

Automatic rice fields mapping by Fusion of in-Decoder CNN and Data Augmentation Techniques on Landsat-8 multi-temporal images

Mahdiyeh Fathi¹, Reza Shah-Hosseini^{2*}

^{1,2}School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

Article history:

Received: 25 May 2020, Received in revised form: 1 August 2021, Accepted: 6 August 2021

ABSTRACT

Rice is the main food for the world's people. Monitoring and mapping rice fields play an important role in agricultural planning. Nowadays, intelligent management of rice fields has improved by remote sensing technology and deep learning algorithms. This research aims to study is the Fusion in-Decoder model and Data Augmentation techniques by using extracted multi-temporal maps of NDVI, LST, and LSWI indices from Landsat-8 images for mapping rice fields at the state of California, in 2020. Therefore, six architectures of Fusion in-Decoder model were designed, after radiometric corrections, atmospheric corrections, and generate multi-temporal maps of NDVI, LST, and LSWI indices, and simulation of different phenologies of rice crop with the shift of multi-temporal indices and PCA algorithm: (1) One Encoder-one Decoder (NDVI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (2) Two Encoders-one Decoder (NDVI-LST) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (3) Two Encoders-one Decoder (NDVI-LSWI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (4) Three Encoders-Decoder (NDVI-LST-LSWI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (5) Three Encoders-one Decoder (NDVI-LST-LSWI) and use of Data Augmentation technique by the shift of multi-temporal indices, and (6) Three Encoders-one Decoder (NDVI-LST-LSWI) without the use of Data Augmentation techniques. The fusion in-decoder and Data Augmentation techniques compared with four classifiers Decision Tree (DT), Logistic Regression (LR), Multi-Layer Perceptron (MLP), and Auto-Encoder (AE). The results showed that the Fusion in-Decoder model with three Encoders-one Decoder (NDVI-LST-LSWI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm performed best with Kappa coefficient (89/85%) for multi-temporal images of months April to August at the state of California. Besides, among the comparison classifiers, AE showed the worst result with Kappa coefficient (31.88%).

KEYWORDS

Rice Identification
Landsat-8
Data Augmentation
Fusion in-Decoder
LST.

1. Introduction

Rice is an important crop and the main food of more than half of the world's people, which needs water and heat to grow (Dong et al., 2016; S. Zhao et al., 2020). For example, in 2020, the value of United States rice exports to the world was \$1.9 billion. After Arkansas, California is the second producer of rice in the United States that generate 41.21

million CW¹ of rice in 2020. Thus, mapping and monitoring rice fields with efficient means such as remote sensing technology is necessary for food security, climate changes, methane greenhouse gas emissions during transplanting and flooding, crop disease, lack of sufficient water sources, etc (Dong et al., 2016). Mapping rice fields with different types of satellite images such as optical images (Landsat, Sentinel-2, SPOT-VGT, NOAA/AVHRR, MODIS, etc.) and radar

¹ CW (centum weight) =45.36 kg
(<https://www.statista.com/statistics/>)

images (ERS-1 and 2, RADARSAT-1 and 2, ENVISAT ASAR, Sentinel-1, etc.) has advantages and disadvantages. The optical images use due to their ability to view the earth's surface in the spectral range 0.4 to 2.5 μm for extracting multi-temporal maps of vegetation indices such as Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), Transformed Vegetation Index (TVI), Rice Growth Vegetation Index (RGVI), etc. Small rice fields are not detected by MODIS images due to low spatial resolution. The advantage of radar images in mapping rice fields is that they are independent of weather conditions such as cloud cover, rain, snow, and solar irradiance. The key factor in the mapping of rice fields using radar images is the temporal variation of radar backscatter over the growing season. The radar images for mapping rice fields are not suitable for reasons such as speckles noise, geometric distortions, low classification accuracy due to the use of single-polarization, and expensive mapping of large-scale areas. Overall to improve the mapping of rice fields uses combining optical and radar images (Mosleh, Hassan, & Chowdhury, 2015).

The extracted multi-temporal maps from vegetation indices are used to mapping rice fields. The extracted phenology from the multi-temporal maps of vegetation indices is called Land Surface Phenology (LSP), which is important for monitoring and mapping rice fields. Phenology is defined as periodic biological phenomena that are correlated with climatic conditions. Phenology parameters such as the beginning of the growing season, the end of the growing season, the maximum values of vegetation index, the length of the growing season, etc. are calculated by increasing and decreasing values of vegetation indices during the planting, holding, and harvesting of rice fields for a period. During the late May to early July period, the rice fields are flooded before planting, thus the values NDVI are less than the values Land Surface Water Index (LSWI) during this period. After the growth, NDVI values increase to reach their maximum values. Then, due to etiolation and senescence, vegetation index values decrease (Wang et al., 2015).

Different methods have been provided to mapping rice fields from satellite images. For example, Phenology-based algorithms have been used for mapping rice fields by setting a threshold for extracted indices from Landsat-7/8, MODIS, Pi-SAR-L2 images, etc (Ding et al., 2020; Liu et al., 2020; Lobell & Asner, 2004; Yonezawa & Watanabe, 2020). Further, Feature-based decision methods have been proposed to take advantage of extracted features such as backscatter difference, the time interval between vegetative growth and maturity stages, backscatter variation rate, average normalized backscatter, and maximum backscatter from Sentinel-1 images (Chang, Chen, Wang, & Chang, 2021). Also, algorithms such as Random forest (RF), Decision tree (DT), Classification And Regression Trees (CART) model,

Conventional Decision Tree method, Support Vector Machine (SVM), Multiclass relevance vector machine (mRVM), K-Nearest Neighbors (KNN), Fourier analysis, Wavelet Analysis, and Dynamic models have been proposed for mapping rice fields using polarimetric features (such as Entropy H, scattering angle, and Anisotropy HH / VV ratio, back-scattering of VH/VV channels), phenology features (such as sowing-transplanting slopes, yearly average, sowing-growing minimum, growing period slopes), and extracted indices from optical images (such as NDVI, EVI, RVI, MNDWI, and LSWI) (Chen et al., 2020; Inoue, Ito, & Yonezawa, 2020; Mansaray, Huang, Zhang, Huang, & Li, 2017; Talema & Hailu, 2020; Wang, Zang, & Tian, 2020; Yang et al., 2017; Zhan, Zhu, & Li, 2021).

According to that machine learning methods such as SVM, KNN, RF, etc. cannot fully extract the spectral and spatial features, deep learning algorithms have been proposed to improve mapping accuracy using high-level features extraction from low-level features in rice fields (Zhao et al., 2020). For example, Convolutional Neural Networks (CNNs) have been proposed for mapping rice fields using STARFM Spatio-temporal fusion techniques of Landsat-8 and MODIS images (Zhang, Lin, Wang, Sun, & Fu, 2018). Also, CNNs have been provided to identify five rice varieties in Australia by using Sentinel-2 MSI, with higher spatial and spectral resolutions than Landsat images (Guo, Jia, & Paull, 2018). One study has been done combining deep learning techniques such as Data Augmentation to overcome the limitations of data using Sentinel-1 images (Jo et al., 2020). The result showed that Data Augmentation had the best performance in improving mapping accuracy in rice fields (Jo et al., 2020). In another study, CNNs algorithms one-dimensional, two-dimensional, and three-dimensional have been developed using Sentinel-2 images. The results showed that two-dimensional Convolution Neural Networks had the best performance in mapping rice fields (Zhang, Liu, Wu, Zhan, & Wei, 2020). Also, combining pre-trained LeNet-5 and Decision Tree have been proposed using H_j1-A/B images (Zhao et al., 2020). This method contains 2 steps as follows: (1) The use of a pre-trained LeNet-5 to classify the crop class from other classes such as a river, forest, etc., and (2) The use of DT model with extracted phenological variables from multi-temporal maps of NDVI index to separate rice paddies from abandoned fields. As well, CNNs have been used to extract features such as abstract, shape, and amplitude from the multi-temporal curve of vegetation indices to identify rice fields at the pixel level using H_j1-A/B images (Jiang, Liu, & Wu, 2018). In a recent study, 1-D CNN and CNN-LSTM have been used to mapping rice fields at the state level using Sentinel-2 and Landsat-8 images. The results of these methods shown that the use of 1-D CNN could successfully identify rice fields in comparison to CNN-LSTM (Rawat, Kumar, Upadhyay, & Kumar, 2021).

In this paper, we present the Fusion in-Decoder networks

and Data Augmentation to an automatic mapping of rice fields using Landsat-8 optical images. Some key contributions of our work are as follows: (1) Multi-temporal maps of NDVI, LST, and LSWI indices were extracted from Landsat-8 multi-temporal images, (2) Data Augmentation techniques were used to simulate various phenologies, (3) Different combinations of multi-temporal images were used as the input of proposed method, (4) Fusion in-Decoder networks were designed, which consists several encoders and a decoder, (5) Thermal growth season characteristics were extracted from multi-temporal maps of Land Surface Temperature (LST) index, and (6) Flooding period characteristics were extracted from multi-temporal maps of Land Surface Water Index (LSWI). LST (surface skin radiative temperature) is an important parameter related to albedo, vegetation, and soil moisture, which is calculated from the thermal infrared band. The Land Surface Water Index (LSWI) that sensitive to the total amount of water in vegetation and its soil is calculated from the shortwave infrared (SWIR) and the NIR bands. Normalized Difference Vegetation Index is the most well-known index to detect

vegetation by using bands of NIR and Red.

INNOVATION. Our work innovations are as follows: (1) The use of Fusion in-Decoder CNN models, (2) Simulation of rice planting time in optical images with a shift of multi-temporal maps, (3) Improvement of edges by using PCA algorithm, and (4) Feature extraction from thermal growth season with Fusion in-Decoder Networks.

2. The Studied areas and Material

2.1. Study area

California is one of the western United States that covers Latitude of 32° 30' N to 42° N and Longitude of 114° 8' W to 124° 24' W. California is the 2nd producer of rice in the US for reasons such as well soil, enough water, and Mediterranean climate in summer (hot days and cool nights) (<https://www.statista.com/statistics/190823/top-us-states-for-rice-production/>). Glenn, Butte, Lake, Colusa, Yuba, Nevada, Sutter, Placer, Yolo, Napa, Sacramento, El Dorado, Placer, and Nevada are the study counties of rice producers in California (Figure 1).

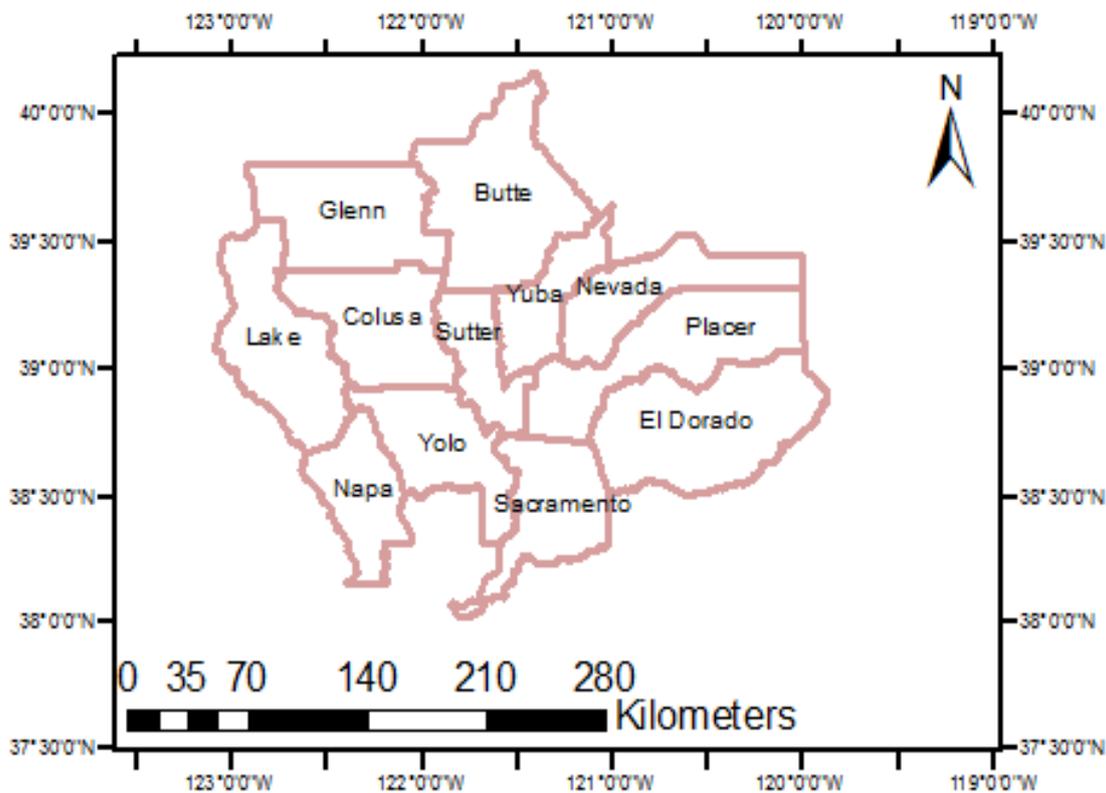


Figure 1. Study counties

2.2. Data Set

In the study, Landsat-8 OLI images were downloaded from USGS². The cloudless images were used from April to October 2020. Landsat-8 images were pre-processed by radiometric calibration and atmospheric correction using the Quick Atmospheric Correction (QUAC) algorithm (Bernstein, Jin, Gregor, & Adler-Golden, 2012). Finally, three indices including Land Surface Temperature (LST) (Artis & Carnahan, 1982), Land Surface Water Index (LSWI) (Chandrasekar, Sesha Sai, Roy, & Dwevedi, 2010), and Normalized Difference Vegetation Index (NDVI) (Rouse, Haas, Schell, & Deering, 1974) were calculated. LST is calculated by using Equations 1-6:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

$$L_{\lambda} = gain \times DN + offset \quad (2)$$

$$BT = \frac{K_2}{\ln(\frac{K_1}{L_{\lambda}} + 1)} - 273.15 \quad (3)$$

$$K_1 = 666.09, K_2 = 1282.71 \quad (4)$$

$$PV = \left(\frac{NDVI - \min_{NDVI}}{\max_{NDVI} - \min_{NDVI}} \right)^2 \quad (4)$$

$$\varepsilon = 0.004 \times PV + 0.98 \quad (5)$$

$$LST = \frac{BT}{1 + \left(\frac{\lambda \times TB}{p} \right) \times \ln \varepsilon} \quad (6)$$

$$p = 1.438 \times 10^{-3} \text{ mk}$$

Where L_{λ} , DN, BT, and ε are spectral radiance, Digital Number, Brightness Temperature, Land Surface Emissivity respectively.

Therefore, a total of 27 extracted features from Landsat-8 images of nine-date were used to mapping rice fields over the study areas, and each date includes NDVI, LSWI, and LST indices (Figure 2). Used images are shown in Table 1.

Table 1. Used images

Number	Date
1	2020/04/03
2	2020/05/21
3	2020/06/06
4	2020/06/22
5	2020/07/08
6	2020/07/24
7	2020/08/09
8	2020/10/12
9	2020/10/28

2.3. Ground truth

Ground truth map with spatial resolution 30 m was downloaded from USDA³. Crop fields were classified based on the Maximum Likelihood classifier method using Landsat TM/ETM satellites, before 2006. But the Decision Tree method is used to classify crop fields by using Landsat 8 sensor, Disaster Monitoring Constellation DEIMOS-1, and

UK2, ISRO ResourceSat-2 LISS-3, and the ESA Sentinel-2 A/B sensors in USDA from 2006. Figure 2 shows the ground truth map.

3. Method

In this study, a new method by using Fusion in-Decoder, Data Augmentation techniques, and Landsat-8 optical images was proposed for automatic mapping of rice fields in the state of California. First, indices (NDVI, LSWI, and LST) were calculated using Landsat-8 images. Then phenology various was simulated by the shift of multi-temporal indices and the PCA algorithm. Finally, the Fusion in-Decoder model with different architectures was used for mapping rice fields. The flowchart of the proposed method is shown in Figure 3.

3.1 Feature selection

The selected features in the study included 3 indices (NDVI, LSWI, and LST) were extracted from the Landsat-8 multi-temporal images. LST and LSWI were used along with NDVI to improve mapping rice fields. Six experimental with different features and architectures were designed to evaluate the mapping of rice fields. The different architectures and models with different features are shown in Table 2.

3.2 Data Augmentation

Data augmentation techniques such as rotation, partitioning, and scaling, are used to overcome the limitations of training data. Also, the data augmentation techniques are applied via temporal differences and photometric differences (Jo et al., 2020). For example, the rice planting season in California varies from late May to early July, depending on the geographical conditions and type of rice. So a shift of multi-temporal indices values and a shift of each pixel value to new pixel values could minimize the phenological differences in various areas. Thus, temporal differences and phenological differences were simulated by moving back and forth between the multi-temporal maps of indices (Figure 4) and a shift of each pixel value to new pixel values by PCA algorithm (Taylor & Nitschke, 2018) (Algorithm 1 and Figure 5) in California.

Algorithm 1

Source indices: $NDVI^t, LST^t, LSWI^t$ and $t = [1, \dots, 7]$
 $M_{NDVI}, M_{LST},$ and $M_{LSWI} \leftarrow$ Create three $N \times 7$ matrices where the columns represent the multi-temporal indices values for each patch image (N: number of pixels)
 PCA is performed on $M_{NDVI}, M_{LST},$ and M_{LSWI} .
for all Pixels $M_{NDVI}(x, y), M_{LST}(x, y),$ and $M_{LSWI}(x, y)$
do
 $[M_{NDVI}_{k_y}^1, \dots, M_{NDVI}_{k_y}^t] \leftarrow$ Add $[p_{NDVI}^1, \dots, p_{NDVI}^t] \times [\alpha \beta_{NDVI}^1, \dots, \alpha \beta_{NDVI}^t]^T$

² <https://earthexplorer.usgs.gov/>

³ <https://nassgeodata.gmu.edu/CropScape/>

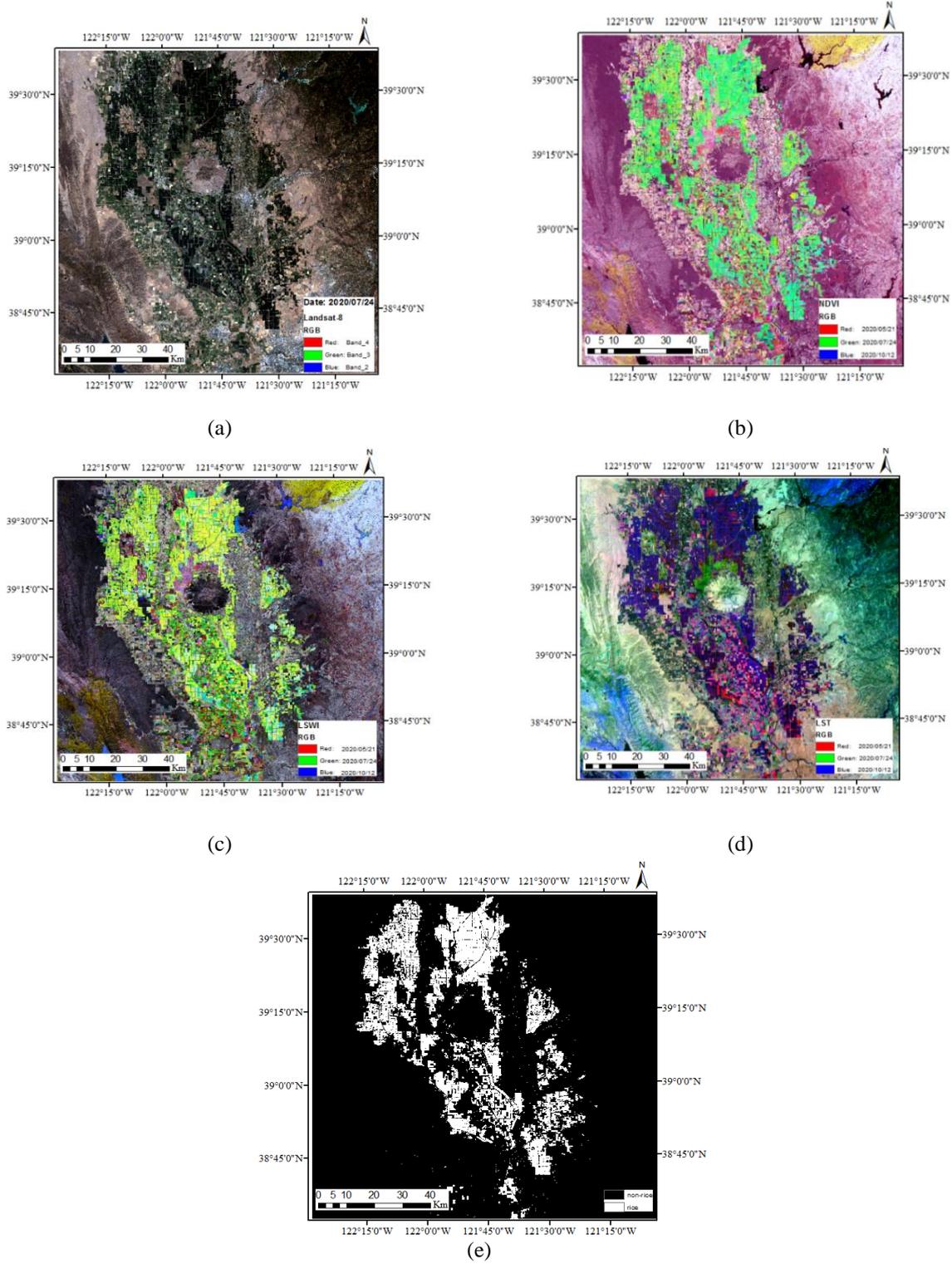


Figure 2. Study area. (a) RGB image, (b) NDVI image, (c) LSWI image (d) LST image, and (e) Ground truth map.

$[M_{LST_{xy}^1}, \dots, M_{LST_{xy}^t}] \leftarrow \text{Add } [p_{LST}^1, \dots, p_{LST}^t] \times$
 $[\alpha \beta_{LST}^1, \dots, \alpha \beta_{LST}^t]^T$
 $[M_{LSWI_{xy}^1}, \dots, M_{LSWI_{xy}^t}] \leftarrow \text{Add } [p_{LSWI}^1, \dots, p_{LSWI}^t] \times$
 $[\alpha \beta_{LSWI}^1, \dots, \alpha \beta_{LSWI}^t]^T$

- β^t is eigenvalue corresponding to the eigenvector p^t .

- Tistranspose.
- α is a random variable that is drawn from a Gaussian with 0 mean and standard deviation 0.1.

end for
return M_{NDVI} , M_{LST} , and M_{LSWI} .

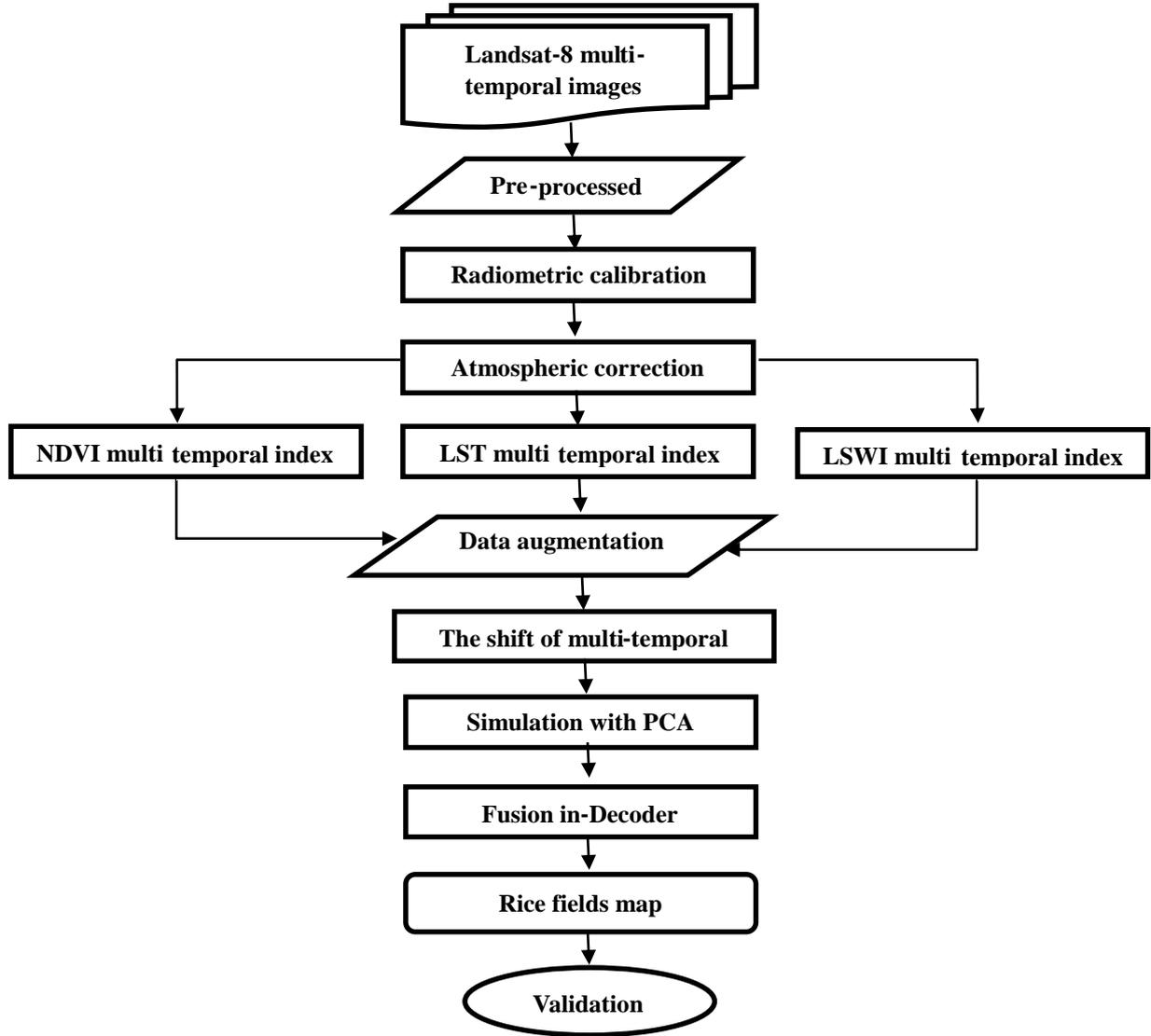


Figure 3. Flowchart of the proposed method.

3.3. Fusion in-decoder CNN

In this paper, we propose Fusion in-Decoder networks by using NDVI, LST, and LSWI multi-temporal indices. Fusion in-Decoder networks are a deep learning algorithm that fusions n Encoder at the end of the Encoding step (Liu et al., 2020). Thus, Fusion in-Decoder networks have two parts: n number Encoder that extracts features from input images, and a Decoder that reconstructs input images in output (Izacard & Grave, 2020). The proposed Fusion in-Decoder networks include two or three Encoders and one Decoder (Depending on the number of indices).

The proposed method is used from image patches with the size of $512 \times 512 \times 7$ or $512 \times 512 \times 9$ as input for each Encoder. For each image patch, feature maps are calculated by using Equations 7-9:

$$F_{NDVI}^l = \text{Pool}_{2 \times 2}(\max(0, F_{NDVI}^{l-1} * W_{NDVI}^l + B_{NDVI}^l)) \quad (7)$$

$$F_{LST}^l = \text{Pool}_{2 \times 2}(\max(0, F_{LST}^{l-1} * W_{LST}^l + B_{LST}^l)) \quad (8)$$

$$F_{LSWI}^l = \text{Pool}_{2 \times 2}(\max(0, F_{LSWI}^{l-1} * W_{LSWI}^l + B_{LSWI}^l)) \quad (9)$$

Where F^{l-1} , F^l , W^l , B^l , L , and $*$ are the input feature map, output feature map, weights, bias, layer number, and 2D Convolution block (Extraction of high-level features from low-level features using spatial relationships between pixels) for each Encoder, respectively (LeCun, Bengio, & Hinton, 2015). In our model, initial values of weight and bias are generated using the Glorot uniform method (Gao, Chai, & Liu, 2017). Also is used ReLU as the activation function (Agarap, 2018).

According to Equations 7-9, a max-pooling (Pool) layer with size 2×2 is performed to prevent over-fitting, reduce feature map dimensions and network parameters after the convolution layer. Then, extracted feature maps from the

corresponding layers of Encoders are concatenated (F^l) and

Table 2a. Experiment with different architectures with different features.

Experimental architectures	features	Feature dimensions
1 Encoder – 1 Decoder	NDVI	7 (with DA by the shift of multi-temporal indices and PCA algorithm)
2 Encoder – 1 Decoder	NDVI, LST	14 (with DA by the shift of multi-temporal indices and PCA algorithm)
2 Encoder – 1 Decoder	NDVI, LSWI	14 (with DA the shift of multi-temporal indices and PCA algorithm)
3 Encoder – 1 Decoder	NDVI, LST, and LSWI	21 (with DA by the shift of multi-temporal indices and PCA algorithm)
3 Encoder – 1 Decoder	NDVI, LST, and LSWI	21 (with DA by the shift of multi-temporal indices)
3 Encoder – 1 Decoder	NDVI, LST, and LSWI	27(without the use of DA)
1 Encoder – 1 Decoder	NDVI, LST, and LSWI	27(without the use of DA)

Table 2b. Experiment with other models (Compared methods).

Decision Tree (DT)	NDVI, LST, and LSWI	27(without the use of DA)
Logistic Regression (LR)	NDVI, LST, and LSWI	27(without the use of DA)
Multi-Layer Perceptron (MLP)	NDVI, LST, and LSWI	27(without the use of DA)
Autoencoder (AE)	NDVI, LST, and LSWI	27(without the use of DA)

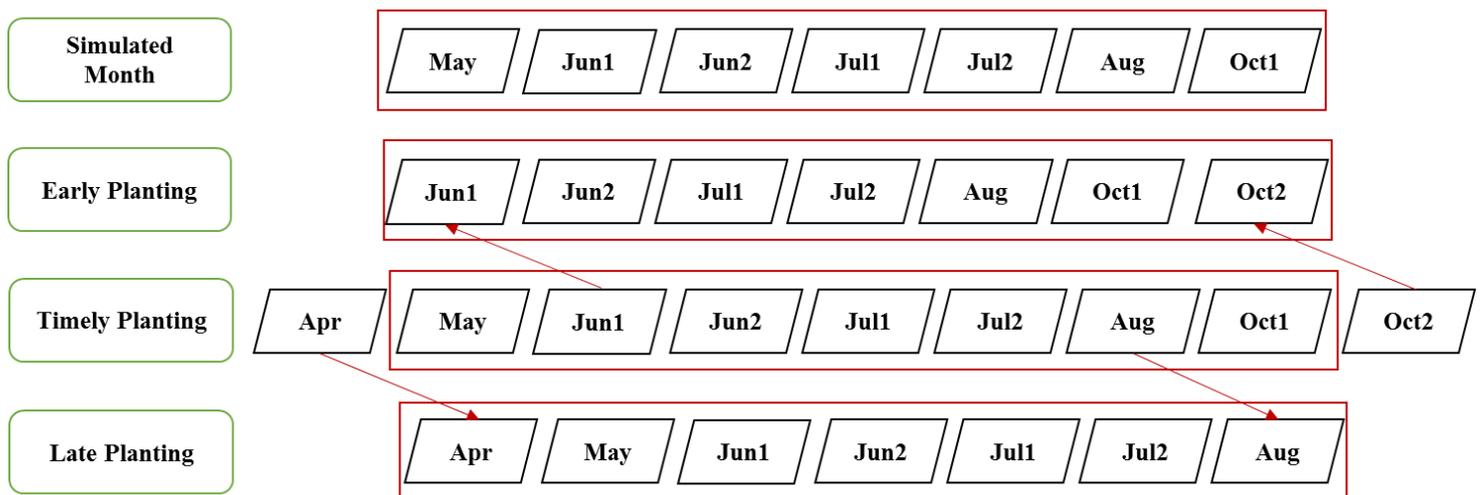


Figure 4. Data augmentation techniques by a shift of multi-temporal maps of indices.

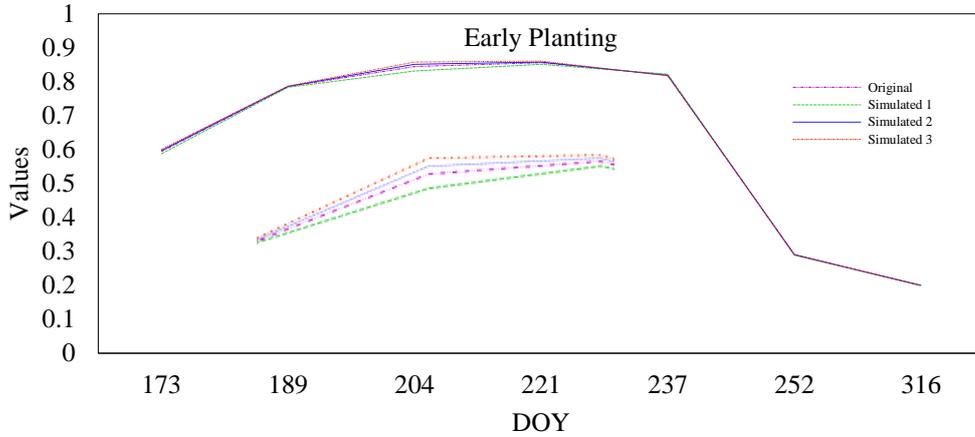


Figure 5. Example of PCA simulation for Early Planting.

the input image is reconstructed by using Equation 10:

$$y = \max(0, F^1 * \tilde{w}l + \tilde{b}l) \quad (10)$$

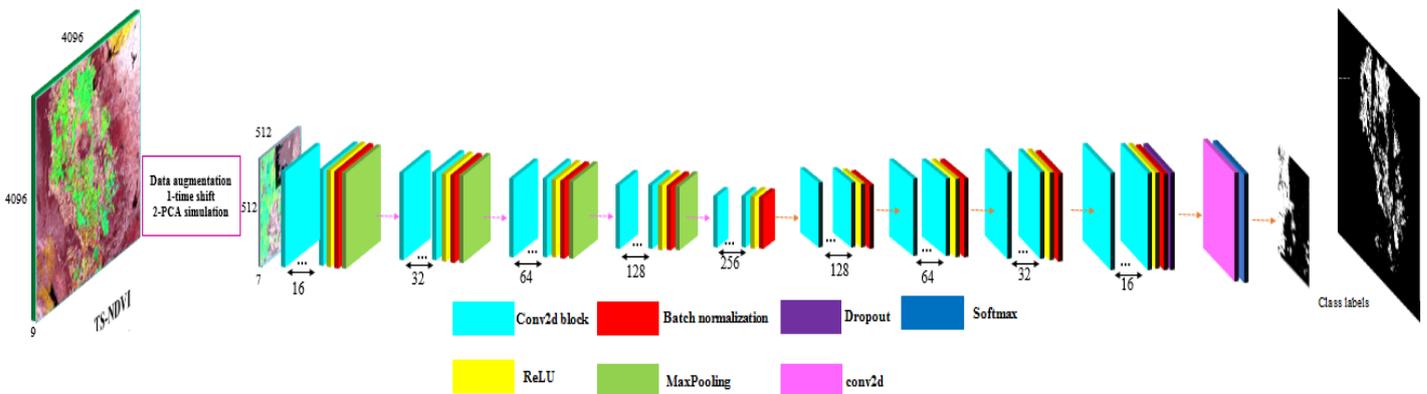
Where $\tilde{w}l$, $\tilde{b}l$, and y are the transposition of B , the transposition of W , and predicted value, respectively (Zhao, Guo, Yue, Zhang, & Luo, 2015). Finally, a 2D Convolution block with a Softmax activation function is used to generate the rice map (Dunne & Campbell, 1997).

In our method, the cross Entropy loss function is used to calculate weight and bias parameters, and the network is trained using the ADAM algorithm for 30 epochs (Wahlberg, Boyd, Annergren, & Wang, 2012; Zhang & Sabuncu, 2018).

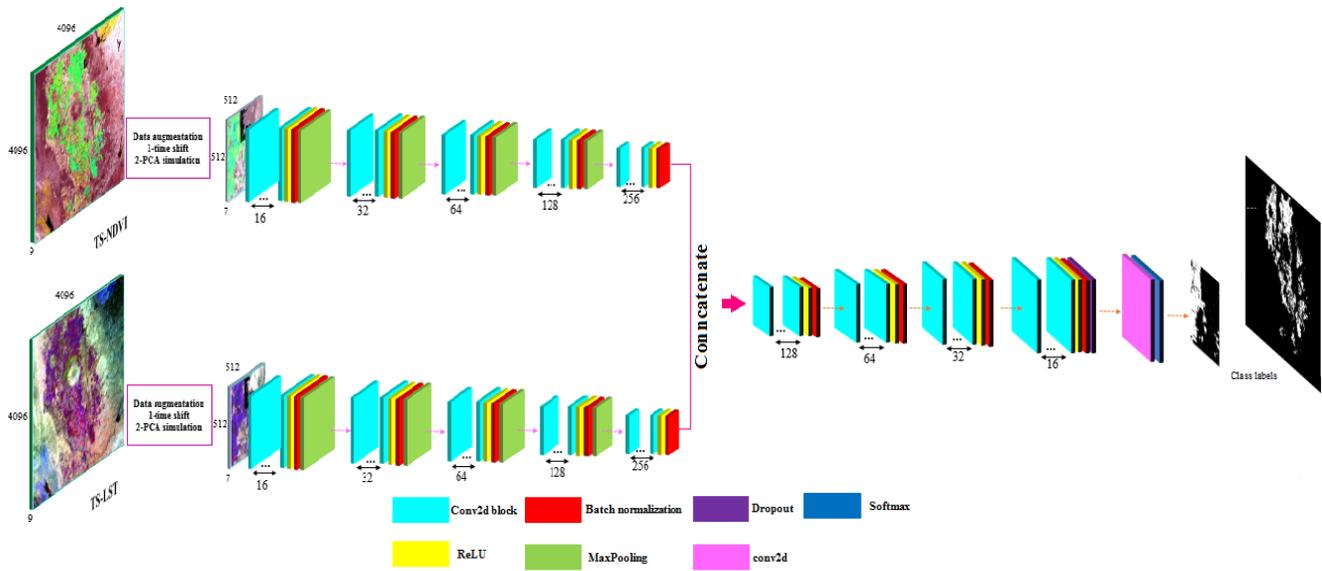
In addition, the Batch Normalization and Dropout techniques are added to improve the network performance. Dropout is applied to reduce overfitting and create different architectures by using removing neurons randomly in the last layer of each Encoder. Batch Normalization is applied to keep the distribution of the input values of each layer and increase the speed of learning (Garbin, Zhu, & Marques, 2020). Parameters input into the Fusion in Decoder model and architectures of Fusion in Decoder are shown in Table 3 and Figure 6.

Table 3. Parameters input into the Fusion in-Decoder (MLP and AE)

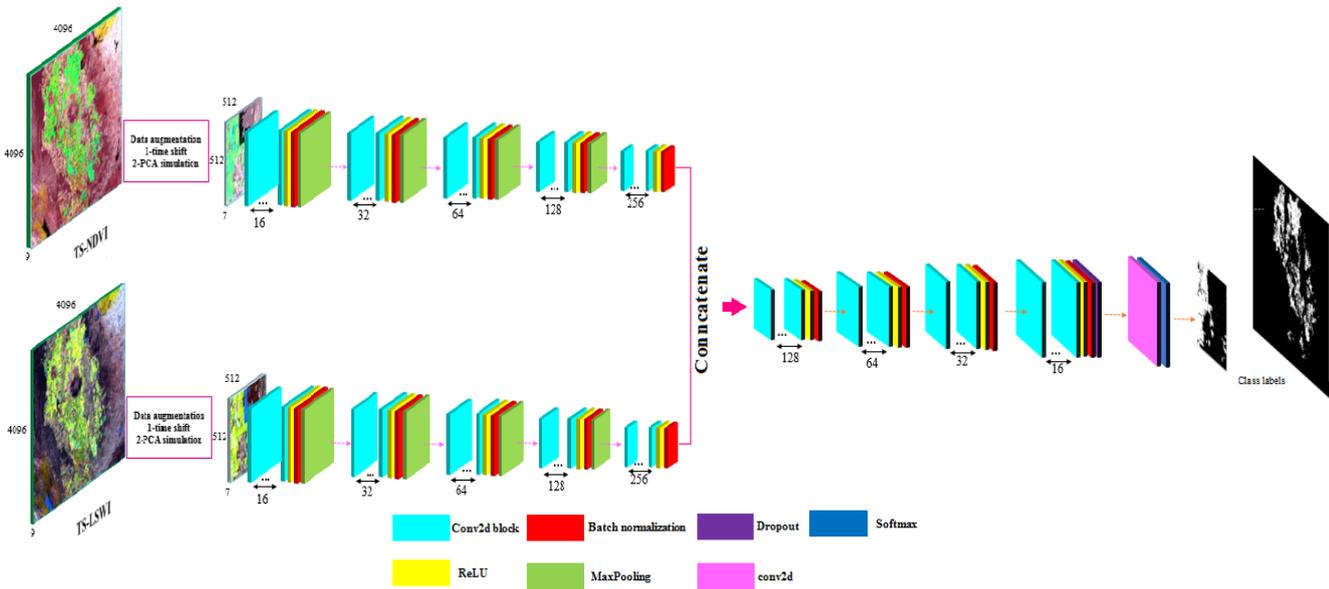
optimization Algorithm	Number of epochs	Patch size	Learning rate	Loss function	Initialization algorithm
ADAM	30	512*512	0.0001	Cross Entropy	Glorot uniform



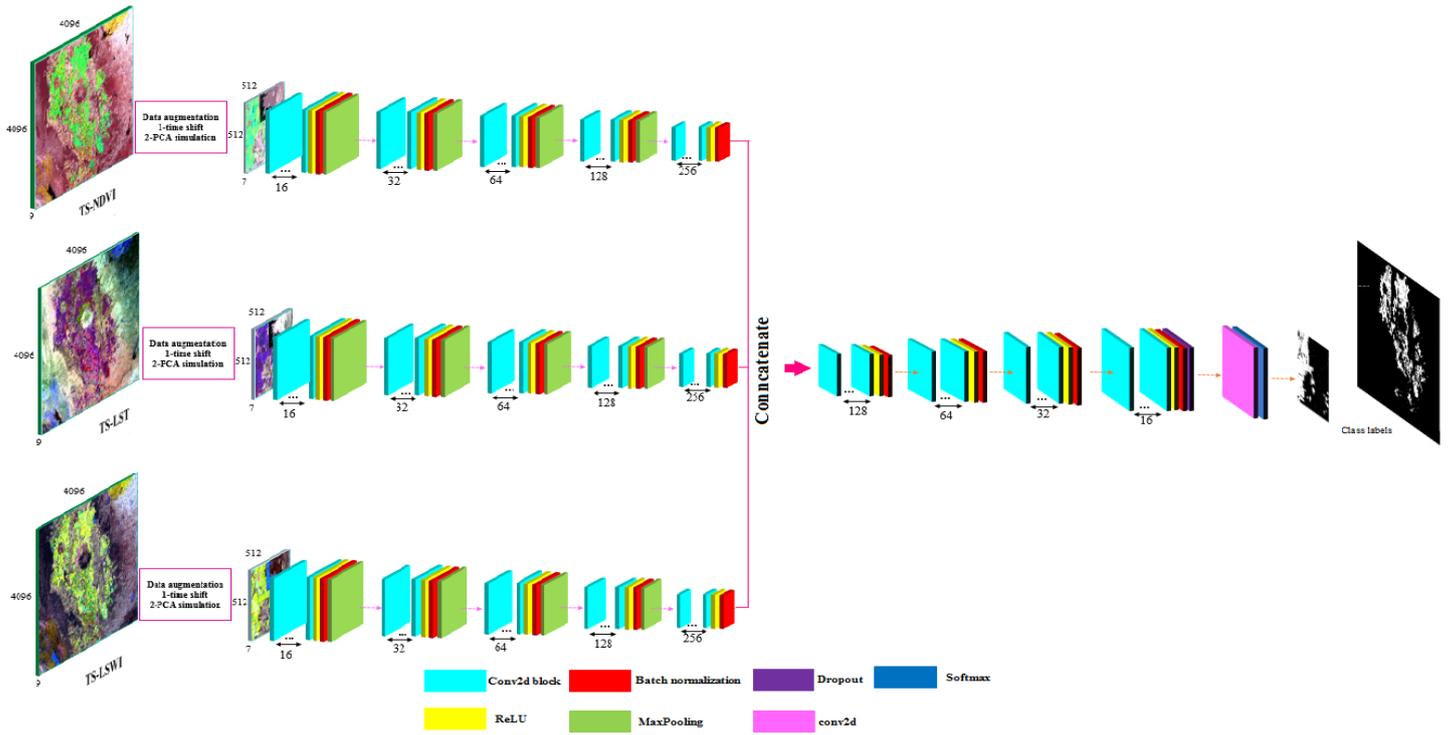
(a)



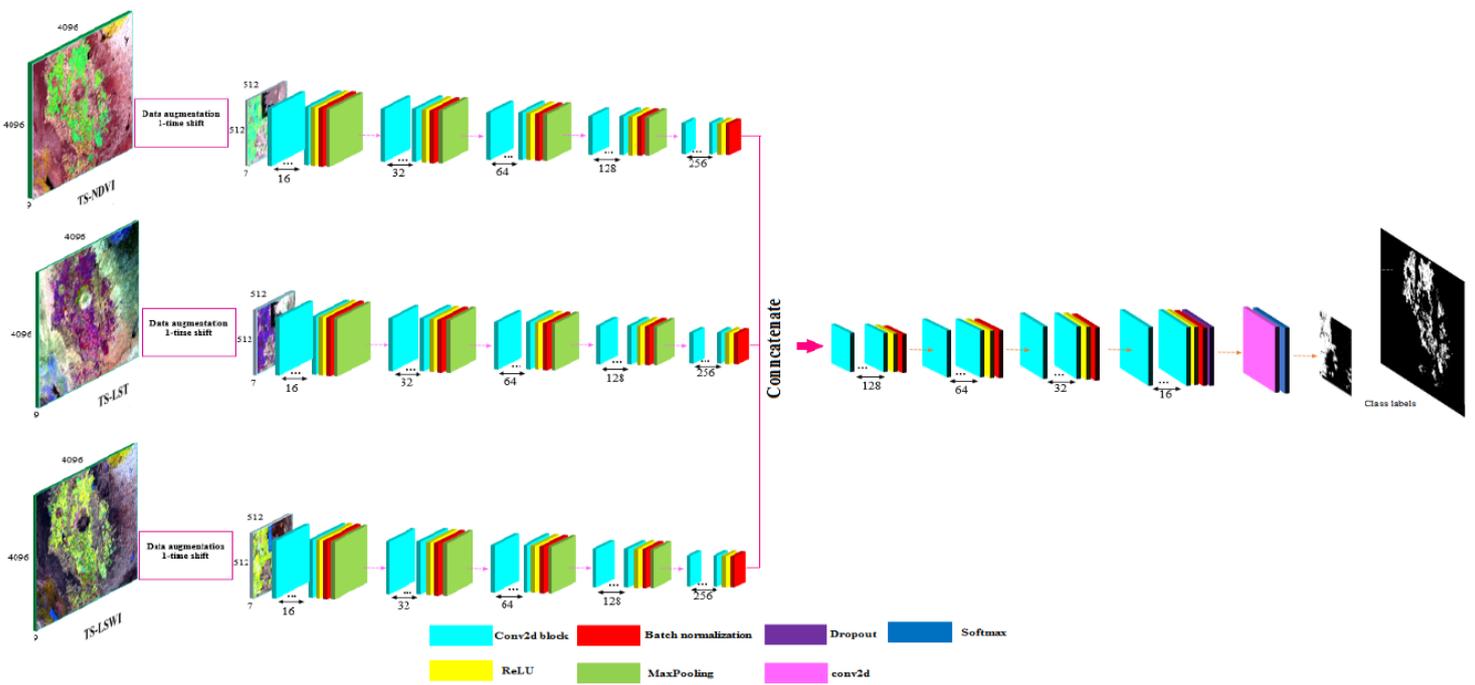
(b)



(c)



(d)



(e)

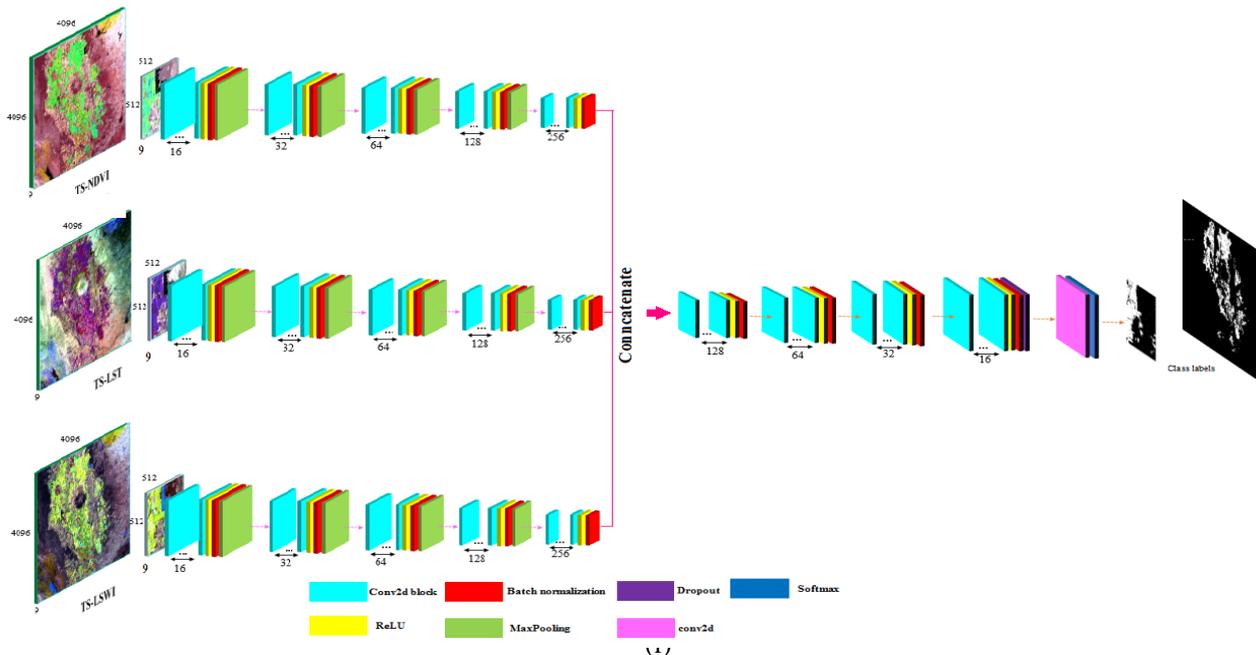


Figure 6. Architectures of Fusion in-Decoder. (a) 1 Encoder-1 Decoder (input feature: NDVI) and use of DA by the shift of multi-temporal indices and PCA, (b) 2 Encoder-1 Decoder (input features: NDVI and LST) and use of DA by the shift of multi-temporal indices and PCA, (c) 2 Encoder-1 Decoder (input features: NDVI and LSWI) and use of DA by the shift of multi-temporal indices and PCA, (d) 3 Encoder-1 Decoder (input features: NDVI, LST, and LSWI) and use of DA by the shift of multi-temporal indices and PCA, (e) 3 Encoder-1 Decoder (input features: NDVI, LST, and LSWI) and use of DA by the shift of multi-temporal indices, and (f) 3 Encoder-1 Decoder (input features: NDVI, LST, and LSWI) without the use of DA.

3.4. Compared Methods

The proposed method was compared with the DT (criterion=entropy, max_depth = 4), LR (C=0.01, solver=liblinear), MLP (Hidden layer1= 60 neuron and Relu activation function, Hidden layer2= 30 neuron and Relu activation function, and output layer= 1 neuron and Sigmoid activation function), and AE methods. Classification accuracies were assessed by Overall Accuracy (OA), Kappa coefficient, and F-score calculated using a confusion matrix.

4. Result

In this study, 1624940 and 738269 number rice pixels were selected to train and validation the Fusion in-Decoder model, respectively. The distribution of the training and validation sites is shown in Figure 7.

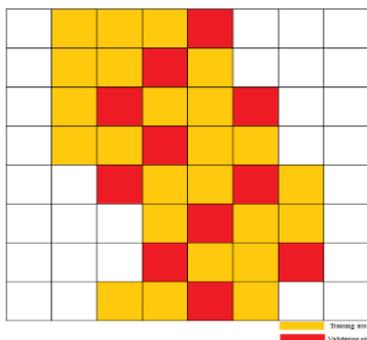


Figure7. Training (yellow) and validation (red) sites.

4.1. Mapping rice fields with proposed Method

Data Augmentation techniques by using the shift of multi-temporal indices and PCA algorithm were applied to training sites. Mapping of rice fields with different architectures and features was performed (Figure 6). To analyze of results, the Fusion in-Decoder model was designed in six different architectures: (1) One Encoder-one Decoder (NDVI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (2) Two Encoders-one Decoder (NDVI-LST), and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (3) Two Encoders-one Decoder (NDVI-LSWI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (4) Three Encoders-one Decoder (NDVI-LST-LSWI) and use of Data Augmentation techniques by the shift of multi-temporal indices and PCA algorithm, (5) Three Encoders-one Decoder (NDVI-LST-LSWI) and use of Data Augmentation technique by the shift of multi-temporal indices, and (6) Three Encoders-one Decoder (NDVI-LST-LSWI) without the use of Data Augmentation techniques. For the Fusion in-Decoder architectures, each Encoder consisted of 5 layers containing 2D Convolution block (with 2 convolution layers), ReLU activation function, Batch Normalization layer, and max pooling layer. The number of filters in each layer of the Encoder was 16, 32, 64, 128, and 256,

respectively. Also, each Decoder consisted of 4 layers containing a 2D Convolution block (with 2 convolution layers), ReLU activation function, and Batch Normalization layer. The number of filters in each layer of the Decoder was 128, 64, 32, and 16, respectively. In addition, a Dropout layer was used with the rate of 0.5 in the last Decoder layer. Following Decoder, the last layer contained 2D Convolution block (with 2 convolution layers) and Softmax activation function. The batch size and numbers of epochs were 5 and 30 in the final models. Training images were divided into three subsets and the values of the weight parameters in each subset were used as the initial values of the next subset. Also,

to analyze the effect of multi-temporal images combinations on the results of each of the Fusion in-Decoder architectures, different combinations of multi-temporal images were used: (1) Multi-temporal images related to dates 06/06, 06/22, 07/08, 07/24, 08/09, 10/12, and 10/28, (2) Multi-temporal images related to dates 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, and 10/12, (3) Multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09, and (4) Multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, 10/12, and 10/28. The proposed method results are shown with different architectures and features in Tables 4-7.

Table 4. Validation results for Multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, 10/12, and 10/28.

Data Augmentation (DA)	features	OA	Kappa	F score
Without DA	NDVI-LST-LSWI	96.39	83.93	87.06

Table 5. Validation results for Multi-temporal images related to dates 06/06, 06/22, 07/08, 07/24, 08/09, 10/12, and 10/28.

Data Augmentation (DA)	features	OA	Kappa	F score
The shift of multi-temporal indices	NDVI-LST-LSWI	97.13	89.07	91.73
The shift of multi-temporal indices & PCA algorithm	NDVI	96.35	84.13	87.45
	NDVI-LST	97.44	89.37	91.67
	NDVI-LSWI	97.09	88.56	91.17
	NDVI-LST-LSWI	97.36	89.69	92.06

Table 6. Validation results for Multi-temporal images related to dates 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, and 10/12.

Data Augmentation (DA)	features	OA	Kappa	F score
The shift of multi-temporal indices	NDVI-LST-LSWI	97.27	89.50	92.00
The shift of multi-temporal indices & PCA algorithm	NDVI	96.34	84.02	87.19
	NDVI-LST	97.56	89.83	91.99
	NDVI-LSWI	97.17	88.76	91.26
	NDVI-LST-LSWI	97.49	89.66	91.88

Table 7. Validation results for Multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09.

Data Augmentation (DA)	features	OA	Kappa	F score
The shift of multi-temporal indices	NDVI-LST-LSWI	96.94	86.21	88.92
The shift of multi-temporal indices & PCA algorithm	NDVI	96.11	83.16	86.47
	NDVI-LST	97.50	89.54	91.75
	NDVI-LSWI	97.09	88.72	91.23
	NDVI-LST-LSWI	97.51	89.85	92.02

According to Table 4, for multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, 10/12, and 10/28, the Kappa coefficient was 83.93% when NDVI, LST, and LSWI multi-temporal indices (three Encoder-one Decoder) without DA were used. According to Table 5, for multi-temporal images related to dates 06/06, 06/22, 07/08, 07/24, 08/09, 10/12, and 10/28, the highest Kappa coefficient was 89.69% when NDVI, LSWI, and LST multi-temporal indices (three Encoders-one Decoder) were used with DA by the shift of multi-temporal indices and PCA algorithm. According to Table 6, for multi-temporal images related to dates 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, and 10/12, the highest Kappa coefficient was 89.83% when NDVI and LST multi-temporal indices (two Encoders-one Decoder) were used with DA by the shift of multi-temporal indices and PCA algorithm. According to Table 7, for multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09, the highest Kappa coefficient was 89.85% when NDVI, LST, and LSWI multi-temporal indices (three Encoders-one Decoder) were used with DA by the shift of multi-temporal indices and PCA algorithm.

Results of Tables 4-7 show that the Fusion in-Decoder model with three Encoders-one Decoder (NDVI-LST-LSWI) and DA by the shift of multi-temporal indices and PCA algorithm for multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09 have higher Kappa coefficient (89.85%).

Accuracy and Loss curves are shown for the six architectures with kernel size 3×3 using training datasets in Figure 8. Generated rice maps by six Fusion in-Decoder architectures for test images are shown in Figures 9-11.

4.1. Mapping rice fields with compared Methods

The classification results of the proposed method were compared with those of other classifiers such as DT, LR, MLP, and AE (with one encoder – one decoder), using the 27 features (NDVI-LST-LSWI). Table 8 shows that the Fusion in-decoder networks and Data Augmentation have higher accuracy (OA=97.51, Kappa=89.85, and F-score=92.02) than the other four classifiers. Among the 4 compared methods, the auto-encoder (AE) method with one encoder- one decoder showed the lowest kappa coefficient (31.88%) that shows the effect of fusion in-decoder than the fused features.

5. Discussion

Remote sensing techniques have been proposed to mapping rice fields at the state and county level. Various sensors (optical and radar sensors) are used to mapping rice fields. The main challenges associated with mapping rice fields are (1) The cloud coverage in rice planting areas when the optical images are used, (2) Not identifying small rice fields when the MODIS images are used Due to low resolution, (3) Not identify of rice fields in the sloping area

the radar images are used, (4) The low accuracy of mapping rice fields with radar images compared to optical images, (5) Separation of rice fields from rainfed crops, (6) The spectral similarity of rice class with other classes, (7) Rice with different varieties, etc (Mosleh et al., 2015).

Machine learning and deep learning algorithms are used for mapping rice fields. In some research, machine learning algorithms are used to generate a map from rice fields. Deep learning methods are high accuracy than machine learning methods because high-level features are extracted from spectral bands and indices using deep learning algorithms. The accuracy of mapping rice fields is improved using deep learning algorithms with phenological parameters. One of the challenges of deep learning methods is to create distortions at the boundary of heterogeneous areas. However, the patch-based CNNs perform better than the pixel-based CNNs for mapping rice fields (Zhang et al., 2018)

One main problem associated with deep learning methods is the lack of enough training datasets Therefore, one solution for overcoming this problem would be Data Augmentation techniques. In this research, we used the shift of multi-temporal indices and PCA algorithm to simulate different phenology. We also used Fusion in-Decoder network with different architectures for mapping rice fields by using Landsat 8 multi temporal images at the several counties level. The numerical results from test data validated the efficiency of the Fusion in Decoder and Data Augmentation for mapping rice fields at the several counties level. A visual interpretation of the results in Figures 9-11 show some key advantages our work that is as follows: (1) The improve mapping results by using PCA algorithm in the edges of rice fields (for example Figure 12), (2) The remove some non-rice class (such as water class) from the rice final map by planting time simulation using the shift of multi-temporal indices (for example Figure 13), (3) The use of LST multi temporal indices for feature extraction of the thermal growing season (to remove rainfed crops from the final), (4) The use of LSWI multi temporal indices for feature extraction of the rice fields flooding, (5) The use of Fusion in-Decoder networks for effective feature extraction from each Index (NDVI, LST, and LSWI), and (6) Short run-time. The numerical results show the efficiency and superiority of the Fusion in Decoder networks with three Encoders one Decoder (NDVI-LST-LSWI) and Data Augmentation in terms of Kappa coefficient (89/85%) for Multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/2, and 08/09, compared with five other Fusion in-Decoder architecture. Some disadvantages of our work are as follows: (1) Non identify some of the rice fields and (Figure 14), and (2) Identify some of the non-rice crops in rice class due to neighbor of rice crop with other crops (Figure 15).

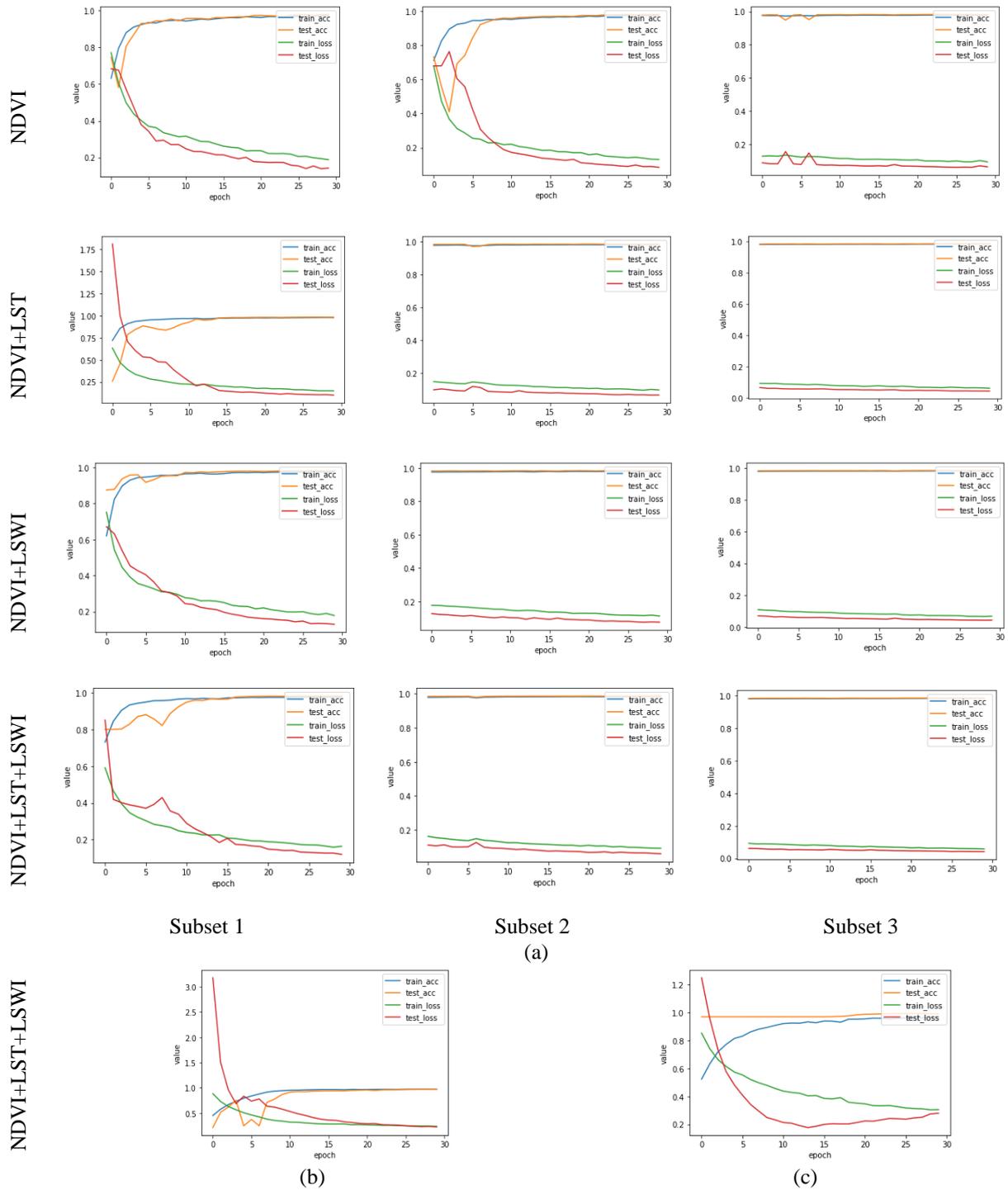
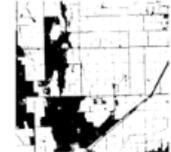
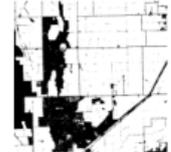
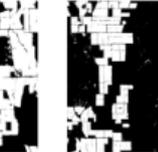
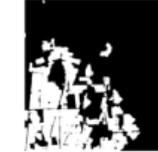
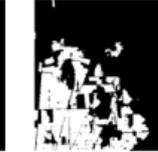
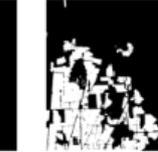
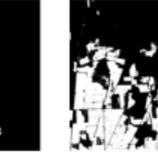
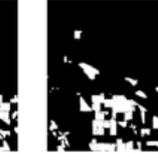
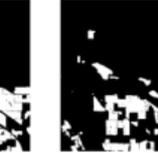
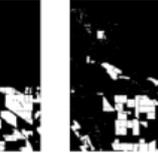
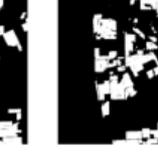


Figure 8. Accuracy and loss curves using training datasets for the Various experiments. (a) Data Augmentation by the shift of multi-temporal indices and PCA algorithm, (b) Data Augmentation by the shift of multi-temporal indices, and (c) Without Data Augmentation.

Multi-temporal images related to dates 06/06, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09

Without DA	DA by the shift of multi-temporal indices	DA by the shift of multi-temporal indices and PCA algorithm				Ground truth
	3Encoder-1Decoder NDVI-LST-LSWI	1Encoder-1Decoder NDVI	2Encoder-1Decoder NDVI-LST	2Encoder-1Decoder NDVI-LSWI	3Encoder-1Decoder NDVI-LST-LSWI	
						
						
						
						
						
						
						
						

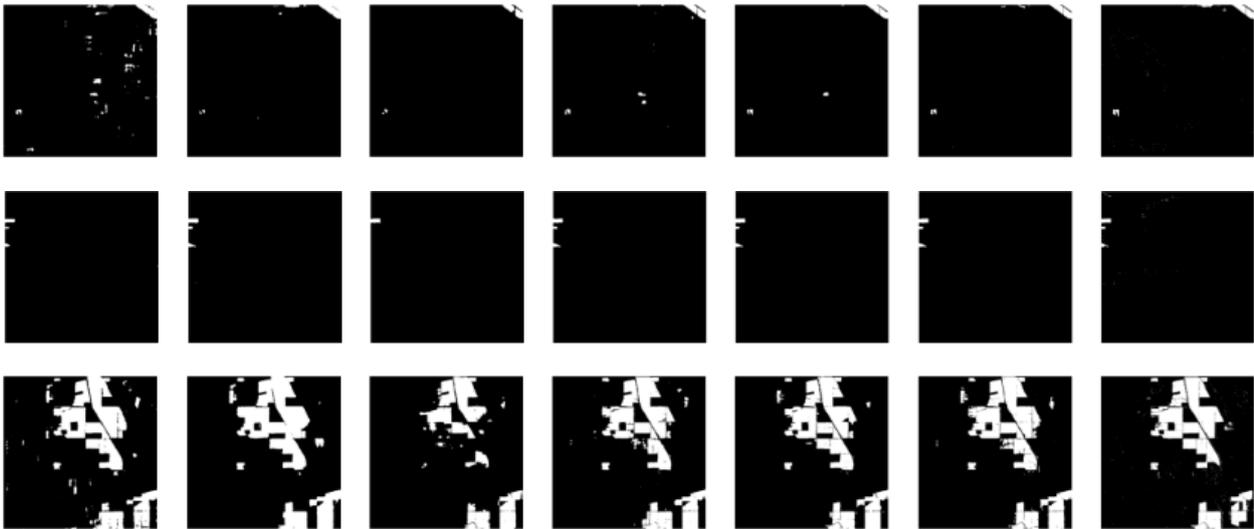


Figure 9. The generated results by Fusion in-Decoder architectures for multi-temporal images related to dates 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, and 10/12.

Multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09						
Without DA	DA by the shift of multi-temporal indices	DA by the shift of multi-temporal indices and PCA algorithm				Ground truth
	3Encoder-1Decoder NDVI-LST-LSWI	1Encoder-1Decoder NDVI	2Encoder-1Decoder NDVI-LST	2Encoder-1Decoder NDVI-LSWI	3Encoder-1Decoder NDVI-LST-LSWI	

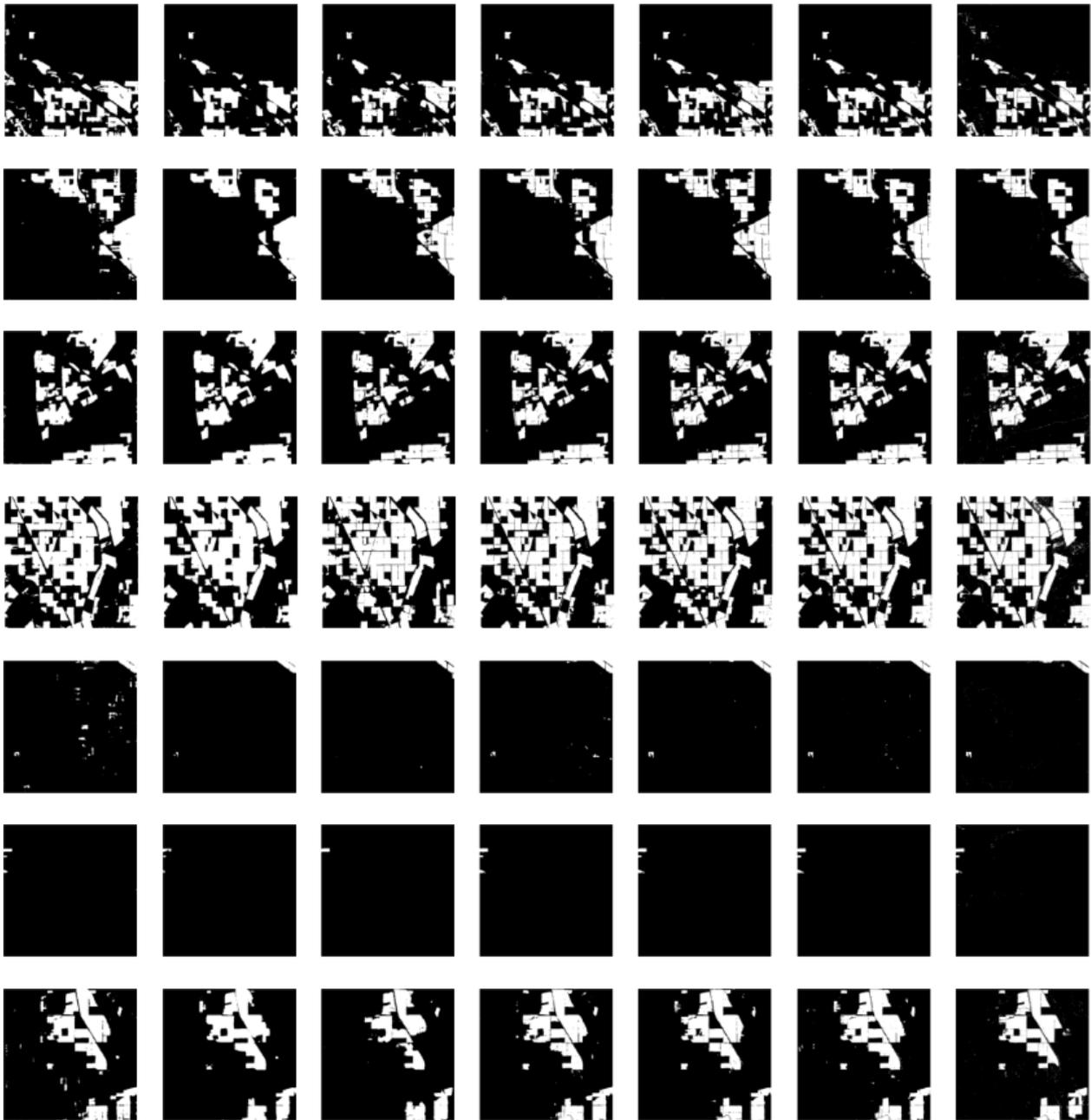


Figure 10. The generated results by Fusion in-Decoder architectures for multi-temporal images related to dates 04/03, 05/21, 06/06, 06/22, 07/08, 07/24, and 08/09.

Multi-temporal images related to dates 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, and 10/12						
Without DA	DA by the shift of multi-temporal indices	DA by the shift of multi-temporal indices and PCA algorithm				Ground truth
	3Encoder-1Decoder NDVI-LST-LSWI	1Encoder-1Decoder NDVI	2Encoder-1Decoder NDVI-LST	2Encoder-1Decoder NDVI-LSWI	3Encoder-1Decoder NDVI-LST-LSWI	

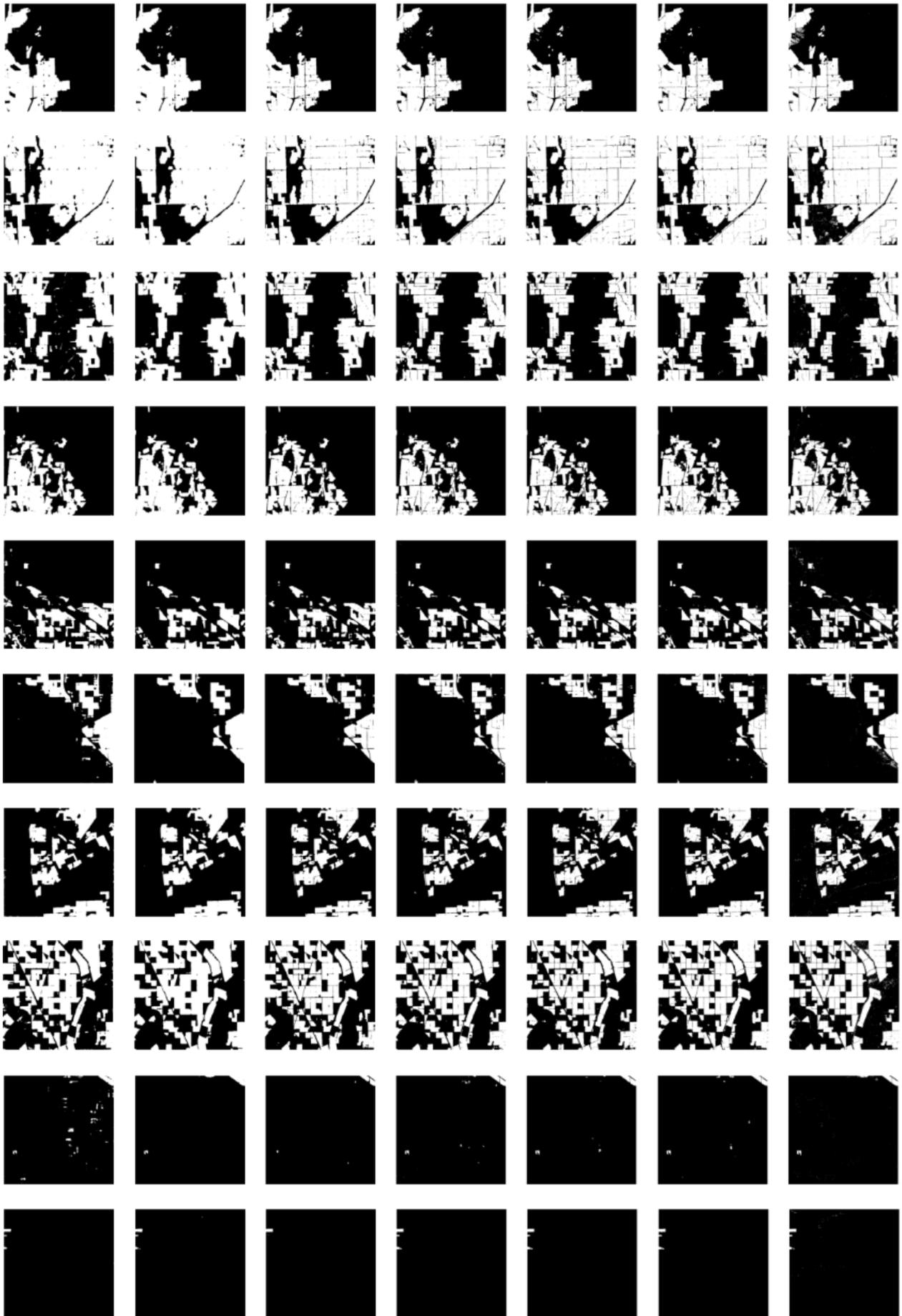




Figure 11. The generated results by Fusion in-Decoder architectures for multi-temporal images related to dates 05/21, 06/06, 06/22, 07/08, 07/24, 08/09, and 10/12.

Table 8. Validation results for compared methods

method	features	OA	Kappa	F score
DT	NDVI-LST-LSWI	96.51	84.91	88.06
LR	NDVI-LST-LSWI	96.78	87.39	90.29
MLP	NDVI-LST-LSWI	93.91	74.92	80.003
AE (Without DA)	NDVI-LST-LSWI	80.84	31.88	39.52

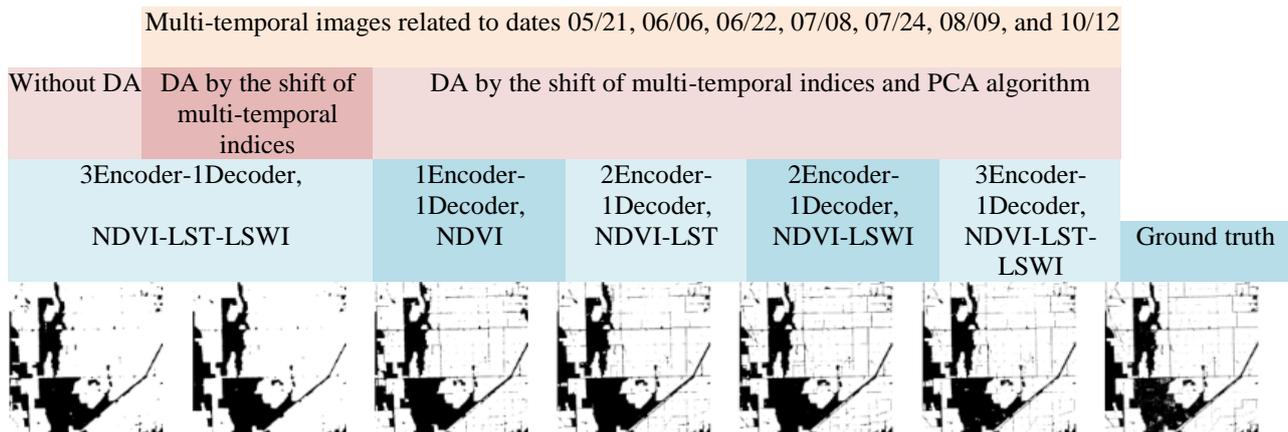


Figure 12. The improved results by using the PCA algorithm in the edges.

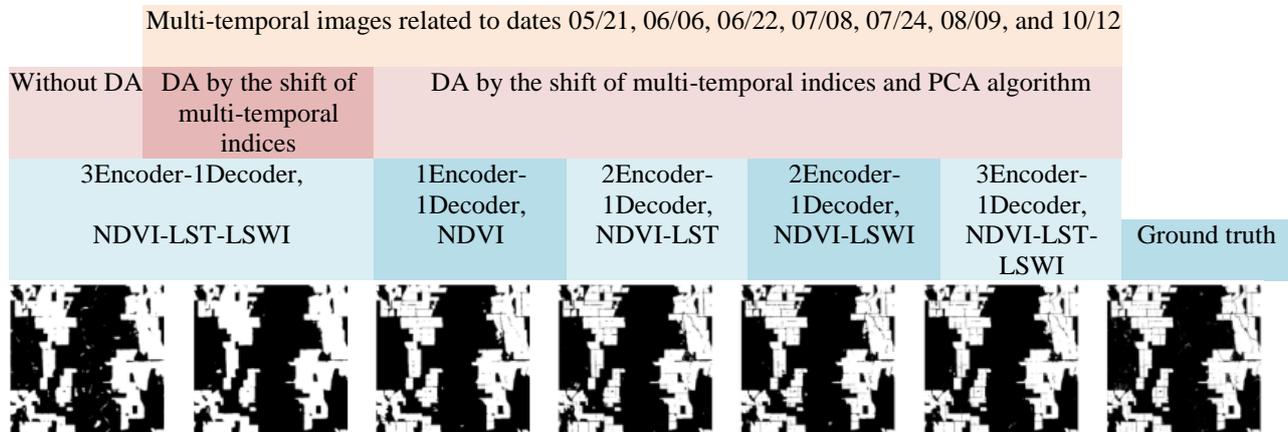


Figure 13. The improve results by using the shift of multi-temporal indices with remove non-rice fields.

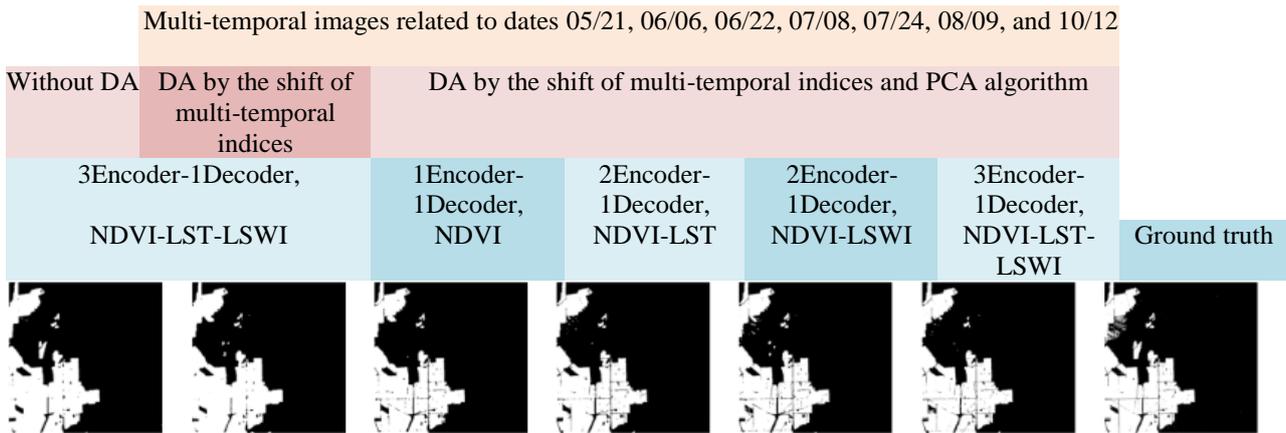


Figure 14. Non-identify some of the rice fields with the proposed method.

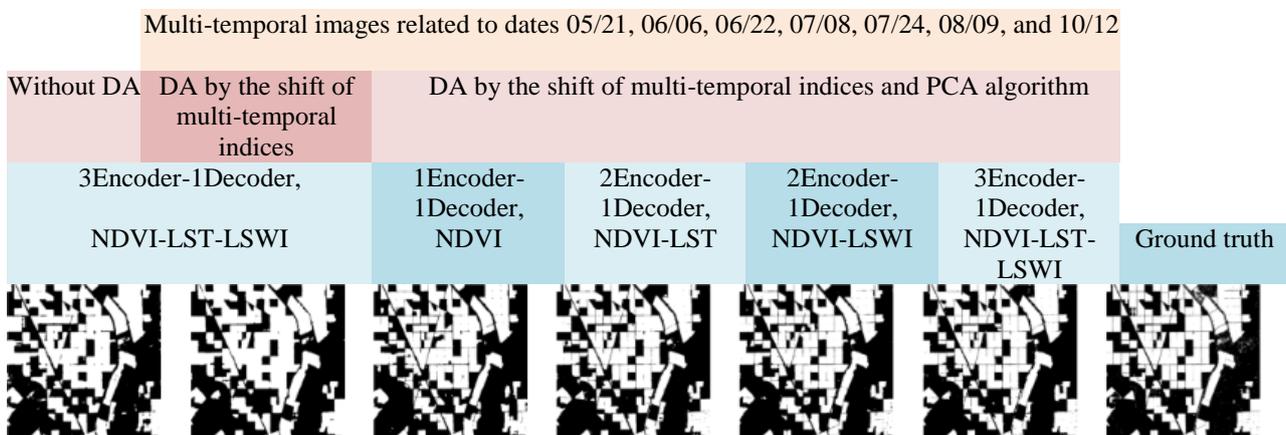


Figure 15. Identify some of the non-rice crops in the rice class with the proposed method.

6. Conclusions

Some researches prove that deep learning-based methods are highly accurate in mapping rice fields than other methods. Thus, in this paper, we proposed a new automatic method for mapping rice fields based on Fusion in-Decoder with different architectures and Data Augmentation techniques by using Landsat-8 multi-temporal images at the California State. The Fusion in-Decoder networks are constructed by fusion of several Encoders before the Decoding stage that learns features from indices such as NDVI, LST, and LSWI as its input, and increases the accuracy of mapping rice fields at the several counties level. This research shows these five results: (1) The improved mapping results using Data augmentation by PCA algorithm in the edges of rice fields, (2) The effect of different combinations of multi-temporal images on Overall Accuracy, (3) High accuracy of the proposed method compared to other methods based on deep learning and machine learning such as Multi-Layer Perceptron, Decision Tree, and Regression Logistic, (4) High Kappa coefficient of

mapping rice fields compared to the proposed method Jo et al., 2020 (86%), and (5) The highest Kappa coefficient (89.85%) for Fusion in-Decoder model with three Encoders-one Decoder (NDVI-LST-LSWI) and Data

Augmentation techniques using multi-temporal images of months April to August. This research focused on the mapping of rice fields based on Fusion in Decoder Networks, Data Augmentation techniques, and Landsat-8 multi-temporal images. Despite the high accuracy of the proposed method for overcoming not identifying rice fields with a smaller area, cloud cover, etc. using other Fusion Based Deep CNN structures using sentinel 1/2 will suggest in future research.

Acknowledgments

The authors thank the United States Department of Agricultural (USDA) and United States Digital Service (USDS) for providing the data to researchers.

References

- Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.
- Artis, D. A., & Carnahan, W. H. (1982). Survey of emissivity variability in thermography of urban areas. *Remote sensing of Environment*, 12(4), 313-329.
- Bernstein, L. S., Jin, X., Gregor, B., & Adler-Golden, S. M. (2012). Quick atmospheric correction code: algorithm description and recent upgrades. *Optical engineering*, 51(11), 111719.
- Chandrasekar, K., Sessa Sai, M. V. R., Roy, P. S., & Dwevedi, R. S. (2010). Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product. *International Journal of Remote Sensing*, 31(15), 3987-4005.
- Chang, L., Chen, Y. T., Wang, J. H., & Chang, Y. L. (2020). Rice-field mapping with sentinel-1A SAR time-series data. *Remote Sensing*, 13(1), 103.
- Chen, N., Yu, L., Zhang, X., Shen, Y., Zeng, L., Hu, Q., & Niyogi, D. (2020). Mapping paddy rice fields by combining multi-temporal vegetation index and synthetic aperture radar remote sensing data using google earth engine machine learning platform. *Remote Sensing*, 12(18), 2992.
- Ding, M., Guan, Q., Li, L., Zhang, H., Liu, C., & Zhang, L. (2020). Phenology-based rice paddy mapping using multi-source satellite imagery and a fusion algorithm applied to the poyang lake plain, Southern China. *Remote Sensing*, 12(6), 1022.
- Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., ... & Moore III, B. (2016). Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote sensing of environment*, 185, 142-154.
- Dunne, R. A., & Campbell, N. A. (1997, June). On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function. In *Proc. 8th Aust. Conf. on the Neural Networks*, Melbourne (Vol. 181, p. 185). Citeseer.
- Gao, T., Chai, Y., & Liu, Y. (2017, November). Applying long short term memory neural networks for predicting stock closing price. In *2017 8th IEEE international conference on software engineering and service science (ICSESS)* (pp. 575-578). IEEE.
- Garbin, C., Zhu, X., & Marques, O. (2020). Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, 79(19), 12777-12815.
- Guo, Y., Jia, X., & Paull, D. (2018, December). Mapping of rice varieties with sentinel-2 data via deep cnn learning in spectral and time domains. In *2018 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-7). IEEE.
- Inoue, S., Ito, A., & Yonezawa, C. (2020). Mapping Paddy fields in Japan by using a Sentinel-1 SAR time series supplemented by Sentinel-2 images on Google Earth Engine. *Remote Sensing*, 12(10), 1622.
- Izcard, G., & Grave, E. (2020). Leveraging passage retrieval with generative models for open domain question answering. arXiv preprint arXiv:2007.01282.
- Jiang, T., Liu, X., & Wu, L. (2018). Method for mapping rice fields in complex landscape areas based on pre-trained convolutional neural network from HJ-1 A/B data. *ISPRS International Journal of Geo-Information*, 7(11), 418.
- Jo, H. W., Lee, S., Park, E., Lim, C. H., Song, C., Lee, H., ... & Lee, W. K. (2020). Deep learning applications on multitemporal SAR (Sentinel-1) image classification using confined labeled data: The case of detecting rice paddy in South Korea. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11), 7589-7601.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Liu, H., Yao, Y., Sun, Z., Li, X., Jia, K., & Tang, Z. (2020). Road segmentation with image-LiDAR data fusion in deep neural network. *Multimedia Tools and Applications*, 79(47), 35503-35518.
- Liu, L., Huang, J., Xiong, Q., Zhang, H., Song, P., Huang, Y., ... & Wang, X. (2020). Optimal MODIS data processing for accurate multi-year paddy rice area mapping in China. *GIScience & Remote Sensing*, 57(5), 687-703.
- Lobell, D. B., & Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93(3), 412-422.
- Mansaray, L. R., Huang, W., Zhang, D., Huang, J., & Li, J. (2017). Mapping rice fields in urban Shanghai, southeast China, using Sentinel-1A and Landsat 8 datasets. *Remote Sensing*, 9(3), 257.
- Mosleh, M. K., Hassan, Q. K., & Chowdhury, E. H. (2015). Application of remote sensors in mapping rice area and forecasting its production: A review. *Sensors*, 15(1), 769-791.
- Rawat, A., Kumar, A., Upadhyay, P., & Kumar, S. (2021). Deep learning-based models for temporal satellite data processing: Classification of paddy transplanted fields. *Ecological Informatics*, 61, 101214.
- Rouse Jr, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation (No. NASA-CR-132982).
- Talema, T., & Hailu, B. T. (2020). Mapping rice crop using sentinels (1 SAR and 2 MSI) images in tropical area: A case study in Fogera wereda, Ethiopia. *Remote Sensing Applications: Society and Environment*, 18, 100290.
- Taylor, L., & Nitschke, G. (2018, November). Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1542-1547). IEEE.
- Wahlberg, B., Boyd, S., Annergren, M., & Wang, Y. (2012). An ADMM algorithm for a class of total variation regularized estimation problems. *IFAC Proceedings Volumes*, 45(16), 83-88.

- Wang, J., Huang, J. F., Wang, X. Z., Jin, M. T., Zhou, Z., Guo, Q. Y., ... & Song, X. D. (2015). Estimation of rice phenology date using integrated HJ-1 CCD and Landsat-8 OLI vegetation indices time-series images. *Journal of Zhejiang University-SCIENCE B*, 16(10), 832-844.
- Wang, Y., Zang, S., & Tian, Y. (2020). Mapping paddy rice with the random forest algorithm using MODIS and SMAP time series. *Chaos, Solitons & Fractals*, 140, 110116.
- Yang, Z., Shao, Y., Li, K., Liu, Q., Liu, L., & Brisco, B. (2017). An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data. *Remote sensing of environment*, 195, 184-201.
- Yonezawa, C., & Watanabe, M. (2020). Analysis of the applicability of multi-temporal full polarimetric airborne L-band SAR scattering to paddy rice field mapping. *International Journal of Remote Sensing*, 41(7), 2500-2516.
- Zhan, P., Zhu, W., & Li, N. (2021). An automated rice mapping method based on flooding signals in synthetic aperture radar time series. *Remote Sensing of Environment*, 252, 112112.
- Zhang, M., Lin, H., Wang, G., Sun, H., & Fu, J. (2018). Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China. *Remote Sensing*, 10(11), 1840.
- Zhang, W., Liu, H., Wu, W., Zhan, L., & Wei, J. (2020). Mapping rice paddy based on machine learning with Sentinel-2 multi-temporal data: Model comparison and transferability. *Remote Sensing*, 12(10), 1620.
- Zhang, Z., & Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31.
- Zhao, S., Liu, X., Ding, C., Liu, S., Wu, C., & Wu, L. (2020). Mapping rice paddies in complex landscapes with convolutional neural networks and phenological metrics. *GIScience & Remote Sensing*, 57(1), 37-48.
- Zhao, W., Guo, Z., Yue, J., Zhang, X., & Luo, L. (2015). On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *International Journal of Remote Sensing*, 36(13), 3368-3379.