Beijing 100048, China
3   National Institute of Natural Hazard, Ministry of Emergency Management of China, Beijing 100085, China
4   Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
5   University of Chinese Academy of Sciences, Beijing 100049, China
*   Correspondence: mi.chen@cnu.edu.cn

Abstract: Timely and effective access to agricultural land-change information is of great significance for the government when formulating agricultural policies. Due to the vast area of Shandong Province, the current research on agricultural land use in Shandong Province is very limited. The classification accuracy of the current classification methods also needs to be improved. In this paper, with the support of the Google Earth Engine (GEE) platform and based on Landsat 8 time series image data, a multiple machine learning algorithm was used to obtain the spatial variation distribution information of agricultural land in Shandong Province from 2016 to 2020. Firstly, a high-quality cloud-free synthetic Landsat 8 image dataset for Shandong Province from 2016 to 2020 was obtained using GEE. Secondly, the thematic index series was calculated to obtain the phenological characteristics of agricultural land, and the time periods with significant differences in terms of water, agricultural land, artificial surface, woodland and bare land were selected for classification. Feature information, such as texture features, spectral features and terrain features, was constructed, and the random forest method was used to select and optimize the features. Thirdly, the random forest, gradient boosting tree, decision tree and ensemble learning algorithms were used for classification, and the accuracy of the four classifiers was compared. The information on agricultural land changes was extracted and the causes were analyzed. The results show the following: (1) the multi-spatial index time series method is more accurate than the single thematic index time series when obtaining phenological characteristics; (2) the ensemble learning method is more accurate than the single classifier. The overall classification accuracy of the five agricultural land-extraction results in Shandong Province obtained by the ensemble learning method was above 0.9; (3) the annual decrease in agricultural land in Shandong Province from 2016 to 2020 was related to the increase in artificial land-surface area and urbanization rate.

Keywords: agricultural land; google earth engine; ensemble learning; random forest; vegetation index

1. Introduction

As a major agricultural province in China, Shandong's total agricultural output value reached USD 140.8 billion in 2020, making it the first province in China to exceed USD 137.9 billion in total agricultural output value. Located in the North China Plain, Shandong has an agricultural land area of more than ten million hectares. More than half of the province's land is agricultural land, and its agricultural value-added has long ranked first in China. Using remote sensing technology to dynamically monitor agricultural land in Shandong Province can obtain agricultural land information in a timely and effective way. Rich and effective information can provide guidance for agricultural development in Shandong Province and allow for the government to carry out scientific macro-control

according to the market economy. This is of great significance to ensure economic benefits to farmers and improve the development of local agricultural resources.

Google Earth Engine (GEE) is a comprehensive, remote sensing, cloud-computing platform developed by Google that integrates scientific analysis and geographic information data visualization. The GEE platform has a very large cloud-storage capacity and powerful cloud-computing abilities. Various Application Programming Interfaces (APIs) provided by GEE can facilitate and quickly view, calculate and process large-scale and long-time-series remote sensing data [1]. Processing and analyzing massive remote sensing data using the GEE platform is an important development direction in the field of agricultural land-use information extraction [2,3]. Relevant scholars used GEE to complete the classification and extraction of cultivated land, crops and land use, and the classification efficiency is significantly better than that of traditional classification methods [4–6].

With the advent of the era of big data, machine learning algorithms have become the mainstream way to deal with big data, such as random forest, gradient lifting tree, support vector machine, decision tree, neural network and other machine learning algorithms that have been widely used in remote sensing classification [7–13]. However, the current research on GEE classification is mainly based on the inherent GEE classification methods, such as random forest and decision tree [14,15]. Classification research is carried out based on a single classifier or a combination of multiple classifiers. Although the classification accuracy is satisfactory, there is still room for further improvements. At the same time, due to the spatial resolution of remote sensing images and the existence of the phenomena of "same object with different spectrum" and "different object with the same spectrum", many misclassifications and missing classifications are prone to occurring, and the classification accuracy needs to be improved [16]; therefore, improving the classification accuracy is the key research direction of current classification research.

Some researchers show that texture features and terrain features are applied to remote sensing classification, and the combination of various features can effectively improve classification accuracy [17–20]. However, it is not that more features will make the classification effect better. Sometimes, more features will make the data redundant and have a negative effect. In addition to adding the classification features that are important to the classification effect, method innovations based on the traditional classification method are also a way to effectively improve classification accuracy. An increasing number of classification methods have been proposed by researchers and have proven the effectiveness of the classification methods [21–23]. However, at present, new classification methods based on the GEE platform are very limited; therefore, it is necessary to innovate the algorithm based on the inherent GEE classification method.

At present, there is little research on the land-use classification in Shandong Province, especially research on the land-use classification of agricultural land. Gao Lin et al. [24] studied the distribution of land-use types in Qingdao in 2005 and 2015 using a supervised classification method. Huang Baohua [25] obtained seven periods of classified land-use results in Shandong Province and analyzed the driving force of land-use changes through a manual visual interpretation of Landsat images from the period 1970–2015. In the existing research results, the majority of land-use classification studies in Shandong Province are based on a small regional scale, and the traditional classification methods are used to analyze the results of the studies at the Shandong Province level. There are also no field sample points to participate in the classification and verification of relevant studies. Therefore, the research on land-use classification and monitoring of changes in Shandong Province needs to be further strengthened.

In view of the shortcomings of the existing research, we made changes to our own research. We improved the classification accuracy by combining different classification features and innovating classification methods, as mentioned above. This paper takes Shandong Province as the research area and uses Landsat 8 OLI image dataset through the GEE platform to study the ensemble learning method, based on multi-feature optimization combined with manual sample points and field sample points, to extract agricultural land

information in Shandong Province. This is then compared with machine learning methods, including random forest, gradient lifting tree and decision tree. Using the ensemble learning algorithm based on multi-feature optimization, information was obtained based on agricultural land-use changes in Shandong Province from 2016 to 2020 and its driving factors were analyzed. The results show that using multi-classification features and the ensemble learning method to integrate the classification results of traditional classifiers can make comprehensive use of the advantages of various classification methods, to improve classification accuracy. The results of this study can provide technical support to the dynamic monitoring of agricultural land and agricultural planning in Shandong Province.

2. Study Area and Datasets

2.1. Study Area

Shandong Province is a coastal province in East China, with latitude 34°22.9′–38°24.01′N and 114°47.5′–122°42.3′E. The mountains in central Shandong are bulging, low-lying and flat in the southwest and northwest, with slow hills in the east, and the terrain is dominated by mountainous hills. The Shandong Peninsula is found in the east, the west and north belong to the North China Plain, and there are mountainous hills in the south-central part. Mountainous hills could be seen as the skeleton, and the crisscrossed landscape of the plain basin belongs to the temperate monsoon climate.

Shandong Province is a large, traditional agricultural province and the main producer of grain and northern fruits in China. The total land area of Shandong Province is 157,126 million hectares, accounting for about 1.63% of the country's total area, ranking 19th in the country. There are 11.566 million hectares of agricultural land, accounting for 73.61% of the total land area and 7.515 million hectares of agricultural land, accounting for 47.8% of the total land area. The grain output is high, and the grain crops are planted in summer and autumn. Summer grain is mainly winter wheat; autumn grain is mainly corn, sweet potato and small grains. Wheat, corn and sweet potato are the three main grain crops in Shandong Province.

2.2. Data and Preprocessing

The main data used in this study include Landsat image data, Shuttle Radar Topography Mission (SRTM) DEM data and sample point data.

2.2.1. Landsat Data

In this paper, the Landsat 8 OLI image data processed by 1C standard from 2016 to 2020 were selected as the data source. The Landsat data include Surface Reflectance (SR) data of Landsat 8. The study area has the weather characteristics of multi-cloud cover. In order to obtain high-quality image data, cloud removal should be carried out first. The Quality Assessment (QA) band was added to the Landsat 8 data released by the United States Geological Survey. The QA band represents the surface, atmosphere and sensor conditions with unsigned shaping data, indicating whether the pixels are affected by instruments or clouds. Cloud marking can quickly be carried out from the QA band, which provides a new cloud recognition method for remote-sensing research, such as vegetation index calculation and land-use change detection. The study area has cloudy weather characteristics, and cloud processing is the primary consideration when obtaining high-quality image data. According to different images, cloud-processing methods are divided into minimum cloud-cover synthesis and the C Function of Mask (CFMASK) algorithm [26]. For Landsat SR data, the CFMASK algorithm was used to complete cloud removal. In GEE, the QA band mask was processed to obtain a cloud-free image for subsequent processing.

The main surface features in the study area can be divided into five categories: water body, agricultural land, artificial surface, and bare land. In the statistical yearbook of Shandong Province, agricultural land includes cultivated land, orchard and grassland. As shown in Table 1, by calculating the Normalized Difference Vegetation Index (NDVI), Land Surface Water Index (LSWI), Normalized Difference Built-up Index (NDBI) and

Enhanced Vegetation Index (EVI), the corresponding spectral characteristic curves are obtained. As is shown in Figure 1, the spectral index characteristics of different feature categories vary greatly between June and October, so June to October can be regarded as the best classification time period. In this paper, 945 pieces of Landsat image data from June 1 to October 1 of each year during 2016–2020 were selected for further study.

Table 1. The vegetation indices computed in this study from Landsat 8.

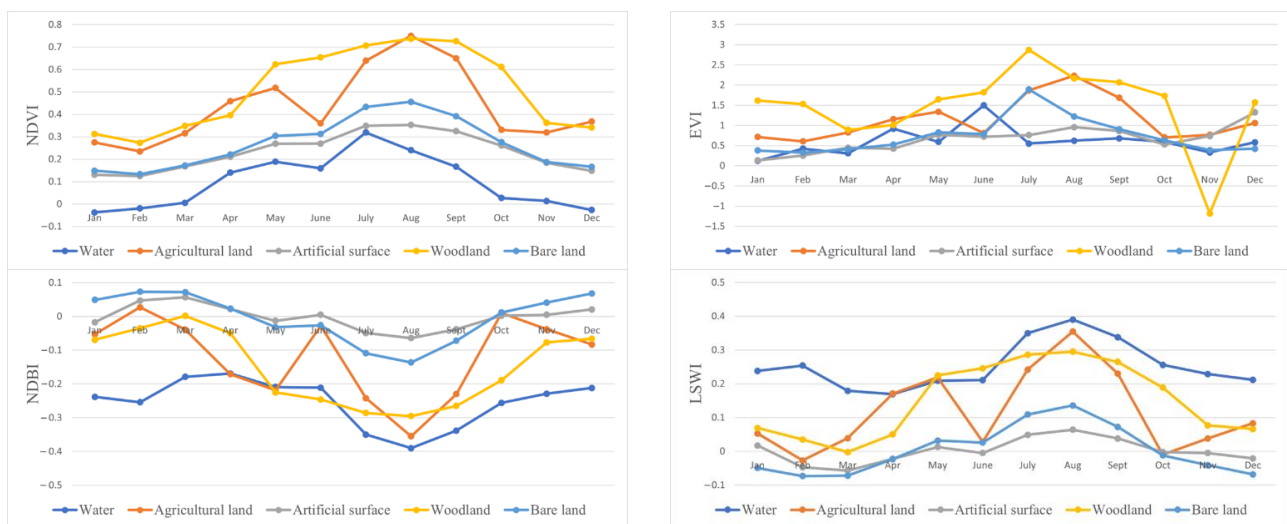| Name	Formula	References
NDVI	$NDVI = \frac{NIR - R}{NIR + R}$	Rousel et al. [27]
LSWI	$LSWI = \frac{NIR - SWIR}{NIR + SWIR}$	Xiao et al. [28]
NDBI	$NDBI = \frac{MIR - NIR}{MIR + NIR}$	Cha et al. [29]
EVI	$EVI = 2.5 \frac{NIR - R}{NIR + 6R - 7.5B + 1}$	Huete [30]

Figure 1. Spectral index characteristic curve.

2.2.2. SRTM DEM Data

In this paper, SRTM DEM data with 30 m resolution provided by NASA are used to obtain topographic features in the study area. These were cut according to the administrative boundary of Shandong Province and resampled.

2.2.3. Sample Point Data

The sample data of this study consist of manually selected data and the field sample points. The distribution information of the sample points is shown in Figure 2. The surface features in the study area were divided into five categories by manual sampling: water body, agricultural land, artificial surface, woodland and bare land. The selection of samples is based on the principle of uniformity and randomness.

The water bodies mainly include river, lake and other water-system features. Agricultural land mainly includes cultivated land, orchards, etc. The artificial surface mainly consists of buildings. Woodland mainly includes woods, forests and other land features. Bare land mainly includes bare ground where no plants grow. Online samples were selected from high-resolution images on the Google Earth platform over the years. The sample points used in this study include the sample points collected by field measurements from 2016 to 2020. The offline sample point data were obtained from sub-meter GPS field measurement, and the offline sample points cover all urban areas of the province.

The sample graph of the field sample points is shown in Figure 3. We collected hundreds of agricultural land sample points, all of which were identified as agricultural

land. The polygon region in Figure 3 is the polygon range of the agricultural land, rather than the crop type. The sample selection information is shown in Table 2, where the number of sample points is the sum of manual sample points and field sample points, and the number of sample points in parentheses is the number of field sample points.
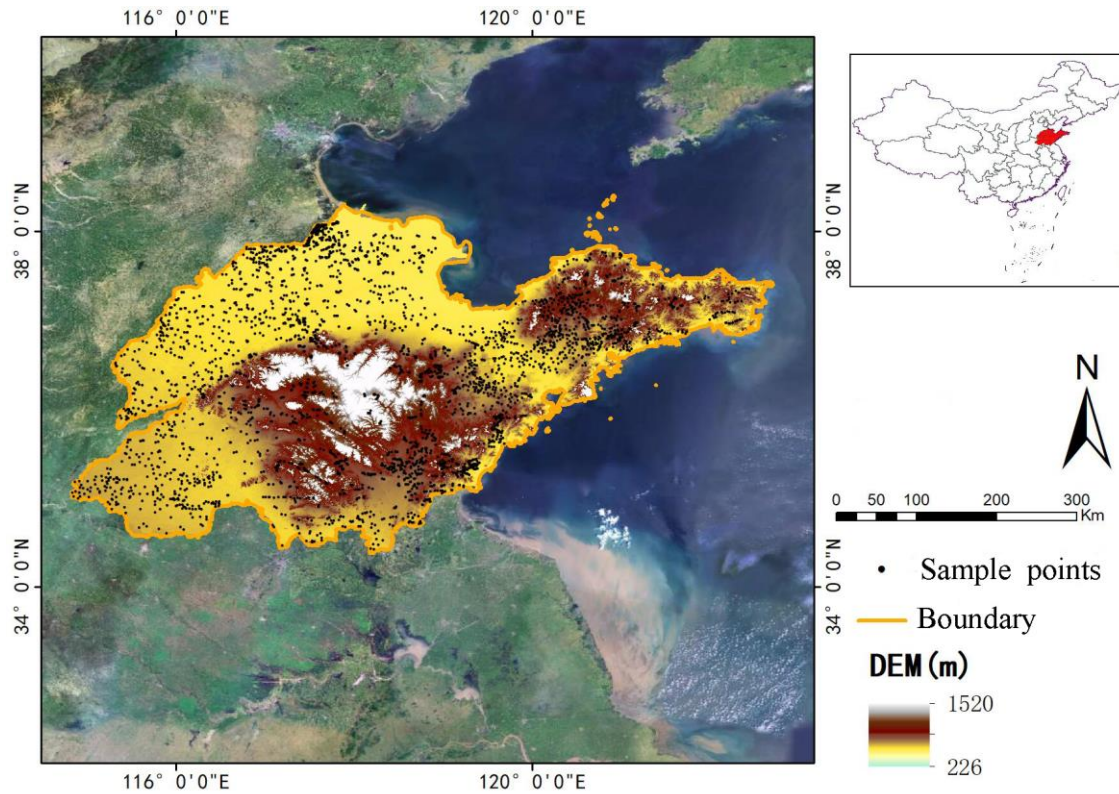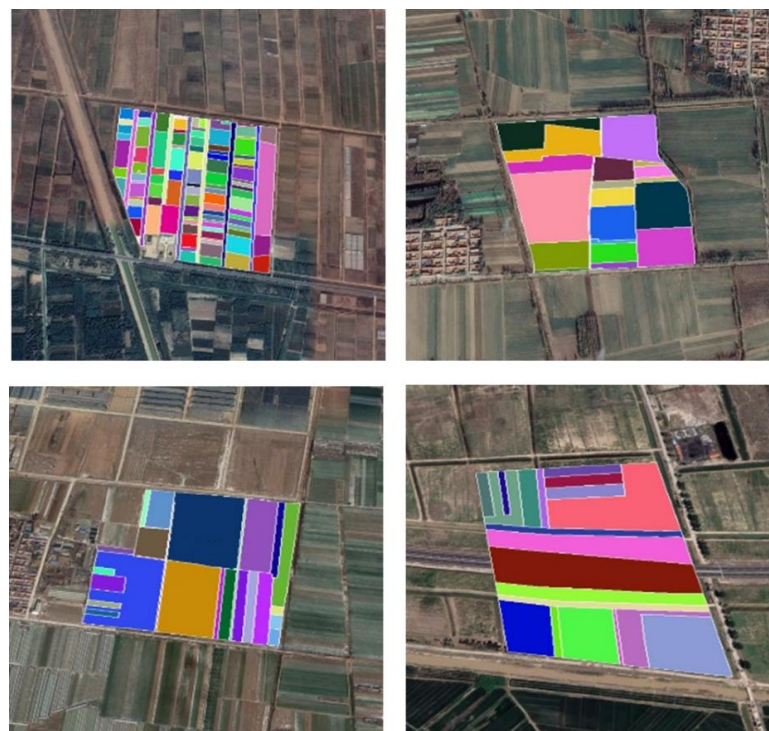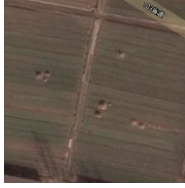

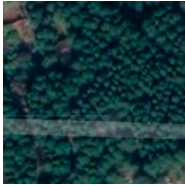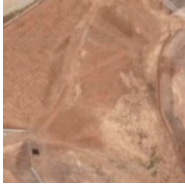**Figure 2.** Scope of study area.

Figure 3. Field sample points.

Table 2. Sample data information.

Class Code	Interpretation Signs	Class Name	Number of Samples of 2016	Number of Samples of 2017	Number of Samples of 2018	Number of Samples of 2019	Number of Samples of 2020
0		water	745 (124)	769 (132)	862 (128)	834 (126)	796 (135)
1		Agricultural land	3001 (264)	3241 (262)	3562 (274)	3620 (272)	3224 (266)
2		Artificial surface	1595	1689	1728	1590	1602
3		Woodland	1416 (138)	1588 (132)	1486 (140)	1564 (136)	1544 (130)
4		Bare land	640 (53)	524 (50)	664 (48)	652 (46)	684 (44)

3. Methods

In this study, with the support of the GEE platform, the time range of images was first determined according to the difference in the spectral index features of ground object classes, and then the Landsat image data with the best classification time phase were selected to quickly complete the image cloud removal, Mosaic and clipping preprocessing. On this basis, the spectral, texture and terrain features were constructed, and the random forest (RF) method was introduced to optimize the features. Then, the machine learning methods, including random forest, gradient lifting tree, classification and regression tree (CART) and ensemble learning were used to classify the images and evaluate the classification accuracy. Finally, the spatial distribution information of agricultural land in Shandong Province from 2016 to 2020 was obtained. The technical flow chart is shown in Figure 4.
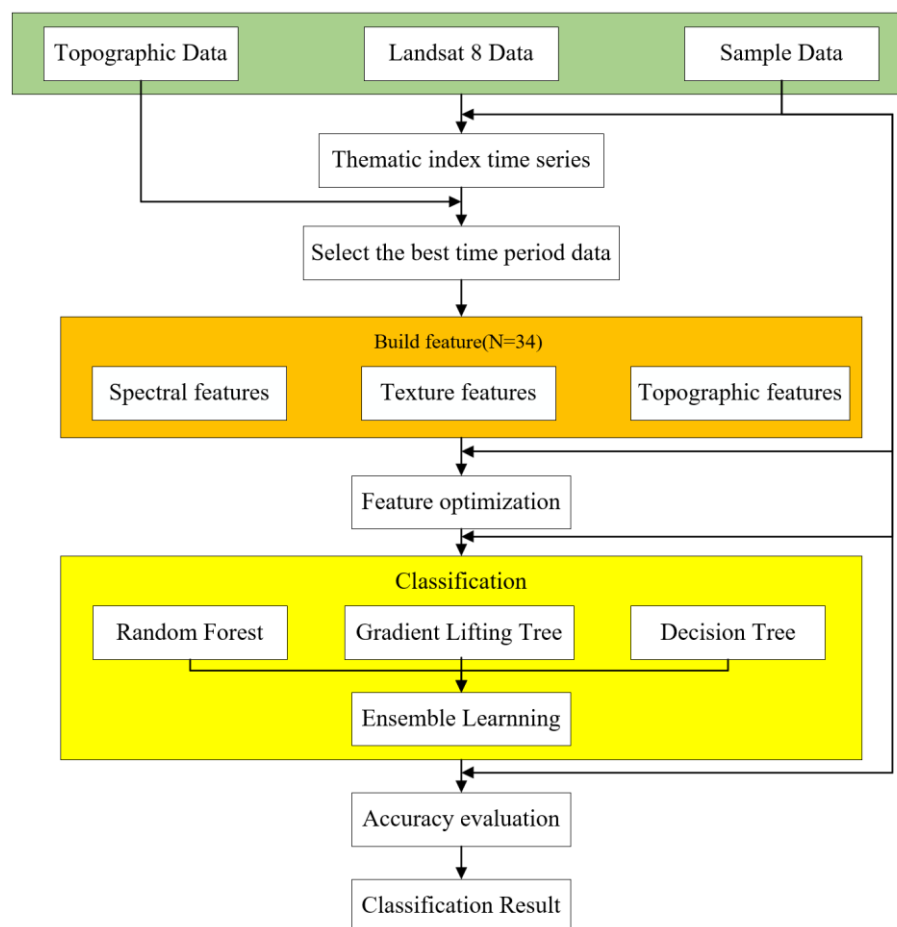
Figure 4. Technical flowchart.

3.1. Feature Construction

The extraction and selection of image features have an important impact on the subsequent image classification. In this study, spectral features, texture features and terrain features were selected as classification features.

3.1.1. Spectral Features

A total of 945 images with the optimal temporal were obtained by preprocessing. The distribution of image data information is shown in Table 3. Each image contains seven original spectral bands. NDVI, EVI, LSWI, NDWI and NDBI were calculated by GEE platform. Each spectral index obtained was added to the original spectral band as an independent band to form 12 image spectral features and complete the construction of spectral features. The spectral bands of Landsat 8 are shown in Table 4.

Table 3. Image data information table.

Year	Number of Images
2016	161
2017	193
2018	194
2019	213
2020	184

Table 4. Landsat image band parameters.

Satellite	Band	The Name of the Band	Resolution (m)
	B2	Blue	30
	B3	Green	30
	B4	Red	30
Landsat 8	B5	NIR	30
	B6	SWIR1	30
	B7	SWIR2	30
	QA	pixel_qa	30

3.1.2. Texture Features

Texture features are the regular distribution of gray value caused by the repeated arrangement of ground objects on the image, which is an important feature in remote sensing image classification. Gray-level co-occurrence Matrix (GLCM) can obtain the co-occurrence matrix and its eigenvalues by calculating the gray image, which can reflect comprehensive information, such as image direction, adjacent interval and change amplitude [31]. The texture features of remote sensing images based on the gray level co-occurrence matrix were calculated in the GEE platform, and 18 texture-feature statistics were selected to complete the construction of texture features. The specific texture statistics are shown in Table 5.

Table 5. GLCM texture statistics information.

Name of Statistic	Description	Name of Statistic	Description
constant_savg	The sum of the average	constant_prom	The clustering process
constant_shade	Clustering of the shadow	constant_svar	The variance in the sum
constant_corr	The correlation	constant_dvar	Differential variance
constant_imcorr1	Correlation information measure 1	constant_var	The variance
constant_asm	Angular second moment	constant_ent	entropy
constant_idm	Deficit moment	constant_diss	The differences
constant_dent	The differential entropy	constant_contrast	contrast
constant_sent	The entropy of the sum	constant_inertia	Moment of inertia
constant_imcorr2	Correlation information measure 2	constant_maxcorr	Maximum correlation coefficient

3.1.3. Topographic Features

The central and southern parts of Shandong Province are mountainous and hilly areas with protruding ridges, while the eastern part contains undulating and gentle wavy hilly areas. Therefore, topographic features are also important features regarding the agricultural-land information extraction for the study area. Four topographic feature components, elevation, slope, aspect and hillshade, were extracted by the GEE platform.

3.2. Feature Optimization

In this study, the spectral features, texture features and topographic features of the images were extracted, and a total of 34 image features were obtained. If all the features were to be used for classification, the classifier burden would be increased, unnecessary time would be spent and the classification accuracy may have been reduced due to the existence of redundant information; therefore, the features needed to be optimized. In this study, we used a random forest approach for feature optimization.

3.3. Classification Method

3.3.1. Random Forest Classifier

Random forest is a supervised learning algorithm that branches each decision regression tree by selecting the optimal features from the subspace of the total feature set [23]. Compared to other classifiers, random forest classification has relatively good accuracy.

The random forest classifier can also be applied to some large datasets. Additionally, the random forest classifier does not require dimensionality reductions when processing the input samples of high-dimensional features, so it can evaluate the importance of each feature in the classification. The random forest method can also obtain good results for the default value problem and ensure the independence and diversity of each decision tree, which can avoid over-fitting to a certain extent.

3.3.2. Gradient Lifting Tree Classifier

The gradient lifting tree is an improvement based on the boost algorithm. It is very different from the traditional boost algorithm [32]. The gradient lifting tree algorithm is an improved algorithm based on the AdaBoost algorithm. Gradient promotion can be regarded as a realization of boosting, which is also an interactive process. However, compared with AdaBoost boosting, gradient lifting tree is a multiple-accumulative model. The core of the gradient lifting algorithm is that each tree learns from the residuals of all previous trees, and the negative gradient value of the loss function in the current model is used as the approximate value of the residuals in the lifting tree algorithm to fit a regression classification tree. This method can deal with both continuous values and discrete values and has strong adaptability to various types of data. It has the advantages of the flexible processing of various types of data, high estimation accuracy, the use of loss function and strong robustness to outliers, and, therefore, can be effectively used for regression estimation.

3.3.3. Decision Tree Classifier

The core of the decision tree is based on the study area and target variables. The core content of the decision tree is to conduct a cyclic analysis of original datasets in the form of a binary tree structure according to learning region variables and target variables. By calculating the Gini coefficient, the decision tree method selects an attribute in the attribute set as the classification attribute of the binary tree and uses this attribute to divide the set of samples for classification into two subsets. This step should be repeated until the set of currently classified samples reaches a leaf node. The CART decision tree algorithm has the characteristics of assuming no statistical distribution of input data, clearly calculating the importance of variables for classification, selecting variables related to classification, simple implementation and a high running speed.

3.3.4. Ensemble Learning

The previous three classifiers are widely used in GEE classification research. Among them, random forest is regarded as an excellent classification method, which is widely used in classification research. Although the effect of random forest is better, this does not mean that the random forest method is superior to the other two methods in all aspects. To collect the advantages of each classifier and further improve the classification accuracy, this study adopts the classification method of ensemble learning to further integrate the classification results of the three classifiers.

Ensemble learning is an ensemble algorithm based on multiple classifiers [33]. The performance of the classifier differs in different regions of the feature space, and a single classifier can easily cause more misclassification and missing phenomena. When we use a classifier to generate incorrect classification results, it is possible to use other classifiers to obtain correct classification results. Each classifier has its own advantages and disadvantages. Combining different classifiers can make up for the disadvantages and disadvantages of a single classifier. In the classification work, the advantages can be combined with other classifiers to achieve complementarity by using a comprehensive analysis of multiple classifiers to improve the classification effect; therefore, the difference between classifiers is the most important factor in the construction of the classifier integration algorithm.

At present, the most common classifier integration methods are mainly based on the selection of training samples. The most common classifier integration methods are the

bagging algorithm and boost algorithm [34,35]. The gradient lifting tree method used above is one of the boost algorithms. Bagging is the same as random forest, where only some of the initial training samples of each base classifier are taken for training, and then the simple voting method is used for classification tasks and the simple average method is used for regression tasks. Random forest uses random attribute selection on the basis of bagging. Generally speaking, the training efficiency of random forest is better than that of bagging.

The ensemble learning algorithm adopted in this study is a parallel ensemble classifier algorithm. This method aims to design a single classifier independently in the parallel integration process, and then fuse the output results of a single classifier according to a certain strategy, train and test its classification accuracy and use this as the weight of each classifier for decision-level integration to obtain the final classification result. The weight voting algorithm sets different weights according to the effect of different classification results, carries out weighted voting and classifies the categories using the maximum number of classifiers as the pixel categories to be classified. This not only reflects the applicability of each classifier but also reduces the probability of the same number of votes being received by the direct voting method. The parallel ensemble learning algorithm considers the output of the member classifier to be independent. Through decision-level integration, we can maximize the preservation of the three classification results that are supported by the majority of classifiers. The flow chart of the ensemble learning method is shown in Figure 5.
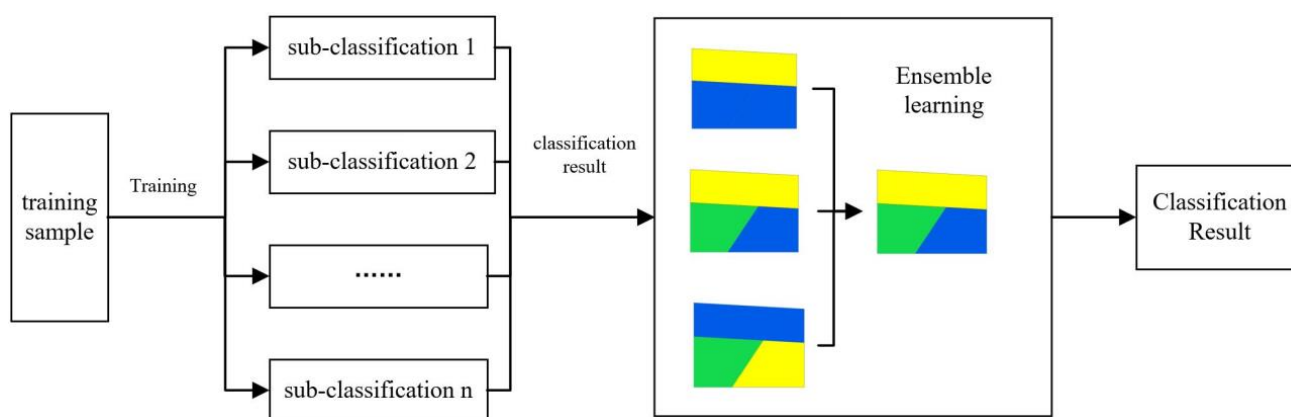

Figure 5. Flow chart of ensemble learning methods.

In this study, random forest, gradient lifting tree and decision tree were used as basic classifiers, and weighted voting was performed using an ensemble learning algorithm to achieve complementary advantages. In GEE, the random forest, gradient lift tree and decision tree classification methods can all be directly used in the GEE platform, while ensemble learning is the algorithm that we built in the GEE platform.

4. Results

4.1. Feature Optimization Results

In this study, the random forest algorithm was used to optimize the features and obtain the importance information of each feature for classification (Figure 6). The top 15 features in the importance ranking were selected as the image classification features, and we finally obtained the 15 best features (Table 6).
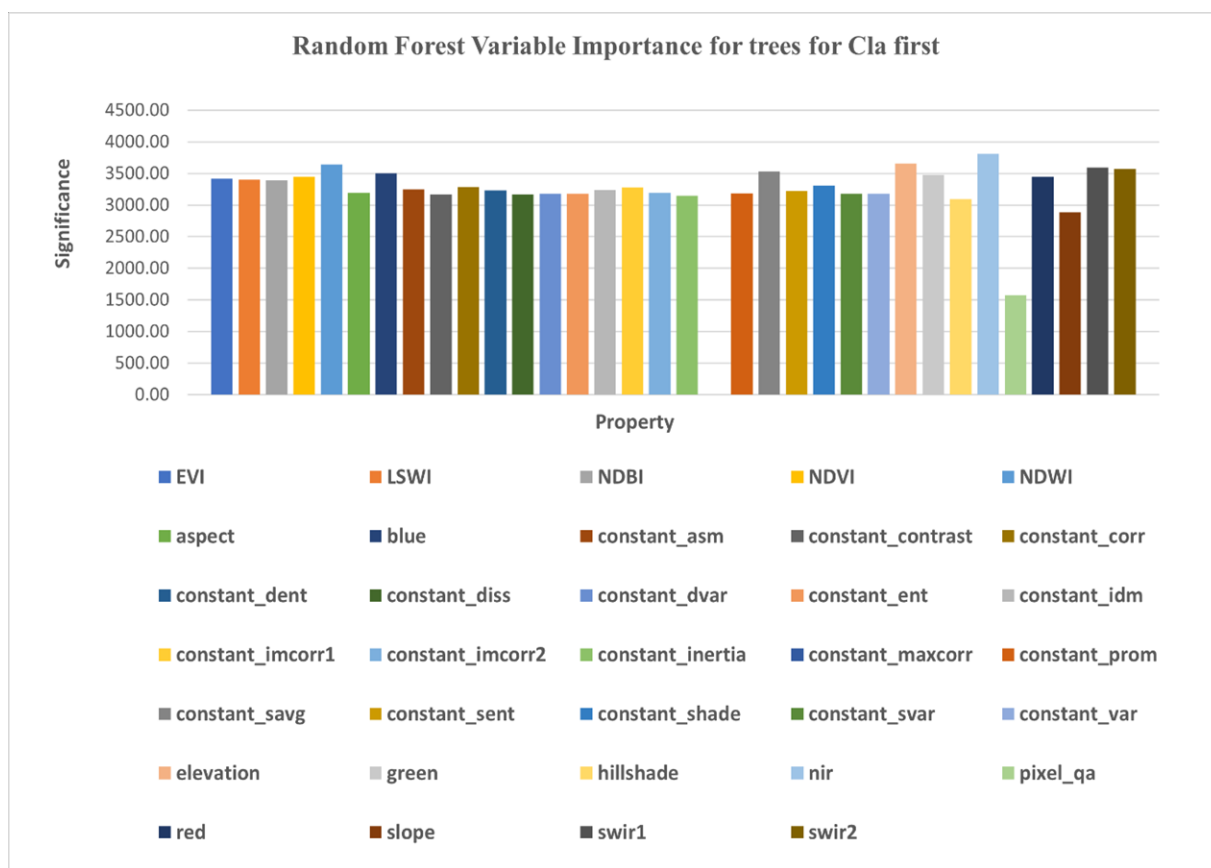
Figure 6. Distribution map of the importance of characteristic variables for classification.

Table 6. Original feature and optimized feature information.

Name of Characteristic	The Original Features	Optimized Feature
Spectral features	Blue, Green, Red, NIR, SWIR1, SWIR2, pixel_qa, EVI, NDVI, LSWI, NDWI, NDBI	NIR, Blue, Red, Green, NDVI, NDBI, EVI, NDWI, SWIR1, SWIR2
Texture features	constant_asm, constant_contrast, constant_corr, constant_var, constant_idm, constant_savg, constant_svar, constant_sent, constant_ent, constant_dvar, constant_dent, constant_imcorr1, constant_imcorr2, constant_maxcorr, constant_prom, constant_diss, constant_inertia, constant_shade	constant_savg, constant_shade, constant_dvar
Terrain features	Elevation, Slope, Aspect, Hillshade	Elevation, Aspect

4.2. Classification Results and Accuracy Evaluation

In this study, random forest, gradient lifting tree, decision tree and ensemble learning classifiers were used to classify the image dataset and extract the distribution information of agricultural land in the study area.

There are two main methods of evaluating classification accuracy. One is to rely on the experience and knowledge of experts in related fields to visually judge and evaluate the classification results, which is a qualitative method. Although the operation of this distribution is simple, it is limited by the subjective factors of expert judgement, and the evaluation results vary from person to person. The other method is based on the calculation of certain evaluation accuracy indicators, such as the overall accuracy of the classification, producer accuracy and user accuracy obtained by the confusion matrix. This quantitative evaluation method can objectively and directly reflect the accuracy of classification. Therefore, in this study, while comparing the classification results with the

optical images, the accuracy evaluation is carried out by calculating the confusion matrix based on the classification results.

According to the distribution of agricultural land in Shandong Province, we selected an area from the classification results and compared the classification of the four classification methods (Figure 7). The study found that the classification effect of the decision tree is poor and there are many misclassification phenomena, while random forest and gradient lifting tree classifications have more misclassification phenomena for agricultural land. Samples were stratified and evaluated through cross-validation. A total of 70% of the sample datasets of each type of ground object were classified, and 30% of the sample datasets were used for accuracy evaluation. Each subset follows the principle of random and uniform hierarchical sampling and the accuracy evaluation index is calculated using the confusion matrix for accurate verification results. Among these, the accuracy evaluation indexes included in the confusion matrix are average cartographic accuracy, average user accuracy and average overall accuracy.
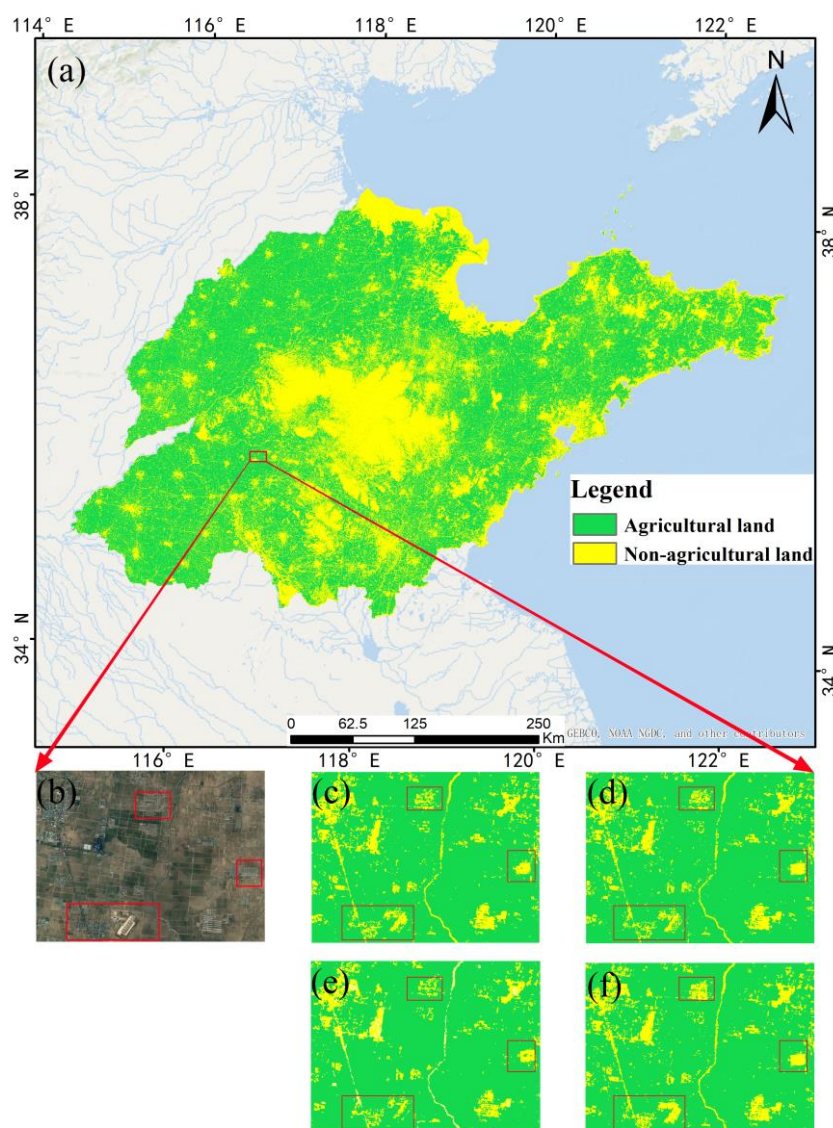
Figure 7. Comparison of different classification methods. In the results, we marked the obvious differences between each classification method with red boxes. (**a**) Winter wheat distribution map in Shandong Province; (**b**) optical image; (**c**) results of random forest classification; (**d**) results of gradient lifting tree classification; (**e**) results of decision tree classification; (**f**) results of ensemble learning classification.

As shown in Figure 8, the classification accuracy evaluation indexes of the four classifiers for five years are counted to generate the classifier accuracy evaluation graph. The mean value of the five years classification indexes for the four classification methods was calculated as the accuracy evaluation result, which is shown in Figure 8. The accuracy evaluation results and the accuracy evaluation diagram of the classifier show that the classification accuracy of the random forest and gradient lifting tree is close, while the classification accuracy of the decision tree is poor. Compared with the other three classifiers, the ensemble learning classifier can significantly improve classification accuracy. Among them, the cartographic accuracy of agricultural land category obtained by the Ensemble Learning method is more than 0.95 in five years.

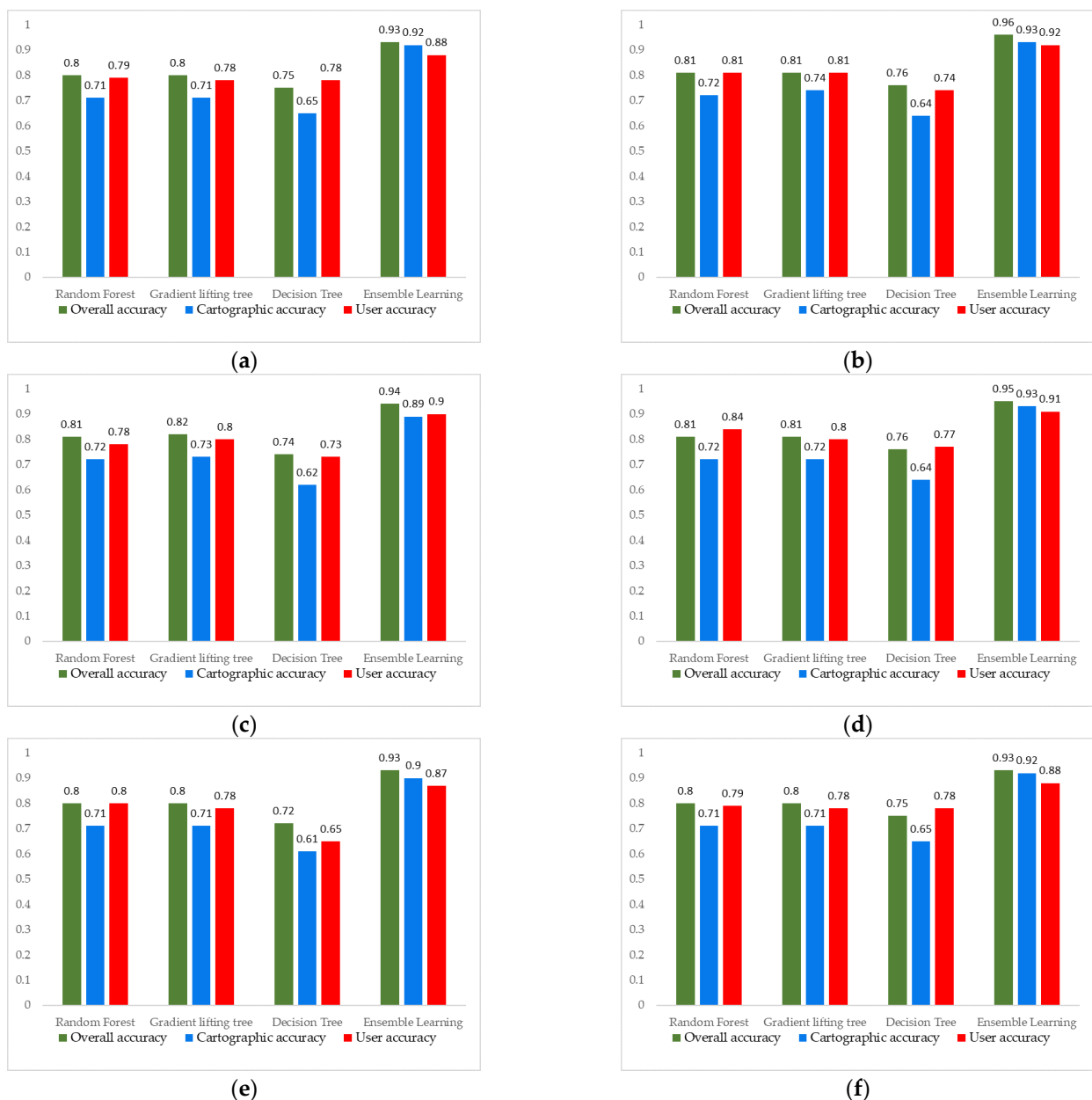**Figure 8.** Accuracy evaluation graph of classifier. (**a**) Classifier accuracy evaluation graph in 2016; (**b**) classifier accuracy evaluation graph in 2017; (**c**) classifier accuracy evaluation graph in 2018; (**d**) classifier accuracy evaluation graph in 2019; (**e**) classifier accuracy evaluation graph in 2020; (**f**) average accuracy evaluation graph of classifier from 2016 to 2020.

The classification map of agricultural land in Shandong Province obtained through the ensemble learning method is shown in Figure 9. From the ensemble learning classification results, it can be seen that Shandong Province is literally a "large province of agricultural land", with a large area of agricultural land that is mainly distributed in the hilly areas in eastern Shandong, mountainous mountains in south-central Shandong and the northwest plains. From the classification results, it can be concluded that the agricultural land area of Shandong Province in 2020 was 120.08 million hectares, accounting for 76% of the total land use area of Shandong Province. Mountain Tai is located in the middle of Shandong Province and has more forest vegetation. As the largest lake in Shandong Province, four southern lakes can be clearly seen in the classification map.

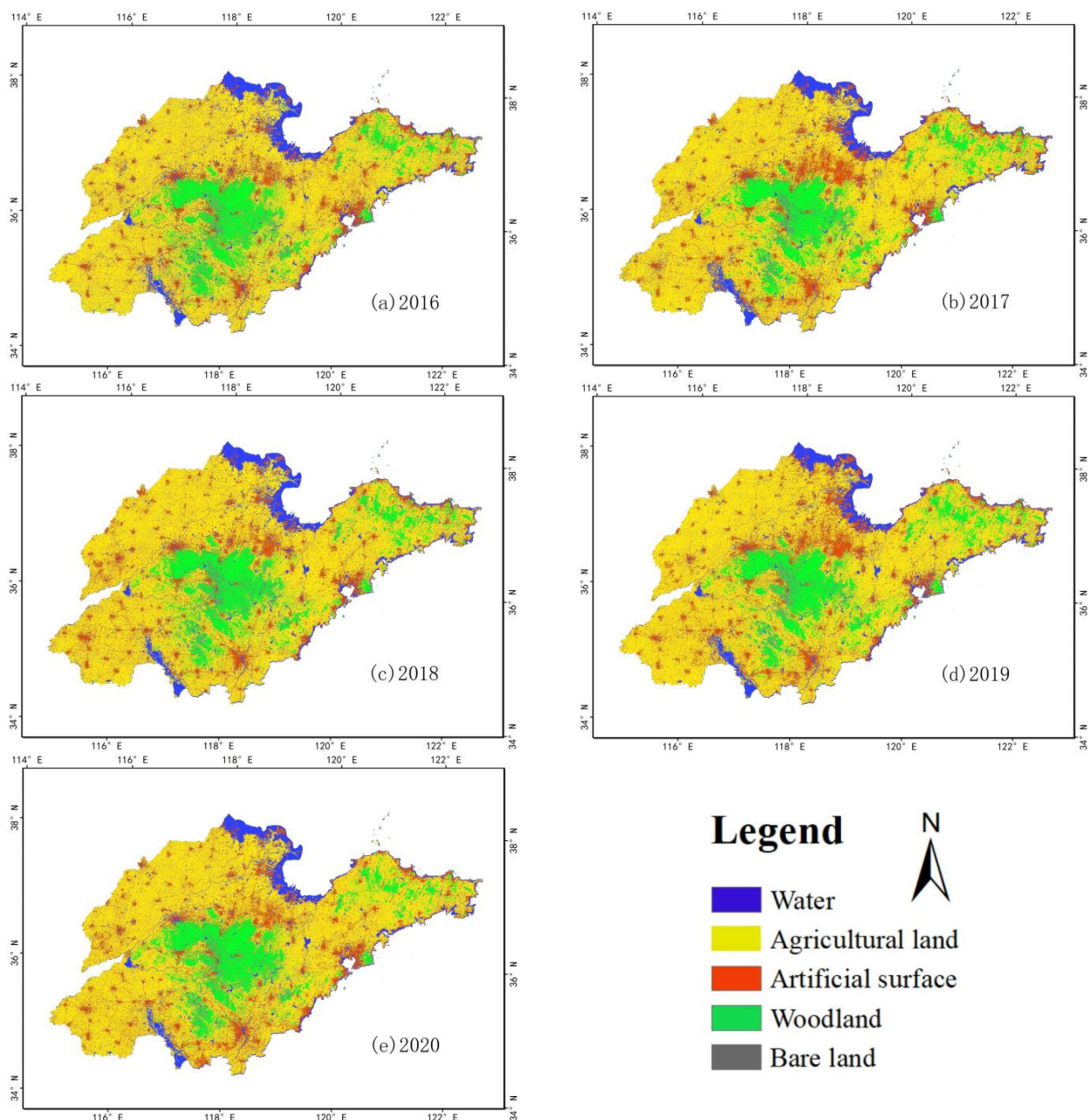**Figure 9.** Agricultural land classification map of Shandong Province in the last five years. (**a**) Land use classification map of Shandong Province in 2016; (**b**) Land use classification map of Shandong Province in 2017; (**c**) Land use classification map of Shandong Province in 2018; (**d**) Land use classification map of Shandong Province in 2019; (**e**) Land use classification map of Shandong Province in 2020.

5. Discussion

5.1. Monitoring the Change of Agricultural Land Area

Obtaining information on agricultural land changes is an important basis of reference for the government to formulate agricultural policies. According to the results of agricultural land classification, a series of monitoring methods, such as the post-classification comparison method and image ratio method, can be used to dynamically monitor agricultural land, which can more intuitively reflect the changes in agricultural land utilization. The transfer matrix is the conversion relationship of land-use types in the same area in different time periods. Based on the extraction results of agricultural land in Shandong Province from 2016 to 2020, this study used the transfer matrix method to obtain the change information on agricultural land use in Shandong Province (Figure 10).
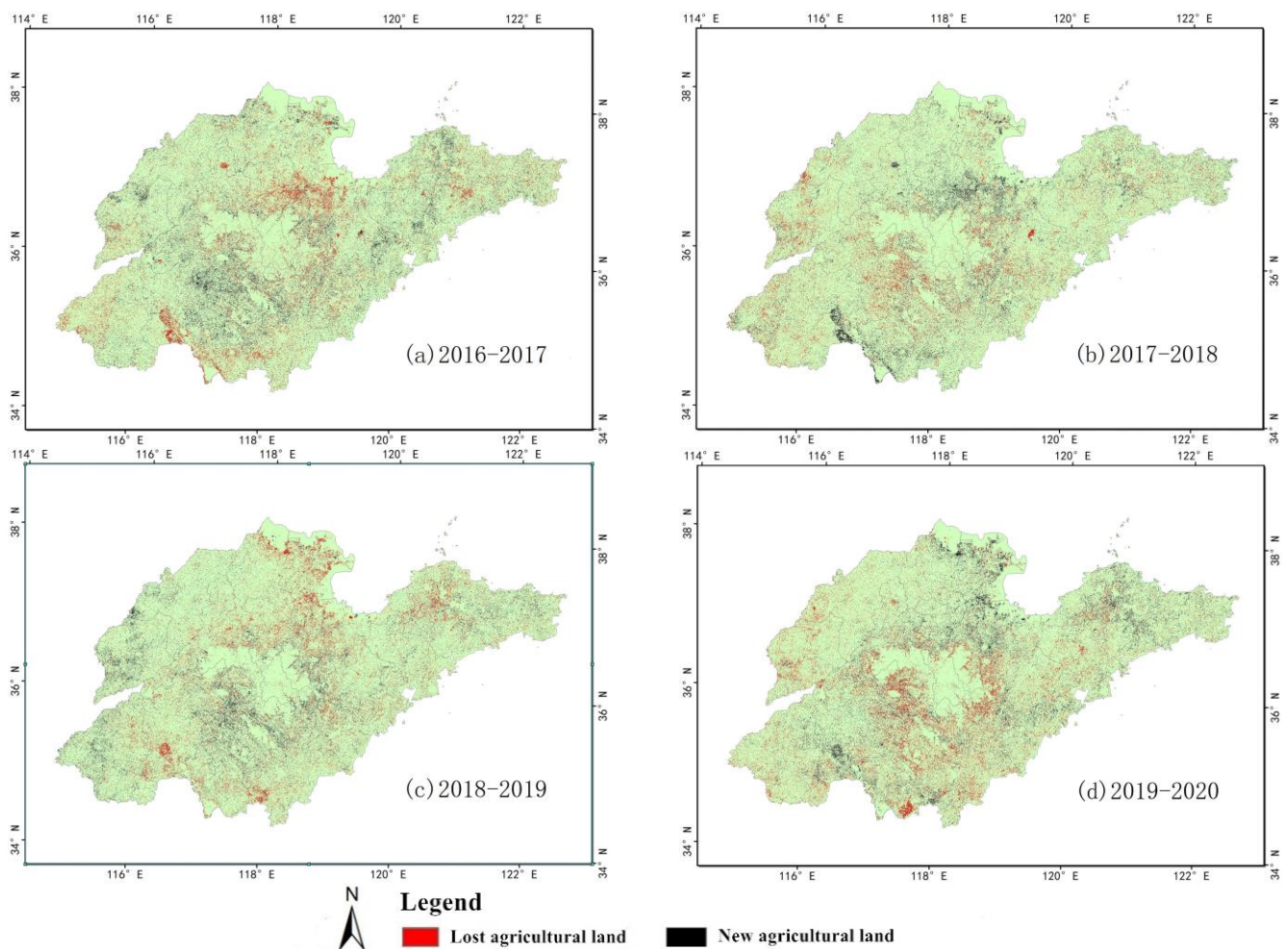

Figure 10. Information on agricultural land use changes in Shandong Province, 2016–2020. (**a**) Monitoring results of agricultural land change in Shandong Province during 2016–2017; (**b**) Monitoring results of agricultural land change in Shandong Province during 2017–2018; (**c**) Monitoring results of agricultural land change in Shandong Province during 2018–2019; (**d**) Monitoring results of agricultural land change in Shandong Province during 2019–2020.

Using the transfer matrix to obtain information on agricultural land changes, the conversion information on agricultural land and non-agricultural land can be obtained to understand the loss of and increases in agricultural land. Meanwhile, to verify the accuracy of agricultural land extraction, we obtained the official statistical yearbook of Shandong Province. The Shandong Provincial Official Statistical Yearbook includes spatial information on land-use types in Shandong Province. By calculating the classified agricultural land area

and comparing this with the statistical yearbook of Shandong Province, the agricultural land area of Shandong Province is shown to be gradually decreasing (Figure 11). Due to the lack of statistical information on agricultural land in the last two years, we can see that our results are consistent with those in the official statistical yearbook by comparing the information on agricultural land in the first three years. As can be seen from the figure, the lost agricultural land was greater than the newly added agricultural land from 2016 to 2020, showing a gradual decline.
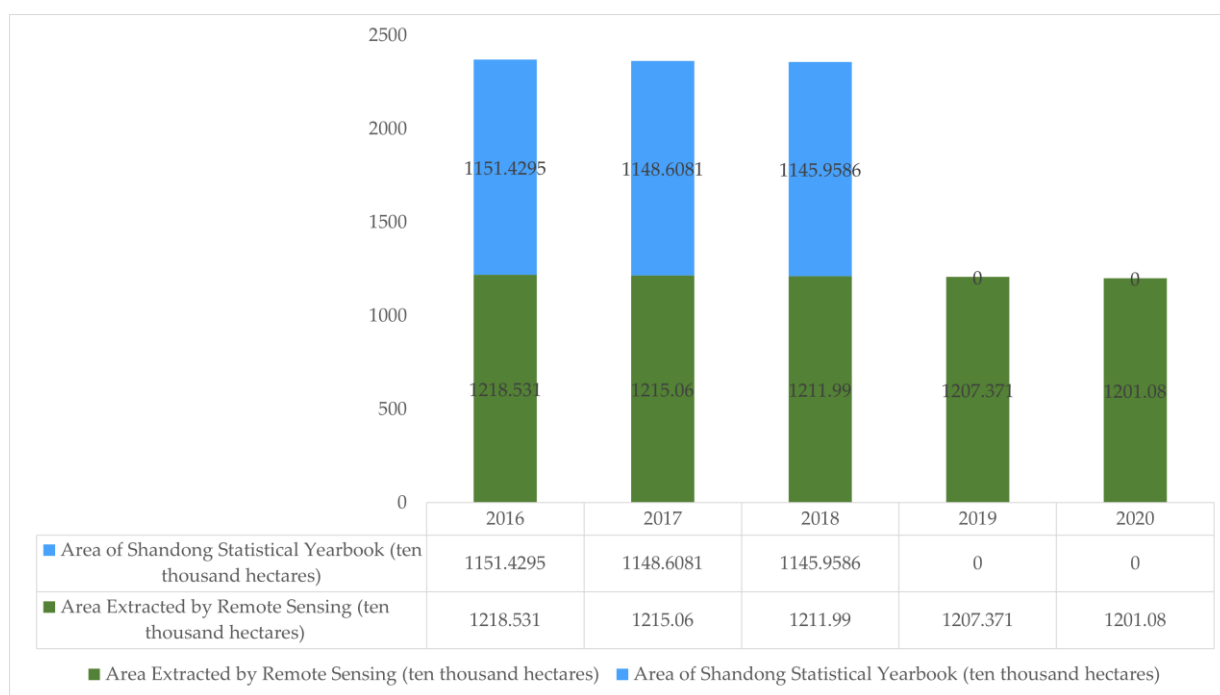

	2016	2017	2018	2019	2020
Area of Shandong Statistical Yearbook (ten thousand hectares)	1151.4295	1148.6081	1145.9586	0	0
Area Extracted by Remote Sensing (ten thousand hectares)	1218.531	1215.06	1211.99	1207.371	1201.08

Figure 11. Classification and extraction of agricultural land area and Shandong official statistical yearbook agricultural land area comparison.

5.2. Monitoring of Agricultural Land Area Change and Analysis of Main Driving Forces

The main reasons for changes in agricultural land include the increase in urbanization rate, the increase in artificial surfaces, and the influence of agricultural policies. Due to the relatively stable agricultural policies in Shandong Province in recent years, this study explored other driving factors affecting agricultural land use, including the urbanization rate and the increase in the artificial surface.

According to the classification results, the area of agricultural land in Shandong Province is decreasing each year, while the area of artificial land is increasing each year. To further study the possible direct relationship between agricultural land and artificial land surface area, a unary linear regression model between agricultural land (X) and artificial land surface (Y) was established (Figure 12): $Y = 7.45 \times 10^2 + 0.39 \times X$ where B1 is a constant term and B2 is the coefficient of agricultural land X. The unary linear regression model of agricultural land (X) and artificial land surface (Y) was obtained through regression analysis where $R^2 = 0.937$ and Pearson correlation value was -0.968, indicating a significant correlation between them.

In addition, relevant studies have shown that agricultural land changes are closely related to the rate of urbanization [36,37]. This study established a unary linear regression model between agricultural land (X) and the urbanization rate (Y) of Shandong Province, published by the Shandong Provincial government (Figure 13): $Y = 2.33 \times 10^2 - 0.14 \times X$ where $R^2 = 0.773$ and Pearson correlation value is -0.879, which also shows a strong correlation. The results show that the decrease in agricultural land in Shandong Province is closely related to the increase in labor construction and urbanization rate each year.
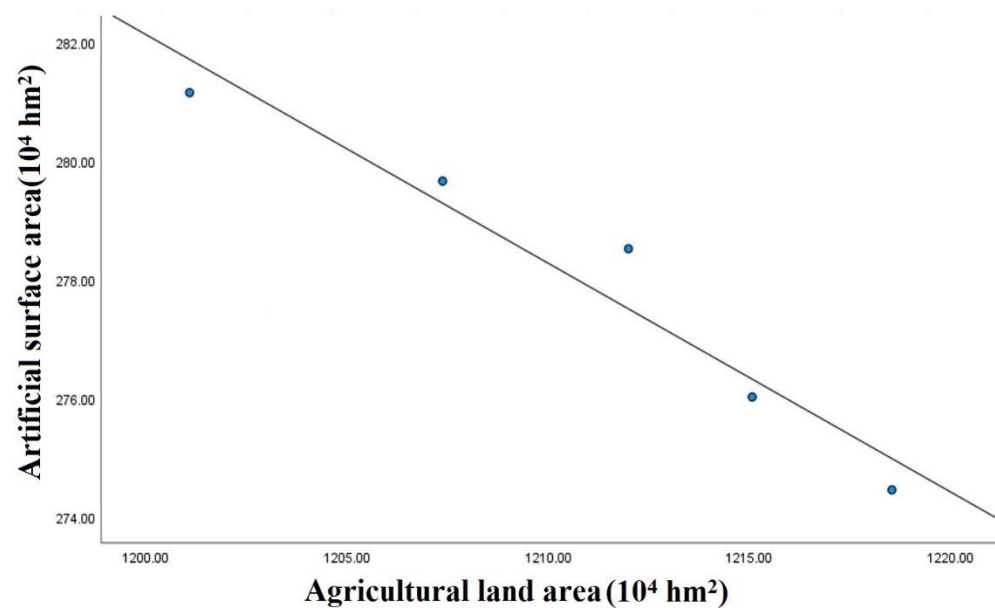
Figure 12. Correlation analysis of agricultural land area and artificial land surface area.
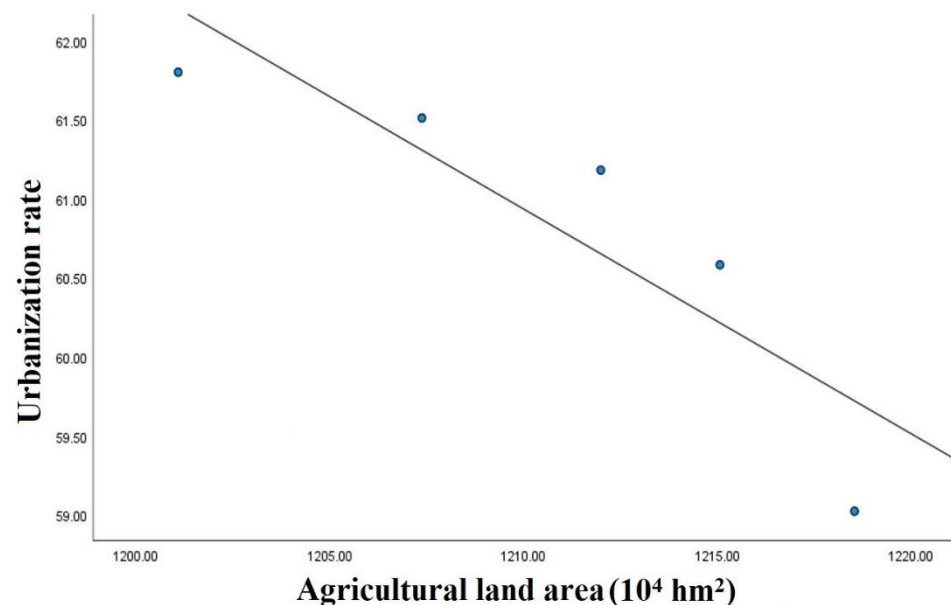

Figure 13. Correlation analysis of agricultural land area and urbanization rate.

6. Conclusions

Using the Google Earth Engine platform, this study quickly obtained remote sensing images of Shandong Province from 2016 to 2020; completed preprocessing such as clipping, mosaicking and cloud removal; and extracted a spectral index to judge the farmland phenology that was suitable for classification. The random forest algorithm was used for feature optimization. The random forest, gradient lifting tree, decision tree and ensemble learning classifiers were used to classify Landsat images of Shandong Province from 2016 to 2020. The spatial location distribution information of cultivated land in Shandong Province was obtained for the last five years, and the annual agricultural land-use area was calculated. The main conclusions are as follows:

1. The GEE platform has unique advantages for processing large-scale data. Through the calculation of the characteristic index, the phenology of the classified ground objects can be understood more clearly, which is convenient for later classification.

On the basis of the spectral features, texture features and terrain features are added, and the random forest algorithm can be used for feature optimization to compress the classification features and retain the most favorable features for classification, which reduces data redundancy and improves classification accuracy.

2. Ensemble learning based on the classification results of a single classifier can greatly improve classification accuracy. Ensemble learning is used to integrate the classification results of three classifiers, and the overall accuracy of the classification results is above 0.9.

3. The analysis of the main driving forces of agricultural land changes in Shandong Province shows that there is a strong correlation between the decrease in agricultural land area and the increase in artificial land surface and the urbanization rate in Shandong Province in the last five years.

In this study, classification based only on Landsat images was shown to have certain limitations. The existing research results show that the fusion of Sentinel-2 images with higher resolution can achieve good classification results. A combination of different classifiers can also improve classification accuracy. Subsequent studies will try to combine high-resolution images or multi-source image data and use different classifiers to combine classification on the GEE platform in order to further improve the accuracy of agricultural land extraction.

Author Contributions: H.L. performed experiments, analyzed the data and prepared the manuscript. M.C. provided crucial guidance and support through the research. Y.L., H.C., C.W. and P.G. significantly contributed to the validation work and data interpretation. C.X. and B.T. provided valuable suggestions for this study. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data supporting the findings of this study are available from the first author (H.L.) upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kumar, L.; Mutanga, O. Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. *Remote Sens.* **2018**, *10*, 1509. [CrossRef]
2. Dong, J.; Wu, W.; Huang, J.; You, N.; He, Y.; Yan, H. Research progress and prospect of remote sensing information extraction for agricultural land use. *J. Geo-Inf. Sci.* **2020**, *22*, 772–783.
3. Wang, L.; Diao, C.; Xian, G.; Yin, D.; Lu, Y.; Zou, S.; Tyler, A. Erickson. A summary of the special issue on remote sensing of land change science with Google earth engine. *Remote Sens. Environ.* **2020**, *248*, 112002. [CrossRef]
4. Buchner, J.; Yin, H.; Frantz, D.; Kuemmerle, T.; Askerov, E.; Bakuradze, T.; Bleyhl, B.; Elizbarashvili, N.; Komarova, A.; Lewińska, K.E.; et al. Land-cover change in the Caucasus Mountains since 1987 based on the topographic correction of multi-temporal Landsat composites. *Remote Sens. Environ.* **2020**, *248*, 111967. [CrossRef]
5. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine. *Remote Sens.* **2017**, *9*, 1065. [CrossRef]
6. Phan, T.N.; Kuch, V.; Lehnert, L.W. Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *Remote Sens.* **2020**, *12*, 2411. [CrossRef]
7. Magidi, J.; Nhamo, L.; Mpandeli, S.; Mabhaudhi, T. Application of the Random Forest Classifier to Map Irrigated Areas Using Google Earth Engine. *Remote Sens.* **2021**, *13*, 876. [CrossRef]
8. McCarty, D.A.; Kim, H.W.; Lee, H.K. Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification. *Environments* **2020**, *7*, 84. [CrossRef]
9. Tassi, A.; Gigante, D.; Modica, G.; Di Martino, L.; Vizzari, M. Pixel- vs. Object-Based Landsat 8 Data Classification in Google Earth Engine Using Random Forest: The Case Study of Maiella National Park. *Remote Sens.* **2021**, *13*, 2299. [CrossRef]
10. Nasiri, V.; Deljouei, A.; Moradi, F.; Sadeghi, S.M.M.; Borz, S.A. Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods. *Remote Sens.* **2022**, *14*, 1977. [CrossRef]
11. Sefrin, O.; Riese, F.M.; Keller, S. Deep Learning for Land Cover Change Detection. *Remote Sens.* **2020**, *13*, 78. [CrossRef]

12. Yu, J.; Zeng, P.; Yu, Y.; Yu, H.; Huang, L.; Zhou, D. A Combined Convolutional Neural Network for Urban Land-Use Classification with GIS Data. *Remote Sens.* **2022**, *14*, 1128. [CrossRef]

13. Cipta Ramadhan Kete, S.; Darma Tarigan, S.; Effendi, H. Land use classification based on object and pixel using Landsat 8 OLI in Kendari City, Southeast Sulawesi Province, Indonesia. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *284*, 012019. [CrossRef]

14. Simón Sánchez, A.-M.; González-Piqueras, J.; de la Ossa, L.; Calera, A. Convolutional Neural Networks for Agricultural Land Use Classification from Sentinel-2 Image Time Series. *Remote Sens.* **2022**, *14*, 5373. [CrossRef]

15. Cherif, E.; Hell, M.; Brandmeier, M. DeepForest: Novel Deep Learning Models for Land Use and Land Cover Classification Using Multi-Temporal and -Modal Sentinel Data of the Amazon Basin. *Remote Sens.* **2022**, *14*, 5000. [CrossRef]

16. Zhang, A.; Jiang, W.; Zhang, Y.; Qian, Z. Remote sensing image change detection based on adaptive interval type 2 fuzzy clustering. *J. Surv. Mapp. Sci. Technol.* **2018**, *35*, 376–382.

17. Matarira, D.; Mutanga, O.; Naidu, M. Google Earth Engine for Informal Settlement Mapping: A Random Forest Classification Using Spectral and Textural Information. *Remote Sens.* **2022**, *14*, 5130. [CrossRef]

18. Saboori, M.; Homayouni, S.; Shah-Hosseini, R.; Zhang, Y. Optimum Feature and Classifier Selection for Accurate Urban Land Use/Cover Mapping from Very High Resolution Satellite Imagery. *Remote Sens.* **2022**, *14*, 2097. [CrossRef]

19. Lu, R.; Wang, N.; Zhang, Y.; Lin, Y.; Wu, W.; Shi, Z. Extraction of Agricultural Fields via DASFNet with Dual Attention Mechanism and Multi-scale Feature Fusion in South Xinjiang, China. *Remote Sens.* **2022**, *14*, 2253. [CrossRef]

20. Zhaoxin, H.; Zhang, M.; Bing-Fang, W.; Qiang, X. Remote sensing Extraction of summer harvest Crops in Jiangsu Province based on Google Earth Engine. *J. Geo-Inf. Sci.* **2019**, *21*, 752–766.

21. Li, X.; Zhang, D.; Jiang, C.; Zhao, Y.; Li, H.; Lu, D.; Qin, K.; Chen, D.; Liu, Y.; Sun, Y.; et al. Comparison of Lake Area Extraction Algorithms in Qinghai Tibet Plateau Leveraging Google Earth Engine and Landsat-9 Data. *Remote Sens.* **2022**, *14*, 4612. [CrossRef]

22. Feng, S.; Li, W.; Xu, J.; Liang, T.; Ma, X.; Wang, W.; Yu, H. Land Use/Land Cover Mapping Based on GEE for the Monitoring of Changes in Ecosystem Types in the Upper Yellow River Basin over the Tibetan Plateau. *Remote Sens.* **2022**, *14*, 5361. [CrossRef]

23. Dubertret, F.; Le Tourneau, F.-M.; Villarreal, M.L.; Norman, L.M. Monitoring Annual Land Use/Land Cover Change in the Tucson Metropolitan Area with Google Earth Engine (1986–2020). *Remote Sens.* **2022**, *14*, 2127. [CrossRef]

24. Lin, G.; Qinghai, D.; Xuexin, G.; Hu, S.; Li, H. Study on dynamic change of urban land use in Qingdao based on remote sensing. *Shandong Land Resour.* **2017**, *33*, 86–91.

25. Huang, B.H. Spatial-temporal change and driving forces of land use in Shandong Province based on remote sensing. *Chin. J. Agric. Sci.* **2021**, *11*, 62–67.

26. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]

27. Rousel, J.; Haas, R.; Schell, J.; Deering, D. *Monitoring Vegetation Systems in the Great Plains with Erts*; NASA Special Publications; NASA: Washington, DC, USA, 1974; p. 351.

28. Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frolking, S.; Li, C.; Salas, W.; Moore, B., III. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sens. Environ.* **2004**, *95*, 480–492. [CrossRef]

29. Zha, Y.; Ni, S.X.; Yang, S. An effective method to automatically extract urban land Use information from TM images. *J. Remote Sens.* **2003**, *1*, 37–40+82.

30. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

31. Mohammadpour, P.; Viegas, D.X.; Viegas, C. Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture Feature—A Case Study for Lousã Region, Portugal. *Remote Sens.* **2022**, *14*, 4585. [CrossRef]

32. Cheng, Q.; Xu, H.; Fei, S.; Li, Z.; Chen, Z. Estimation model of summer Maize Cover based on Stacking Ensemble Learning. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 195–202.

33. Huang, Z.; Qi, H.; Kang, C.; Su, Y.; Liu, Y. A rapid extraction method of urban new construction land based on ensemble learning. *Environ. Monit. Early Warn.* **2019**, *11*, 39–45.

34. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

35. Freund, Y.; Schapire, R.E. Experiments with a new Boosting algorithm. *Mach. Learn. Proc. Int. Conf.* **1996**, *96*, 148–156.

36. Si, R. Study on the relationship between urbanization level and cultivated land area change: A case study of Shandong Province. *Hubei Agric. Sci.* **2019**, *58*, 146–149.

37. Zhao, Z.; Peng, P.; Zhang, F.; Wang, J.; Li, H. The Impact of the Urbanization Process on Agricultural Technical Efficiency in Northeast China. *Sustainability* **2022**, *14*, 12144. [CrossRef]