

Article

Mapping Paddy Rice Using Weakly Supervised Long Short-Term Memory Network with Time Series Sentinel Optical and SAR Images

Mo Wang ^{1,2} , Jing Wang ^{3,*} and Li Chen ^{1,2}

¹ Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China; wangmo@caas.cn (M.W.); chenli02@caas.cn (L.C.)

² Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081, China

³ China Center for Information Industry Development, Beijing 100086, China

* Correspondence: wangj.13b@igsrr.ac.cn

Received: 4 October 2020; Accepted: 16 October 2020; Published: 19 October 2020



Abstract: Rice is one of the most important staple food sources worldwide. Effective and cheap monitoring of rice planting areas is demanded by many developing countries. This study proposed a weakly supervised paddy rice mapping approach based on long short-term memory (LSTM) network and dynamic time warping (DTW) distance. First, standard temporal synthetic aperture radar (SAR) backscatter profiles for each land cover type were constructed on the basis of a small number of field samples. Weak samples were then labeled on the basis of their DTW distances to the standard temporal profiles. A time series feature set was then created that combined multi-spectral Sentinel-2 bands and Sentinel-1 SAR vertical received (VV) band. With different combinations of training and testing datasets, we trained a specifically designed LSTM classifier and validated the performance of weakly supervised learning. Experiments showed that weakly supervised learning outperformed supervised learning in paddy rice identification when field samples were insufficient. With only 10% of field samples, weakly supervised learning achieved better results in producer's accuracy (0.981 to 0.904) and user's accuracy (0.961 to 0.917) for paddy rice. Training with 50% of field samples also presented improvement with weakly supervised learning, although not as prominent. Finally, a paddy rice map was generated with the weakly supervised approach trained on field samples and DTW-labeled samples. The proposed data labeling approach based on DTW distance can reduce field sampling cost since it requires fewer field samples. Meanwhile, validation results indicated that the proposed LSTM classifier is suitable for paddy rice mapping where variance exists in planting and harvesting schedules.

Keywords: paddy rice mapping; dynamic time warping; LSTM; weakly supervised learning; cropland mapping

1. Introduction

Rice is an important staple food source for more than half of the global population [1]. In Asia in particular, the rice planting area was 146 million ha in 2018, accounting for 88% of the world's total (FAOSTAT, 2020). The demand for rice over the next 30 years is anticipated to increase by 90% in Asia [2]. Monitoring rice planting is of great importance to global food security and informed policymaking.

Remote sensing has been providing timely and efficient means for agricultural monitoring for decades. Paddy rice is the only crop that grows under wetland conditions [3]. Its growing cycle comprises three phases: vegetative phase, reproductive phase, and ripening phase, which are detailly divided into 10 growth stages. The growing cycle of paddy rice features unique optical reflectance and

radar scattering characteristics compared to other crops due to the abundant water conditions in the early stages. Many study cases have employed optical (e.g., Landsat, Moderate Resolution Imaging Spectroradiometer (MODIS), SPOT, and Sentinel-2), microwave (e.g., RADARSAT, Phased Array Type L-band Synthetic Aperture Radar (PALSAR), and Sentinel-1), or a combination of the two remote sensing data sources to map rice-growing area at regional or continental scales. Dong and Xiao [4] sorted out rice mapping approaches into four categories, i.e., (1) reflectance data and image statistic-based approaches, (2) vegetation index (VI) data and enhanced image statistic-based approaches, (3) VI or radar backscatter-based temporal analysis approaches, and (4) phenology-based approaches through remote sensing recognition of key growth phases. The essence of rice mapping approaches is classification models using a single (category 1 and 2) or multiple temporal (category 3 and 4) remote sensing images. A single temporal image has its limitations in discriminating paddy rice from other land cover types due to similar remote sensing signals at certain growing stage and mixed pixels [5]. Approaches based on multi-temporal and phenology-based classifiers are drawing more research interests for paddy rice mapping and other cropland mappings [5–9].

Flooding and transplanting signals are unique features for paddy rice. They are often used as key signatures to identify the rice field. Xiao et al. [7,10] generated rice maps for China and Southeast Asia with time-series MODIS data by identifying flooding and transplanting phases using the relationship between normalized difference vegetation index (NDVI) (enhanced vegetation index (EVI)) and land surface water index (LSWI). Yin et al. [11] applied a time series vegetation index to detect flooding and transplanting phases using fusion data from MODIS and Landsat images. Besides vegetation index and water index, synthetic aperture radar (SAR) datasets were also used to detect flooding and transplanting signals. Torbick et al. [12] and Zhang et al. [13] used PALSAR data to identify rice growing stages. Other phenology-based approaches with SAR data sources include European Remote Sensing Satellite-1 (ERS-1) [14] and RADARSAT-2 [15]. More recently, the novel European Space Agency (ESA) Sentinel-1 has been frequently used as a SAR data source for paddy rice mapping [8,9,16,17], as it provides higher spatial resolution (10 m) and more frequent revisit (6 days). It should be noted that optical remote sensing data is often limited to construct an ideal time series dataset due to cloud cover in rice-growing regions. Therefore, radar remote sensing (e.g., SAR) is more suited for time series analysis since it is independent of weather conditions.

Many multi-temporal or phenology-based paddy rice mapping studies treated all rice fields as synchronized in phenology and input multi-temporal data into classifiers simultaneously. Such a manner ignores the planting schedule difference of farmers and variation in rice-growing periods of different rice cultivars. For example, Le Toan et al. [18] observed a shift of up to 6–8 weeks between fields within a region in Indonesia. This could lead to classification error for regional rice mapping with medium and high spatial resolution images. Some studies [19–21] attempted to address this problem with dynamic programming algorithms, such as dynamic time warping (DTW). The authors classified cropland on the basis of the DTW distance of NDVI from each pixel to samples using a threshold or a simple classifier. To obtain a complete NDVI time series dataset, they either had to reconstruct the data from cloud cover or use existing data products, which often are not available. The reconstruction of data for cloudy days could introduce uncertainties to the final time series dataset. Radar remote sensing data, on the other hand, has greater potential in finding phenological alignment using dynamic programming algorithms. Yet, few cropland mapping studies exist that have used such a strategy, except for Li and Bijker [22] who focused on vegetable classification.

Being consistent with general land cover classification development, various supervised and unsupervised classification algorithms have been applied for paddy rice mapping, such as maximum likelihood [23,24], support vector machines (SVM) [25–27], random forest (RF) [27,28], and decision trees (DT) [29]. In recent years, deep learning methods have shown great capability in fields such as computer vision and remote sensing image classification [30,31]. Convolutional neural network (CNN) and recurrent neural network (RNN) are the most adopted algorithms for remote sensing classification tasks, including rice mapping. Compared to shallow learning methods (such as SVM, RF, and DT),

deep learning models rely more on abundant training samples to make reliable predictions and to avoid over-fitting. Kussul et al. [31] presented one of the early attempts of using deep learning models for cropland classification. The authors proposed a 1-D CNN model from multiple Sentinel-1 and Landsat images sensed in the growing season. A total of 547 polygons of training samples were selected by extensive field surveys. Zhang et al. [32] devised a CNN-based approach that classifies small patches for each pixel with fused multi-temporal Landsat and MODIS NDVI data, along with a synthetic phenological variable. They acquired training samples on the basis of existing land use/land cover (LULC) maps and Google Earth. Sun et al. [33] proposed a long short-term memory (LSTM) recurrent neural network (RNN) model for yearly cropland classification of North Dakota, USA, using Landsat imagery. More than 10 million samples were selected to train the LSTM model on the basis of the Cropland Data Layer (CDL). All these deep learning-based cropland classification approaches rely on either a trustworthy labeled training data source or a laborious field survey, which might not be available or may be too expensive for some projects.

With concerns of variance in rice cultivation schedule and abundant training samples needed for deep learning models, we propose a paddy rice mapping approach that labels weak samples using DTW distance with time series SAR data and then feeds them into an LSTM network classifier. First, standard SAR backscatter profiles for each land cover type were constructed on the basis of a relatively small number of paddy rice pixel samples. The polarization of SAR data was carefully selected for that which best reflected flooding and transplanting signals of rice growing stages. DTW distances from each pixel of the study area to standard profiles were then computed. The k -nearest pixels were selected as weak samples. Using field samples and weak samples as a training set, we trained an LSTM model using time series multispectral and SAR bands as input. This weakly supervised deep learning approach could benefit from the strong learning ability of RNN models on time series data with less field sampling cost.

2. Materials and Methods

2.1. Study Area and Materials

2.1.1. Study Area

The study area is Tongxiang County of China's Zhejiang Province. It was chosen because it is one of China's major rice-growing regions. The county is located in the Hangjiahu Plain of East China, with an area of 727 km² (Figure 1). The annual number of days $\geq 10^\circ\text{C}$ is 217–222. The accumulated temperature in degree-days is 4772–4960 °C, featured with a subtropical monsoon climate. Rainfall and river network of the region provide adequate irrigation for single/double-cropping systems. Motivated by the agricultural economy, only single-cropping of rice is practiced in Tongxiang. Rice growing season of the study area is from May to November. However, rice harvest could happen as early as in October and as late as in December.

The study area meets the requirements to test our method for several reasons. First, it represents the typical geography of rice cultivation regions in East China with small to medium field size. Second, the weather of the region is an example of a scenario wherein optical remote sensing data are often not applicable due to cloud cover during the rice-growing season. Third, the size of the region is fit for rice mapping with high to medium resolution images that are to be trained with deep learning models.

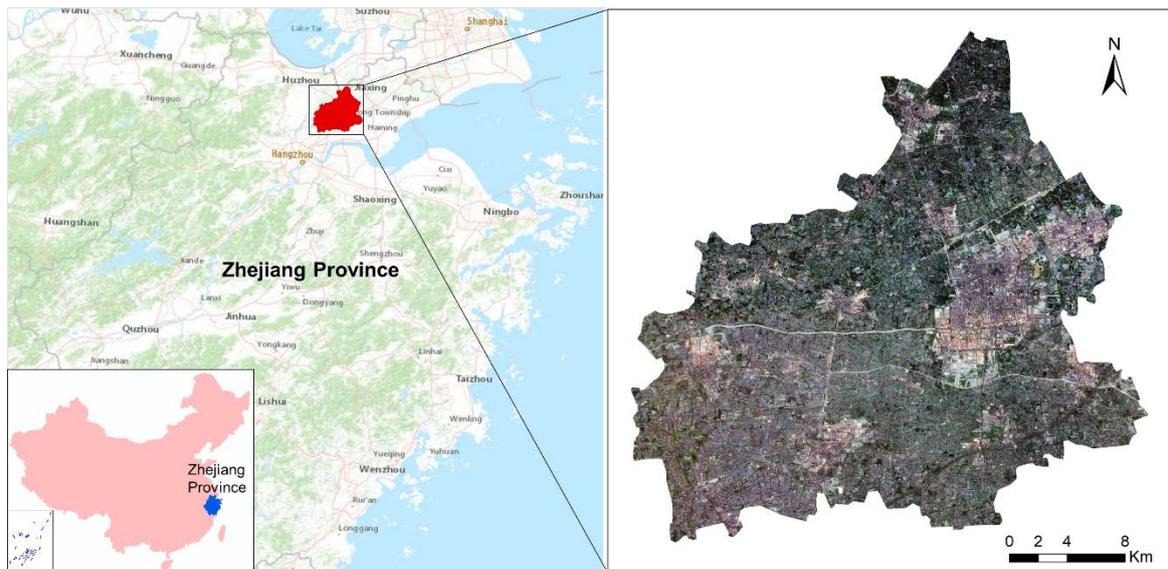


Figure 1. Location of the study area.

2.1.2. Datasets and Preprocessing

Time Series SAR Data

Time series SAR data were used to compute similarities between pixels to land cover samples on the basis of sequence alignment. Time coverage of the SAR data was from April to December, which was extended by a month for the beginning and the end of the period compared with the normal cultivation period (May to November). This was because abnormal cultivation schedules exist for some farmers, for example, early rice planting (e.g., before May) and late harvest (e.g., after November). A total of 23 images of Sentinel-1A Level-1 Ground Range Detected (GRD) product with an interval of 12 days from April to December 2018 were accessed from ESA Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). Sentinel-1B data product was omitted since it was not provided by ESA for the study area. The images were sensed with C-band in interferometric wide (IW) mode, and in dual of vertical transmitted; vertical received (VV); and vertical transmitted, horizontally received (VH) channels. Attributes of SAR images are listed in Table 1. The radar backscatter data were used to detect key phenology stages of paddy rice, especially flooding signals of early growing stages. According to Clement et al. [34], VV polarization slightly outperforms VH polarization while delineating flooding. Hence, VV bands were chosen to construct the SAR time series.

Several preprocessing steps were carried out using Sentinel Toolboxes software SNAP. First, subsets of original images were generated by cropping extend to the study area. The images were then calibrated to sigma naught (σ_0) backscatter coefficients in a linear unit. Terrain correction was applied subsequently using Range Doppler Terrain Correction with Shuttle Radar Topography Mission (STRM) 3 sec grid. We generated 10×10 m pixel spacing images, which were reprojected to the Universal Transverse Mercator (UTM) zone 51 N. Then, a multi-temporal speckle filtering was applied on the terrain-corrected images using a 7×7 gamma map filter with 3 looks to reduce the noise in the images. Finally, the linear intensity values of the VV bands were converted to the dB unit.

Sentinel-2 Time Series Images

Sentinel-2 multispectral images covering rice growing season were obtained from ESA Copernicus Open Access Hub. Screened by cloud cover, only 11 cloud-free dates were usable. Top-of-atmosphere level 1C data products of those dates were downloaded. Some preprocessing steps were also required for the level 1C product. First, atmospheric, terrain, and cirrus correction were performed using the Sen2Cor v2.8 tool provided by ESA to achieve bottom-of-atmosphere level 2A product. Among the

13 spectral bands of Sentinel-2 images, we selected 10 bands (B2-B8, B8a, B11, and B12) as input features, including visible, near-infrared (NIR), and short-wave infrared (SWIR) bands. The reason to include SWIR bands is that they are sensible to leaf water content [35], which could help differentiate rice fields from other crops. B2-B4 and B8 were in a resolution of 10 m. The remaining 6 bands (20 m resolution) were resampled to 10 m by nearest neighbor interpolation using the processing module in SNAP. Table 2 lists the attributes of used Sentinel-2 images.

Table 1. Attributes of synthetic aperture radar (SAR) images.

SAR Images	
Sensor	Sentinel-1A
Data level	Level-1 ground range detected
Spatial resolution	10 m
Wavelength	C band
Polarization	VV
Pass	Ascending
Acquisition mode	IW
Acquisition date (23 successive images with 12-day interval)	4 April 2018
	16 April 2018
	28 April 2018

	24 December 2018

Table 2. Attributes of optical images.

Multispectral Optical Images	
Sensor	Sentinel-2A, Sentinel-2B
Data level	Level-1C
Spatial resolution	10 m, 20 m
Band	B2, B3, B4, B5, B6, B7, B8, B8a, B11, B12
Acquisition date (11 images)	9 April 2018
	19 April 2018
	4 May 2018
	9 May 2018
	18 June 2018
	18 July 2018
	23 July 2018
	28 July 2018
	7 August 2018
	1 October 2018
	10 November 2018

Field Sampling Data

Field sampling was conducted to collect samples belonging to broad land cover categories, i.e., paddy rice fields, water bodies (rivers, lakes, and reservoirs), construction (buildings, roads, bare surfaces, and paved surfaces), vegetation (forest, shrubs, orchards, and other planted area), and other croplands. A total of 506 sample sites (points of interest, POIs) were selected through scattering the study area for the five land cover types. The sample sites were selected in a homogeneous spatial distribution principle to cover cultivation variation between sub-regions. A geo-referenced field photo was taken for each POI in November 2018. Locations of geo-referenced photos were converted

into kmz format and imported to Google Earth. Polygons of sample sites (areas of interest, AOIs) were then delineated with the assistance of Google Earth and reference photos. Boundaries of polygons can be distinguished with high-resolution satellite images of 2018 from Google Earth. The land cover type of an AOI was assigned with the land cover type of the field photo where it was taken (as shown in Figure 2).

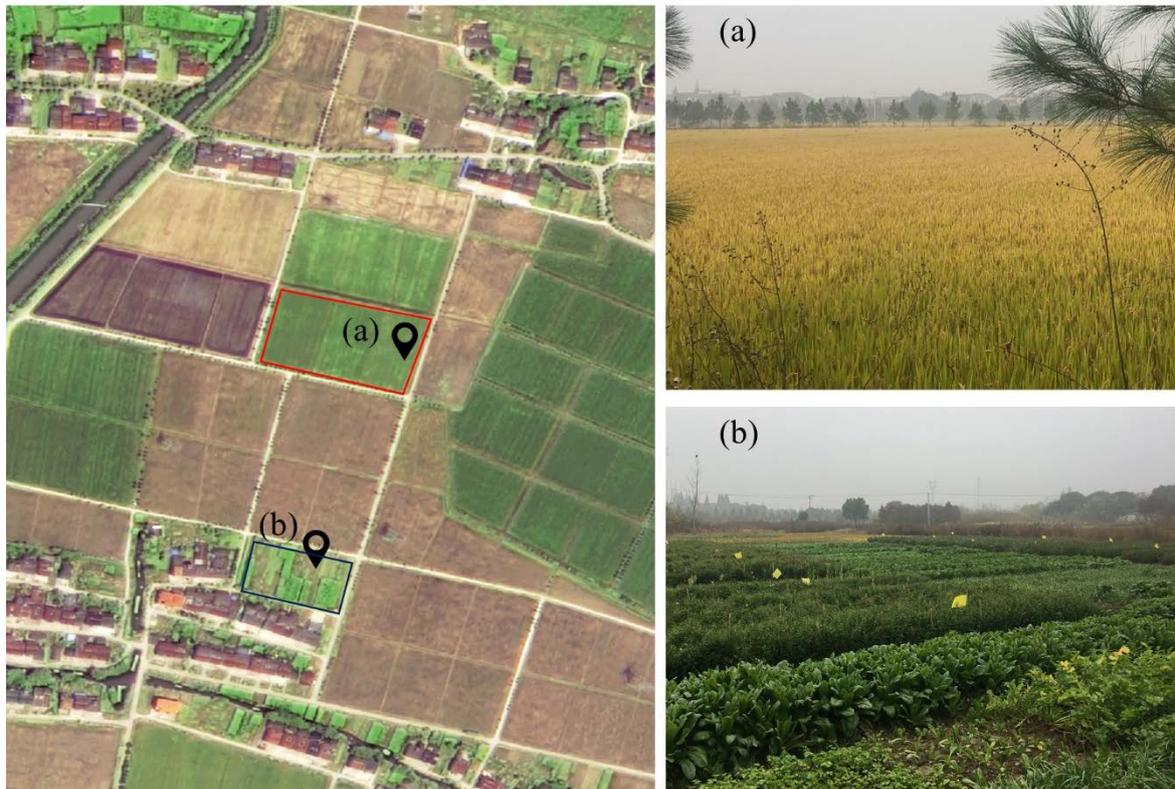


Figure 2. Illustration of the field sampling process. Image (a) (paddy rice field) and (b) (other croplands) are example photos of points of interest (POIs). Their locations are shown in the Google Earth image on the left. Boundaries of fields (a,b) were delineated with the Google Earth image. Likewise, areas of interest (AOIs) for the other three land cover types were also created. Note that the dates of the Google Earth image and the field photos are not the same but were taken in the same year.

A total number of 506 AOIs (one for each sample site) were created and then rasterized into a GeoTIFF image with 10 m resolution. Final AOIs include rice field (4569 pixels), water body (4805 pixels), vegetation (4372 pixels), construction (6094 pixels), and other croplands (4592 pixels).

2.2. Methodology

We integrated time series optical and radar remote sensing data to best capture phenological characteristics of paddy rice. The time series data were to be fed into a RNN to cope with non-linear relationship and temporal correlation. The framework of this study is shown in Figure 3. First, a labeling strategy was developed. Following the preprocessing steps, standard SAR backscatter profiles for each land cover type were created with a relatively small number of field samples. Pairwise DTW distances to the standard SAR backscatter profile for each land cover type were computed for the whole study area. Top- k nearest pixels were then screened out as weakly labeled samples with the highest confidence. A specifically designed LSTM network was trained with integrated input features of optical and SAR data, using field samples (supervised) or optionally combined with DTW-labeled samples (weakly supervised). Finally, a supervised or weakly supervised approach was to be adopted on the basis of experiment results to generate a paddy rice map.

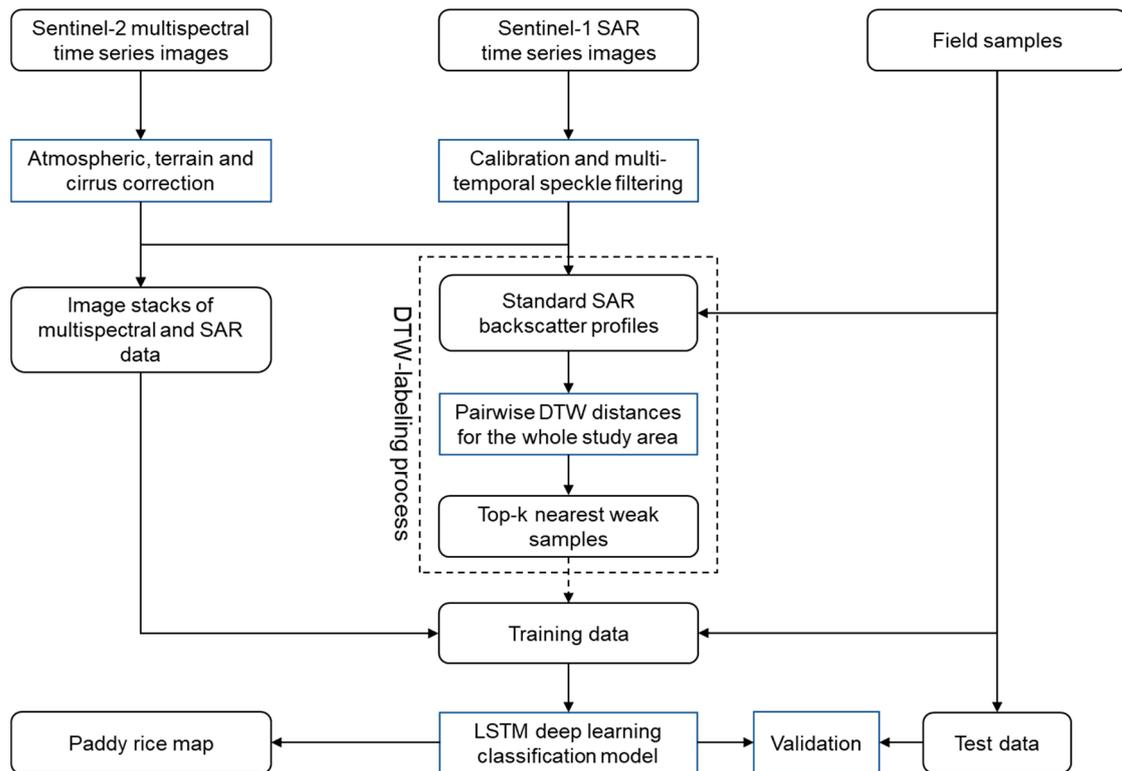


Figure 3. The framework of mapping paddy rice with dynamic time warping (DTW)-labeling strategy and weakly supervised classification with long short-term memory (LSTM) deep learning classifier.

2.2.1. Standard Time Series SAR Backscatter Profiles

To simulate the scenario wherein only a small number of field samples are available, we took 10% of field sample pixels of each land cover type as a base to compute standard SAR backscatter profiles. SAR backscatter values of each date were extracted from the time series of the 23 Sentinel-1 SAR backscatter coefficient images from April to December. For each date and land cover type, the standard SAR backscatter value was computed by averaging on the selected base samples (Figure 4). The five land cover types had different levels of SAR backscatter in dB. Waterbody had a distinctly low radar backscatter profile with litter temporal variance. In contrast, construction had the highest radar backscatter value. Vegetation and cropland were similar in the radar backscatter level. Compared to other cropland, rice fields showed a greater variance due to temporal change in the structure of the canopy and that of the underlying soil surface.

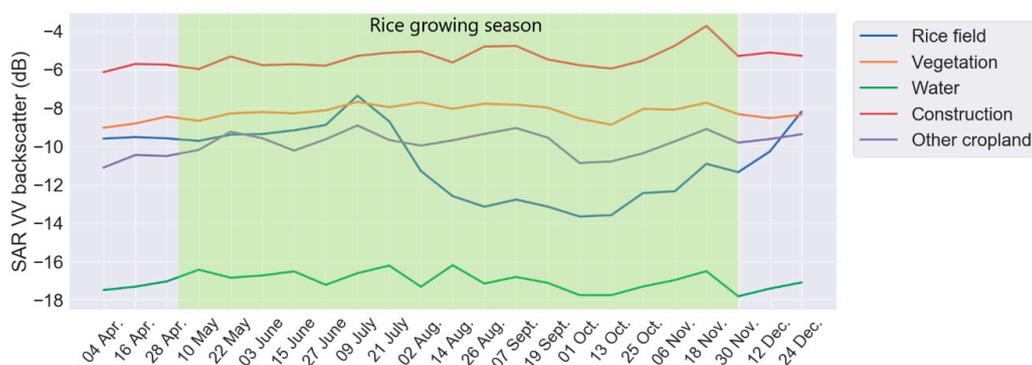


Figure 4. Time series SAR backscatter profiles for the five land cover types.

2.2.2. DTW Distance-Based Sampling

The dynamic time warping (DTW) algorithm uses dynamic programming techniques to find the optimal alignment and the minimal distance of two temporal sequences. The algorithm was originally developed for speech recognition to cope with different speaking speeds [36], and was introduced in time series remote sensing image analysis in recent years [19,20,37,38]. The alignment of two sequences is through being warped non-linearly by stretching or shrinking the time dimension. In our case, DTW can deal with temporal distortion in rice-growing phases of different rice fields or rice varieties. Mathematical principle and a detailed procedure to implement DTW was elaborated by Berndt and Clifford [39]. The basic idea is illustrated in Figure 5, where A and B are two SAR backscatter time series of rice field samples. The DTW algorithm finds the best alignment that allows minimum accumulated distance for each time step. In this example, the planting date of B was later than that of A. DTW algorithm aligns the phases between these two SAR time series by shifting with the corresponded time gap. As a result, the similarity measure in the SAR time series curve is justified, regardless of planting date difference and rice varieties.

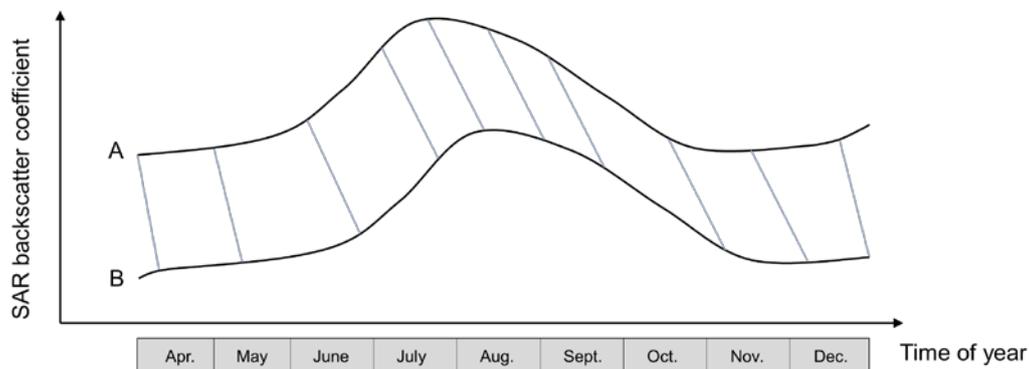


Figure 5. Illustration of DTW alignment between two time series SAR backscatter curves.

For each land cover type, pairwise DTW distances to its standard SAR time series were computed. The shorter the distance, the more likely a pixel belongs to the corresponding land cover type. Top-*k* nearest strategy was applied to select weak samples, which means they had the highest confidence belonging to a certain land cover type. The decision on the number of screened samples was made on the basis of the rough size of the rice planting area in the study area. A larger number of samples may enhance the training of the deep learning model. However, it could bring more uncertainty in the sample label, i.e., lower sample quality. The number of labeled samples should be a small portion of the total rice planting area. In practice, a DTW Python module by Pierre Rouanet (<https://github.com/pierre-rouanet/dtw>) was adopted.

2.2.3. LSTM Deep Learning Classifier

LSTM [40] is an RNN architecture designed to avoid the long-term dependency problem. Compared to feed-forward neural networks, RNN has a feedback connection that deals with temporal dynamic behavior. RNNs use an internal state (memory) from the previous time step as one of the inputs for the current time step. Traditional RNN often encounters vanishing gradient problem while processing long-sequence data. LSTM, on the other hand, can avoid this problem by adding gates to control the cell state. A common LSTM unit has an input gate, an output gate, and a forget gate. A standard RNN unit and LSTM unit are illustrated in Figure 6.

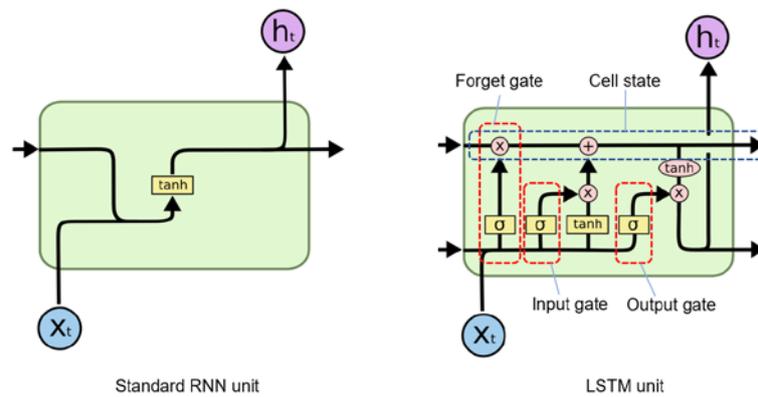


Figure 6. The architecture of standard recurrent neural network (RNN) cell and LSTM cell [41].

Gates in LSTM cells use a sigmoid activation function to decide whether to keep a feature. “0” means the gates are blocking inputs. “1” means gates are allowing inputs to pass through it. Equations of the three gates are

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (3)$$

where i_t , f_t , and o_t denote input gate, forget gate, and output gate, respectively; σ denotes sigmoid function; and w_x , h_{t-1} , x_t , and b_x denote the weight for respective gate neurons, the output of the previous LSTM block, input at the current timestamp, and biases for the respective gates, respectively.

The cell state and final output are defined by equations

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_i) \quad (4)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (5)$$

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

where c_t and \tilde{c}_t denote cell state and candidate for cell state at time t , and h_t denotes the final output (hidden state) of the LSTM cell.

Considering the large span of the time series data, we developed an LSTM network to identify paddy rice from other land cover types. The core of the model is a stack of two LSTM layers, one fully connected layer, and two dropout layers, as shown in Figure 7. Output layers used the softmax activation function to predict the likelihood that a pixel belongs to one land cover type. Categorical cross-entropy was chosen as the loss function. The choice was made on the basis of several concerns. First, it has been proven that the use of cross-entropy losses made great performance improvements in models with softmax outputs [42]. Second, the training samples for each class were fairly balanced, which makes the network unbiased towards each class. Therefore, other loss functions were neglected, for instance, focal loss, which is suitable to deal with the imbalance problem in datasets. Parameters used to train the network are shown in Table 3. The model was implemented with Python 3.7 language and TensorFlow library.

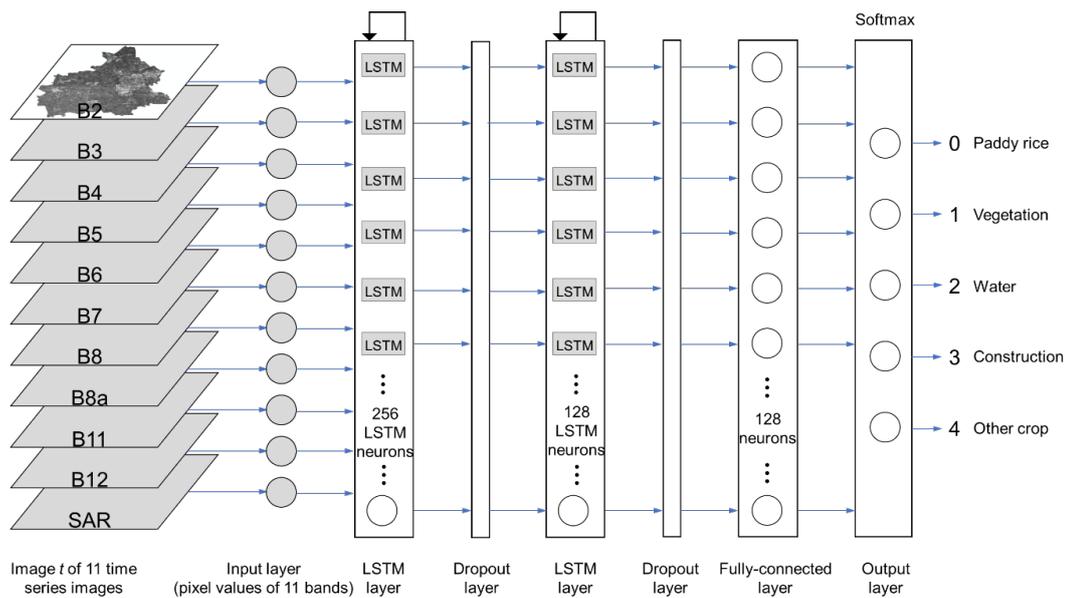


Figure 7. LSTM network structure. The inputs of the classifier are pixel values of 11 bands (10 multispectral bands and 1 SAR band). At time t , the classifier is fed with the band values of image t , and the LSTM layers receive cell state and hidden state from the previous timestamp $t - 1$. The final prediction is made by the output layer of the last timestamp.

Table 3. Parameters of proposed LSTM classifier.

Learning Rate	Dropout Rate	Batch Size	Loss Function	Optimizer
0.001	0.5	64	Categorical cross-entropy	Adam algorithm

The input multispectral features (10 bands) of the classification model were carefully selected from optical imagery, as described within Section 2.1.2. Meanwhile, SAR backscatter signals provided valuable information to discriminate paddy rice fields. The SAR backscatter coefficient was therefore added as an input feature. It should be noted that the dates of Sentinel-1 SAR images did not coincide with that of Sentinel-2 optical images. The SAR images with the closest sensing date to corresponding optical images were selected. This strategy should be feasible since the time difference between the two data sources was minor (less than 5 days). The SAR data were normalized to the same range (0, 1) as optical bands. The final input features comprised 10 optical bands and 1 SAR band.

2.2.4. Experiment Design

We designed 3 experiment schemes on the basis of the number of field samples. In each scheme, supervised and weakly supervised learning were both to be tested. Scheme 1 and 2 are two comparative experiments that were to be validated by a fixed test set with a sufficient sample number. First, the test set was randomly selected from the rest of the field samples, apart from those for computing standard SAR backscatter profiles. The sample number of the test set accounted for 50% of all field samples. In scheme 1, 10% of field samples and 5000 DTW-labeled samples each type were used for training. The 10% was the same set of samples for computation of the standard SAR backscatter profiles in Section 2.2.1. Scheme 1 is designed to test for the case that only a small number of field samples were available. In scheme 2, training samples were the rest of the half field samples, apart from the test set plus 2000 DTW-labeled samples of each type. Scheme 3 utilized 80% of field samples and 2000 DTW-labeled samples of each type for training. It was designed to compare weakly supervised with supervised learning in a condition of relatively adequate field samples. The details of experiment schemes are listed:

- Scheme 1: Supervised learning training on 10% of field samples compares with weakly supervised learning training on (10% of field samples + 5000 DTW-labeled samples for each land cover type).
- Scheme 2: Supervised learning training on 50% of field samples compares with weakly supervised learning training on (50% of field samples + 2000 DTW-labeled samples for each land cover type).
- Scheme 3: Supervised learning training on 80% of field samples compares with weakly supervised learning training on (80% of field samples + 2000 DTW-labeled samples for each land cover type).

3. Results

3.1. DTW Distance-Based Sampling Results

Using the SAR time series of each pixel in the study area and the five standard SAR profiles, we generated five pairwise DTW distance maps (Figure 8). Each map depicted the likelihood that image pixels belong to the corresponding land cover type. The lower the value, the more likely the pixel belonged to the corresponding land cover type. For water body and construction ((b) and (d), respectively, in Figure 8), their DTW distance maps show visibly different patterns than the other three. This can be explained by the SAR backscatter differences between vegetation and non-vegetation. In some parts of the map, DTW distances for paddy rice ((a) in Figure 8) and other crops ((e) in Figure 8) show a similar pattern of low value. Those pixels are likely to be agricultural land whose SAR profile lies between the standard paddy rice profile and other-crop profile. This ambiguity was avoided by the top- k nearest strategy to pick out the weak samples with the highest confidence.

For each land cover type, the top- k (5000 or 2000 for experiments) pixels with the lowest DTW distance were selected as weak samples. No conflict of labels was found in the selected weak samples since a relatively small amount (to the pixel number of the whole study area) of weak samples were picked.

3.2. LSTM Classification Results

We evaluated the experiments with four metrics, namely, overall accuracy (OA), producer's accuracy (PA) for paddy rice, user's accuracy (UA) for paddy rice, and Kappa coefficient. UA evaluates how often the class in the classification map will be present on the ground, and PA evaluates how often real land cover types in the study area are correctly shown on the classification map. They can be derived from the confusion matrix. PA and UA for paddy rice were of most concern since our goal was solely focusing on paddy rice mapping. The evaluation of three experimental schemes is shown in Table 4 (PA and UA for paddy rice are emphasized in bold and in tables thereafter).

Paddy rice PA and UA for supervised learning in scheme 1 were 0.904 and 0.917, respectively, whereas for weakly supervised learning were 0.981 (improved by 0.77) and 0.961 (improved by 0.44), respectively. However, supervised learning surpassed weakly supervised in OA (0.937 to 0.854) and Kappa (0.921 to 0.817), which indicated that weak samples introduced errors in land cover types other than paddy rice. Further analysis of classification results for scheme 1 was listed in Tables 5 and 6. Compared with supervised learning, PA and OA for paddy rice were notably improved. Meanwhile, PA and OA for vegetation and other crops were depressed. The expanded weak samples might introduce errors in labels for vegetation and other crops since they could have similar time series SAR profiles (Figure 4). Scheme 1 proves that with a small number of field samples for training, the proposed weakly supervised approach had evident improvement in paddy rice identification, regardless of possibly worse identification of other land cover types.

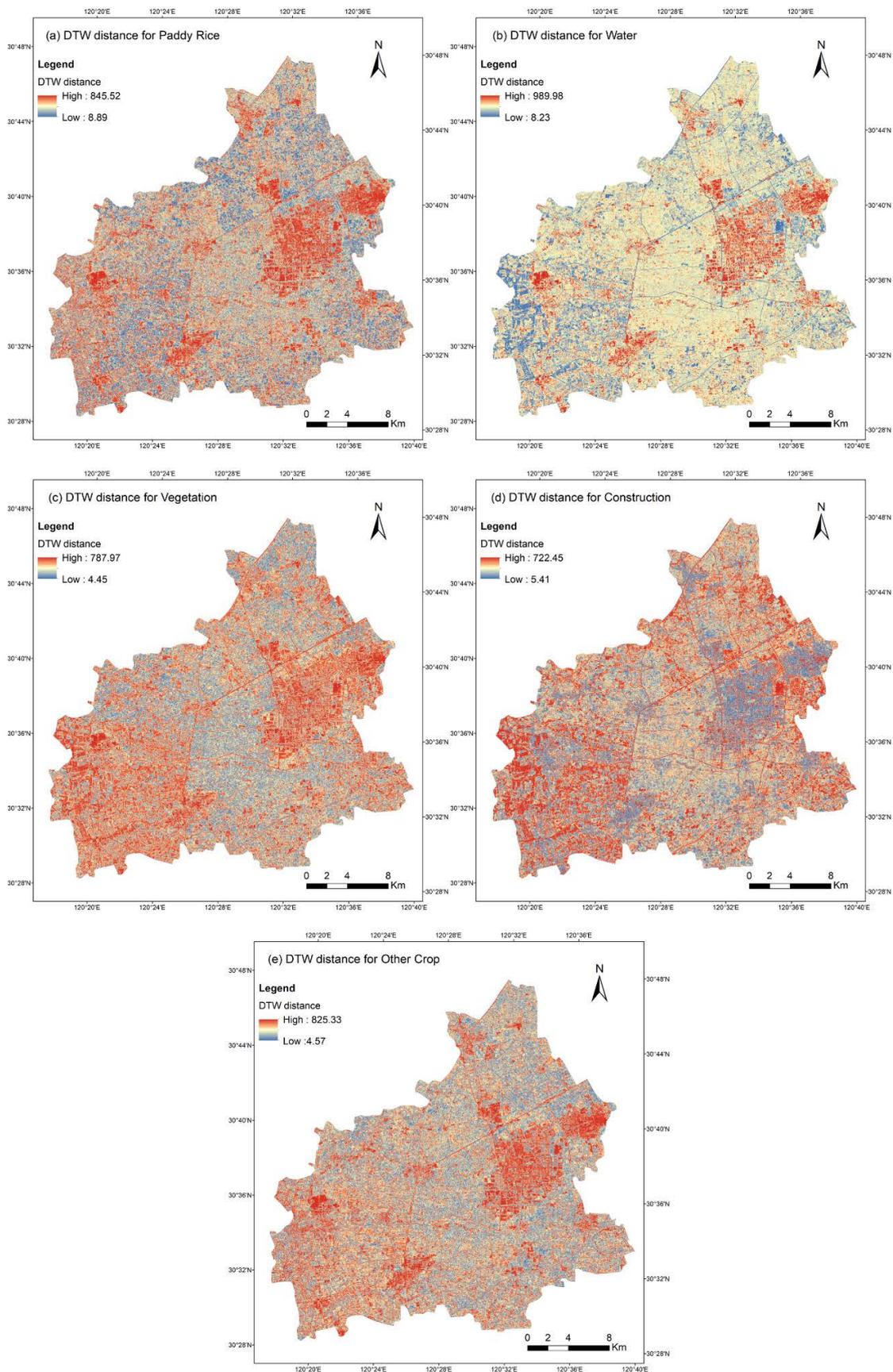


Figure 8. DTW distances of SAR time series to standard SAR backscatter profiles. (a–e) are five DTW distance maps for each land cover type respectively.

Table 4. Evaluation of three experiment schemes.

Experiment		OA	Paddy Rice PA	Paddy Rice UA	Paddy Rice Support	Kappa
Scheme 1	Supervised	0.937	0.904	0.917	2281	0.921
	Weakly supervised	0.854	0.981	0.961	2281	0.817
Scheme 2	Supervised	0.973	0.968	0.972	2281	0.966
	Weakly supervised	0.986	0.985	0.993	2281	0.982
Scheme 3	Supervised	0.982	0.988	0.987	913	0.978
	Weakly supervised	0.989	0.996	0.984	913	0.986

Table 5. Confusion matrix for supervised learning in scheme 1.

	Paddy Rice	Vegetation	Water	Construction	Other Crops
Paddy rice	2092	54	0	23	112
Vegetation	42	2276	3	4	66
Water	5	0	2118	24	0
Construction	3	0	21	3015	2
Other crops	173	158	7	72	1885
PA	0.904	0.915	0.986	0.961	0.913
UA	0.917	0.948	0.986	0.99	0.821
OA			0.937		
Kappa			0.921		

Table 6. Confusion matrix for weakly supervised learning in scheme 1.

	Paddy Rice	Vegetation	Water	Construction	Other Crops
Paddy rice	2193	12	0	6	70
Vegetation	6	2066	0	58	261
Water	6	0	2091	16	34
Construction	0	50	2	2742	247
Other crops	30	775	39	163	1288
PA	0.981	0.712	0.981	0.919	0.678
UA	0.961	0.864	0.974	0.902	0.561
OA			0.854		
Kappa			0.817		

Scheme 2 used more field samples for training than scheme 1. The experiment results also showed improvement with weakly supervised learning in paddy rice PA and UA, although not as much as scheme 1. PA was improved by 0.017 from 0.968 to 0.985, and UA improved by 0.021 from 0.972 to 0.993. This can be explained by the classification model. The proposed classification model was fairly fitted by training with 50% of field samples. No evident improvement was observed in scheme 3, while 80% of field samples were used for training. Paddy rice PA was slightly improved. Considering the relatively small sample number of the test set, this improvement was neglectable.

3.3. Paddy Rice Map

On the basis of the model evaluation, we implemented the weakly supervised approach in scheme 3 to generate a paddy rice map (Figure 9). The resulting map shows that the majority of paddy rice cultivation was in the north and west of the county. The total area of mapped paddy rice fields was 13,998.8 ha, slightly less than the government-published figure of 14,666 ha. With closer examination of four selected areas (Figure 10), we found that the resulting map depicts paddy rice fields with fair accuracy. Meanwhile, the precise boundaries of contiguous rice fields failed to be delineated due to the

resolution of the data source. Boundaries between rice fields are generally 1–3 m in width, and hence are unfeasible to be mapped with 10-m resolution images. The narrow boundaries were also counted as paddy rice fields in the produced map.

As we were focusing on paddy rice mapping, the classification accuracy of other land cover types was less regarded. With visual interpretation, water and construction type were well identified, while vegetation and other crops were mixed up in some areas. In some cases, these two broad categories could have a similar time series pattern in terms of spectral and radar signals.

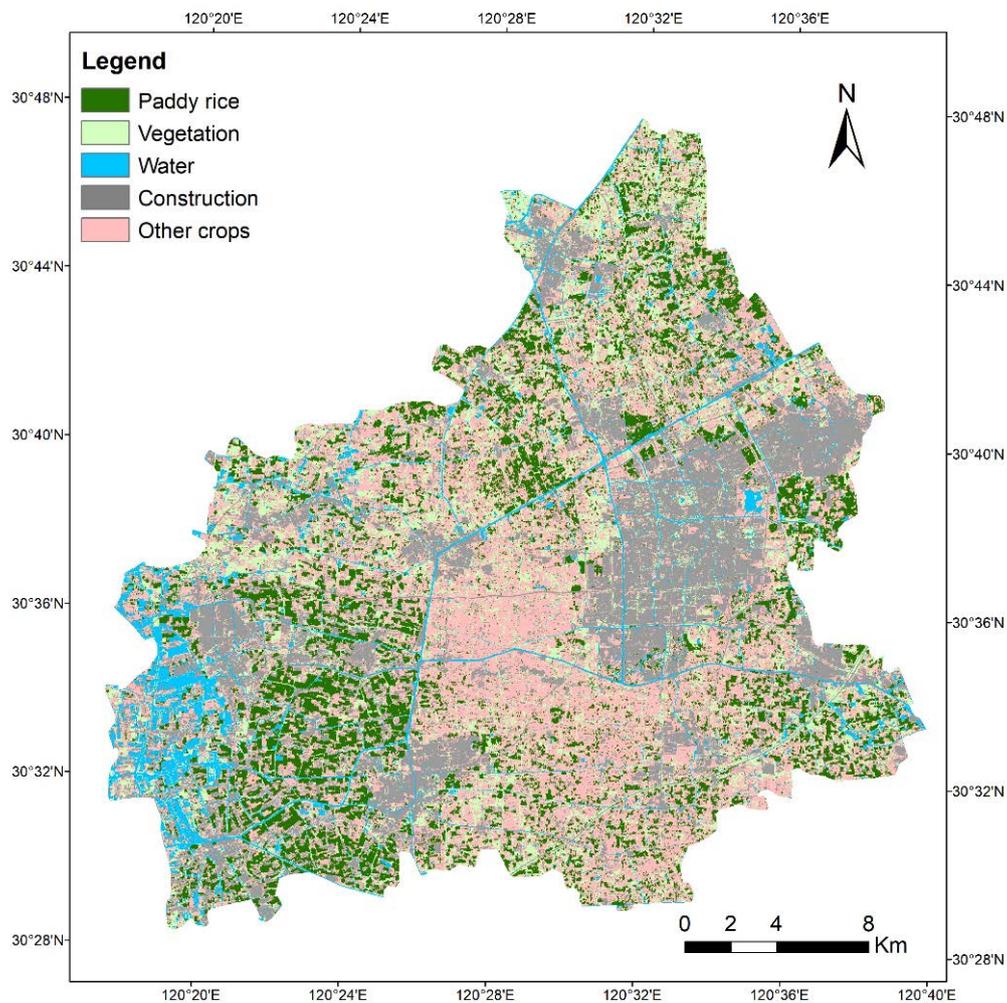


Figure 9. Classification map for paddy rice with weakly supervised LSTM deep learning classifier.

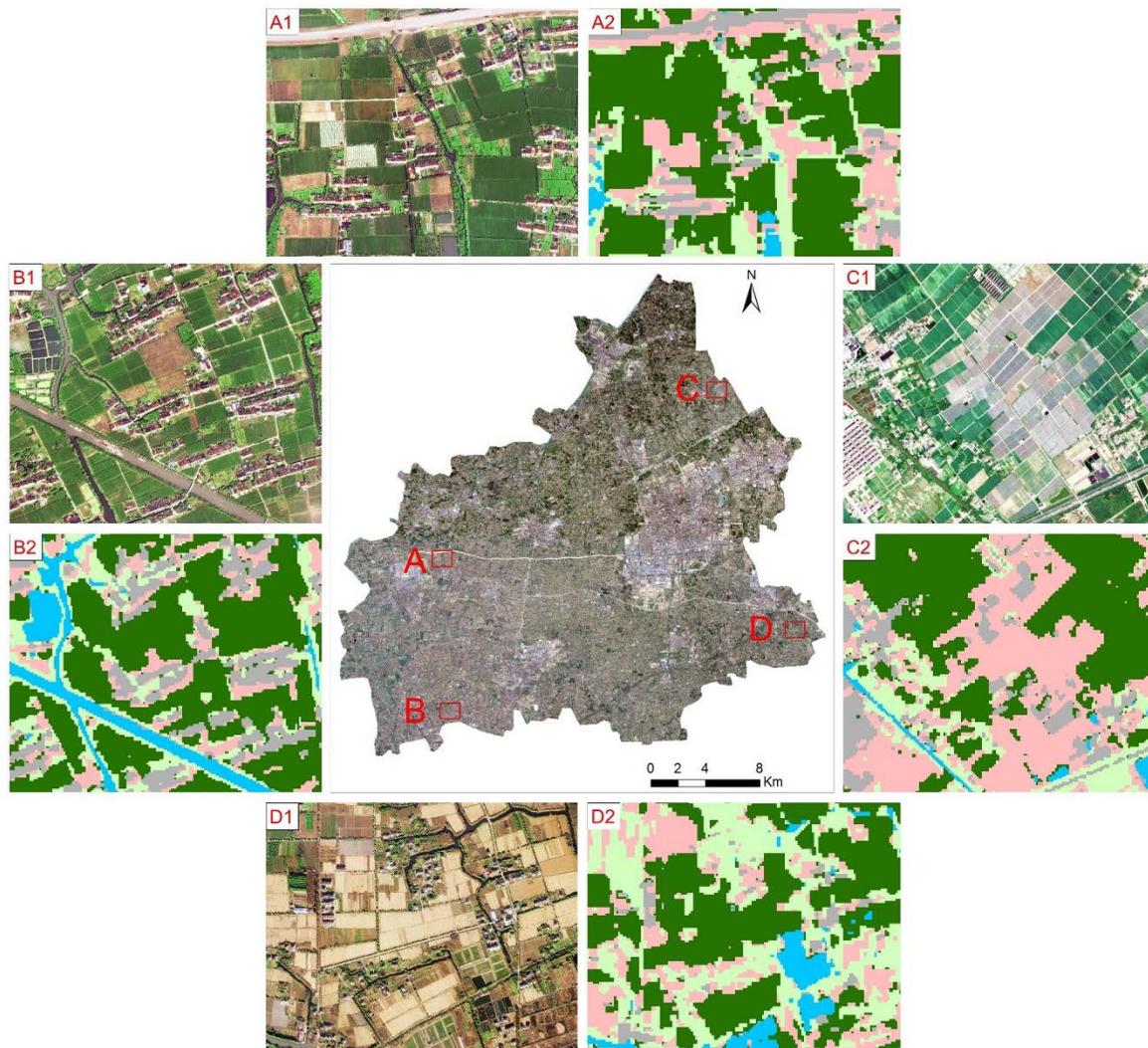


Figure 10. Classification result inspection on selected areas with high-resolution satellite images. **A1** is the high-resolution remote sensing image and **A2** is its classification result, vice versa for **B–D**. Images (**A1,B1**) were sensed in May 2018, (**C1**) was sensed in July 2018, and (**D1**) was sensed in December 2018. By visual interpretation, paddy rice fields in (**A1,B1,C1**) are displayed in green, while in (**D1**), paddy rice fields are displayed in yellow since the crop was harvested.

4. Discussion

This study mapped paddy rice fields using a weakly supervised LSTM deep learning network with time series optical and SAR remote sensing data. The motivation of this study was to show that deep learning classifiers have a strong capability in crop mapping, but often require a larger number of training samples. To solve this problem, a DTW distance-based sampling method was developed to deal with the time alignment in rice cultivation and label weak samples. Experiments on different numbers of training samples showed that the DTW distance-based sampling improved classification results for paddy rice, whereas the field samples were insufficient.

Previous studies, e.g., Zhong et al. [43] and Sun et al. [33], have applied RNNs in crop mapping, although not specifically for rice mapping. The effectiveness of using RNNs to classify croplands has been proven. However, the experiment results of those studies are not comparable to our study since different datasets and experiment schemes were used. Those studies relied on high-quality field survey data to provide a large number of training samples. Our experiments showed that with weakly supervised learning, only 10% of field samples as training data achieved comparable results (PA 0.981,

UA 0.961) to supervised learning on 50% of field samples (PA 0.968, UA 0.972), and weakly supervised learning on 50% field samples (PA 0.985, UA 0.993) had similar performance to supervised learning on 80% field samples (PA 0.988, UA 0.987). Although in scheme 1, weakly supervised learning with weak samples impaired the differentiation of vegetation and other crops, the goal of paddy rice mapping was promoted. The results imply that DTW-labeled paddy rice, water, and construction samples had high confidence and contributed significantly to the training of the RNN classifier.

Many factors could have an impact on the identification and mapping of paddy rice with multi-temporal remote sensing data. First, the resolution of satellite images creates the bottleneck of the precision of paddy rice mapping. The average size of paddy rice fields determines the appropriate image resolution to be used. The typical length of crop field sides is tens of meters in the study area. The boundaries between crop fields are narrow. Among the 11 input feature bands, 5 of them were in 10-m resolution and 6 of them were resampled to 10 m from 20 m. A mixed pixel problem still exists for small land parcels such as boundaries between croplands. Image segmentation on high-resolution images as supplementary data is a possible solution to alleviate this problem. The second factor is the continuity of time series data. Optical remote sensing data were critically confined by weather conditions in the study area. Sentinel-2 images of only 11 dates (less than a quarter of sensed images) from April to November were qualified for generating cloud-free data. With more complete time series images, the RNN classifier would yield more reliable results. This is a common problem for many studies using optical remote sensing data. The third factor is feature selection. Multi-temporal profile analysis of paddy rice widely utilizes temporal analysis of vegetation indexes (VIs) and water indexes (WIs) [19,44,45]. Some evident characteristics can be observed with VIs or WIs at a certain phase of paddy rice growth. For instance, in the heading phase, normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) show an increase and decrease variation, respectively, with a peak; land surface water index (LSWI) shows a jump in the transplanting phase due to water inundation [46]. As for the deep learning classifier in this study, we avoided VIs or WIs as input features due to the following concern. The range of VIs or WIs is different from that of spectral bands. If the VIs or WIs are rescaled to the range of spectral reflectance, the principle of the indexes is no longer valid. On the other hand, the neural networks would learn an implicit representation of raw input data, which is analogous to manually created features such as VIs and WIs.

Flooding signals are key features at early growing phases of paddy rice to differentiate from other crop types. Before transplanting, radar backscatter is low due to flooded fields. Radar backscatter increases during the vegetative phase, reaching a maximum at the heading stage [47]. We utilized the SAR backscatter coefficient as an input to the LSTM classifier. However, the power of SAR data as an input feature to identify the rice field was restrained by the number of valid optical images. Only a small portion of SAR images was used.

Most studies on rice mapping using multi-temporal data ignored planting schedule differences. Inputting all multi-temporal features as a whole into shallow machine learning classifiers, such as SVM and random forest, could impair the classifier's predictive ability. RNNs have strong adaptability to avoid this problem through learned memory control. Accordingly, many land cover classification studies using deep learning classifiers achieved better results compared to shallow classifiers. In the meantime, a larger number of training samples were required for the training of deep learning models. The proposed labeling method was especially designed for paddy rice, whose radar signal pattern through the growing season is unique. The experiment was conducted for a single-cropping regime in East China. However, the DTW-labeling strategy is valid for the multiple-cropping regime and other regions of the world, since it only relies on all-weather SAR remote sensing data. If applied to a multiple-cropping regime, each cropping regime should be categorized as an independent land cover type. For instance, single/double/triple-cropping rice fields should be separate land cover types. With enough labeled samples, state-of-the-art deep learning classifiers (non-sequential or sequential as in this study) could be chosen to adapt to a specific rice-growing region in the world.

5. Conclusions

This study proposed a paddy rice mapping approach based on a weakly supervised LSTM network with DTW distance-based sampling strategy. The purpose of this study was to enable a good fit for deep learning classifiers in the circumstance that field samples were not sufficient or to reduce the cost of field sampling. Apart from field samples, a larger number of weak samples were labeled on the basis of DTW distance to the standard SAR profiles of each land cover type. A specifically designed LSTM classifier was trained on combined optical spectral bands and SAR backscatter data as input features. A paddy rice map was finally generated with the best training scheme.

A few conclusions can be made on the basis of experiment results. First, the weakly supervised approach has a positive impact on paddy rice mapping when limited field samples are available. The DTW distance-based sampling strategy effectively reduced field sampling costs since fewer field samples were required. Second, the LSTM network classifier achieved high precision on paddy rice mapping. The mechanism of LSTM networks is feasible for paddy rice mapping where variance exists in terms of planting and harvesting schedules.

Author Contributions: Conceptualization, M.W.; data curation, L.C.; methodology, M.W.; project administration, J.W.; resources, L.C.; software, M.W.; writing—original draft, M.W.; writing—review and editing, J.W. and L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the China Central Public-Interest Scientific Institution Basal Research Fund (no. Y2020PT14), Basal Research Fund of AII CAAS (no. JBYW-AII-2020-17), Open Research Fund Program of Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, and Special Project of National Science and Technology Library (no. 2020QBW008).

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

References

- Sarris, A. Rice in Global Markets. In Proceedings of the FAO Rice Conference 2004, Rome, Italy, 12–13 February 2004.
- Muthayya, S.; Sugimoto, J.D.; Montgomery, S.; Maberly, G.F. An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1324*, 7–14. [[CrossRef](#)]
- Bouman, B. How much water does rice use. *Management* **2009**, *69*, 115–133.
- Dong, J.; Xiao, X. Evolution of regional to global paddy rice mapping methods: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 214–227. [[CrossRef](#)]
- Jin, C.; Xiao, X.; Dong, J.; Qin, Y.; Wang, Z. Mapping paddy rice distribution using multi-temporal Landsat imagery in the Sanjiang Plain, northeast China. *Front. Earth Sci.* **2016**, *10*, 49–62. [[CrossRef](#)] [[PubMed](#)]
- 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](#)]
- Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frolking, S.; Li, C.; Salas, W.; Moore, B., III. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sens. Environ.* **2005**, *95*, 480–492.
- Bazzi, H.; Baghdadi, N.; El Hajj, M.; Zribi, M.; Minh, D.H.T.; Ndikumana, E.; Courault, D.; Belhouchette, H. Mapping paddy rice using Sentinel-1 SAR time series in Camargue, France. *Remote Sens.* **2019**, *11*, 887. [[CrossRef](#)]
- Onojeghuo, A.O.; Blackburn, G.A.; Wang, Q.; Atkinson, P.M.; Kindred, D.; Miao, Y. Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1A and Landsat data. *Int. J. Remote Sens.* **2018**, *39*, 1042–1067. [[CrossRef](#)]
- Xiao, X.; Boles, S.; Frolking, S.; Li, C.; Babu, J.Y.; Salas, W.; Moore, B., III. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sens. Environ.* **2006**, *100*, 95–113. [[CrossRef](#)]
- Yin, Q.; Liu, M.; Cheng, J.; Ke, Y.; Chen, X. Mapping Paddy Rice Planting Area in Northeastern China Using Spatiotemporal Data Fusion and Phenology-Based Method. *Remote Sens.* **2019**, *11*, 1699. [[CrossRef](#)]
- Torbick, N.; Salas, W.A.; Hagen, S.; Xiao, X. Monitoring rice agriculture in the Sacramento Valley, USA with multitemporal PALSAR and MODIS imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *4*, 451–457. [[CrossRef](#)]

13. Zhang, Y.; Wang, C.; Wu, J.; Qi, J.; Salas, W.A. Mapping paddy rice with multitemporal ALOS/PALSAR imagery in southeast China. *Int. J. Remote Sens.* **2009**, *30*, 6301–6315. [[CrossRef](#)]
14. Aschbacher, J.; Pongsrihadulchai, A.; Karnchanasutham, S.; Rodprom, C.; Paudyal, D.; Le Toan, T. Assessment of ERS-1 SAR data for rice crop mapping and monitoring. In Proceedings of the 1995 International Geoscience and Remote Sensing Symposium, IGARSS'95, Florence, Italy, 10–14 July 1995; pp. 2183–2185.
15. Wu, F.; Wang, C.; Zhang, H.; Zhang, B.; Tang, Y. Rice crop monitoring in South China with RADARSAT-2 quad-polarization SAR data. *IEEE Geosci. Remote Sens. Lett.* **2010**, *8*, 196–200. [[CrossRef](#)]
16. Clauss, K.; Ottinger, M.; Künzer, C. Mapping rice areas with Sentinel-1 time series and superpixel segmentation. *Int. J. Remote Sens.* **2018**, *39*, 1399–1420. [[CrossRef](#)]
17. Nguyen, D.B.; Gruber, A.; Wagner, W. Mapping rice extent and cropping scheme in the Mekong Delta using Sentinel-1A data. *Remote Sens. Lett.* **2016**, *7*, 1209–1218. [[CrossRef](#)]
18. Le Toan, T.; Ribbes, F.; Wang, L.-F.; Floury, N.; Ding, K.-H.; Kong, J.A.; Fujita, M.; Kurosu, T. Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 41–56. [[CrossRef](#)]
19. Guan, X.; Huang, C.; Liu, G.; Meng, X.; Liu, Q. Mapping rice cropping systems in Vietnam using an NDVI-based time-series similarity measurement based on DTW distance. *Remote Sens.* **2016**, *8*, 19. [[CrossRef](#)]
20. Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [[CrossRef](#)]
21. Guan, X.; Liu, G.; Huang, C.; Meng, X.; Liu, Q.; Wu, C.; Ablat, X.; Chen, Z.; Wang, Q. An Open-Boundary Locally Weighted Dynamic Time Warping Method for Cropland Mapping. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 75. [[CrossRef](#)]
22. Li, M.; Bijker, W. Vegetable classification in Indonesia using Dynamic Time Warping of Sentinel-1A dual polarization SAR time series. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 268–280. [[CrossRef](#)]
23. Chen, C.-E.; Son, N.-T.; Chen, C.-R.; Chang, L.-Y. Wavelet filtering of time-series moderate resolution imaging spectroradiometer data for rice crop mapping using support vector machines and maximum likelihood classifier. *J. Appl. Remote Sens.* **2011**, *5*, 053525. [[CrossRef](#)]
24. McCloy, K.; Smith, F.; Robinson, M. Monitoring rice areas using Landsat MSS data. *Int. J. Remote Sens.* **1987**, *8*, 741–749. [[CrossRef](#)]
25. Clauss, K.; Yan, H.; Kuenzer, C. Mapping paddy rice in China in 2002, 2005, 2010 and 2014 with MODIS time series. *Remote Sens.* **2016**, *8*, 434. [[CrossRef](#)]
26. Küçük, C.; Taskın, G.; Erten, E. Paddy-rice phenology classification based on machine-learning methods using multitemporal co-polar X-band SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2509–2519. [[CrossRef](#)]
27. Park, S.; Im, J.; Park, S.; Yoo, C.; Han, H.; Rhee, J. Classification and mapping of paddy rice by combining Landsat and SAR time series data. *Remote Sens.* **2018**, *10*, 447. [[CrossRef](#)]
28. Teluguntla, P.; Thenkabail, P.S.; Oliphant, A.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K.; Huete, A. A 30-m Landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 325–340. [[CrossRef](#)]
29. Gumma, M.K.; Nelson, A.; Thenkabail, P.S.; Singh, A.N. Mapping rice areas of South Asia using MODIS multitemporal data. *J. Appl. Remote Sens.* **2011**, *5*, 053547. [[CrossRef](#)]
30. Wang, J.; Yang, Y.; Mao, J.; Huang, Z.; Huang, C.; Xu, W. CNN-RNN: A unified framework for multi-label image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26–30 June 2016; pp. 2285–2294.
31. 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](#)]
32. Zhang, M.; Lin, H.; Wang, G.; Sun, H.; Fu, J. Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China. *Remote Sens.* **2018**, *10*, 1840. [[CrossRef](#)]
33. Sun, Z.; Di, L.; Fang, H. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *Int. J. Remote Sens.* **2019**, *40*, 593–614. [[CrossRef](#)]
34. Clement, M.; Kilsby, C.; Moore, P. Multi-temporal synthetic aperture radar flood mapping using change detection. *J. Flood Risk Manag.* **2018**, *11*, 152–168. [[CrossRef](#)]

35. 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](#)]
36. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **1978**, *26*, 43–49. [[CrossRef](#)]
37. Maus, V.; Câmara, G.; Cartaxo, R.; Sanchez, A.; Ramos, F.M.; De Queiroz, G.R. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3729–3739. [[CrossRef](#)]
38. Petitjean, F.; Inglada, J.; Gancarski, P. Satellite image time series analysis under time warping. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3081–3095. [[CrossRef](#)]
39. Berndt, D.J.; Clifford, J. Using dynamic time warping to find patterns in time series. In Proceedings of the KDD Workshop, Seattle, WA, USA, 31 July 1994; pp. 359–370.
40. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
41. Understanding LSTM Networks. Available online: <https://colah.github.io/posts/2015-08-Understanding-LSTMs> (accessed on 16 June 2020).
42. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. Deep Learning. MIT press Cambridge: Cambridge, MA, USA, 2016; pp. 226–227.
43. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
44. Qiu, B.; Li, W.; Tang, Z.; Chen, C.; Qi, W. Mapping paddy rice areas based on vegetation phenology and surface moisture conditions. *Ecol. Indic.* **2015**, *56*, 79–86. [[CrossRef](#)]
45. Zhang, G.; Xiao, X.; Dong, J.; Kou, W.; Jin, C.; Qin, Y.; Zhou, Y.; Wang, J.; Menarguez, M.A.; Biradar, C. Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data. *Isprs J. Photogramm. Remote Sens.* **2015**, *106*, 157–171. [[CrossRef](#)]
46. Xiao, X.; Boles, S.; Froking, S.; Salas, W.; Moore, B., III; Li, C.; He, L.; Zhao, R. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. *Int. J. Remote Sens.* **2002**, *23*, 3009–3022. [[CrossRef](#)]
47. Chen, C.; McNairn, H. A neural network integrated approach for rice crop monitoring. *Int. J. Remote Sens.* **2006**, *27*, 1367–1393. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).