

Article

Incorporating Multi-Temporal Remote Sensing and a Pixel-Based Deep Learning Classification Algorithm to Map Multiple-Crop Cultivated Areas

Xue Wang ^{1,2,3,4}, Jiahua Zhang ^{3,4,*} , Xiaopeng Wang ³, Zhenjiang Wu ⁴ and Foyez Ahmed Prodhan ⁵ 

- ¹ Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, China; wangxue233@mailsucas.ac.cn
- ² Key Laboratory of Earth Observation of Hainan Province, Hainan Aerospace Information Research Institute, Sanya 572029, China
- ³ Remote Sensing Information and Digital Earth Center, College of Computer Science and Technology, Qingdao University, Qingdao 266071, China; wxp@qdu.edu.cn
- ⁴ College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China; wuzhenjiang22@mailsucas.ac.cn
- ⁵ Department of Agricultural Extension and Rural Development, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh; foyez@bsmrau.edu.bd
- * Correspondence: zhangjh@radi.ac.cn; Tel.: +86-10-82178122

Abstract: The accurate monitoring of crop areas is essential for food security and agriculture, but accurately extracting multiple-crop distribution over large areas remains challenging. To solve the above issue, in this study, the Pixel-based One-dimensional convolutional neural network (PB-Conv1D) and Pixel-based Bi-directional Long Short-Term Memory (PB-BiLSTM) were proposed to identify multiple-crop cultivated areas using time-series NaE (a combination of NDVI and EVI) as input for generating a baseline classification. Two approaches, Snapshot and Stochastic weighted averaging (SWA), were used in the base-model to minimize the loss function and improve model accuracy. Using an ensemble algorithm consisting of five PB-Conv1D and seven PB-BiLSTM models, the temporal vegetation index information in the base-model was comprehensively exploited for multiple-crop classification and produced the Pixel-Based Conv1D and BiLSTM Ensemble model (PB-CB), and this was compared with the PB-Transformer model to validate the effectiveness of the proposed method. The multiple-crop cultivated area was extracted from 2005, 2010, 2015, and 2020 in North China by using the PB-Conv1D combine Snapshot (PB-CDST) and PB-CB models, which are a performance-optimized single model and an integrated model, respectively. The results showed that the mapping results of the multiple-crop cultivated area derived by PB-CDST (OA: 81.36%) and PB-BiLSTM combined with Snapshot (PB-BMST) (OA: 79.40%) showed exceptional accuracy compared to PB-Transformer combined with Snapshot and SWA (PB-TRSTSA) (OA: 77.91%). Meanwhile, the PB-CB (OA: 83.43%) had the most accuracy compared to the pixel-based single algorithm. The MODIS-derived PB-CB method accurately identified multiple-crop areas for wheat, corn, and rice, showing a strong correlation with statistical data, exceeding 0.7 at the municipal level and 0.6 at the county level.

Keywords: multiple-crop classification; multi-temporal data; deep learning; remote sensing; snapshot; SWA; ensemble classifier



Citation: Wang, X.; Zhang, J.; Wang, X.; Wu, Z.; Prodhan, F.A. Incorporating Multi-Temporal Remote Sensing and a Pixel-Based Deep Learning Classification Algorithm to Map Multiple-Crop Cultivated Areas. *Appl. Sci.* **2024**, *14*, 3545. <https://doi.org/10.3390/app14093545>

Academic Editor: Krzysztof Koszela

Received: 11 March 2024

Revised: 14 April 2024

Accepted: 16 April 2024

Published: 23 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Crop area identification plays a significant role in food security [1,2]. Timely and accurate crop spatial distribution maps are essential for agricultural production, crop planting structures, and the formulation of national food policies [3,4]. The traditional approaches for acquiring information on crop planting areas relied on manual measurement

and statistical sampling, resulting in time consumption, the wastage of human resources, and inefficiencies [5].

Over the past few years, remote sensing (RS) technology has proven effective in monitoring the dynamics of crop-growing areas on a regional or global scale owing to its advantages in rapidity and accuracy [6–8]. Many researchers have utilized RS images with temporal features of crop growth to map crop planting areas [9–11]. The temporal features derived from time series captured by Moderate Resolution Imaging Spectroradiometer (MODIS) data, such as normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI), include different stages of growth and development in crops [12,13]. Time series of vegetation indices (VI) have provided rich and valuable information in detecting crop growth variations and spatial planting distribution [14]. Based on the temporal data derived from MODIS images, He et al. [15] and Zhang et al. [16] combined the NDVI and EVI with Attention-based Long Short-term Memory networks (A-LSTM) and phenological information to identify the winter-wheat area in Huanghuaihai region and corn in Northeast China, respectively. Wang et al. [17] inserted NaE (a combination of NDVI and EVI) into the model and found that the classification results were better than the data from the use of either the NDVI or the EVI. The NaE inserts both the NDVI and the EVI into the model simultaneously, which will effectively avoid data collection errors due to MODIS sensors, clouds, or rain, thereby in turn effectively avoiding classification errors caused by data loss or measurement errors and allowing the model to extract more features. Therefore, the NaE was used as input data in this study.

Various methods have processed data in multi-temporal RS to extract temporal information or phenological metrics for crop classification [18,19]. The traditional methods for crop identification are simple statistics, threshold-based equations, and pre-defined mathematical equations [20]. Luo et al. [21] combined inflection and threshold-based methods to detect the key phenological stages of maize, wheat, and rice for crop distribution across China. Xun et al. [22] used a fused representation-based algorithm for cotton area identification and obtained the best accuracy result of 79.46%. Although the above approaches in temporal feature extraction provide many options for crop classification, there are some problems. Firstly, the threshold-based methods relied on manual work, and the model design depended on human experience and domain knowledge. This is a time-consuming and ineffective process [20]. The temporal feature extracted from time series based on human experts' experience will result in an information loss in granular data and the incomplete utilization of interval features. Secondly, pre-defined mathematical functions inevitably constrain flexibility in handling temporal data; selecting a suitable function for all crop types presents challenges, especially in diverse crop classification studies [14]. The dynamic field of deep learning offers solutions to address these weaknesses.

In recent years, using deep learning classification methods for identifying crop distribution has achieved positive outcomes [23–25]. One advantage of the deep learning method is its flexibility, as it does not confine itself to predefined models [20]. The models learn feature representations from data in an end-to-end form, avoiding the need for manual feature engineering based on human experience and prior knowledge [26]. Convolution neural network (CNN), recurrent neural network (RNN), and self-attention network are three renowned neural network architectures adept at effectively processing time series data [27]. CNN can extract high-dimensional features from RS data's spatial and temporal domains for crop recognition [28–30]. Zhong et al. [20] extracted the temporal feature from MODIS NDVI time series data using CNN and LSTM classifiers to identify the 14 categories of crops in the San Joaquin Valley, California. RNN observes the connected neurons to learn the time sequence relationship and performs better in processing time series data. LSTM is a variant of RNN that shows long-term temporal dependence by incorporating a gate control mechanism [31]. To further improve the temporal feature extraction of the LSTM model, the bi-directional flow method is introduced to form BiLSTM [32]. Xu et al. [33] combined ARD data, Landsat 7 Enhanced Thematic Mapper Plus, Landsat8 Operational Land Imager data, and the DCM classifier to identify the main crops in six

study sites. Based on the deep learning algorithm, Kumar et al. [29] and Mou et al. [34] combined the multi-temporal images to extract features with 1D convolution and RNN classification algorithms for land cover classification in Ukraine and crop classification, respectively. In 2017, the self-attention mechanism was proposed for analyzing sequences in natural language processing [35]. It facilitates parallel feature extraction from super-long sequences and achieves excellent recognition results. Rußwurm and Körner [36] used the self-attention classification method to identify the crop type. With the aim of improving the efficacy of deep learning models, many strategies have emerged. Based on the CIFAR-10 and CIFAR-100 datasets, Huang et al. [37] and Izmailov et al. [38] combined the Snapshot and SWA strategies with the DenseNet and ResNet-164 network for image identification, and the results showed that, compared to the single model, Snapshot's performance was improved by 0.34% and 1.84% and, compared with the SGD optimization algorithm, SWA's performance was enhanced by 0.61% and 1.86%, respectively.

The combination of multi-temporal satellite data with a deep learning model for identifying crop types has been widely used, and favorable results have been obtained. Luo et al. [39] combined NDVI time series, textural, and phenological features from Sentinel-2 data with an artificial neural network (ANN) for multiple-crop identification in North China; However, the accuracy and kappa coefficient failed to exceed 80% and 0.7, respectively. Li et al. [40] proposed an object-based convolutional neural network (TS-OCNN) combined with UAVSAR images for multiple-crop classification at two stations in the Sacramento Valley, with an accuracy of 81.63% and 85.88%, respectively. Wang et al. [7] used EVI time series with phenology information for corn identification, achieving an accuracy of over 85% and an R^2 value of 0.81 at the city level. However, there are still two problems in the above research. Firstly, the study with higher accuracy only identified a single crop or small-large area. In addition, the model's accuracy is worse with multiple-crop classification. Hence, the potential and performance of this method for identifying multiple-crop areas have not been fully exploited. Developing deep learning models suitable for multiple-crop areas by comparing them to PB-Transformer is necessary to assess the latter's potential and capabilities. Overall, mapping multiple-crop spatial distribution is crucial for national food security. Nevertheless, most existing studies on multiple-crop identification were conducted with single crops or unsatisfying results.

This study aims to develop a pixel-based deep learning algorithm that is adaptable to the multi-temporal data from the MODIS satellite for classifying multi-crop cultivation areas in North China. The objectives were: (1) to propose two pixel-based deep learning models named PB-Conv1D and PB-BiLSTM, comparing them with PB-Transformer; (2) to introduce Snapshot and SWA techniques to enhance model accuracy in deep learning; (3) to develop the PB-CB model through ensemble learning techniques from PB-Conv1D and PB-BiLSTM; (4) to generate multiple-crop distribution maps in North China using the time series Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data from MODIS satellite imagery combined with the PB-CB model for the years 2005, 2010, 2015, and 2020; (5) to compare the accuracy of various deep learning methods.

2. Materials and Methods

The overall workflow, as depicted in Figure 1, comprises five distinct parts. Firstly, training samples were selected through agricultural sites, field surveys, and visual interpretation, while test samples were masked from GLC_FCS30 data to extract the cropland distribution. The annual time series NDVI and EVI data of the sample sites were extracted from the processed MOD13Q1 data for the sample points. Secondly, two DL models named PB-Conv1D and PB-BiLSTM were developed and compared with the pixel-based Transformer (PB-Transformer) model, followed by the adoption of Snapshot and SWA to improve the classification accuracy of the models. Thirdly, the integrated classifier was obtained by combining DL models based on the ensemble learning method, comparing all the above models to obtain the optimal classification model. Finally, the proposed optimal model mapped the crop cultivation area for 2005, 2010, 2015, and 2020.

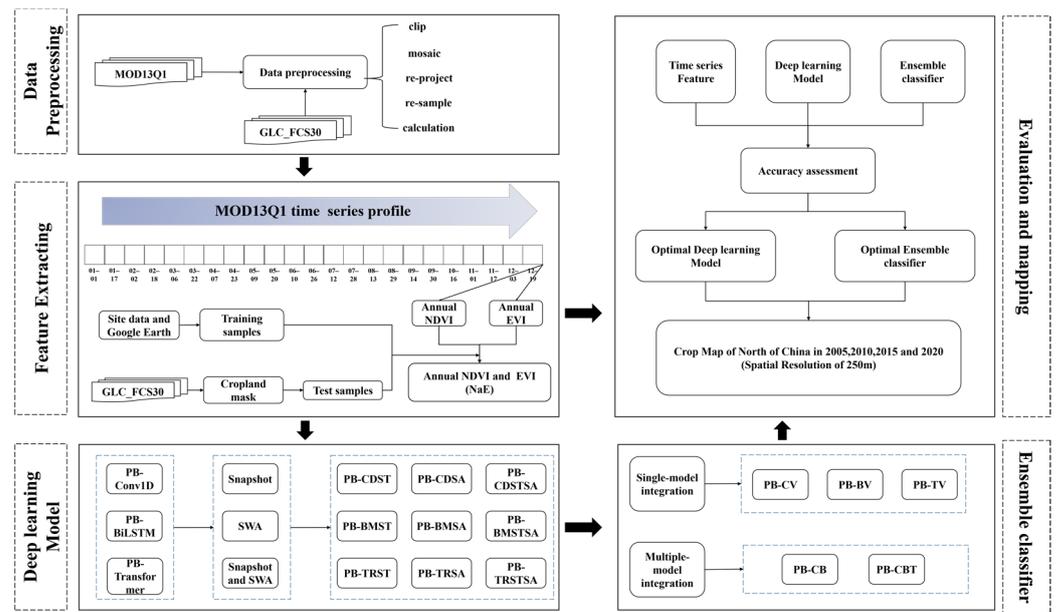


Figure 1. Flowchart illustrating the process of crop cultivated area identification in this study.

2.1. Study Area

In this study, a total of seven provinces—Hebei, Beijing, Tianjin, Henan, Anhui, Shandong, and Jiangsu—were selected as the study area (32°–40° N, 114°–121° E), which is also known as North China (Figure 2).

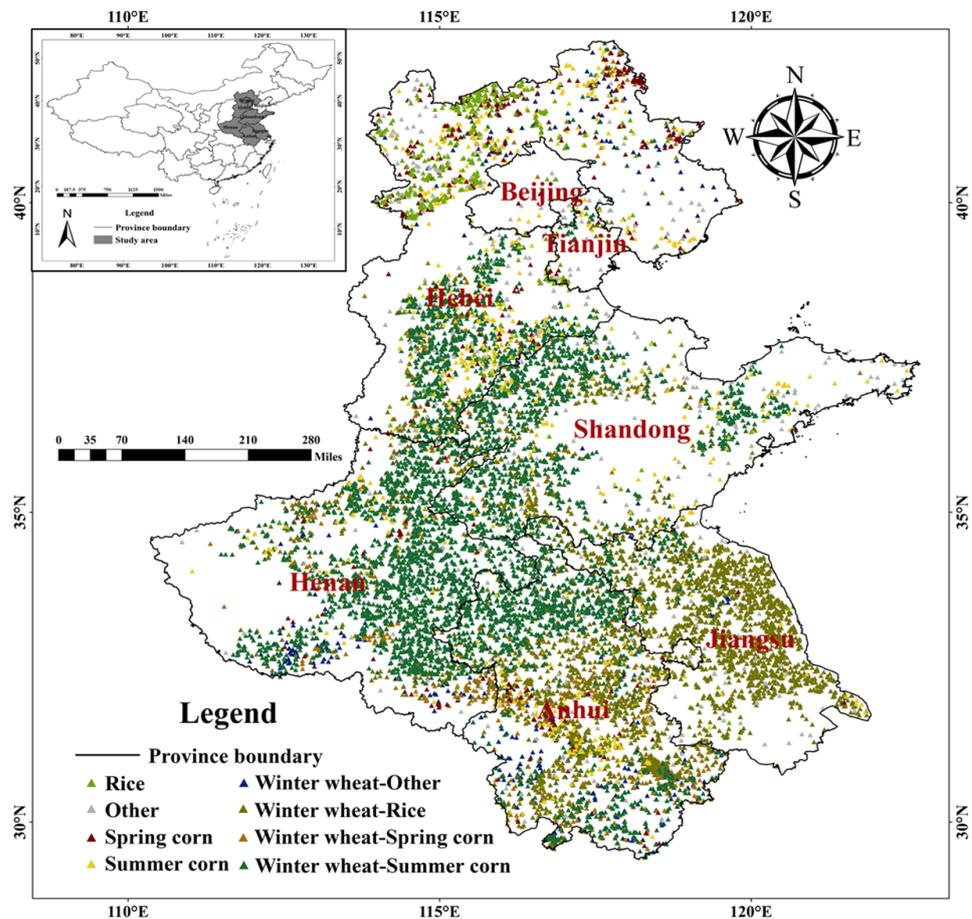


Figure 2. The spatial distribution map and sample points of this study.

The North China Plain is an important agricultural production area in China, mainly planting rice, wheat, corn, etc. The region experiences a temperate monsoon and subtropical monsoon climate, characterized by hot and rainy summers and cold and dry winters, with an annual precipitation range of approximately 500 to 800 mm [7]. Wheat, corn, and rice are the main crops cultivated in North China. For climatic reasons, North China is suitable for producing single and double cropping, primarily winter wheat and other crops in rotation. The growing season of corn is different, comprising May to August and June to September, named spring corn and summer corn, respectively (Figure 3). This study selected the winter wheat–summer corn (WT–SU), winter wheat–spring corn (WT–SP), winter wheat–rice (WT–RI), winter wheat–other (WT–OT), summer corn (SU), spring corn (SP), rice (RI), and other (OT), a total of eight classes for identification.

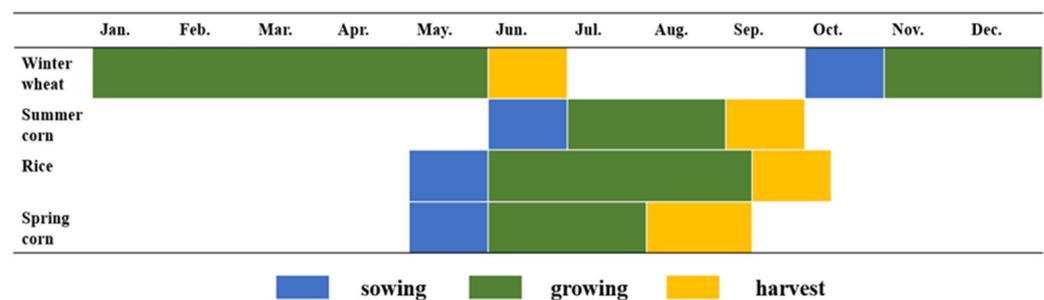


Figure 3. Crop calendars in North China. Blue represents seeding, green represents crop growth and development, and yellow represents maturity. Blank indicates the time of planting for crops that did not fall under any category.

2.2. Data

2.2.1. MODIS Images and Preprocessing

The MODIS onboard the Terra and Aqua satellites were developed by the Earth observation system (EOS) and can capture images with a spatial resolution ranging from 250 to 1000 m and a revisiting period of 16 days [41]. The MOD13Q1 satellite data spatial resolution is 250 m, which provides 23 annual temporal images and contains two VI types, including the NDVI and EVI. The VI datasets were obtained from the Google Earth Engine (GEE) from 2003 to 2015 and 2020, with 323 images used for crop identification.

2.2.2. Reference Samples

In this study, the samples were obtained from China Meteorological Science Data Sharing Service (CMSDSS) (<http://data.cma.cn/>, accessed on 15 March 2023), which includes crop type labels, longitude, and latitude and provides 196 agricultural weather stations, commonly utilized as a reference map for crop identification [22]. Moreover, the team conducted field surveys in Hebei Province in 2012 and 2014, selecting sample sites for the winter wheat–summer corn class and using a handheld Global Positioning System (GPS) to determine the latitude and longitude coordinates of the sample. According to the agricultural station, field survey, and visual interpretation, a total of 12,750 sample pixels were obtained from 2003 to 2015 (Figure 2).

The pixel sample points were randomly divided into three datasets: training, validation, and test, following the ratios of 60%, 20%, and 20% (Table 1). The training datasets were utilized to train the classification algorithms. The optimal hyper-parameters were selected with the validation datasets. The test datasets were used to assess the model's ability to generalize and the performance of the model for unseen data.

Table 1. The classification was conducted using a total of eight categories in this study. The first seven categories corresponded to wheat, corn, and rice, while the remaining category encompassed other crops.

Category Code	Abbreviations	Training Set	Validation Set	Test Set	Total
Summer corn	SU	499	166	167	832
Rice	RI	274	91	92	457
Spring corn	SP	241	80	81	402
Other	OT	764	255	255	1274
Winter wheat-Other	WT-OT	281	94	94	469
Winter wheat-Rice	WT-RI	2221	741	741	3703
Winter wheat-Spring corn	WT-SP	473	158	158	789
Winter wheat-Summer corn	WT-SU	2894	965	965	4824
Total		7647	2550	2553	12,750

2.2.3. Cropland Mask

MODIS data were masked by rained cropland and irrigated cropland in order to exclude interference from non-cultivated areas (i.e., cover, forest, shrubland, grassland, vegetation, wetlands, and others). The cropland mask for the study area was generated using the Global Land Cover with Fine Classification System at 30 m (GLC_FCS30dataset [42]), and the original spatial resolution (30 m) was resampled to 250 m for the years 2005, 2010, 2015, and 2020. The masked images are shown in Figure 4.

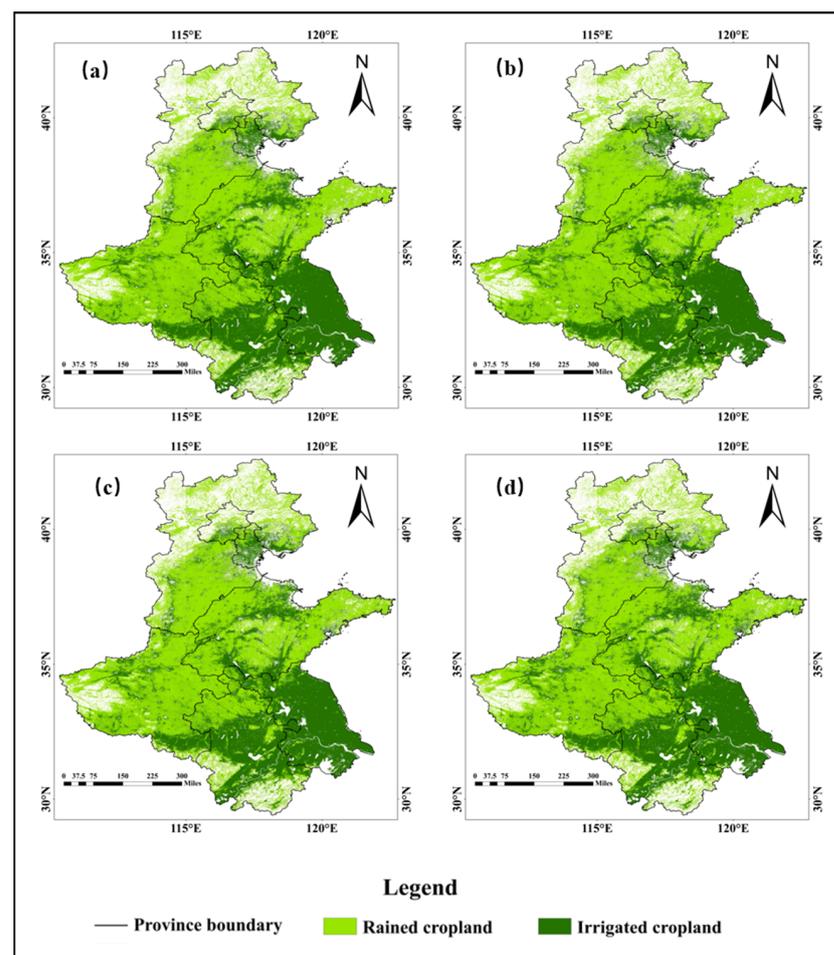


Figure 4. GLC_FCS30 datasets masked by the rained cropland and irrigated cropland in North China and resampled to 250 m in (a) 2005, (b) 2010, (c) 2015, and (d) 2020.

2.3. Methodology

2.3.1. Pixel-Based One-Dimensional Convolutional Neural Network (PB-Conv1D)

The one-dimensional convolutional neural network (Conv1D) effectively extracts sequential features using one-dimensional filters to capture temporal features from feature maps [26]. Convolutional layers can capture the local features in the lower layers' feature map and summarize the global features in the upper layers [43].

The optimization hyper-parameters of the PB-Conv1D classifier were more complex. Due to the characteristic variability, there is no standard method to search for the optimal configuration of hyper-parameters and network structure. In this study, the PB-Conv1D classifier was implemented in combination with Conv1D, Dropout, Batch Normalization, Pooling, and Fully Connected layers. Dropout is an optimization method that can randomly drop some neurons; therefore, the model does not rely on a few neurons, which will avoid over-fitting [44]. Max-pooling layers were consistently used to compress the dimensionality of the feature map and speed up calculation. The last layer is Fully Connected, which collects feature information and determines the output classes. The model hyperparameters are shown in Table 2.

Table 2. PB-Conv1D hyperparameters.

Model Hyperparameters	Hyperparameters Value
Convolutional filters	3, 5, 7
Convolutional channel	32, 64, 128, 256
Convolutional layer	1, 2, 3, 4, 5, 6, 7, 8, 9
dropout	0, 10%, 20%, 30%, 40%, 50%, 60%

The diagram illustrating the architecture of PB-Conv1D is depicted in Figure 5. The model contains four parts. The Input Module represented the time series data. The Convolution Module used convolution to extract features of time series data. The Fully-Connected Module merged temporal features extracted by the Convolution Module. The Output Module used softmax activation to obtain the final crop category. The optimal hyper-parameters were chosen through experimental selection.

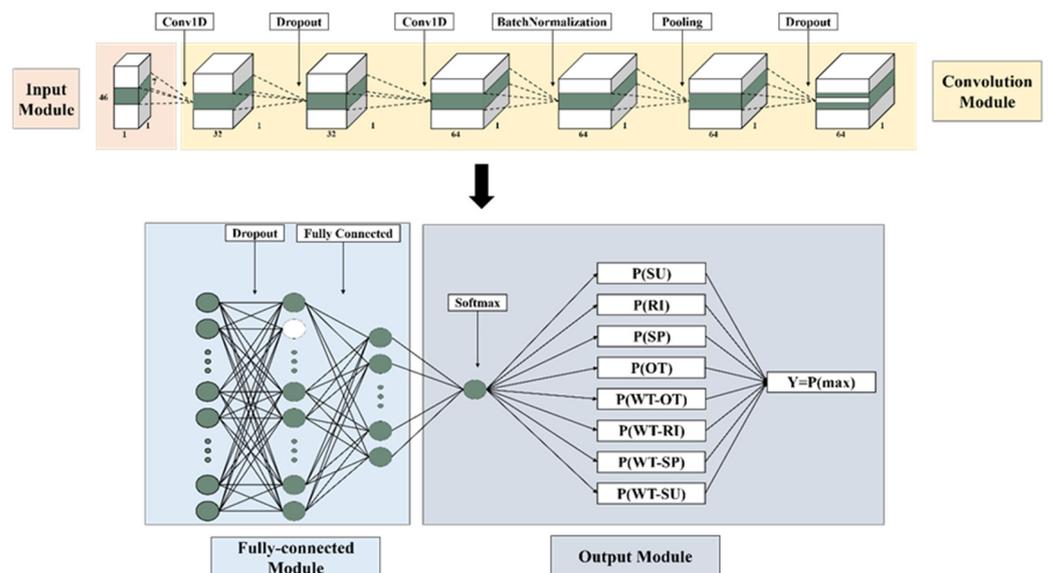


Figure 5. The proposed PB-Conv1D methodology. The green parts of the input and convolution modules symbolic of the path of the sample data flowing through the model, marking the process of data processing. In the Fully-connected layer, the circles represent neurons, where the green circles show neurons in the active state, while the white circles indicate those neurons that were temporarily removed from the network by the dropout mechanism during the training process.

2.3.2. Pixel-Based Bi-Long Short-Term Memory (PB-BiLSTM)

Recurrent neural networks (RNN) are frequently considered a choice for capturing the temporal relationship within time series [45]. However, when the time series data cover an excessively long period, it can result in many model parameters, leading to a slowdown in training speed. Moreover, issues like gradient explosion and disappearance can diminish the information processing capacity of RNN. To solve the phenomenon of “long-term dependence” in RNN, the Long Short-Term Memory (LSTM) model came into being [46]. LSTM improves the memory capacity of the model by introducing forget, input, and output gates [47]. When updating the state of a recursive cell, the input gate determines which new input information can be stored. The forget gate determines which information to discard from the current cell, which allows the forgetting gate to control the influence of historical information on the current memory cell state value. The output gate determines which information is available for output [48]. Due to the addition of the gate mechanism, the LSTM model mitigates gradient explosion and disappearance compared to the RNN model [49]. The operation of LSTM is shown below in Equations (1)–(5).

$$f_t = \sigma(W_{fh}h_{t-1} + W_{fx}X_t + b_f) \quad (1)$$

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}X_t + b_i) \quad (2)$$

$$S_t = f_t \times S_{t-1} + i_t \times \tan h(W_{sh}h_{t-1} + W_{sx}X_t + b_s) \quad (3)$$

$$O_t = \sigma(W_{oh}h_{t-1} + W_{ox}X_t + b_o) \quad (4)$$

$$h_t = O_t \times \tanh(S_t) \quad (5)$$

where f_t , i_t , and O_t represent the values of “forget gate”, “input gate”, and “output gate”, respectively, S_t represents the state value of the memory cell, and h_t represents the value of the hidden layer. σ and \tanh represent the activation functions sigmoid(x) and tanh(x). O_t represents the entire cell’s output value at time t, l represents the current layer, and l – 1 represents the previous layer.

LSTM has a problem in the modeling process, namely that it can only capture the data information relationship from front-to-back and not encode the dependency from back-to-front. In some studies, the current output depends not only on the previous state but also on the future state. BiLSTM extracts high-level abstract temporal features from input sequences hierarchically [33].

The construction process of the PB-BiLSTM model is similar to the PB-Conv1D. The BiLSTM layer is used to extract time-dependent features. To prevent overfitting, each layer of BiLSTM is connected to the Dropout layer, and BatchNormalization is performed to normalize the data after four layers of BiLSTM. Then, partial data is dropped, and the Fully-Connected layer is used to obtain the output results (Figure 6) and the model hyperparameters, as shown in Table 3.

Table 3. PB-BiLSTM hyperparameters.

Model Hyperparameters	Hyperparameters Value
BiLSTM layer	1, 2, 3, 4, 5, 6, 7, 8, 9
Dropout	0, 10%, 20%, 30%, 40%, 50% 60%.

2.3.3. Pixel-Based Transformer (PB-Transformer)

Crop classification uses the encoder part of the Transformer model [35], which is called PB-Transformer. The PB-Transformer model comprises an input module, encoder module, fully connected layer, and output module. The encoder module realizes feature extraction by stacking multiple multi-head attention mechanisms [36,50]. ShortCut was used to obtain better results, and Layer Normalization was used to avoid overfitting. Finally, a fully connected layer is used for crop identification.

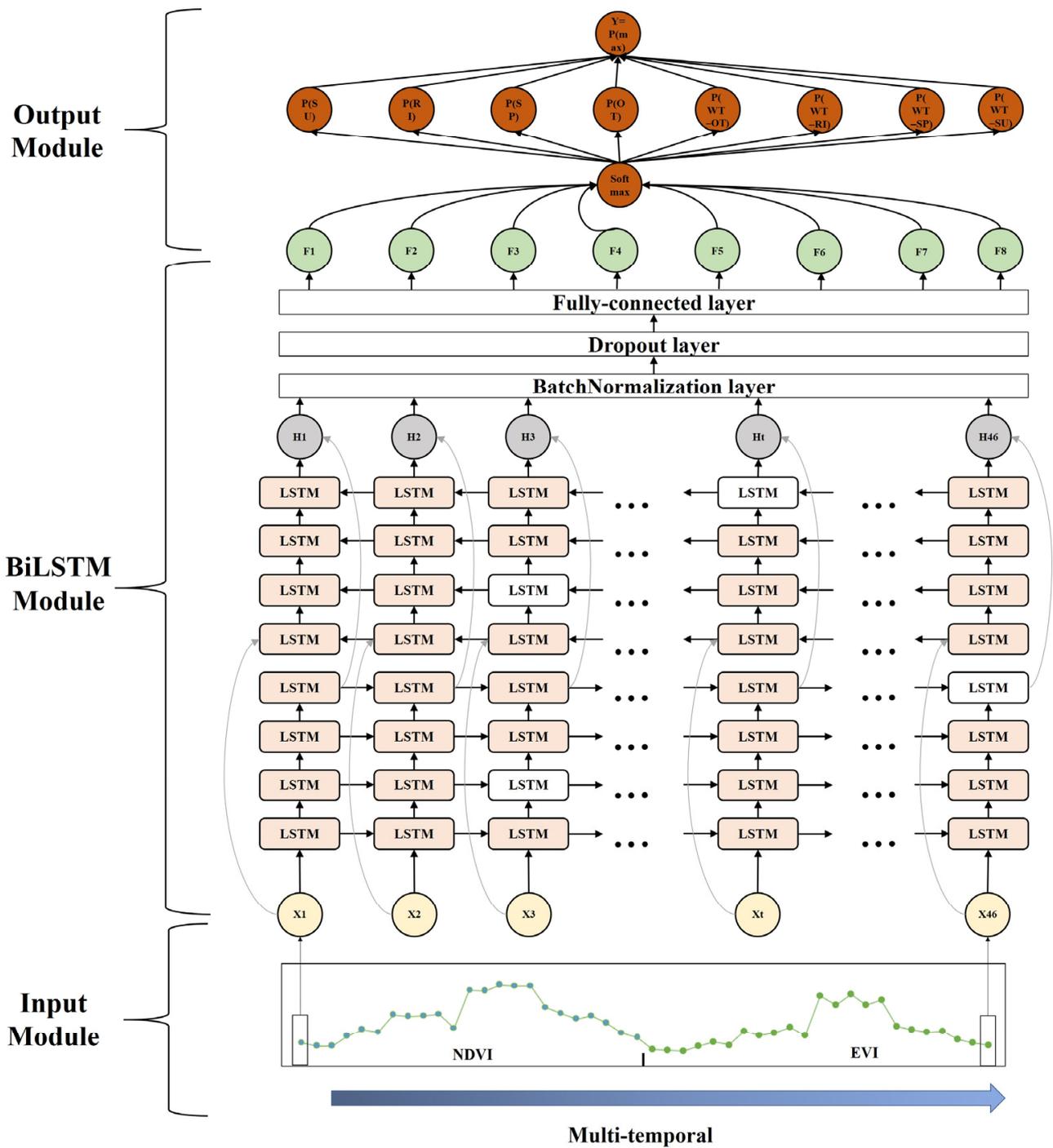


Figure 6. The proposed PB-BiLSTM methodology.

In order to optimize the hyper-parameters of PB-Transformer’s model, we conducted a search to determine the optimal number of self-attention heads from the options 2, 4, 6, and 8, the self-attention layers from the choices of 2, 4, 6, and 8, and the fully-connected layers within the range of 1024 and 2048.

2.3.4. Snapshot Ensemble Cyclic Learning Rate

Snapshot technology is a cyclic learning rate with an annealing strategy. Traditional deep learning methods find the minimum point of the loss function by attenuating the learning rate once. However, due to the irregularity of the loss function, existing techniques

cannot find its minimum [51]. Therefore, a single decay in the learning rate may find a non-local minimum of the loss function or the saddle point, resulting in bad results for the model. Snapshot technology will carry out multiple learning rate decay to avoid this phenomenon. After each learning rate decay is completed and the minimum value of the loss function of this epoch is found, the learning rate will be increased to the original value and will carry out the learning rate decay, which will repeat this process iteratively until the maximum number of iterations is reached.

The traditional learning rate curve saves the corresponding model at the end of each cycle and then makes predictions so that the model can only be reserved once. Because Snapshots are recurrent learning rates, multiple-model training can be performed using single or multiple models to make predictions [37].

$$a(t) = \frac{a_0}{2} \left(\cos \left(\frac{\pi \text{mod}(t-1, [T/M])}{[T/M]} \right) + 1 \right) \quad (6)$$

where $a(t)$ represents the learning rate for the t -th epoch, a_0 is the maximum learning rate, T is the total number of iterations, M is the cycle number, and the square brackets indicate the rounding down operation.

In this study, the PB-Conv1D, PB-BiLSTM, and PB-Transformer models are trained using 0.0001 and 0.0005 learning rates combined with Snapshot technology to obtain a better classification model. Because of the models' differences, the iteration numbers for model training vary; the specific parameters are shown in Table 4.

Table 4. Number of iterations, epochs, and cycles of the learning rate of different models using Snapshot.

Model	Iteration	Epochs	Cycles
PB-Conv1D	600	40	15
PB-BiLSTM	800	80	10
PB-Transformer	800	80	10

2.3.5. Stochastic Weighted Averaging (SWA)

The stochastic weighted averaging (SWA) strategy tends to keep the loss function to a global minimum. It was observed that the local minimum of the loss function obtained at the end of each learning rate cycle tends to accumulate at the region's boundary on the loss space. A local minimum loss can be obtained by averaging these low-loss values. When Izmailov experimented in the training data, the SWA algorithm received a loss value at the 125th iteration, higher than the SGD method, and the loss obtained using the SGD was higher than with the SWA method when performing the 125th iteration in the test dataset [38]. This indicates that SWA will achieve better generalization than traditional training methods and finding the local minimum loss is possible.

In this study, PB-Conv1D, PB-BiLSTM, and PB-Transformer used Adam optimization algorithms to minimize the loss function. Therefore, after 200 training iterations, SWA strategies will be introduced for weight averaging.

2.3.6. Ensemble Classifier

The ensemble learning method integrates the base model to improve performance [52]. In this study, two methods are used to model the ensemble. One is to integrate the single model, and the other is to integrate multiple models.

The single-model and multiple-model integrations are categorized into three parts (Figure 7). The first is training DL (base model). They are then combined with the DL model, and three strategies (Snapshot, SWA, Snapshot + SWA) and the top K , N , and M models are selected. Finally, the top- $K + N + M$ model is randomly integrated to obtain the optimal ensemble model. Multi-model integration is a similar process to single-model integration.

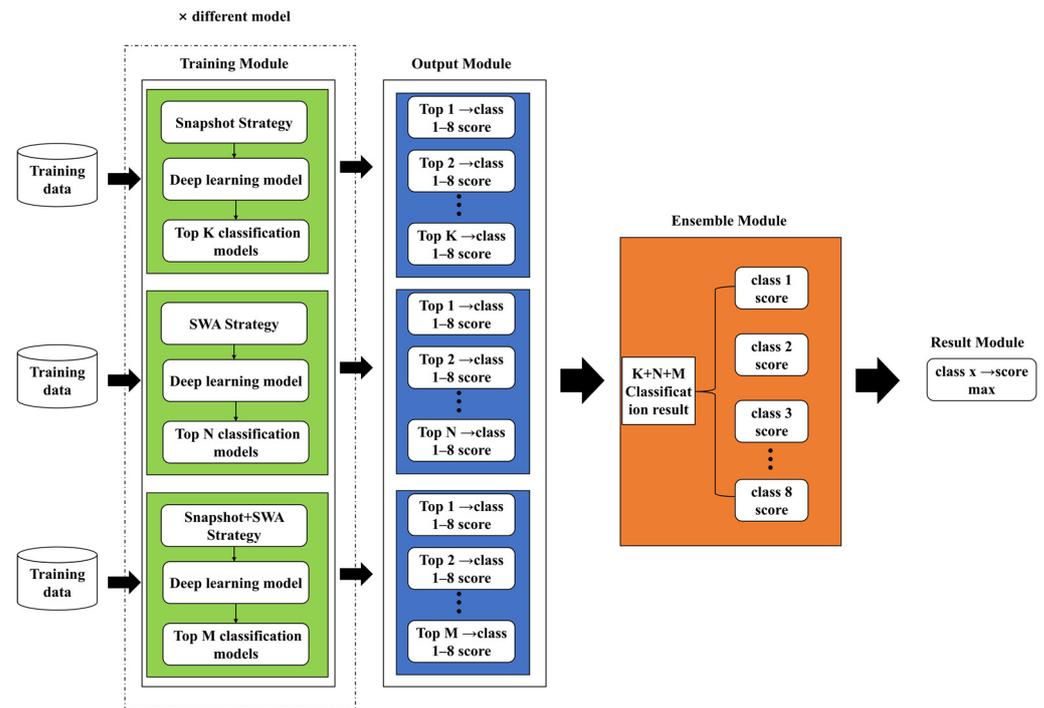


Figure 7. Architecture of the ensemble classifier.

The models were constructed using Python 3.8 with the PyCharm Community 2023.2.1 software for debugging, and the environment configuration is based on Anaconda 23.7.2, where TensorFlow 2.3 is installed as the deep learning framework.

2.3.7. Accuracy Assessment

The spatial identity of the crop map was evaluated with test samples that would calculate the overall accuracy (OA), precision (UA), recall (PA), and Weighted F1 Score [53]. When the sample points are unbalanced, it is not suitable to use the same weight to calculate the F1 Score, and the number of each class should be used as the weight value [17].

The OA indicates the proportion of correctly predicted samples by the model out of all samples. Precision shows the proportion between the number of samples correctly predicted by the model and the total number of samples predicted by the model for a specific category. Recall represents the ratio of correct samples predicted by the model in the actual samples of a certain class. F1 Score reflects the relationship between the precision and recall of the model. The Kappa coefficient signifies the balance of the confusion matrix and is employed for assessing consistency [54]. R^2 and RMSE were used to calculate statistical correlations.

$$\text{overall accuracy} = \frac{TP + FN}{TP + FP + FN + TN} \tag{7}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{9}$$

$$F1 \text{ Score}_{\text{class}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{10}$$

$$\text{Kappa coefficient} = \frac{P_o - P_e}{1 - P_e} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2} \tag{12}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \tag{13}$$

The meanings of the parameters in Table 5 are shown below: TP indicates the correct classification of the positive class, FN represents the misclassification of the positive class as the negative class, FP signifies the misclassification of the negative class as the positive class, and TN indicates the correct classification of the negative class. P_0 represents the overall accuracy, and P_e is the ratio of the total number of cells in the actual cluster multiplied by the total number of cells on the diagonal of the confusion matrix to the square of the number of samples. y_i indicates true value, \bar{y}_i denotes the average of true values, and \hat{y}_i represents predicted value.

Table 5. Confusion matrix diagram illustrating the predicted and actual classifications.

		Predict	
		Positive	Negative
Actual	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

3. Results

3.1. Evaluate a Single Classification Model

As the PB-Conv1D had the shortest running time of the three models, the best analysis of its hyperparameters, and the best results generated during the procedure. Other models were not repeated. The PB-Conv1D model and PB-Conv1D combine different strategies and have been experimented with several times; the model’s parameters are shown in Table 6. Although the foundational structure of the base model remains consistent, the parameters of the models for different strategies are different.

Table 6. The learning rate and dropout parameters of the optimal model obtained based on the PB-Conv1D.

Classifier	Dropout/%	Learning Rate
PB-Conv1D	40	0.0001
PB-Conv1D combine Snapshot (PB-CDST)	40	0.0005
PB-Conv1D combine SWA (PB-CDSA)	40	0.0005
PB-Conv1D combine Snapshot and SWA (PB-CDSTSA)	10	0.0005

Figure 8 shows the accuracy and loss function changes of different methods. The PB-Conv1D model added an early stopping strategy when the loss value of multiple iterations did not decrease during the training process, stopping the model training, and the model was iterated 140 times. The number of iterations for the other methods was fixed. Figure 8(Ac,Bc) shows that PB-CDSA achieved a high accuracy of nearly 95%, with a decreasing loss function on the training set; however, the accuracy was slightly above 80%, and the loss function showed a decreasing and then increasing pattern, which indicated the limited generalization ability of the model. The accuracy function of PB-CDST was cyclical because it used the cycle learning rate, and the learning rate would be relearned from the initial learning rate at the end of each cycle. Combining A and B, the method’s accuracy decreased, and the loss value increased in the validation set. However, each model would be saved after 40 iteration cycles, which obtained models with a better generalization ability.

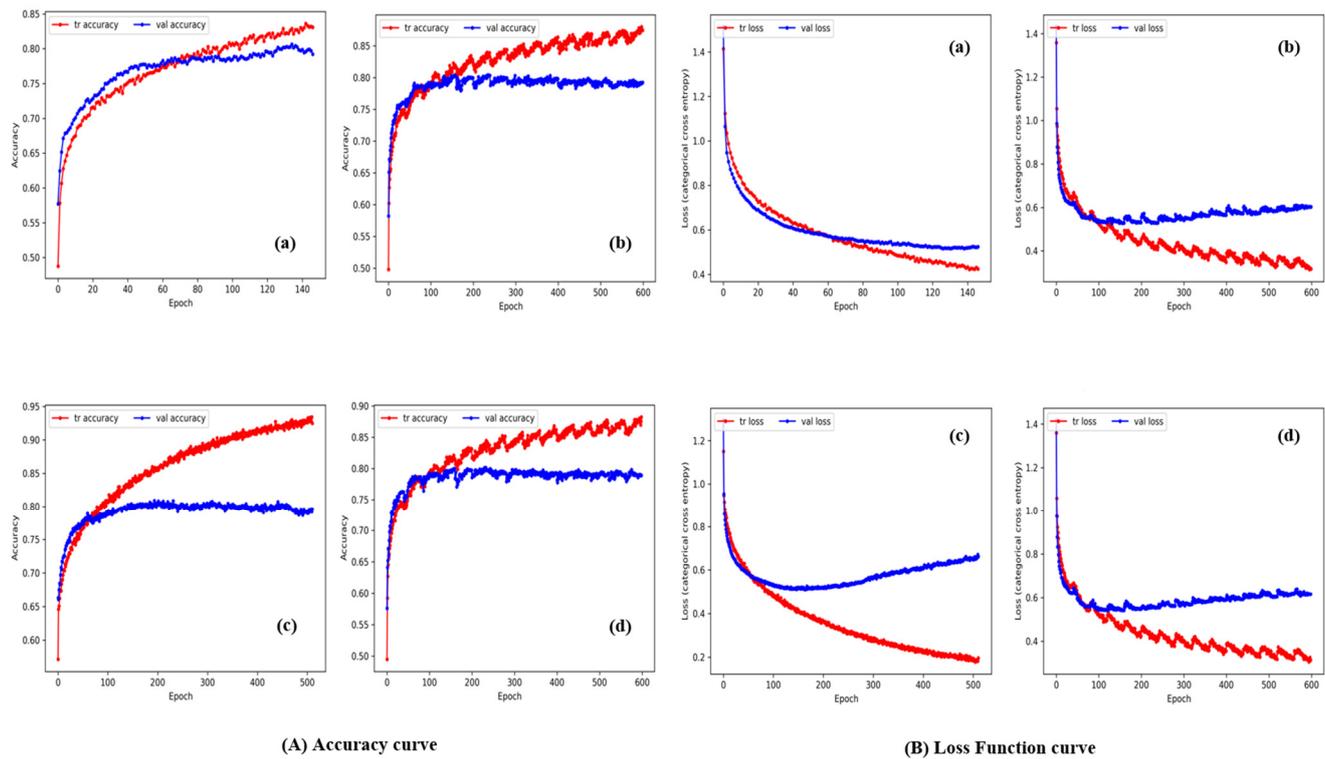


Figure 8. Different strategies obtained the (A) accuracy curve and the (B) loss function curve: (a) PB-Conv1D (b) PB-CDST (c) PB-CDSA (d) PB-CDSTSA.

Figure 9 illustrates the effectiveness of the combined strategy compared to the base-model. Among the three models, two achieved optimal results after incorporating the Snapshot strategy, while the third model combined both the Snapshot and SWA strategies to achieve the best results. With the combination of Table A1, the results show that the PB-CDST method had the higher OA among all PB-Conv1D-based models, with an improvement of 0.79%. PB-BMST achieved the best results, resulting in an accuracy improvement of 1.81% compared to the other PB-BiLSTM-based algorithm. The PB-Transformer combined SWA (PB-TRSA) method could not significantly improve the classification accuracy, with an accuracy decrease of 5.99%. Compared with the PB-Transformer method, PB-TRSTSA improved the accuracy by 0.63%. The optimal hyperparameter results for PB-BiLSTM and PB-Transformer are shown in Table 7.

Table 7. Optimal hyperparameters of the PB-BiLSTM and PB-Transformer model.

Classifier	Dropout/%	Learning Rate
PB-BiLSTM combine Snapshot (PB-BMST)	10	0.0005
PB-Transformer combine Snapshot and SWA (PB-TRSTSA)	60	0.0005

Comparing the metrics of PB-Conv1D, PB-BiLSTM, and PB-Transformer, the PB-Conv1D model is more effective for crop identification. According to the folded line of Figure 9, it is concluded that the accuracy of the PB-Conv1D model fluctuates less after the integration of the strategy, which means that the parameters of PB-Conv1D change slightly in the Loss function space without causing significant improvement in the model. The folded line of PB-BiLSTM showed that the accuracy fluctuates slightly after adopting the strategy. The more extensive floating folded line of the PB-Transformer indicates that the PB-TRSA has severe deviations in the loss space, with the parameters obtaining inferior loss values on average.

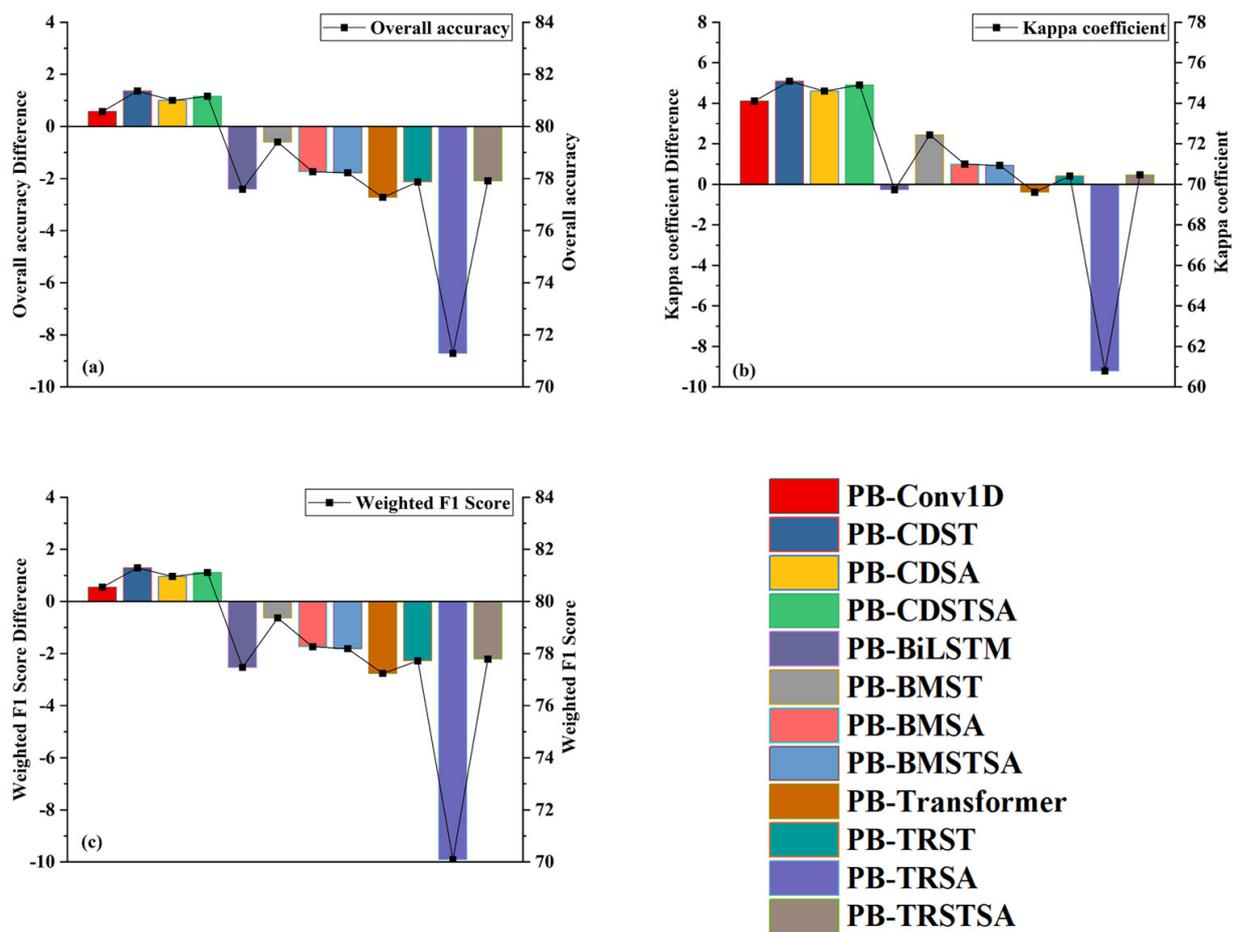


Figure 9. The primary Y-axis represents the performance metrics for the various models, which include (a) overall accuracy (at 80%), (b) Kappa coefficient (at 0.7), and (c) weighted F1 score (at 0.8); negative variances indicate that the performance metrics are below the baseline values. The secondary Y-axis indicates the actual value of the performance metric. The horizontal axis indicates the model type.

Figure 10 indicates that the crop recognition from different methods of the PB-Conv1D model was relatively stabilized. PB-BiLSTM enhanced the recognition of OT after the adoption of the strategy. PB-Transformer significantly reduced the recognizability of WT-OT and WT-SP after the addition of the SWA strategy. This indicates that the strategy is ineffective in classifying these categories, with the minimum loss value to be found when the parameters are averaged. Further analysis of Figure 10 shows that the WT-SU category had a more concentrated distribution of all models, which means that the three models could distinguish it better. WT-SU was misclassified as WT-SP, whereas SU, WT-SP, WT-RI, and WT-OT were misclassified as WT-SU. WT-SU and WT-RI are rotation crops; due to the similar growth cycles of SU and RI, they are easily confused. WT-SP can be easily misclassified as WT-SU due to the similarity between summer corn’s and spring corn’s growth curves. WT-OT contains information on several crops, such as cotton and soybean, which are difficult to distinguish from WT-SU since there are similarities between the phenological information of these crops. The distinguishing ability of the three models for SU was weak. SU was misclassified into OT and WT-OT for many sample points due to the multiple information contained in OT sample points. The model had the insufficient ability to distinguish this information. More details are given in Table A1.

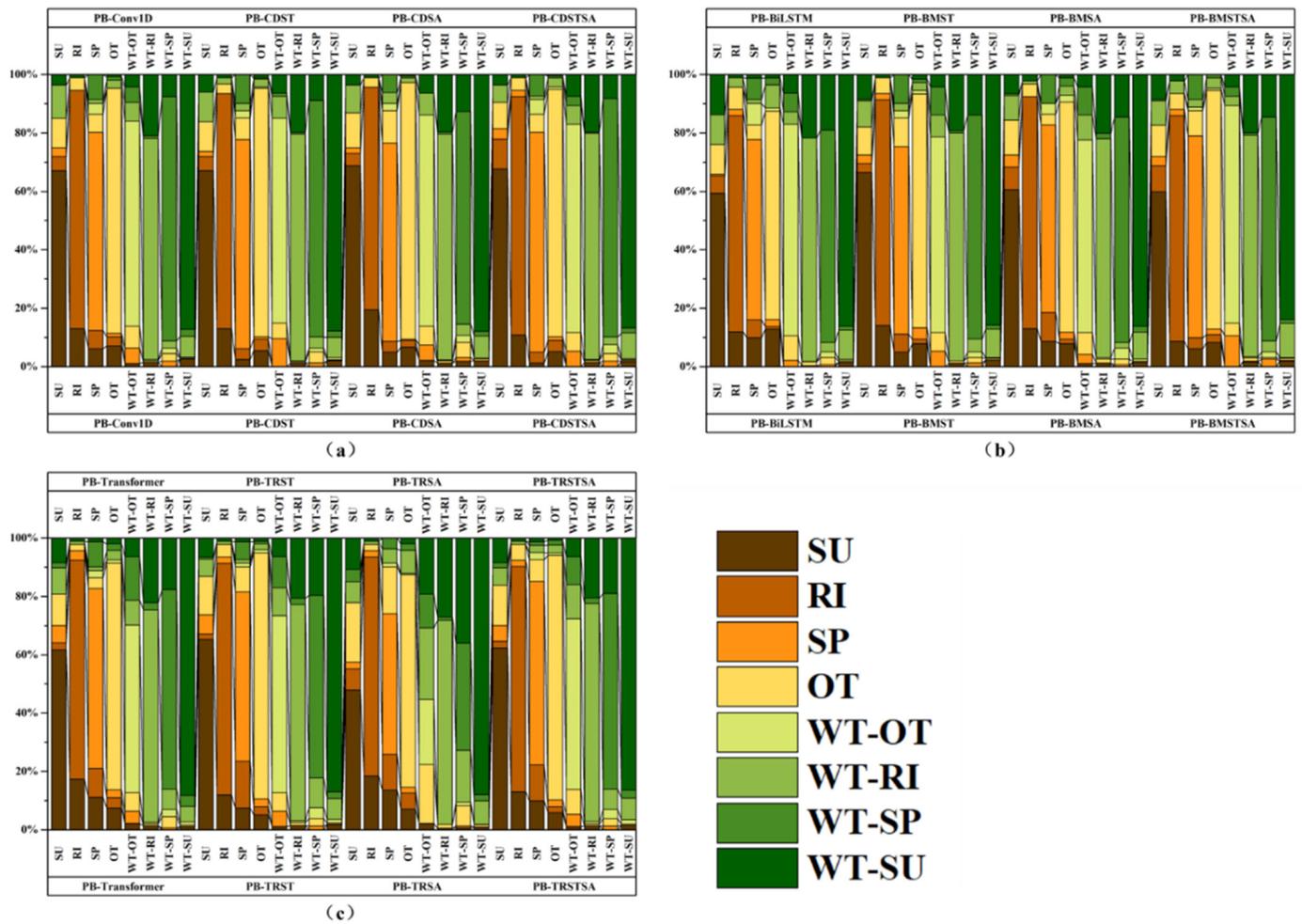


Figure 10. The confusion proportions for each class in (a) PB-Conv1D, (b) PB-BiLSTM, and (c) PB-Transformer.

3.2. Evaluation of the Ensemble Classifier

In this study, to obtain a more effective classification model, original models and voting algorithms were combined to obtain a better ensemble algorithm. The top-k models with better performance than the individual models were integrated to obtain the final ensemble model. The results after many experiments are shown in Table 8.

Table 8. Description of integrated model construction.

Model Name	Description	k-Value
PB-CV	The top K best-performing PB-Conv1D, PB-CDST, PB-CDSA, and PB-CDSTSA models were voting	5
PB-BV	The top K best-performing PB-BiLSTM, PB-BMST, PB-BMSA, and PB-BMSTSA models were voting	7
PB-TV	The top K best-performing PB-Transformer, PB-TRST, PB-TRSA, and PB-TRSTSA models were voting	3
PB-CB	The PB-CV and PB-BV were voting	12
PB-CBT	The PB-CV, PB-BV, and PB-TV were voting	15

Figure 11 shows that the classification effect of the PB-BV algorithm was best and the PB-TV algorithm was worst in single-model integration. When the single model was classified, the performance of the PB-CDST method was excellent, with its accuracy increased by 0.74% after integration. The performance of the TB-TRST was less than the

PB-CDST method, with a difference of 1.96%, and the accuracy of the model increased by 2.84% after integrating PB-BiLSTM, which is a considerable improvement, exceeding the PB-CV model. Compared with the best-performing single model ensemble (PB-BV), the PB-CBT model's accuracy improved by 1.21%. The performance of PB-CBT (83.47%) was slightly higher than that of the PB-CB (83.43%) classifier. Although the PB-Transformer model was added to PB-CBT, the accuracy of the PB-CBT did not significantly improve, and the running time of the PB-CBT was greatly increased. This underscores the fact that not integrating all models will improve the classification effect and may even increase the classification time. The PB-CB model can be used as a candidate algorithm for classification studies in subsequent experiments.

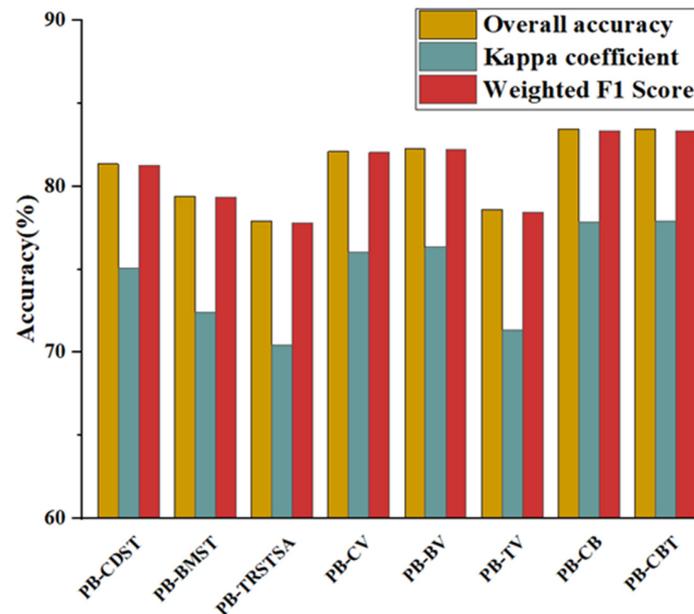


Figure 11. Comparison of the results of each classifier.

Figure 12 shows that the PB-TV model obtained inferior classification results for each crop type. Since the PB-Transformer model had the worst discrimination, the integrated learning could not modify the temporal information from the model. Compared with the PB-CDST (Figure 10), PB-CV improved the discrimination of SU. PB-BV enhanced the recognition of WT-OT. PB-CB promoted the identification of RI, WT-OT, WT-RI and WT-SP. Compared to PB-CB, PB-CBT improved the recognition of OT, WT-RI, and WT-SU, illustrating that PB-Transformer can effectively discriminate among three categories. The results for SP and OT were better in a single classifier, indicating that model integration may have decreased the recognition effect of some classes.

With the combination of Table A1, the F1 score of SU did not exceed 70% after model integration. It indicated that the models performed weakly for SU. Analyzing Figure 11, the F1 scores obtained by integrating the models were significantly improved and far higher than the highest F1 scores of single models. The PB-CV model was superior to the PB-BV model in identifying SU, RI, and SP. The PB-BV model was more effective in identifying OT and WT-OT than the PB-CV model. For other rotational crops, both were identified effectively, with F1 scores above 80%. The PB-BV model made it easier to confuse RI and SP, and the PB-CV model made it easier to classify WT-OT into SP, WT-RI, and WT-SU, which made the results consistent with those obtained from the analysis of the single model (Figure 10). This indicates that ensemble learning is a combination of results from simple voting and without feature learning. Further observation of Figure A1m,n reveals that the PB-CV model demonstrated better recognition on most non-rotation crops along the diagonal of the confusion matrix, while the PB-BV model exhibited a better recognition effect on most rotational crops.

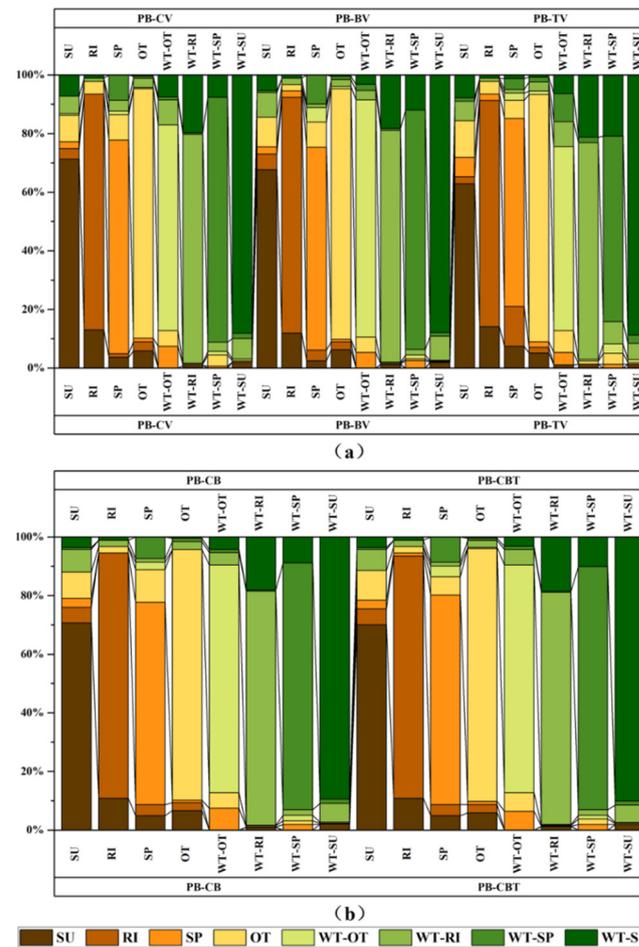


Figure 12. The confusion proportions for each class in (a) PB-CV, PB-BV, PB-TV, (b) PB-CB, and PB-CBT.

Figure 12 showed that the PB-CB and PB-CBT models had inferior classification abilities for SU and SP and superior classification abilities for the other categories. SU misclassified as OT as WT-RI and WT-SU, RI, SP, and OT misclassified as SU produced more samples, which is the same situation as in other experiments. It means that the results obtained from the multiple model ensembles are consistent with the single model and that the accuracy will be slightly changed by integrating the crop classes with a worse classification effect into the single model while the accuracy of the crop classes with a better classification effect will be significantly improved by integrating them. For non-rotational crops, F1 scores decreased with both the PB-CB and PB-CBT models compared to the PB-CV model and increased with the PB-BV model (Table A1). Since the number of PB-Conv1D and PB-BiLSTM models was seven and five, respectively, PB-BiLSTM showed numerical advantages for integration, and the model results were similar to those with the PB-BV model. By observing the F1 scores for each crop in PB-CB and PB-CBT, it is found that integration with the PB-Transformer model caused the F1 scores for some classes to decrease, indicating that the PB-Transformer model was ineffective in identifying these classes.

3.3. Multiple-Crop Classification Map

Through an analysis of the spatial distribution of crops in the study area from 2005 to 2020, Figure 13 suggests that there has been minimal variation in cropping patterns within the study area over the span of 15 years. The main planting patterns in the middle and

south of the study area were WT-SU and WT-RI. The main planting patterns in the north and east of the study area were single cropping, mainly spring corn and summer corn.

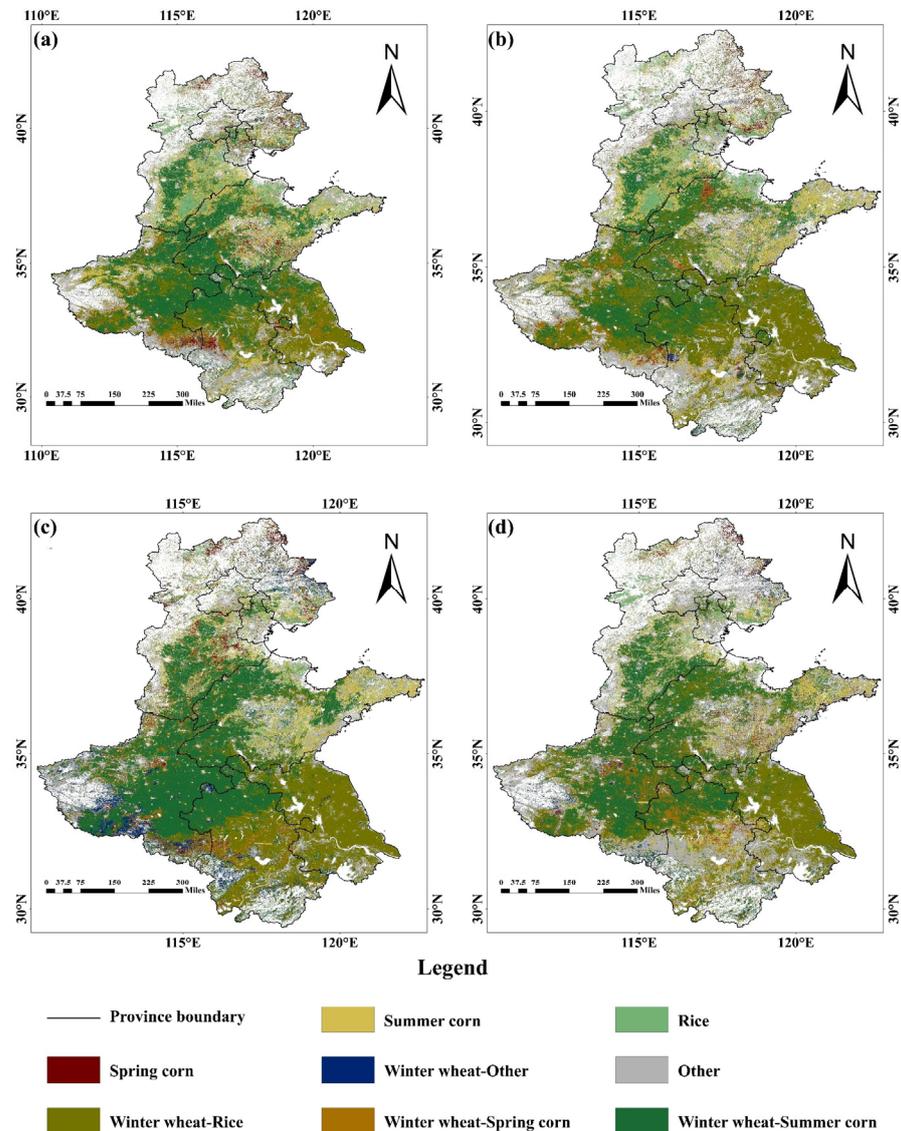


Figure 13. Crop spatial distribution map in North China using PB-CB in 2005 (a), 2010 (b), 2015 (c), 2020 (d).

From observation, WT-RI planting areas in Anhui Province showed a downward trend from 2015 to 2020, consistent with the results of the National Statistical Yearbook. In 2020, Anhui Province suffered from floods, which caused severe damage to crops and decreased RI by 6632.133 km². As a result, large areas in the crop distribution map are OT, reflecting the accuracy of the crop planting map sideways.

3.4. Accuracy Assessment with Statistical Data

The crop cultivation areas obtained from the MODIS data were summarized at the municipal level. The accuracy assessment was conducted by calculating the R^2 and RMSE using municipal statistical data. The area of wheat, corn, and rice detected from MODIS images using the PB-CB algorithm demonstrated a positive correlation with municipal statistics. The R^2 for wheat was 0.76, with an associated RMSE of 83.82 Kha; for corn, the R^2 was 0.72, accompanied by an RMSE of 72.83 Kha; and for rice, the R^2 was 0.70, with an RMSE of 69.64 Kha, while considering all data pairs for 2005, 2010, 2015, and 2020, as shown

in Figure 14A. These results show excellent agreement between the area data derived from MODIS and the statistical data. To further analyze the results, the crop cultivation area for 2015 derived from the PB-CB algorithm was compared with county-level statistics and still showed a positive correlation. The R^2 reached 0.64 and RMSE of 14.60 Kha for wheat, R^2 attained 0.63 and RMSE of 14.01 Kha for corn, and R^2 reached 0.65 and RMSE of 12.92 Kha for rice (Figure 14B). This further confirms the feasibility of using the PB-CB model combined with NaE data for crop cultivation area extraction.

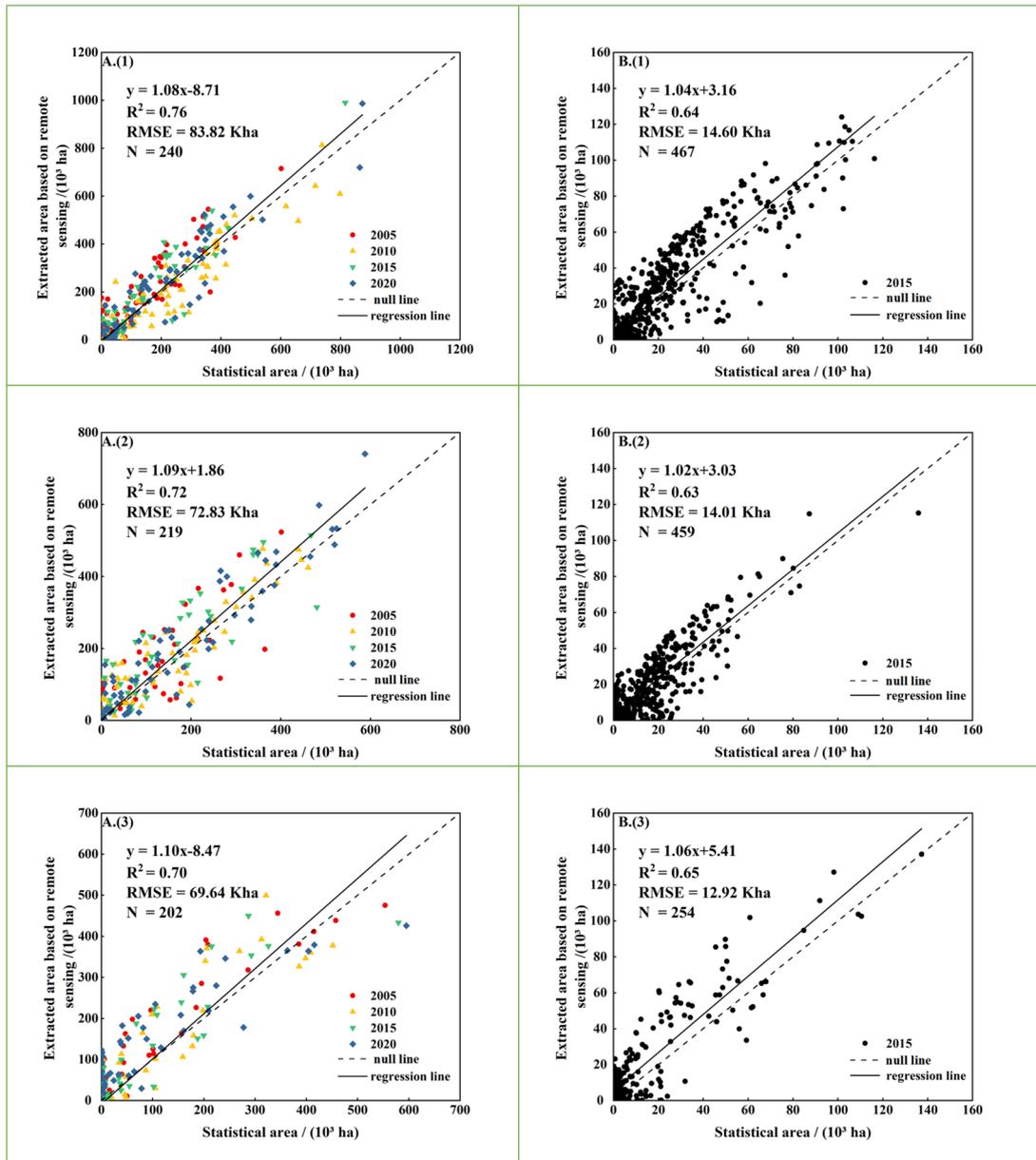


Figure 14. (A) A municipal-level comparison of (1) wheat, (2) corn, and (3) rice cultivation areas detected from the MODIS data in 2005, 2010, 2015, and 2020 combining the PB-CB algorithm. (B) County-level comparisons of (1) wheat, (2) corn, and (3) rice planted detected from the MODIS data in 2015 combining the PB-CB algorithm.

4. Discussion

4.1. Advantage of the Ensemble Model

Timely and accurate crop spatial distribution and yield estimation monitoring are significant to national food security and strategic arrangements [55]. Deep learning algorithms

capable of mapping multiple crops’ spatial distribution have proven effective in recent years [15,17]. However, most methods for extracting multiple-crop data in North China have lower accuracy. Xun et al. [56] used sparse-representation-based algorithms combined with the Leaf Area Index (LAI) to classify multiple crops in North China, and the final accuracy reached 76.05%. Wang et al. [17] combined NaE and Stacking (SVM, RF, KNN) for multiple-crop classification in North China, achieving an overall accuracy of 77.12%. Mapping the crop area in regions with a large and multiple-crop type remains challenging. The development of a method capable of identifying multiple-crop types within a large geographical area is in demand.

To this end, in this study a new PB-CB network is proposed, combining it with time series NDVI and EVI (NaE) data to generate crop planting distribution maps. The results show that PB-CB can effectively conduct crop classification research. Compared to existing methods, the PB-CB method offers several advantages. Firstly, deep learning models follow the end-to-end principle without needing expert experience and prior knowledge, which can automatically obtain crop results [20]. Secondly, Snapshot and SWA strategies combined with PB-Conv1D and PB-BiLSTM models can simplify the feature extraction process and obtain higher accuracy. Finally, combining multiple models is equivalent to fusing numerous features, resulting in better feature results, which is particularly well-suited for crop classification research.

To better demonstrate the advantages of the PB-CB model, we conducted a comparison between the PB-CB model and two other models: the PB-CDSA classification model, which demonstrated the best performance in the single model, and the PB-BV model, which achieved the highest performance among the single model ensemble. Table 9 shows the difference between the confusion matrices of the PB-CB and PB-CDSA models, with positive values on the diagonal and negative values on the non-diagonal, indicating the number of samples correctly classified by the PB-CB model than PB-CDSA. The total showed that the PB-CDSA model outperformed PB-CB only in classifying SP. The PB-CB model indicated more vital classification ability when dividing rotation crops, especially WT-RI and WT-SU. By calculation, the number of samples correctly classified by the PB-CB than the PB-CDSA model accounted for 4.15% of the total test samples, which is enough to illustrate the advantages of the PB-CB model.

Table 9. Difference between the confusion matrix of the PB-CB classifier and PB-CDSA classifier. Values are calculated as Figure A1p minus Figure A1b.

Reference Classes	Classified								Total
	SU	RI	SP	OT	WT-OT	WT-RI	WT-SP	WT-SU	
SU	6	1	2	−2	0	−4	1	−4	12
RI	−2	3	0	−1	0	0	0	0	6
SP	2	0	−2	3	0	−1	−2	0	−4
OT	3	−3	0	1	−1	0	2	−2	2
WT-OT	0	0	−2	0	7	−3	0	−2	14
WT-RI	−2	1	0	0	−2	17	−1	−13	34
WT-SP	0	0	1	−4	1	−3	5	0	10
WT-SU	0	0	0	0	−4	−5	−7	16	32
Total	5	4	−3	5	13	33	12	37	106

Table 10 shows the difference between the confusion matrix of the PB-CB model and that of the PB-BV model, and the meaning is consistent with Table 7. The total indicates that the PB-BV model only showed a stronger classification ability than PB-CB when classifying WT-OT. This is because the PB-BV model had a better effect on WT-OT, while the PB-CB model integrates the PB-BV and PB-CV models, and the PB-CV model had a weak classification ability for WT-OT than integrated PB-BV and PB-CV, which reduced the classification ability of the PB-BV model. The two models exhibit similar classification capabilities for SP and OT. When dividing WT-SU, the PB-CB model showed a strong

classification ability. According to calculations, the number of samples correctly classified by the PB-CB model compared to the PB-BV model accounted for 2.35% of the test samples, indicating that PB-CB has certain advantages over PB-BV.

Table 10. Difference between the confusion matrix of the PB-CB classifier and that of the PB-BV classifier. Values are calculated as Figure A1p minus Figure A1n.

Reference Classes	Classified								Total
	SU	RI	SP	OT	WT-OT	WT-RI	WT-SP	WT-SU	
SU	5	0	1	−2	0	−1	0	−3	10
RI	−1	3	−2	0	0	0	0	0	6
SP	2	0	0	2	−2	0	−2	0	0
OT	1	0	0	0	−2	1	0	0	0
WT-OT	0	0	2	0	−3	1	−1	1	−6
WT-RI	−2	0	0	0	−1	6	−2	−1	12
WT-SP	0	0	−1	1	1	0	4	−5	8
WT-SU	1	−1	0	0	1	−18	2	15	30
Total	4	4	0	−1	0	23	7	23	60

4.2. Uncertainty and Potential Refinements

The PB-CB method used in this study has dramatically improved the results of previous studies. However, certain factors and uncertainties still continue to influence the accuracy of the mapping results. Firstly, GLC_FS30 is used as auxiliary data for land mask processing and has some errors that may lead to propagation errors in the final crop classification map. In addition, the MOD13Q1 data has a spatial resolution of 250 m, which will cause the problem of mixed pixels and lead to errors in crop recognition.

Although the work of this study is carried out in North China, the methods used in this study can also be applied in other regions. Several improved ways may be more applicable to other crop identification. First of all, time series was used for crop identification, and these data still face the challenge of intra-class variation. Multi-source data fusion can be used to classify crops [57]. Crops are affected not only by phenology but also by longitude, dimension, sowing time, and climate change [58,59]. Therefore, the phenological, longitude, and latitude information can be adopted in the recognition algorithm of crop area [60]. In addition, medium spatial resolution data have limited acquisition and coverage capabilities, and rough spatial resolution, such as MODIS, still has some problems in mixed-pixel recognition [61–64]. In future research, higher spatial resolution data can be used for the crop recognition.

5. Conclusions

Accurate and timely crop distribution access is important to food production management. This study proposed a new method for crop identification by using multi-temporal MODIS remote sensing images in North China. Firstly, the time series were combined as inputs to the PB-Con1D and PB-BiLSTM models for producing the original classification results. Then, the Snapshot and SWA strategies were introduced in the deep learning models to gradually improve crop classification accuracy and obtain the best-performing PB-CDST (OA: 81.36%) and PB-BMST (OA: 79.37%) models. They were compared with the PB-TRSTSA (OA: 77.79%) model to show the advantages of our proposed models for multiple-crop classification. Subsequently, the better-performing PB-Conv1D and PB-BiLSTM were combined with ensemble learning to obtain the PB-CB model. The proposed PB-CB (OA: 83.43%) algorithm combines the temporal features learned from the PB-Conv1D and PB-BiLSTM models to classify multiple-crop areas, ensuring classification accuracy. Additionally, the results of the correlation of multiple-crop cultivated areas generated by MODIS combined with the PB-CB method to statistical data showed that the R^2 of wheat, corn, and rice exceeded 0.7 and 0.6 at the municipal and county levels, respectively. Therefore, the newly presented PB-CB model is an effective method for multiple-crop area

identification from time series MODIS remotely sensed imagery. Meanwhile, the PB-CB model is available for other crops (e.g., soybean and cotton) and consequently has strong prospects for application. Future work can investigate the method’s feasibility for the mapping of other crops and the production of higher-spatial-resolution images.

Author Contributions: X.W. (Xue Wang): Methodology, Investigation, Software, Data curation, Validation, Writing—original draft. J.Z.: Conceptualization, Investigation, Validation, Resources, Supervision, Project administration, Funding acquisition, Writing—review & editing. X.W. (Xiaopeng Wang): Software, Investigation, Data curation, Visualization, Writing—review & editing. Z.W.: Software, Investigation, Data curation, Visualization, Writing—review & editing. F.A.P.: Investigation, Visualization, Writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was jointly supported by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (No. KF-2021-06-081), the Finance Science and Technology Project of Hainan Province (No. ZDYF2021SHFZ063), the Shandong Key Research and Development Project (No. 2018GNC110025, ZR2017ZB0422), The 2023 Zhonglou District of Changzhou City Science and Technology Research Project (No. JBGS2023011), and the Science and Technology Support Plan for Youth Innovation of Colleges and Universities of Shandong Province of China (No. 20235KJ232).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Acknowledgments: We thank Shawkat Ali and Hidayat Ullah for editing English text.

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

Appendix A. Confusion Matrix

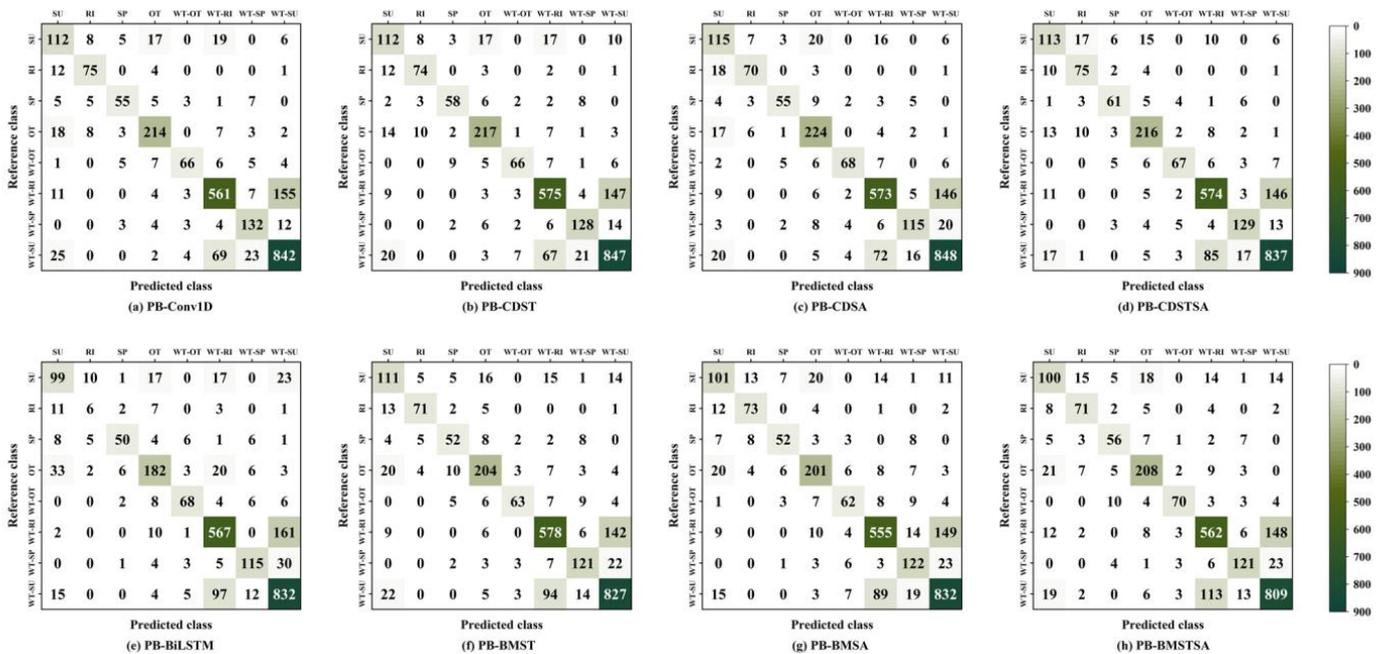


Figure A1. Cont.

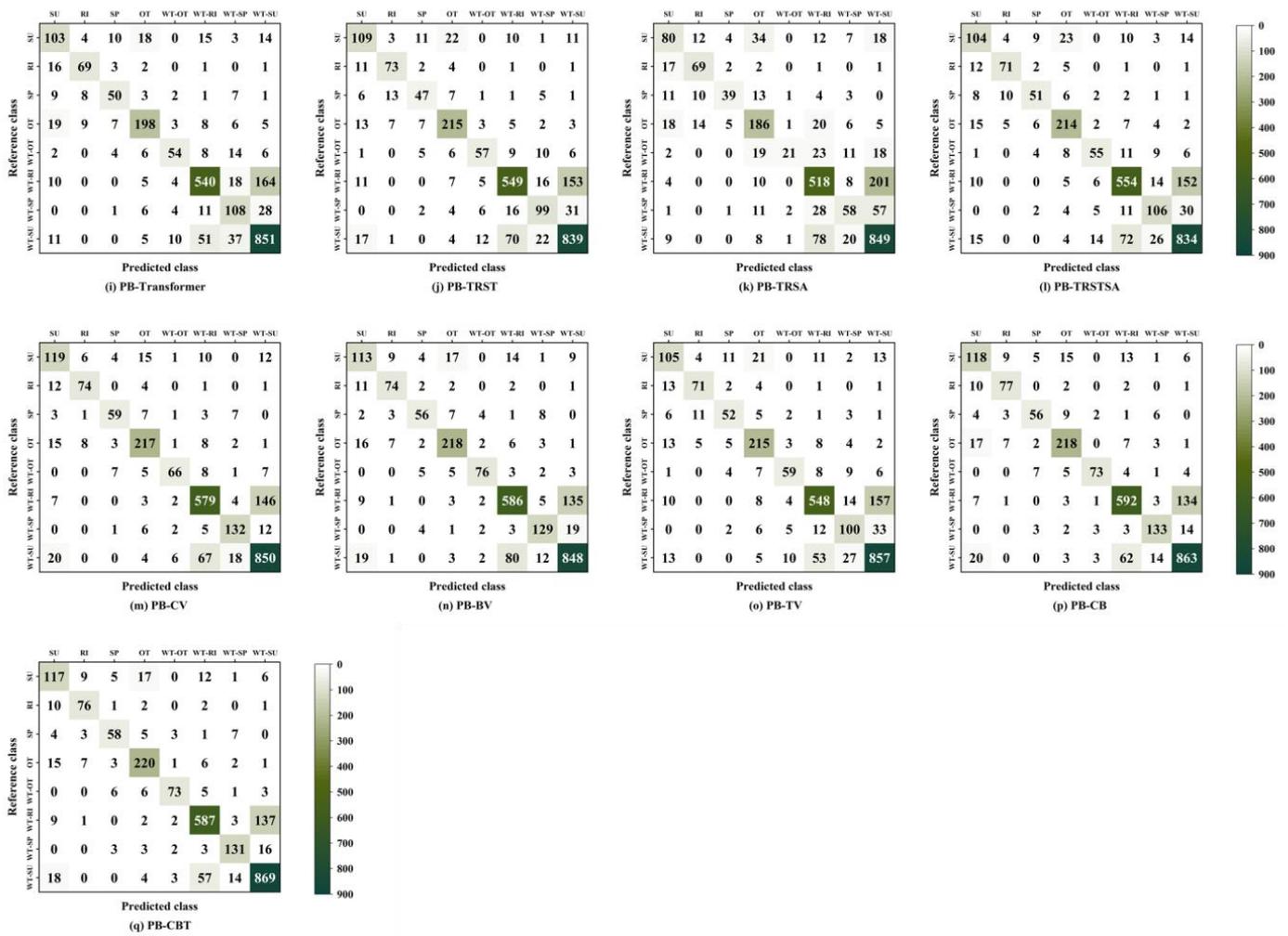


Figure A1. Confusion matrix performance of different classifiers in the test dataset.

Appendix B. Performance Matrix of Various Model

Table A1. The performance of the different classifiers in the test set, including the OA, Kappa coefficient, weighted average F1 score for model with PA, UA and F1 scores for each category.

	Unit/%	SU	RI	SP	OT	WT-OT	WT-RI	WT-SP	WT-SU	OA.	Kappa	Weighted F1
PB-Conv1D	P.A.	67.07	81.52	67.9	83.92	70.21	75.71	83.54	87.25	80.57	74.11	80.55
	UA.	60.87	78.13	77.46	83.27	83.54	84.11	74.58	82.39			
	F1	63.82	79.79	72.37	83.59	76.30	79.69	78.81	84.75			
PB-CDST	P.A.	67.07	80.43	71.6	85.1	70.21	77.6	81.01	87.78	81.36	75.09	81.29
	UA.	66.27	77.89	78.38	83.46	81.48	84.19	78.53	82.39			
	F1	64.79	78.65	74.83	83.58	78.16	80.59	76.41	85.1			
PB-CDSA	P.A.	68.86	76.09	67.90	87.84	72.34	77.33	72.78	87.88	81.00	74.60	80.96
	UA.	61.17	81.4	83.33	79.72	85.00	84.14	80.42	82.49			
	F1	66.67	79.14	74.84	84.27	75.43	80.76	79.75	85.00			
PB-CDSTSA	P.A.	67.66	81.52	75.31	84.71	71.28	77.46	81.65	86.74	81.16	74.90	81.11
	UA.	68.48	70.75	76.25	83.08	80.73	83.43	80.63	82.79			
	F1	68.07	75.76	75.78	83.88	75.71	80.34	81.13	84.72			
PB-BiLSTM	P.A.	59.28	79.31	61.73	71.37	72.34	76.52	72.78	86.22	77.59	69.73	77.47
	UA.	58.93	80.00	80.65	77.12	79.07	79.41	79.31	78.71			
	F1	59.10	76.84	69.93	74.13	75.56	77.94	75.91	82.29			

Table A1. Cont.

	Unit/%	SU	RI	SP	OT	WT-OT	WT-RI	WT-SP	WT-SU	OA	Kappa	Weighted F1
PB-BMST	PA.	66.47	77.17	64.20	80.00	67.02	78.00	76.58	85.70	79.40	72.44	79.37
	UA.	62.01	83.53	68.42	80.63	85.14	81.41	74.69	81.56			
	F1	64.16	80.23	66.24	80.32	75.00	79.67	75.63	83.58			
PB-BMSA	PA.	60.48	79.35	64.20	78.82	65.96	74.90	77.22	86.22	78.26	70.99	78.26
	UA.	61.21	74.49	75.36	80.08	70.45	81.86	67.78	81.25			
	F1	60.84	76.84	69.33	79.45	68.13	78.22	72.19	83.66			
PB-BMSTSA	PA.	59.88	77.17	69.14	81.57	74.47	75.84	76.58	83.83	78.22	70.93	78.19
	UA.	60.60	71.00	68.29	80.93	85.37	78.82	78.57	80.90			
	F1	60.24	73.96	68.71	81.25	79.55	77.30	77.56	82.34			
PB-Transformer	PA.	61.68	75.00	61.73	77.65	57.45	72.87	68.35	88.19	77.28	69.61	77.24
	UA.	60.59	76.67	66.67	81.48	70.13	85.04	55.96	79.53			
	F1	61.13	75.82	64.10	79.52	63.16	78.49	61.54	83.64			
PB-TRST	PA.	65.27	79.35	58.02	84.31	60.64	74.09	62.66	86.94	77.87	70.41	77.72
	UA.	64.88	75.26	63.51	79.93	67.86	83.06	63.87	80.29			
	F1	65.07	77.25	60.65	82.06	64.04	78.32	63.26	83.48			
PB-TRSA	PA.	47.9	75.00	48.15	72.94	22.34	69.91	36.71	87.98	71.29	60.79	70.09
	UA.	56.34	65.71	76.47	65.72	80.77	75.73	51.32	73.89			
	F1	51.78	70.05	59.09	67.14	35.00	72.70	42.80	80.32			
PB-TRSTSA	PA.	62.28	77.17	62.96	83.92	58.51	74.76	67.09	86.42	77.91	70.47	77.79
	UA.	63.03	78.89	68.92	79.55	65.48	82.93	65.03	80.19			
	F1	62.65	78.02	65.81	81.68	61.80	78.64	66.04	83.19			
PB-CV	PA.	71.26	80.43	72.84	85.10	70.21	78.14	83.54	88.08	82.10	76.07	82.05
	UA.	67.61	83.15	79.73	83.14	83.54	85.02	80.49	82.60			
	F1	69.39	81.77	76.13	84.11	76.30	81.43	81.99	85.26			
PB-BV	PA.	67.66	80.43	69.14	85.49	80.85	79.08	81.65	87.88	82.26	76.34	82.21
	UA.	66.47	77.89	76.71	85.16	86.36	84.32	80.63	83.46			
	F1	67.06	79.14	72.73	85.32	83.52	81.62	81.14	85.61			
PB-TV	PA.	62.87	77.17	64.20	84.31	62.77	73.95	63.29	88.81	78.61	71.35	78.45
	UA.	65.22	78.02	68.42	79.34	71.08	85.36	62.89	80.09			
	F1	64.02	77.60	66.24	81.75	66.67	79.25	63.09	84.23			
PB-CB	PA.	70.66	83.70	69.14	85.49	77.66	79.89	84.18	83.49	83.43	77.87	83.34
	UA.	67.05	79.38	76.71	84.82	89.02	86.55	82.61	84.36			
	F1	68.80	81.48	72.73	85.16	82.95	83.09	83.39	86.82			
PB-CBT	PA.	70.06	82.61	71.60	86.27	77.66	79.22	82.91	90.05	83.47	77.92	83.34
	UA.	67.63	79.17	76.32	84.94	86.90	87.22	82.39	84.12			
	F1	68.82	80.85	73.89	85.60	82.02	83.03	82.65	86.99			

References

- Gao, Z.; Guo, D.; Ryu, D.; Western, A.W. Training sample selection for robust multi-year within-season crop classification using machine learning. *Comput. Electron. Agric.* **2023**, *210*, 107927. [\[CrossRef\]](#)
- Zhang, H.K.; Roy, D.P. Using the 500m MODIS land cover product to derive a consistent continental scale 30m Landsat land cover classification. *Remote Sens. Environ.* **2017**, *197*, 15–34. [\[CrossRef\]](#)
- Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *GIScience Remote Sens.* **2017**, *54*, 918–938. [\[CrossRef\]](#)
- Jia, K.; Li, Q.; Tian, Y.; Wu, B.; Zhang, F.; Meng, J. Crop classification using multi-configuration SAR data in the North China plain. *Int. J. Remote Sens.* **2012**, *33*, 170–183. [\[CrossRef\]](#)
- Xun, L.; Zhang, J.; Cao, D.; Yang, S.; Yao, F. A novel cotton mapping index combining Sentinel-1 SAR and Sentinel-2 multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 148–166. [\[CrossRef\]](#)
- Skakun, S.; Franch, B.; Vermote, E.; Roger, J.C.; Becker-Reshef, I.; Justice, C.; Kussul, N. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. *Remote Sens. Environ.* **2017**, *195*, 244–258. [\[CrossRef\]](#)
- Wang, X.; Zhang, S.; Feng, L.; Zhang, J.; Deng, F. Mapping maize cultivated area combining MODIS EVI time series and the spatial variations of phenology over huanghuaihai plain. *Appl. Sci.* **2020**, *10*, 2667. [\[CrossRef\]](#)
- Wu, Z.; Zhang, J.; Deng, F.; Zhang, S.; Zhang, D.; Xun, L.; Javed, T.; Liu, G.; Liu, D.; Ji, M. Fusion of gf and modis data for regional-scale grassland community classification with evi2 time-series and phenological features. *Remote Sens.* **2021**, *13*, 835. [\[CrossRef\]](#)

9. Bradley, B.A.; Jacob, R.W.; Hermance, J.F.; Mustard, J.F. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sens. Environ.* **2007**, *106*, 137–145. [\[CrossRef\]](#)
10. Rhif, M.; Ben Abbes, A.; Martinez, B.; Farah, I.R. A deep learning approach for forecasting non-stationary big remote sensing time series. *Arab. J. Geosci.* **2020**, *13*, 1174. [\[CrossRef\]](#)
11. Sakamoto, T.; Van Nguyen, N.; Ohno, H.; Ishitsuka, N.; Yokozawa, M. Spatio-temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers. *Remote Sens. Environ.* **2006**, *100*, 1–16. [\[CrossRef\]](#)
12. Li, R.; Xu, M.; Chen, Z.; Gao, B.; Cai, J.; Shen, F.; He, X.; Zhuang, Y.; Chen, D. Phenology-based classification of crop species and rotation types using fused MODIS and Landsat data: The comparison of a random-forest-based model and a decision-rule-based model. *Soil Tillage Res.* **2021**, *206*, 104838. [\[CrossRef\]](#)
13. Belgiu, M.; Bijker, W.; Csillik, O.; Stein, A. Phenology-based sample generation for supervised crop type classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102264. [\[CrossRef\]](#)
14. Zhong, L.; Hawkins, T.; Biging, G.; Gong, P. A phenology-based approach to map crop types in the San Joaquin Valley, California. *Int. J. Remote Sens.* **2011**, *32*, 7777–7804. [\[CrossRef\]](#)
15. He, T.; Xie, C.; Liu, Q.; Guan, S.; Liu, G. Evaluation and Comparison of Random Forest and A-LSTM Networks for Large-scale Winter Wheat Identification. *Remote Sens.* **2019**, *11*, 1665. [\[CrossRef\]](#)
16. Zhang, J.; Feng, L.; Yao, F. Improved maize cultivated area estimation over a large scale combining MODIS-EVI time series data and crop phenological information. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 102–113. [\[CrossRef\]](#)
17. Wang, X.; Zhang, J.; Xun, L.; Wang, J.; Wu, Z.; Henchiri, M.; Zhang, S.; Zhang, S.; Bai, Y.; Yang, S.; et al. Evaluating the Effectiveness of Machine Learning and Deep Learning Models Combined Time-Series Satellite Data for Multiple Crop Types Classification over a Large-Scale Region. *Remote Sens.* **2022**, *14*, 2341. [\[CrossRef\]](#)
18. Antonijević, O.; Jelić, S.; Bajat, B.; Kilibarda, M. Transfer learning approach based on satellite image time series for the crop classification problem. *J. Big Data* **2023**, *10*, 54. [\[CrossRef\]](#)
19. Rußwurm, M.; Courty, N.; Emonet, R.; Lefèvre, S.; Tuia, D.; Tavenard, R. End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 445–456. [\[CrossRef\]](#)
20. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [\[CrossRef\]](#)
21. Luo, Y.; Zhang, Z.; Li, Z.; Chen, Y.; Zhang, L.; Cao, J.; Tao, F. Identifying the spatiotemporal changes of annual harvesting areas for three staple crops in China by integrating multi-data sources. *Environ. Res. Lett.* **2020**, *15*, 074003. [\[CrossRef\]](#)
22. Xun, L.; Zhang, J.; Cao, D.; Wang, J.; Zhang, S.; Yao, F. Mapping cotton cultivated area combining remote sensing with a fused representation-based classification algorithm. *Comput. Electron. Agric.* **2021**, *181*, 105940. [\[CrossRef\]](#)
23. Chen, S.W.; Tao, C.S. Multi-temporal PolSAR crops classification using polarimetric-feature-driven deep convolutional neural network. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017; pp. 1–4.
24. Gadiraju, K.K.; Ramachandra, B.; Chen, Z.; Vatsavai, R.R. Multimodal Deep Learning Based Crop Classification Using Multispectral and Multitemporal Satellite Imagery. In Proceedings of the KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, 23–27 August 2020; pp. 3234–3242.
25. Karim, F.; Majumdar, S.; Darabi, H.; Harford, S. Multivariate LSTM-FCNs for time series classification. *Neural Netw.* **2019**, *116*, 237–245. [\[CrossRef\]](#) [\[PubMed\]](#)
26. LeCun, Y.; Bengio, Y. Convolutional Networks for Images, Speech, and Time-Series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1995.
27. Yuan, Y.; Lin, L.; Zhou, Z.G.; Jiang, H.; Liu, Q. Bridging optical and SAR satellite image time series via contrastive feature extraction for crop classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 222–232. [\[CrossRef\]](#)
28. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [\[CrossRef\]](#)
29. Kumar, S.; Shukla, S.; Sharma, K.K.; Kumar Singh, K.; Akbari, A.S. Classification of Land Cover and Land Use Using Deep Learning. *Lect. Notes Electr. Eng.* **2021**, *796*, 321–327.
30. Marcos, D.; Volpi, M.; Kellenberger, B.; Tuia, D. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 96–107. [\[CrossRef\]](#)
31. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y. Remote Sensing of Environment Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [\[CrossRef\]](#)
32. Rußwurm, M.; Korner, M. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 129. [\[CrossRef\]](#)
33. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [\[CrossRef\]](#)
34. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning spectral-spatial features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 924–935. [\[CrossRef\]](#)

35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5999–6009.
36. Rußwurm, M.; Körner, M. Self-attention for raw optical Satellite Time Series Classification. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 421–435. [[CrossRef](#)]
37. Huang, G.; Li, Y.; Pleiss, G.; Liu, Z.; Hopcroft, J.E.; Weinberger, K.Q. Snapshot ensembles: Train 1, get M for free. *arXiv* **2017**, arXiv:1704.00109.
38. Izmailov, P.; Podoprikin, D.; Garipov, T.; Vetrov, D.; Wilson, A.G. Averaging weights leads to wider optima and better generalization. *arXiv* **2018**, arXiv:1803.05407.
39. Luo, K.; Lu, L.; Xie, Y.; Chen, F.; Yin, F.; Li, Q. Crop type mapping in the central part of the North China Plain using Sentinel-2 time series and machine learning. *Comput. Electron. Agric.* **2023**, *205*, 107577. [[CrossRef](#)]
40. Li, H.; Tian, Y.; Zhang, C.; Zhang, S.; Atkinson, P.M. Temporal Sequence Object-based CNN (TS-OCNN) for crop classification from fine resolution remote sensing image time-series. *Crop J.* **2022**, *10*, 1507–1516. [[CrossRef](#)]
41. Sakamoto, T.; Gitelson, A.A.; Arkebauer, T.J. Near real-time prediction of U.S. corn yields based on time-series MODIS data. *Remote Sens. Environ.* **2014**, *147*, 219–231. [[CrossRef](#)]
42. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: Global land-cover product with fine classification system at 30m using time-series Landsat imagery. *Earth Syst. Sci. Data* **2021**, *13*, 2753–2776. [[CrossRef](#)]
43. Picon, A.; Seitz, M.; Alvarez-Gila, A.; Mohnke, P.; Ortiz-Barredo, A.; Echazarra, J. Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Comput. Electron. Agric.* **2019**, *167*, 105093. [[CrossRef](#)]
44. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Phys. Lett. B* **2014**, *299*, 345–350.
45. Bermúdez, J.D.; Achanccaray, P.; Sanches, I.D.; Cue, L.; Happ, P.; Feitosa, R.Q. Evaluation of Recurrent Neural Networks for Crop Recognition from Multitemporal Remote Sensing Images. In Proceedings of the Anais do XXVII Congresso Brasileiro de Cartografia, Rio de Janeiro, Brazil, 6–9 November 2017; pp. 800–804.
46. Xu, J.; Yang, J.; Xiong, X.; Li, H.; Huang, J.; Ting, K.C.; Ying, Y.; Lin, T. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sens. Environ.* **2021**, *264*, 112599. [[CrossRef](#)]
47. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
48. Su, H.; Zhang, T.; Lin, M.; Lu, W.; Yan, X.H. Predicting subsurface thermohaline structure from remote sensing data based on long short-term memory neural networks. *Remote Sens. Environ.* **2021**, *260*, 112465. [[CrossRef](#)]
49. Cao, C.; Dragičević, S.; Li, S. Short-term forecasting of land use change using recurrent neural network models. *Sustainability* **2019**, *11*, 5376. [[CrossRef](#)]
50. Sainte Fare Garnot, V.; Landrieu, L.; Giordano, S.; Chehata, N. Satellite image time series classification with pixel-set encoders and temporal self-attention. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 12322–12331.
51. Kawaguchi, K. Deep learning without poor local minima. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 586–594.
52. Chellamy, M.; Ferré, T.P.A.; Greve, M.H. Evaluating an ensemble classification approach for crop diversity verification in Danish greening subsidy control. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 10–23. [[CrossRef](#)]
53. Kent, M.; Parkinson, T.; Kim, J.; Schiavon, S. A data-driven analysis of occupant workspace dissatisfaction. *Build. Environ.* **2021**, *205*, 108270. [[CrossRef](#)]
54. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**, arXiv:2010.16061.
55. Ashourloo, D.; Shahrabi, H.S.; Azadbakht, M.; Aghighi, H.; Matkan, A.A.; Radiom, S. A Novel Automatic Method for Alfalfa Mapping Using Time Series of Landsat-8 OLI Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4478–4487. [[CrossRef](#)]
56. Xun, L.; Wang, P.; Li, L.; Wang, L.; Kong, Q. Identifying crop planting areas using Fourier-transformed feature of time series MODIS leaf area index and sparse-representation-based classification in the North China Plain. *Int. J. Remote Sens.* **2019**, *40*, 2034–2052. [[CrossRef](#)]
57. Chen, H.; Qiu, Y.; Yin, D.; Chen, J.; Chen, X.; Liu, S.; Liu, L. Stacked spectral feature space patch: An advanced spectral representation for precise crop classification based on convolutional neural network. *Crop J.* **2022**, *10*, 1460–1469. [[CrossRef](#)]
58. Wang, Z.; Chen, J.; Xing, F.; Han, Y.; Chen, F.; Zhang, L.; Li, Y.; Li, C. Response of cotton phenology to climate change on the North China Plain from 1981 to 2012. *Sci. Rep.* **2017**, *7*, 6628. [[CrossRef](#)] [[PubMed](#)]
59. Song, X.P.; Potapov, P.V.; Krylov, A.; King, L.A.; Di Bella, C.M.; Hudson, A.; Khan, A.; Adusei, B.; Stehman, S.V.; Hansen, M.C. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sens. Environ.* **2017**, *190*, 383–395. [[CrossRef](#)]
60. Zhu, Z.; Gallant, A.L.; Woodcock, C.E.; Pengra, B.; Olofsson, P.; Loveland, T.R.; Jin, S.; Dahal, D.; Yang, L.; Auch, R.F. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 206–221. [[CrossRef](#)]
61. Yang, L.; Jin, S.; Danielson, P.; Homer, C.; Gass, L.; Bender, S.M.; Case, A.; Costello, C.; Dewitz, J.; Fry, J.; et al. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 108–123. [[CrossRef](#)]

62. Massey, R.; Sankey, T.T.; Congalton, R.G.; Yadav, K.; Thenkabail, P.S.; Ozdogan, M.; Sánchez Meador, A.J. MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types. *Remote Sens. Environ.* **2017**, *198*, 490–503. [[CrossRef](#)]
63. Jin, S.; Yang, L.; Danielson, P.; Homer, C.; Fry, J.; Xian, G. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* **2013**, *132*, 159–175. [[CrossRef](#)]
64. Jin, S.; Yang, L.; Zhu, Z.; Homer, C. A land cover change detection and classification protocol for updating Alaska NLCD 2001 to 2011. *Remote Sens. Environ.* **2017**, *195*, 44–55. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.