Crop Classification Method Based on Optimal Feature Selection and Hybrid CNN-RF Networks for Multi-Temporal Remote Sensing Imagery

Shuting Yang 1, Lingjia Gu 1,*, Xiaofeng Li 2, Tao Jiang 2 and Ruizhi Ren 1

1 College of Electronic Science & Engineering, Jilin University, Changchun 130012, China; yst18@mails.jlu.edu.cn (S.Y.); rrz@jlu.edu.cn (R.R.)
2 Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China; lixiaofeng@iga.ac.cn (X.L.); jiangtao@iga.ac.cn (T.J.)

* Correspondence: gulingjia@jlu.edu.cn

Received: 26 August 2020; Accepted: 21 September 2020; Published: 23 September 2020

Abstract: Although efforts and progress have been made in crop classification using optical remote sensing images, it is still necessary to make full use of the high spatial, temporal, and spectral resolutions of remote sensing images. However, with the increasing volume of remote sensing data, a key emerging issue in the field of crop classification is how to find useful information from massive data to balance classification accuracy and processing time. To address this challenge, we developed a novel crop classification method, combining optimal feature selection (OFSM) with hybrid convolutional neural network-random forest (CNN-RF) networks for multi-temporal optical remote sensing images. This research used 234 features including spectral, segmentation, color, and texture features from three scenes of Sentinel-2 images to identify crop types in the Jilin province of northeast China. To effectively extract the effective features of remote sensing data with lower time requirements, the use of OFSM was proposed with the results compared with two traditional feature selection methods (TFSM): random forest feature importance selection (RF-FI) and random forest recursive feature elimination (RF-RFE). Although the time required for OFSM was 26.05 s, which was between RF-FI with 1.97 s and RF-RFE with 132.54 s, OFSM outperformed RF-FI and RF-RFE in terms of the overall accuracy (OA) of crop classification by 4% and 0.3%, respectively. On the basis of obtaining effective feature information, to further improve the accuracy of crop classification we designed two hybrid CNN-RF networks to leverage the advantages of one-dimensional convolution (Conv1D) and Visual Geometry Group (VGG) with random forest (RF), respectively. Based on the selected optimal features using OFSM, four networks were tested for comparison: Conv1D-RF, VGG-RF, Conv1D, and VGG. Conv1D-RF achieved the highest OA at 94.27% as compared with VGG-RF (93.23%), Conv1D (92.59%), and VGG (91.89%), indicating that the Conv1D-RF method with optimal feature input provides an effective and efficient method of time series representation for multi-temporal crop-type classification.

Keywords: feature selection; convolutional neural network; crop classification; multi-temporal remote sensing images; data fusion

1. Introduction

In recent years, with the increase in satellites at different spatial, temporal, radiometric, and spectral resolutions, remote sensing techniques have emerged as optimal tools to identify crop types over larger areas. Timely and accurate crop-type classification is essential for estimating crop yields, strengthening crop production management, and crop insurance [1]. Currently, crop-type classification methods...
in remote sensing have been developed based on the following two strategies [2]. The first is to use the original spectral information of a crop or to aggregate spectral bands into vegetation indices that represent the physical characteristics of vegetation [3], such as the normalized difference vegetation index (NDVI) or enhanced vegetation index (EVI) [4]. This approach involves classifying different types of land cover with distinctive spectral characteristics of high-resolution remote sensing images. However, in the peak-growing period of some crops, the spectral characteristics of different crops may be similar, which makes it difficult to classify crops accurately. The second strategy is to use both the spectral and temporal information of crops during the growing seasons [5]. This approach extracts features from time-series data to retrieve useful information, utilizing the wealth of seasonal patterns and sequential relationships.

When obtaining high volumes of spectral and temporal information data, considering information redundancy and the processing time required, feature selection preprocessing is used to improve the speed and accuracy of crop classification. As a consequence, there has been increasing interest in feature selection methods such as random forest (RF) [6], extreme gradient boosting (XGBoost) [7], and recursive feature elimination (RFE) [8], which extract features to obtain useful information for classification. Feature selection is generally considered from two aspects. One is whether the feature diverges or converges, and the other is the correlation between the feature and the target. For example, Hao et al. [9] used the Random Forest (RF) algorithm to calculate the importance scores for all features, and classified crop types with an 88.81% accuracy. Yin et al. [10] sorted all features by the global separability index of each crop, and eliminated redundant features according to accuracy changes when adding new features. Liu et al. [11] used the maximum likelihood classification–recursive feature elimination (MLC-RFE) method and the support vector machine-recursive feature elimination (SVM-RFE) method to select the optimal bands, and further analyzed the classification results of different methods. However, these introduced methods still have problems in the process of feature selection, such as computational cost and high correlation between features.

Machine-learning methods have been applied to various agricultural applications, such as crop classification and regression. Machine-learning methods, such as SVM [12], K-nearest neighbors [13], the maximum likelihood method [14], extreme gradient boosting (XGBoost) [15], decision trees [16], RF [17], and artificial neural networks [18], belong to a traditional domain of classification techniques that can effectively distinguish the vegetation type in remote sensing images. For example, Li et al. [19] achieved a 90.50% crop classification accuracy with a random forest algorithm based on full-year fully-polarpolarimetric L-band Unmanned Aerial Vehicle Synthetic Aperture Radar (UAVSAR) data. Zhang et al. [20] achieved a 90.7% accuracy with vegetation-type classification in the Dzungarian Basin by applying an extreme gradient boosting (XGBoost) classifier to vegetation data. In recent years, deep-learning methods have been widely used as promising classification techniques because of their high learning efficiency [21]. Castelluccio et al. [22] used GoogLeNet and CaffeNet to classify crop types with accuracies of 91.83% and 90.94%, respectively. Xu et al. [23] utilized convolutional neural networks (CNN) to classify multi-source remote sensing images and obtained land-cover classification with a 97.92% accuracy. However, these networks were not designed to process sequential data or represent temporal features. That is, they ignored temporal dependency. Recurrent neural networks and long short-term memory (LSTM) [24] were developed for sequential data analysis. Rußwurm et al. [25] used LSTM to extract dynamic temporal features from sequential images to classify crop types with a 74.3% accuracy. Additionally, one-dimensional convolution (Conv1D) has great potential in temporal feature representation [26]. Guidici et al. [27] applied a multi-temporal Conv1D model to classify land cover in hyperspectral images with an accuracy of 89.9%, slightly higher than that of SVM (89.5%) and much higher than that of RF (82.2%).

There are primarily two different classification strategies: single classifier (SC) and multiple classifiers (MC). In traditional pattern recognition, a single classifier [28,29] is commonly used to determine which category a given pattern belongs to. However, in many cases, the classification accuracy of crops can be improved through an ensemble of classifiers or multiple classifiers [30,31], which allows
individual classifiers to support each other for decision-making. Although previous studies have obtained satisfactory results using single classifier approaches, the complementary effect of multiple classifiers should theoretically obtain better classification results than a single classifier. Data fusion breaks through the constraints of using a single classification method and effectively leverages the advantages of multiple classifiers’ complementary classifications, thus providing opportunities to achieve a more accurate and comprehensive crop-type classification. The earliest integrated algorithm to combine the classification advantages of multiple classifiers was the “minority obeying majority” approach, which has certain limitations [32]. Other methods mainly present a hybrid model that leverages the synergy of CNN and SVM, with the CNN classifier used for feature extraction and the SVM classifier used for classification [33]. Furthermore, the weights are assigned to multiple classifiers to make a final decision. These fusion methods can achieve more accurate classification results compared with a single classifier.

This study aims to develop a novel crop classification method based on optimal feature selection (OFSM) and hybrid CNN-RF networks for multi-temporal remote sensing images. We first propose OFSM, considering the efficiency of information extraction and processing, noting that the results of feature selection significantly influence classification accuracy. In addition, to make full use of the complementary advantages of multiple classifiers, hybrid classification models are designed. We compare the classification results of the proposed models with those from some leading deep-learning classifiers.

The major contributions of this study include:

1. One of the main innovations of this paper is OFSM, which is different from traditional feature selection methods, including filter, embedded, wrapper, and hybrid. The filter selection method selects features regardless of the model used and is commonly robust in overfitting and effective in computation time. The wrapper method performs evaluation on multiple subsets of the features and chooses the best subset of features that gives the highest accuracy to the model. Since the classifier needs to be trained multiple times, the computation time using the wrapper method (e.g., RFE) is usually much larger than that using the filter method. The embedded method (e.g., RF and XGBoost) can interact with the classifier and is less computationally intensive than the wrapper method, but it ignores the correlation between multiple features. OFSM is a hybrid method of filter, embedded, and wrapper, and has advantages in processing time and recognition accuracy. Considering the correlation between the multi-features and the processing time during the feature selection process, the features selected by OFSM are independent of each other and the time required for processing is acceptable. The experimental results demonstrate that OFSM performs optimally and the accuracy of the selected features for crop classification is higher than that of the original image directly sent to the classifier. Thus, we show that the preprocessing of feature selection is critical prior to classification.

2. Considering the advantages of multiple classifiers, we propose two hybrid CNN-RF networks to integrate the advantages of Conv1D and Visual Geometry Group (VGG) with RF, respectively. A traditional CNN uses an FC layer to make the final classification decision, and there is usually overfitting, especially with inadequate samples, which is not sufficiently robust and is computationally intensive. The use of RF instead of the FC layer to make the final decision can effectively alleviate the occurrence of overfitting. At the same time, we are committed to providing a reasonable scheme for the selection of a CNN network structure in crop mapping based on multi-temporal remote sensing images, and selecting the optimal hyperparameters for the CNN network can further improve the identification accuracy of crops. The results demonstrate that the proposed hybrid networks can integrate the advantages of the two classifiers and achieve more optimal crop classification results than the original deep-learning networks. In particular, the combination of temporal feature representation network (Conv1D) and RF achieves the optimal crop classification results. Compared with the mainstream networks (e.g., LSTM-RF, ResNet, and U-Net), the proposed Conv1D-RF still obtains better crop recognition
results, indicating that the Conv1D-RF framework can mine more effective and efficient time series representations and achieve more accurate identification results for crops in multi-temporal classification tasks.

The remainder of this paper is organized as follows. Section 2 introduces the study area and data used in this work. Section 3 details the specific workflow of research, including OFSM and traditional feature selection methods (TFSM), classification methods based on hybrid CNN-RF networks and original deep-learning networks, and evaluation. Section 4 compares various classification results using the proposed method and other traditional methods. Discussions and conclusions are presented in Sections 5 and 6, respectively.

2. Data Resources

2.1. Study Area

The study area was in the Jilin province of northeast China (Figure 1), and is a major area for agricultural production. The climate is characterized by a temperate continental semi-humid monsoon pattern, is warm and rainy in the summer and cold and humid in the winter, and has an annual mean temperature of 7 °C. The field investigation was conducted from June to September in 2017, which is the main growing season for crops. In the study area, there were 83 experimental measurements of four land cover types, including rice, urban, corn, and soybean, as shown in Table 1.

Table 1. The number of experimental measurements for each land cover type.

| ID | Type | 1 | 2 | 3 | 4 | Total |
|----|------|---|---|---|---|-------|
|    | Rice | 23| 17| 26| 17| 83    |
|    | Urban|    |   |   |   |       |
|    | Corn |    |   |   |   |       |
|    | Soybean | |   |   |   |       |

2.2. Data

Sentinel-2 imagery (L2C-level), was selected and processed by radiometric calibration and atmospheric correction. Sentinel-2 delivers high-resolution optical images for land monitoring, emergency response, and security services. The imagery provides a versatile set of 13 spectral bands spanning from visible, red-edge, and near-infrared (NIR) to shortwave infrared (SWIR), featuring four spectral bands (B2, B3, B4, B8) at a 10 m spatial resolution, six bands (B5, B6, B7, B8A, B11, B12) at a 20 m spatial resolution, and three bands (B1, B9, B10) at a 60 m spatial resolution, as listed in Table 2. In this study, three scenes of Sentinel-2 imagery (28 June 2017, 18 July 2017, 11 September 2017) without cloud cover were collected in the crop growing season. Excluding the three “atmospheric” bands with
a spatial resolution of 60 m, the other 10 bands of each scene were selected and resampled to a 10 m spatial resolution [34,35].

| Band Name                      | Central Wavelength (um) | Resolution (m) |
|--------------------------------|-------------------------|----------------|
| B1-Coastal aerosol            | 0.443                   | 60             |
| B2-Blue                       | 0.49                    | 10             |
| B3-Green                      | 0.56                    | 10             |
| B4-Red                        | 0.665                   | 10             |
| B5-Vegetation Red Edge        | 0.705                   | 20             |
| B6-Vegetation Red Edge        | 0.74                    | 20             |
| B7-Vegetation Red Edge        | 0.783                   | 20             |
| B8-NIR                        | 0.842                   | 10             |
| B8A-Vegetation Red Edge       | 0.865                   | 20             |
| B9-Water vapor                | 0.945                   | 60             |
| B10-SWIR-Cirrus               | 1.375                   | 60             |
| B11-SWIR                      | 1.61                    | 20             |
| B12-SWIR                      | 2.19                    | 20             |

In Figure 2, by combining the field investigation samples with higher-resolution optical remote sensing images, four labels were created for the corresponding land cover types to form the reference dataset [36] for training, validation, and testing. Applying the training dataset, the individual classification model was trained by setting the parameters of the classifier. The validation dataset was used to select the optimal parameters for the model. The test dataset was used to evaluate the performance of the final classification. Here, the training, validation, and test datasets were independent of each other and randomly assigned by a ratio of 25%:25%:50%. There were 45,136 pixels in the training dataset, 44,312 in the validation dataset, and 87,898 in the test dataset.

3. Methodology

Figure 3 presents a flowchart outlining the method used in this study. First, five types of features from the spatial and spectral information of multi-temporal remote sensing images are extracted as introduced in Section 3.1. Then, TFSM and the proposed OFSM are described in Section 3.2. In Section 3.3, two hybrid CNN-RF networks (Conv1D-RF and VGG-RF) are designed in detail. Finally, six parameters are introduced to evaluate the performance of the crop classification methods.
3.1. Feature Extraction

Feature extraction transforms the original features into a group of features with obvious physical or statistical significance. In this study, the raw spectral features, color features, segmentation features, spectral index features, and texture features are extracted. These features can effectively cooperate with the spectral and spatial information of land cover, and their combination can greatly improve the crop recognition ability and accuracy in remote sensing images. As shown in Figure 4, a total of 234 features were extracted from three scenes of Sentinel-2 images employed to identify crop types. Among them, we selected 30 raw spectral features, 30 segmentation features using a graph-based segmentation algorithm [37], 9 color features extracted from HSI color space [38], 45 spectral index features [39] listed in Table 3, and 120 texture features [40] listed in Table 4 using three scenes of Sentinel-2 imagery. In addition, “seg” represents segmentation feature, “H” represents hue, “S” represents saturation, “I” represents intensity, “CON” represents contrast, “ENT” represents entropy, “ASM” represents angular second moment, and “HOM” represents homogeneity in this study.

![Figure 3. The flowchart of the research methodology.](image_url)

![Figure 4. Details of feature extraction.](image_url)

### Table 3. Spectral indices

| Spectral Index Calculation Formula |
|------------------------------------|
| GNDVI = (NIR - Red) / (NIR + Red) |
| GBC = Green Band Contrast |
| NDVI = (NIR - Red) / (NIR + Red) |
| NBR = (NIR - Red) / (NIR + Red) |
| NDWI = (NBR - Red) / (NBR + Red) |
| EVI = 2.5 * ((NIR - Red) / (NIR + Red)) |
| EVI2 = (2.5 * (NIR - Red)) / (NIR + Red) |
| EVI3 = (2.5 * (NIR - Red)) / (NIR + Red) |
| EVI4 = (2.5 * (NIR - Red)) / (NIR + Red) |

*Note: NIR and Red represent near-infrared and red bands, respectively.*
Table 3. Spectral indices.

| Spectral Index | Calculation Formula |
|----------------|---------------------|
| NDVI           | \( \frac{(B8 - B4)}{(B8 + B4)} \) |
| DVI            | \( B8 - B4 \) |
| RDVI           | \( \frac{(B8 - B4)}{\sqrt{B8 + B4}} \) |
| NDI            | \( \frac{(B3 - B8)}{(B3 + B8)} \) |
| RVI            | \( \frac{B8}{B4} \) |
| EVI            | \( 2.5 \times \frac{B8 - B4}{B8 + B4} \) |
| TVI            | \( 0.5 \times [120 \times (B8 - B3) - 200 \times (B4 - B3)] \) |
| TCARI          | \( 3 \times [(B8 - B4) - 0.2 \times (B8 - B3) \times (B8/B4)] \) |
| G1             | \( B3/B4 \) |
| VIgreen        | \( \frac{(B3 - B4)}{(B3 + B4)} \) |
| VARGreen       | \( \frac{(B3 - B4)}{(B3 + B4 - B2)} \) |
| GARI           | \( \frac{B8 - (B3 + B4)}{B8 - (B3 + B4 - B2)} \) |
| GDVI           | \( B8 - B3 \) |
| SAVI           | \( 1.5 \times \frac{(B8 - B4)}{B8 + B4 - 0.5} \) |
| SIFI           | \( \frac{(B8 - B2)}{(B8 - B4)} \) |

Table 4. Texture features.

| Texture Features | Statistical Characteristics |
|------------------|-----------------------------|
| Homogeneity: HOM | \( \sum_{i} \sum_{j} \frac{f(i,j)}{1+(i-j)^2} \) | Measure local homogeneity |
| Contrast: CON    | \( \sum_{i} \sum_{j} (i-j)^2f(i,j) \) | Measure the difference between the maximum and minimum values in the neighborhood |
| Entropy: ENT     | \( -\sum_{i} \sum_{j} f(i,j) \log[f(i,j)] \) | Measuring image disorder |
| Angular Second Moment: ASM | \( \sum_{i} \sum_{j} (f(i,j))^2 \) | Describe local stationarity |

3.2. Feature Selection

The basic types of approaches exploited in feature selection and reduction include filter, wrapper, embedded, and hybrid, respectively. First, two traditional feature selection methods (TFSM) are introduced, including RF-FI and RF-RFE, where RF-FI is an embedded method and RF-RFE is a hybrid method of embedded and wrapper. On the basis of TFSM, OFSM is proposed as a hybrid method of filter, embedded, and wrapper.

3.2.1. Traditional Feature Selection Methods (TFSM)

(A) Random Forest Feature Importance Selection (RF-FI):

In this method [41], the features are firstly sorted according to their importance score, then the unimportant features are eliminated. Here, we use the prediction performance of RF to realize the quantification of the feature importance, including the out of bag (OOB) error and the quantification of feature importance (FI), which are the key elements of the feature selection strategy. For example, these features are ranked by sorting the FI in descending order, and the less important features are eliminated under the given threshold. Here, \( M \) denotes the number of retained features with an averaged FI exceeding this threshold.

(B) Random Forest Recursive Feature Elimination (RF-RFE):

The RF-RFE selection method [42] is basically a recursive process that ranks features according to the measure of feature importance (FI) given by RF. At each iteration, the less important feature is eliminated according to the measurement result of feature importance (FI). The recursive is necessary during the stepwise elimination process because the relative importance of each feature could change substantially when evaluating over a different subset of features. The final ranking is constructed by
the inverse order of feature elimination. The feature selection process itself only includes the first $M$ features from this ranking.

3.2.2. Optimal Feature Selection Method (OFSM)

In order to gain the advantages provided by the different feature selection methods, this study developed a hybrid method to increase the efficiency and provide a higher accuracy. The implementation steps of OFSM are as follows and the structure of OFSM is shown in Figure 5.

Step 1: Calculate the Spearman rank correlation coefficient \cite{43,44} between the input features and the labels. The Spearman rank correlation coefficient tests the direction (negative or positive) and strength of the relationship between two variables. First, the measurements of feature and labels are assigned corresponding ranks according to their average descending position in the total measurements. Then, the Spearman correlation coefficient $\rho_{sFL}$ between the $F$th feature data ($F$) and labels ($L$) is calculated by:

$$\rho_{sFL} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2\right]^{1/2}}, \quad (1)$$

where $n$ is the number of measurements in each of the two variables (the feature and label) in the correlation, $x_i$ and $y_i$ represent the rank of the $i$th measurements for the two variables, and $\overline{x}$ and $\overline{y}$ represent the average rank of the two variables, respectively. $\rho_{sFL}$ is between 1.0 (a perfect positive correlation) and $-1.0$ (a perfect negative correlation). The larger the Spearman correlation coefficient, the stronger the monotonicity between the two variables. Then, we rank the features by sorting the $\rho_{sFL}$ in descending order, and eliminate the less relevant features whose $\rho_{sFL}$ value is smaller than a given threshold $T_1$. By setting the threshold $T_1$, the features with a strong monotonicity with the labels are retained. Here, $M_1$ denotes the number of retained features after Step 1.

Step 2: The $M_1$ features are ranked by sorting the $\rho_{sFL}$ in descending order, and we further calculate the rank correlation coefficient between these features. The measurements of the two features are firstly assigned corresponding ranks based on their average descending position in the total measurements. Then, the Spearman correlation coefficient $\rho_{sFF'}$ between any two features is calculated by:

$$\rho_{sFF'} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(x_i' - \overline{x}')}{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (x_i' - \overline{x}')^2\right]^{1/2}}, \quad (2)$$

where $n$ is the number of correlation measurements of the two selected features ($F$ and $F'$), $x_i$ and $x_i'$ represent the rank of the $i$th measurements for the two selected features, and $\overline{x}$ and $\overline{x}'$ represent the average rank of the two selected features, respectively. The nested loop of $\rho_{sFF'}$ between the baseline features ($F$) and other features ($F'$) is constructed to eliminate the strongly relevant features satisfying the condition that the $\rho_{sFF'}$ value is higher than a given threshold $T_2$. Here, $M_2$ denotes the number of retained features after Step 2.

Step 3: Given the number of final retained features $M$, we construct the nested collection of RF models involving the $K$ features for $K = M_2$ down to $M$, and eliminate the feature involved in the model leading to the smallest feature importance (FI) during each iteration. Finally, the remaining $M$ features compose the optimal feature combination.

3.3. Deep-Learning Classification

A traditional CNN usually uses an FC layer as the final decision. This section introduces two hybrid CNN-RF networks. The designed networks use original deep-learning networks to extract high-dimensional features and combine the advantages of RF instead of the FC layer to make the final classification decision.
3.3.1. Visual Geometry Group Combined with Random Forest (VGG-RF)

Figure 6 shows the architecture of the VGG-RF and VGG [45], including the convolutional layers, pooling layers, fully connected layers, and dropout in detail. This paper studies a deep-learning method of crop classification based on pixels, which are limited by the number of bands, so the width of the convolutional filter was set to 2. Continuous 2 × 2 convolution kernels were selected to replace the larger convolution kernel to ensure network depth improvement under the same perception field. In the network structure of the VGG combined with random forest (VGG-RF), we tested the hyperparameters and selected the optimal hyperparameter values for training the network. The channel numbers of the first convolution layer were measured as 32, 64, and 128, respectively. The optimal value is 64, which is the same as that in [46]. During the training process, the pooling layers were fixed to “max-pooling” with a window size of 2 × 2. Dropout is a regularization technique that randomly drops some neurons. The proportion of dropped neurons was set to 50%. VGG contains three fully connected layers at the output end, and the last fully connected layer contains four neurons, corresponding to the probability of the four classes: rice, urban, corn, and soybean. The 1024 × 1 feature vector of the Fc8 layer output was extracted and put into random forest (RF) for classification. As a hybrid CNN-RF network, the designed VGG-RF used the high-dimensional features extracted by VGG and combined the advantages of RF to replace the fully connected (FC) layer to make the final decision.

3.3.2. One-Dimensional Convolution Combined with Random Forest (Conv1D-RF)

The hybrid Conv1D-RF network uses the high-dimensional features extracted by Conv1D and leverages the advantages of RF to replace the fully connected (FC) layer to make the final decision. Similar to the hybrid CNN-SVM [33] network, the CNN is used for feature extraction and the SVM classifier is used for classification. The proposed Conv1D-RF network uses one-dimensional convolution (Conv1D), which has great potential in temporal feature representation. Thus, we aim to combine the high-dimensional features extracted by the FC1 layer of Conv1D with RF. A traditional CNN uses an FC layer to make the final classification decision, and there is usually overfitting, especially
with inadequate samples, which is not sufficiently robust and is computationally intensive. The use of the RF instead of the FC layer to make the final decision can effectively alleviate the occurrence of overfitting [47]. Due to the low requirement of the number of samples, the RF classifier can still obtain satisfactory decision results when the sample size is insufficient. Similarly, other studies have attempted to replace the FC layer with other structures, such as the convolutional layer used to replace the FC layer in FCN [48]. As shown in Figure 7, Conv1D is a special form of CNN that is implemented by pooling layers, fully connected layers, and dropout. The convolutional filter width was set to 3. The number of channels in the first convolution layer was set to 64 and the channel number increased with depth. The proportion of dropped neurons was set to 40% and 50%. We designed and tested the inception module to concentrate convolutional and pooling layers with different sizes. The input of this module had three branches, including two convolutional layers with filter widths of 3 and 5, respectively, and one max-pooling layer with a filter width of 2. Convolution was simultaneously performed on multiple scales, and features at different scales were extracted. The obtained multi-scale features were concentrated to make the subsequent classification decisions more accurate. Conv1D contains two fully connected layers at the output end, and the last fully connected layer contains four neurons, corresponding to the probability of the four classes. The 512 × 1 feature vector of the first fully connected layer output was extracted and put into random forest (RF) for classification. As a hybrid CNN-RF network, the designed Conv1D-RF used the high-dimensional features extracted by Conv1D and combined the advantages of RF to replace the FC layer to make the final decision.

![Figure 7. Architecture of the one-dimensional convolution combined with random forest (Conv1D-RF) and Conv1D model.](image)

### 3.4. Evaluation

The crop classification accuracy of each network was evaluated on the test dataset. We applied six parameters of pixel-based evaluation, overall accuracy (OA), K coefficient, precision, recall, F1 score, and intersection-over-union (IoU) to assess the crop classification accuracy.

OA is expressed as:

\[
OA = \frac{\sum_{i=1}^{n} p_{i,j}}{\sum_{j=1}^{n} \sum_{i=1}^{n} p_{i,j}}, \tag{3}
\]

where \( p_{i,j} \) represents the total number of pixels that belong to class \( i \) and are assigned to class \( j \) and \( n \) represents the number of categories.

The K coefficient is:

\[
K = \frac{N^2 \times OA - \sum_{i=1}^{n} a_{i} b_{i}}{N^2 - \sum_{i=1}^{n} a_{i} b_{i}}, \tag{4}
\]

where \( N \) denotes the total number of samples, \( a_{1}, a_{2}, \ldots, a_{n} \) are the numbers of real samples in each type, and \( b_{1}, b_{2}, \ldots, b_{n} \) are the numbers of samples predicted for each type.
By comparison with ground truth (GT), true positive (TP), false positive (FP), and false negative (FN) represent the number of correctly extracted classes, incorrectly extracted classes, and missing classes, respectively. Using these counts, precision and recall are defined as:

\[
\text{Recall} = \frac{TP}{TP + FN},
\]

\[
\text{Precision} = \frac{TP}{TP + FP}.
\]

F1 score is a representation of the harmonic mean of precision and recall, and is calculated by:

\[
F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\]

IoU describes the overlap rate between the crop classification result and GT, and is calculated by:

\[
\text{IoU} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall} - \text{Precision} \times \text{Recall}}.
\]

4. Result

4.1. Feature Selection Comparison

In order to effectively compare the OFSM with TFSM, these feature selection methods used the same classifier and importance evaluation criteria. The random forest (RF) classifier was used and the parameters were obtained by grid search, including n_estimators, max_depth, min_samples_leaf, min_samples_split, and max_features.

4.1.1. Features from OFSM

The OFSM presented in Section 3.2 is used to select the optimal feature combination from the input 234 features. First, the Spearman rank correlation coefficient \(\rho_{FL}^{s}\) was used to calculate the correlation between the input features and the labels (e.g., rice, corn, soybean, urban). Then, we ranked the features by sorting the \(\rho_{FL}^{s}\) in descending order, and eliminated the unimportant features satisfying \(\rho_{FL}^{s} < 0.2\). There were 88 features retained after this step. The top 88 important correlation coefficients between the input features and labels are shown in Figure 8a. Through comparative analysis, there was a stronger correlation between the features in June and the labels, which indicated that the features in June contribute greatly to identifying different crops. Moreover, the features with a stronger correlation were concentrated in the raw spectral features and segmentation features of the red-edge bands (e.g., B7 and B8A), SWIR (B11 and B12), and NIR band (B8), indicating that these bands can provide special spectral information to improve the identification of crops. The correlation coefficient between the 88 features is shown in Figure 8b. The results demonstrated that there was still much redundant information existing in the remaining features.
results demonstrated that there was still much redundant information existing in the remaining 33 features by sorting the FI in descending order, and eliminated the feature with the smallest FI until 16 features were retained. Table 5 shows the optimal feature combination selected by OFSM. From the results of the feature combination, it can be seen that segmentation features contribute greatly to the classification of crops. OFSM also selected the saturation feature in September as an important feature to effectively identify crops. Similar to the raw spectral feature selection results using RF-RFE, the red-edge band centered at 705 nm (B5) and SWIR band centered at 2190 nm (B12) in OFSM were also found to be the most important spectral bands for identifying crops. This showed that SWIR and red-edge bands can indeed provide effective spectral information to identify crops. OFSM and RF-RFE simultaneously selected the spectral index of Green Atmospherically Resistant Vegetation Index (GARI) in June as an important feature. These results demonstrate that the combination of spatial, spectral, and color information is of great significance for the classification of crops; however, the contribution of texture information is not obvious.

We calculated the correlation \( \rho_{FF} \) between the 88 features and constructed the nested loops to continuously eliminate the strongly relevant features satisfying \( \rho_{FF} > 0.9 \). Thus, there were 33 features retained at the end, which greatly reduced the time consumption of subsequent processing. The correlation between the 33 features is shown in Figure 9. It can be seen that the redundancy between the retained 33 features was greatly reduced.

We computed the FI of the RF of the nested models starting from the 33 features, ranked the features by sorting the FI in descending order, and eliminated the feature with the smallest FI until 16 features were retained. Table 5 shows the optimal feature combination selected by OFSM. From the results of the feature combination, it can be seen that segmentation features contribute greatly to the classification of crops. OFSM also selected the saturation feature in September as an important feature to effectively identify crops. Similar to the raw spectral feature selection results using RF-RFE, the red-edge band centered at 705 nm (B5) and SWIR band centered at 2190 nm (B12) in OFSM were also found to be the most important spectral bands for identifying crops. This showed that SWIR and red-edge bands can indeed provide effective spectral information to identify crops. OFSM and RF-RFE simultaneously selected the spectral index of Green Atmospherically Resistant Vegetation Index (GARI) in June as an important feature. These results demonstrate that the combination of spatial, spectral, and color information is of great significance for the classification of crops; however, the contribution of texture information is not obvious.
Table 5. Feature selection based on OFSM.

| Raw Spectral Feature | Segmentation Feature | Spectral Index Feature | Color Feature   |
|----------------------|----------------------|------------------------|----------------|
| B5 and B11, June; B5 and B12, September | B2, B4 and B5, June; B6, July; B2, B5, B6 and B12, September | GARI, June; NDWI and VARIgreen, July | Saturation, September |

4.1.2. Methods Comparison

Figure 10 shows the Spearman rank correlation coefficient between the 16 features of the three feature selection methods. The features are arranged in the order of raw spectral features, segmentation features, spectral index features, color features, and texture features. It can be seen that the 16 features of TFSM, including RF-FI and RF-RFE, were still highly correlated, meaning that there was redundant information between the selected features. Compared with the RF-FI method, there was more redundant information in the features selected by the RF-RFE method. Moreover, we found that the correlation between the segmentation features and the spectral index features was high using the RF-RI method, and the correlation between the raw spectral features and the segmentation features was high using the RF-RFE method. By contrast, the features of OFSM were relatively independent, and the redundant information between the selected features was relatively small.

![Figure 10](image)

**Figure 10.** The Spearman rank correlation coefficient between the 16 features of the three feature selection methods. (a) Random forest feature importance selection (RF-FI), (b) Random forest recursive feature elimination (RF-RFE), (c) OFSM.

Table 6 lists the time consumption of the three feature selection methods. It can be seen that the time consumption of OFSM was in the intermediate level, while that of RF-FI and RF-RFE were the smallest and largest, respectively. According to the designed feature selection strategy, OFSM greatly reduced the time consumption. In summary, the features selected by OFSM were independent of each other and the time consumption was acceptable.

| Method       | RF-FI   | RF-RFE  | OFSM   |
|--------------|---------|---------|--------|
| Time Consumption | 1.97 s  | 132.54 s| 26.05 s|

Software: Anaconda3-2018.12 Python 3.7.1
Computer configuration: Windows 10 x64, i5-8300H CPU @ 2.30GHz, 8G RAM

4.2. Deep-Learning Network Hyperparameter Selection

The hyperparameter settings of deep-learning networks usually affect the training results. In order to select the optimal hyperparameters for the hybrid CNN-RF networks, we tested the common hyperparameters, including num_filter1, convolution kernel_size, pooling kernel_size, learning_rate,
dropout, max_iterations and batch_size. The tested and optimal hyperparameters of VGG-RF and Conv1D-RF are listed in Table 7. Based on the experimental results, when the optimal parameters were selected the training efficiency and the accuracy of the hybrid CNN-RF networks were the best.

Table 7. Tested and optimal hyperparameters of VGG-RF and Conv1D-RF.

| Hyperparameter Name (Description) | Tested Values | Optimal Values |
|----------------------------------|---------------|----------------|
| num_filter1 (number of filters in the first convolutional layer) | 32, 64, 128 | 64 | 64 |
| convolution kernel_size (the filter size of convolutional layers) | 2 × 2, 3 × 2, 3 × 3 | 2 × 2 | 3/5 |
| pooling kernel_size (the filter size of pooling layers) | 2 × 2, 3 × 2, 3 × 3 | 2, 3, 4 | 2 × 2 |
| learning_rate (learning rate) | 0.1, 0.01, 0.001 | 0.1, 0.01, 0.001 | 0.001 | 0.001 |
| dropout (dropout rate in hidden layers) | 40%, 50%, 60%, 70%, 80% | 40%, 50%, 60%, 70%, 80% | 40%/ 50% |
| max_iterations (maximum number of iterations) | 5000, 10000, 15000 | 10000 | 5000 |
| batch_size (number of samples for each training) | 50, 60, 80, 100 | 50, 60, 80, 100 | 80 | 80 |

4.3. Classification and Accuracy Assessment

Based on the selected optimal features using OFSM and TFSM, four networks were tested for comparison: Conv1D-RF, VGG-RF, Conv1D, and VGG. Since the most satisfactory classification results were achieved by the designed hybrid Conv1D-RF network, we further compared Conv1D-RF with three mainstream networks, including LSTM-RF, ResNet, and U-Net.

4.3.1. Comparison of the Hybrid CNN-RF Networks with the Original Deep-Learning Networks

In this study, the training and validation datasets were used to train and select the optimal parameters for the deep-learning networks. Classification results by Conv1D and VGG were used to represent the performance of popular deep-learning algorithms. Furthermore, Conv1D-RF and VGG-RF, as presented in Section 3.3, were proposed to represent the fusion performance of deep-learning and machine-learning algorithms. Table 8 shows the classification results of the hybrid CNN-RF networks and original deep-learning networks based on the selected features using OFSM and TFSM. The Conv1D and VGG classifiers performed worse than the Conv1D-RF and VGG-RF, with many areas with a higher “speckle” (i.e., more heterogeneity) of classes across the landscape. Compared to the TFSM results, there was less noise in the OFSM classification results, and some improvements were made regarding the misclassification and omission of the indistinguishable corn and soybean. The use of Conv1D-RF based on OFSM particularly improved the classification results of crop types compared with the other three networks.

To evaluate the effectiveness of the hybrid CNN-RF networks and original deep-learning networks, comparisons were conducted based on the test dataset. As shown in Table 9, the OA and K coefficient of OFSM were superior to those of TFSM. The classification results of the hybrid CNN-RF networks were better than those of the original deep-learning networks with softmax. The OA of Conv1D-RF was higher than that of Conv1D by 1.7%, and OA of VGG-RF was higher than that of VGG by 1.3%. The features extracted by the deep-learning networks were further input into random forest (RF) classifier for analysis, which can better extract and identify useful crop information. The experimental results demonstrate that the hybrid networks can make full use of the advantages of the two classifiers to effectively identify crops.
Table 8. Comparisons of crop classification of the hybrid convolutional neural network-random forest (CNN-RF) networks and the original deep-learning networks using traditional feature selection method (TFSM) and OFSM.

| Method | OA/K Coefficient |
|--------|------------------|
|        | Conv1D-RF | VGG-RF | Conv1D | VGG |
| OFSM   | 94.27%/0.917 | 93.23%/0.894 | 92.59%/0.894 | 91.89%/0.884 |
| RF-RFE | 94.01%/0.914 | 92.81%/0.897 | 92.33%/0.890 | 91.58%/0.880 |
| RF-FI  | 90.97%/0.871 | 90.13%/0.853 | 87.47%/0.824 | 86.74%/0.814 |

Although the time consumption of RF-RFE was the largest, the OA of RF-RFE was obviously higher than that of RF-FI. The OA of OFSM was slightly higher than that of RF-RFE by 0.3%, and much higher than that of RF-FI by 4%. Conv1D and VGG showed a distinct capability. Conv1D employed one-dimensional filters to capture the temporal pattern or shape features of the input sequence. The OA of Conv1D-RF was higher than that of VGG-RF by 1%, and the OA of Conv1D was higher than that of VGG by 0.7% for multi-temporal features. The Conv1D-RF based on OFSM had the highest accuracy (94.27%) and K coefficient (0.917) among all types of networks, so this hybrid network is assumed to be the best choice in this study.

Table 9. Comparisons of the OA and K coefficient of the hybrid CNN-RF networks and original deep-learning networks using the three methods.

| Method | OA/K Coefficient |
|--------|------------------|
|        | Conv1D-RF | VGG-RF | Conv1D | VGG |
| RF-FI  | 90.97%/0.871 | 90.13%/0.853 | 87.47%/0.824 | 86.74%/0.814 |
| RF-RFE | 94.01%/0.914 | 92.81%/0.897 | 92.33%/0.890 | 91.58%/0.880 |
| OFSM   | 94.27%/0.917 | 93.23%/0.903 | 92.59%/0.894 | 91.89%/0.884 |

4.3.2. Comparison of Conv1D-RF with Mainstream Networks

This study compared the hybrid Conv1D-RF network with popular deep-learning based networks, including LSTM-RF [49], ResNet [50], and U-Net [51]. LSTM-RF uses RF instead of the FC layer to make the final classification decision. ResNet uses global average pooling (GAP) instead of the FC layer and U-Net is a fully convolutional network without FC layers. Table 10 shows the overall accuracy (OA) and Kappa (K) coefficient of the four deep-learning networks using the three feature selection methods. The proposed hybrid Conv1D-RF network combined with OFSM achieved the highest OA of 94.27% compared with ResNet (93.55%), LSTM-RF (92.91%), and U-Net (91.92%).
Table 10. Comparisons of the OA and K coefficient of the four deep-learning networks using the three feature selection methods.

| Method   | OA/K Coefficient |
|----------|------------------|
|          | Conv1D-RF | LSTM-RF | ResNet | U-Net |
| RF-FI    | 90.97%/0.871 | 91.16%/0.874 | 84.76%/0.789 | 84.33%/0.777 |
| RF-RFE   | 94.01%/0.914 | 92.84%/0.896 | 92.14%/0.887 | 91.89%/0.884 |
| OFSM     | 94.27%/0.917 | 92.91%/0.899 | 93.55%/0.905 | 91.92%/0.885 |

In order to further analyze the superiority of Conv1D-RF, we applied the four parameters of pixel-based evaluation, precision, recall, F1 score, and IoU to evaluate the performance of each class using the four deep-learning networks and three feature selection methods. Figure 11 shows the analysis of the evaluation parameters of the four deep-learning networks using OFSM and TFSM. The experimental results demonstrated that Conv1D-RF had an outstanding performance compared with other three networks for the same feature combinations obtained by OFSM and TFSM. The performance of the four deep-learning networks for rice recognition was similar, but for corn and soybean, which were difficult to distinguish, Conv1D-RF achieved better recognition results for the two crop types in the four networks. In particular, the combination of Conv1D-RF and OFSM obtained the best crop recognition results and performed well in the four evaluation parameters.

![Figure 11. Cont.](image)
5. Discussion

5.1. Analysis of Feature Selection Using OFSM

Current studies often input the raw bands of remote sensing images into deep-learning models. For example, Kussul et al. [52] input the raw bands of multi-temporal Landsat-8 and Sentinel-1A images into 1D CNN and 2D CNN networks for training and learning. Ji et al. [3] input the raw bands of multi-temporal GF-2 images into a 3D CNN network for crop recognition. We compare the crop recognition accuracy and processing time for the networks’ training processes for optimal feature combination and raw spectral bands to illustrate the necessity of feature selection processing.

The 30 raw spectral bands of multi-temporal Sentinel-2, and the 16 feature bands selected by OFSM were put into the Conv1D-RF and VGG-RF models, respectively. The OA and time consumption during training of the two hybrid networks are shown in Table 11. The features selected by OFSM were superior to the raw spectral bands in both time consumption and OA, indicating that this proposed method can be applied for crop identification with a high efficiency and accuracy.
Two important agricultural commodities, corn and soybeans, are commonly difficult to distinguish due to their phenological similarity. In recent years, many studies have been carried out based on the mapping of corn and soybeans [36]. Zhong et al. [53] used a decision tree classifier and vegetation phenology information to distinguish corn and soybeans, and achieved an overall classification accuracy of more than 88% was achieved [54]. The main factors affecting the accuracy of crop classification are commonly caused by mixed pixels. For example, the bias of mixed pixel classification accuracy and K coefficient of 87.2% and 0.804, respectively, for crop mapping in the state of Paraná, Brazil, for the crop year 2012. The results showed that some corn was mistakenly classified as soybeans. When the data used for training the classifier was the same as the mapping year, a high level of accuracy was achieved.

The characteristics of the Conv1D-based network can be inspected by visualizing the feature maps on different layers. For example, Zhong et al. [26] inspected the behavior of the Conv1D-based model by visualizing the activation on different layers. The shallow layers of the Conv1D model could capture local feature variations, while the higher layers focused on the overall feature patterns. The Conv1D layers were used as a multi-level feature extractor in crop classification tasks, which automatically extract features from the input time series during the training process. We use visualization techniques to examine what the deep-learning network model learns and how it understands the input optimal features from the time series. Figure 12 visualizes the output feature maps obtained from the training dataset of the four classes using Conv1D, including the first Conv1D layer, the inception module, the second Conv1D layer, and the third Conv1D layer. The output feature map size of the first Conv1D layer is $14 \times 64$, the inception module is $29 \times 128$, the second Conv1D layer is $27 \times 128$, and the third Conv1D layer is $25 \times 256$, where 14, 29, 27, and 25 refer to the feature size, and 64, 128, and 256 refer to the number of channels. As shown in Figure 12, there are significant differences between the features extracted from various classes (urban, corn, rice, and soybean) in the same neural network layer. Conv1D layers can be stacked so that lower layers focus on certain temporal patterns, whereas higher layers can aggregate simple patterns into complex shapes. Therefore, the neurons in the shallow layers of the network can extract low-level features, with increasing network depth, the neural network can still efficiently extract more holistic features from the higher layers.

5.2. Conv1D Feature Map Visualization

The characteristics of the Conv1D-based network can be inspected by visualizing the feature maps on different layers. For example, Zhong et al. [26] inspected the behavior of the Conv1D-based model by visualizing the activation on different layers. The shallow layers of the Conv1D model could capture local feature variations, while the higher layers focused on the overall feature patterns. The Conv1D layers were used as a multi-level feature extractor in crop classification tasks, which automatically extract features from the input time series during the training process. We use visualization techniques to examine what the deep-learning network model learns and how it understands the input optimal features from the time series. Figure 12 visualizes the output feature maps obtained from the training dataset of the four classes using Conv1D, including the first Conv1D layer, the inception module, the second Conv1D layer, and the third Conv1D layer. The output feature map size of the first Conv1D layer is $14 \times 64$, the inception module is $29 \times 128$, the second Conv1D layer is $27 \times 128$, and the third Conv1D layer is $25 \times 256$, where 14, 29, 27, and 25 refer to the feature size, and 64, 128, and 256 refer to the number of channels. As shown in Figure 12, there are significant differences between the features extracted from various classes (urban, corn, rice, and soybean) in the same neural network layer. Conv1D layers can be stacked so that lower layers focus on certain temporal patterns, whereas higher layers can aggregate simple patterns into complex shapes. Therefore, the neurons in the shallow layers of the network can extract low-level features, with increasing network depth, the neural network can still efficiently extract more holistic features from the higher layers.

5.3. Crop Distribution Analysis

Two important agricultural commodities, corn and soybeans, are commonly difficult to distinguish due to their phenological similarity. In recent years, many studies have been carried out based on the mapping of corn and soybeans [36]. Zhong et al. [53] used a decision tree classifier and vegetation...
phenology information to distinguish corn and soybeans, and achieved an overall accuracy and K coefficient of 87.2% and 0.804, respectively, for crop mapping in the state of Paraná, Brazil, for the crop year 2012. The results showed that some corn was mistakenly classified as soybeans. When the data used for training the classifier was the same as the mapping year, a classification accuracy of more than 88% was achieved [54]. The main factors affecting the accuracy of crop classification are commonly caused by mixed pixels. For example, the bias of mixed pixel classification will be affected by the complexity of the terrain, and sometimes even a small amount of sub-pixel natural vegetation will also affect the phenological detection of crops.

In 2018, the cultivated land areas of corn and soybean planted in the study area were approximately 906 ha (44.8% of the area) and 259 ha (12.80%), respectively. We analyzed the effects of different feature selection methods and networks on crop distribution, especially corn and soybeans, which are difficult to distinguish. The comparison of crop-type distribution is shown in Table 12. In the blue columns, the Conv1D-RF network is used with the features selected by OFSM and TFSM (RF-RFE and RF-FI), and the last column shows the crop-type distribution using VGG-RF with the optimal features selected by OFSM. Compared with TFSM, the classification results based on OFSM were much closer to the reference dataset for the distribution of crop-type, which further illustrates the effectiveness of the proposed method. Conv1D-RF based on OFSM more accurately mapped areas of corn and soybean than the VGG-RF based on OFSM.

Table 12. Comparison of the crop-type distribution.

| Land Cover Type | Reference Dataset | OFSM+ Conv1D-RF | RF-RFE+ Conv1D-RF | RF-FI+ Conv1D-RF | OFSM+ VGG-RF |
|-----------------|-------------------|-----------------|------------------|-----------------|--------------|
| Rice            | 473.15/23.40%     | 478.60/23.67%   | 477.80/23.63%    | 479.82/23.73%   | 476.18/23.55% |
| Corn            | 905.86/44.80%     | 874.52/43.25%   | 865.42/42.80%    | 840.95/41.59%   | 859.96/42.53% |
| Soybean         | 258.81/12.80%     | 266.50/13.18%   | 282.68/13.98%    | 309.97/15.33%   | 284.70/14.08% |
| Urban           | 384.18/19.00%     | 402.38/19.90%   | 396.10/19.59%    | 391.26/19.35%   | 401.16/19.84% |

Figure 13 more clearly shows the differences in the crop classification results obtained by the four methods. Compared with the reference dataset, the changes in rice distribution obtained by all the methods were not significantly different, while the distribution changes for corn and soybeans varied greatly. A negative percentage change in the corn area indicates that some corn is underestimated, while a positive percentage change in the soybean area indicates that the soybeans are overestimated. Table 13 shows the confusion matrix of the classification results of Conv1D-RF based on OFSM. The number of missing pixels of corn was the highest, and most of the missing corn was mistakenly classified as soybeans (5903 pixels), resulting in a relatively high commission for soybeans. The omitted pixels for rice and urban were relatively small. Part of the missing soybeans was mistakenly classified as corn (1205 pixels), and the other part was mistakenly classified as urban (1078 pixels). According to the field survey, the corn planting in the study area is relatively regular, while the soybean planting is more scattered and some areas are mixed with corn, causing the mixed pixel phenomenon. In addition, the reflectance of soybean in the spectrum is high, which may make the reflectance of mixed pixels shift toward that of soybeans, resulting in some corn being mistakenly classified as soybeans.
VGG-RF, a 1.7% greater OA than Conv1D, and a 2.4% greater OA than VGG. Since the hierarchical architecture of Conv1D uses time series as a classification input, it can effectively extract features of crop-growth dynamics during model training. In summary, the proposed hybrid CNN-RF network based on OFSM is a promising approach for utilizing the advantages of two classifiers by complementary classification which can achieve a higher accuracy of crop identification and a lower time consumption. The method of applying the hybrid deep-learning models to classify crops based on their capability of identifying individual crop types, Conv1D-RF had a 1% greater OA than VGG-RF, a 1.7% greater OA than Conv1D, and a 2.4% greater OA than VGG. Since the hierarchical architecture of Conv1D uses time series as a classification input, it can effectively extract features of crop-growth dynamics during model training. In summary, the proposed hybrid CNN-RF network based on the features selected by OFSM is a promising approach for utilizing the advantages of two classifiers by complementary classification which can achieve a higher accuracy of crop identification and a lower time consumption. The method of applying the hybrid deep-learning models to classify remote sensing imagery is still in the stage of continuous practice and exploration. Given sufficient data, however, the hybrid deep-learning models can be used to learn the most appropriate band combination for a specific task, possibly eliminating the input of redundant bands. Therefore, the kind of information needed and how to transform the information for classification by deep-learning models is worth exploring. To yield a higher accuracy in future applications, model architectures...
based on three-dimensional (3D) spatiotemporal convolution should be considered. Since Sentinel-2 imagery can be disturbed by clouds, future work could focus on developing a multi-source remote sensing imagery fusion approach for crop classification. The combination of optical images and synthetic-aperture radar images may improve the accuracy of crop classification, better evaluating the distribution of crops. In addition, to improve the accuracy of agricultural mapping, it would be helpful to identify the components of mixed pixels in heterogeneous regions for the modeling and inversion process of agricultural remote sensing methods which would support the strategic needs of agricultural sustainable development.

Author Contributions: Validation, T.J.; investigation, X.L.; data curation, R.R.; writing—original draft preparation, S.Y.; writing—review and editing, L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 41871225, 41871248 and 41771400.

Acknowledgments: The authors would like to thank the editors and reviewers for their suggestions and revisions.

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

References

1. Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. For. Meteorol. 2013, 173, 74–84. [CrossRef]
2. Cai, Y.; Guan, K.; Peng, J.; Wang, S.; Seifert, C.; Wardlow, B.; Li, Z. 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. 2018, 210, 35–47. [CrossRef]
3. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sens. 2018, 10, 75. [CrossRef]
4. Wardlow, B.D.; Egbert, S.L.; Kastens, J.H. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. Remote Sens. Environ. 2007, 108, 2903–2910. [CrossRef]
5. Chang, J.; Hansen, M.C.; Pittman, K.; Carroll, M.; DiMiceli, C. Corn and soybean mapping in the United States using MODIS time-series data sets. Agron. J. 2007, 99, 1654–1664. [CrossRef]
6. Duro, D.C.; Franklin, S.E.; Dubé, M.G. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. Int. J. Remote Sens. 2012, 33, 4502–4526. [CrossRef]
7. Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. Energies 2017, 10, 1168. [CrossRef]
8. Yan, K.; Zhang, D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens. Actuators B Chem. 2015, 212, 353–363. [CrossRef]
9. Hao, P.; Zhan, Y.; Wang, L.; Niu, Z.; Shakir, M. Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA. Remote Sens. 2015, 7, 5347–5369. [CrossRef]
10. Yin, L.; You, N.; Zhang, G.; Huang, J.; Dong, J. Optimizing Feature Selection of Individual Crop Types for Improved Crop Mapping. Remote Sens. 2020, 12, 162. [CrossRef]
11. Liu, H.; An, H. Preliminary tests on the performance of MLC-RFE and SVM-RFE in Landsat-8 image classification. Arab. J. Geosci. 2020, 13, 130. [CrossRef]
12. Mathur, A.; Foody, G.M. Crop classification by support vector machine with intelligently selected training data for an operational application. Int. J. Remote Sens. 2008, 29, 2227–2240. [CrossRef]
13. Ahmad, I.; Siddiqui, M.H.; Fatima, I.; Lee, S.; Lee, Y.K. Weed classification based on Haar wavelet transform via k-nearest neighbor (k-NN) for real-time automatic sprayer control system. In Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, Seoul, Korea, 21–23 February 2011; p. 17.
14. Murthy, C.; Raju, P.; Badrinath, K. Classification of wheat crop with multi-temporal images: Performance of maximum likelihood and artificial neural networks. Int. J. Remote Sens. 2003, 24, 4871–4890. [CrossRef]
15. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y. Xgboost: Extreme Gradient Boosting. R Package Version 0.6-4. Available online: Cran.fhcrc.org/web/packages/xgboost/vignettes/xgboost.pdf (accessed on 1 January 2017).
16. Friedl, M.A.; Brodley, C.E. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 1997, 61, 399–409. [CrossRef]

17. Tatsumi, K.; Yamashiki, Y.; Torres, M.C.; Taipe, C.R. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Comput. Electron. Agric.* 2015, 115, 171–179. [CrossRef]

18. Kumar, P.; Gupta, D.K.; Mishra, V.N.; Prasad, R. Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data. *Int. J. Remote Sens.* 2015, 36, 1604–1617. [CrossRef]

19. Li, H.; Zhang, C.; Zheng, S.; Atkinson, P.M. Crop classification from full-year fully-polarimetric L-band UAVSAR time-series using the Random Forest algorithm. *Int. J. Appl. Earth Obs.* 2020, 87, 102032. [CrossRef]

20. Zhang, H.; Eziz, A.; Xiao, J.; Tao, S.; Wang, S.; Tang, Z.; Fang, J. High-Resolution Vegetation Mapping Using Extreme Gradient Boosting Based on Extensive Features. *Remote Sens.* 2019, 11, 1505. [CrossRef]

21. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* 2016, 54, 1349–1362. [CrossRef]

22. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. Available online: http://arxiv.org/abs/1508.00092 (accessed on 14 August 2015).

23. Xu, X.; Li, W.; Ran, Q. Multisource remote sensing data classification based on convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 937–949. [CrossRef]

24. Rußwurm, M.; Körner, M. Multi-Temporal Land Cover Classification with Long Short-Term Memory Neural Networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2017, 42, 551. [CrossRef]

25. Rußwurm, M.; Körner, M. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1496–1504.

26. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 2019, 221, 430–443. [CrossRef]

27. Guidici, D.; Clark, M. One-Dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. *Remote Sens.* 2017, 9, 629. [CrossRef]

28. Ko, A.H.R.; Sabourin, R. Single Classifier-based Multiple Classification Scheme for weak classifiers: An experimental comparison. *Expert Syst. Appl.* 2013, 40, 3606–3622. [CrossRef]

29. Debeir, O.; Van Den Steen, I.; Latinne, P.; Van Ham, P.; Wolff, E. Textural and contextual land-cover classification using single and multiple classifier systems. *Photogramm. Eng. Remote Sens.* 2002, 68, 597–606.

30. Briem, G.J.; Benediktsson, J.A.; Sveinsson, J.R. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* 2002, 40, 2291–2299. [CrossRef]

31. Du, P.; Xia, J.; Zhang, W.; Tan, K.; Liu, Y.; Liu, S. Multiple classifier system for remote sensing image classification: A review. *Sensors* 2012, 12, 4764–4792. [CrossRef]

32. Zhou, Z.H. Ensemble learning. In *Encyclopedia of Biometrics*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 411–416.

33. Leng, J.; Li, T.; Bai, G.; Dong, Q.; Dong, H. Cube-CNN-SVM: A novel hyperspectral image classification method. In Proceedings of the 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), San Jose, CA, USA, 6–8 November 2016; pp. 1027–1034.

34. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* 2016, 8, 166. [CrossRef]

35. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* 2018, 72, 122–130. [CrossRef]

36. Sharma, V.; Irmak, S.; Kilic, A.; Sharma, V.; Gilley, J.E.; Meyer, G.E.; Marx, D. Quantification and mapping of surface residue cover for maize and soybean fields in south central Nebraska. *Trans. ASABE* 2016, 59, 925–939.

37. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* 2004, 59, 167–181. [CrossRef]

38. Connolly, C.; Fleiss, T. A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. *IEEE Trans. Image Process.* 1997, 6, 1046–1048. [CrossRef] [PubMed]
39. Peña-Barrağán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. Remote Sens. Environ. 2011, 115, 1301–1316. [CrossRef]
40. Ruiz, L.A.; Fdez-sarra, A.; Recio, J.A. Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. Int. Arch. Photogramm. Remote Sens. 2004, XXXV, 1682–1750.
41. Genuer, R.; Poggi, J.M.; Tuleau-Malot, C. VSURF: An R Package for variable selection using random forests. R J. 2015, 7, 19–33. [CrossRef]
42. Granitto, P.M.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemom. Intell. Lab. Syst. 2006, 83, 83–90. [CrossRef]
43. Zar, J.H. Significance testing of the spearman rank correlation coefficient. J. Am. Stat. Assoc. 1972, 62, 578–580. [CrossRef]
44. Khare, S.; Bhandari, A.; Singh, S.; Arora, A. ECG arrhythmia classification using spearman rank correlation and support vector machine. In Proceedings of the International Conference on Soft Computing for Problem Solving (SoCProS 2011), Roorkee, India, 20–22 December 2011; pp. 591–598.
45. Sengupta, A.; Ye, Y.; Wang, R.; Liu, C.; Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. Front. Neurosci. 2019, 13. [CrossRef]
46. Mateen, M.; Wen, J.; Song, S.; Huang, Z. Fundus image classification using VGG-19 architecture with PCA and SVD. Symmetry 2019, 11, 1. [CrossRef]
47. Dong, L.; Du, H.; Mao, F.; Han, N.; Li, X.; Zhou, G.; Liu, T. Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2019, 13, 113–128. [CrossRef]
48. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sens. 2017, 9, 498. [CrossRef]
49. Punia, S.; Nikolopoulos, K.; Singh, S.P.; Madaan, J.K.; Litsiou, K. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. Int. J. Prod. Res. 2020, 1–16. [CrossRef]
50. Liu, S.; Tian, G.; Xu, Y. A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. Neurocomputing 2019, 338, 191–206. [CrossRef]
51. Rakhlin, A.; Davydow, A.; Nikolenko, S. Land Cover Classification from Satellite Imagery with U-Net and Lovasz-Softmax Loss. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
52. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 2017, 14, 778–782. [CrossRef]
53. Zhong, L.; Hu, L.; Yu, L.; Gong, P.; Biging, G.S. Automated mapping of soybean and corn using phenology. ISPRS J. Photogramm. Remote Sens. 2016, 119, 151–164. [CrossRef]
54. Zhong, L.; Gong, P.; Biging, G.S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. Remote Sens. Environ. 2014, 140, 1–13. [CrossRef]